Statistical models of appearance for medical image analysis and computer vision

T.F. Cootes and C.J. Taylor

Imaging Science and Biomedical Engineering, University of Manchester, UK

ABSTRACT

Statistical models of shape and appearance are powerful tools for interpreting medical images. We assume a training set of images in which corresponding 'landmark' points have been marked on every image. From this data we can compute a a statistical model of the shape variation, a model of the texture variation and a model of the correlations between shape and texture. With enough training examples such models should be able to synthesize any image of normal anatomy. By finding the parameters which optimize the match between a synthesized model image and a target image we can locate all the structures represented by the model. Two approaches to the matching will be described. The Active Shape Model essentially matches a model to boundaries in an image. The Active Appearance Model finds model parameters which synthesize a complete image which is as similar as possible to the target image. By using a 'difference decomposition' approach the current difference between target image and synthesized model image can be used to update the model parameters, leading to rapid matching of complex models. We will demonstrate the application of such models to a variety of different problems.

Keywords: Shape Models, Appearance Models, Model Matching

1. INTRODUCTION

Many problems in medical image interpretation involve the need for an automated system to 'understand' the images with which it is presented - that is, to recover image structure and know what it means. This necessarily involves the use of models which describe and label the expected structure of the world. Real applications are also typically characterised by the need to deal with complex and variable structure and with images that provide noisy and possibly incomplete evidence - it is often impossible to interpret a given image without prior knowledge of anatomy.

Model-based methods offer potential solutions to all these difficulties. Prior knowledge of the problem can, in principle, be used to resolve the potential confusion caused by structural complexity, provide tolerance to noisy or missing data, and provide a means of labelling the recovered structures. We would like to apply knowledge of the expected shapes of structures, their spatial relationships, and their grey-level appearance to restrict our automated system to 'plausible' interpretations. Of particular interest are generative models - that is, models sufficiently complete that they are able to generate realistic images of target objects. An example would be a face model capable of generating convincing images of any individual, changing their expression and so on. Using such a model, image interpretation can be formulated as a matching problem: given an image to interpret, structures can be located and labelled by adjusting the model's parameters in such a way that it generates an 'imagined image' which is as similar as possible to the real thing.

Because real applications often involve dealing with classes of objects which are not identical we need to deal with variability. This leads naturally to the idea of deformable models - models which maintain the essential characteristics of the class of objects they represent, but which can deform to fit a range of examples. There are two main characteristics we would like such models to possess. First, they should be general - that is, they should be capable of generating any plausible example of the class they represent. Second, and crucially, they should be specific - that is, they should only be capable of generating 'legal' examples - because, as we noted earlier, the whole point of using a model-based approach is to limit the attention of our system to plausible interpretations. In order to obtain specific models of variable objects, we need to acquire knowledge of how they vary.

A powerful approach is to learn the variation from a suitably annotated training set of typical images. We describe below how statistical models can be constructed to represent both the shape and the 'texture' (the pattern of pixel intensities) of examples of structures of interest. These models can generalise from the training set and be used to match to new images, locating the structure in the images. Two approaches are summarised. The first, the 'Active Shape Model', concentrates on matching a shape model to an image, typically matching the model to boundaries

Medical Imaging 2001: Image Processing, Milan Sonka, Kenneth M. Hanson, Editors,
Proceedings of SPIE Vol. 4322 (2001) © 2001 SPIE · 1605-7422/01/$15.00

of the target structure. The second approach, the 'Active Appearance Model', attempts to synthesize the complete appearance of the target image, choosing parameters which minimise the difference between the target image and an image generated from the model. Both algorithms have proved to be fast, accurate and reliable.

In the remainder of this paper we outline our approach to modelling shapes, spatial relationships and grey-level appearance, show how these models can be used in image interpretation, describe practical applications of the approach in medical image interpretation, discuss the strengths and weaknesses of the approach, and draw conclusions.

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 model to constrain solutions to be valid examples of the structure(s) modelled. Of more fundamental importance is the fact that, once a model 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[1,2] We give here a brief review covering some of the more important points.

Model matching algorithms can be crudely classified as 'shape based', in which a deformable model represents, and matches to, boundaries or other sparce features, and 'appearance based', in which the model represents the whole image region covered by the structure.

2.1. Shape Based Approaches

Various approaches to modelling variability have been described previously. The most common general approach is to allow a prototype to vary according to some physical model. Kass and Witkin[3] describe 'snakes' which deform elastically to fit shape contours. Park et al[4] and Pentland and Sclaroff[5] both represent prototype objects using finite element methods and describe variability in terms of vibrational modes. Alternative approaches include representation of shapes using sums of trigonometric functions with variable coefficients[6,7] and parameterised models, hand-crafted for particular applications[8,9] Grenander et al[10] and Dryden and Mardia[11] described statistical models of shape. These were, however, difficult to use in automated image interpretation. Goodall[12] and Bookstein[13] have used statistical techniques for morphometric analysis. Subsol *et. al.* [14] extract crest-lines, which they use to establish landmark-based correspondence. They use these to perform morphometrical studies and to match images to atlases.

2.2. Appearance Based Approaches

The simplest form is that of using correllation to match a 'golden' image of an object to a new target. Image registration[2] uses an extension of this general idea, in which a single image is matched to a new image either rigidly or allowing non-rigid deformations. In this case typically the texture is fixed but the shape is allowed to vary.

An extension is to match a model image (or anatomical atlas) to a target image, in order to interpret the latter. For instance Bajcsy and Kovacic[15] describe a volume model (of the brain) that also deforms elastically to generate new examples.

In later work, Bajcsy *et. al.* describe an image-based atlas that deforms to fit new images by minimising pixel/voxel intensity differences.[16] 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[17,18]

Kirby and Sirovich[19] have described statistical modelling of grey-level appearance (particularly for face images) but did not address shape variability.

Nastar *et. al.*[20] describe a model of shape and intensity variations by using a 3D deformable model of the intensity landscape. They used a closest point surface matching algorithm to perform the fitting, which tends to be sensitive to the initial position. Jones and Poggio use a model capable of synthesizing faces and describe a stochastic optimisation method to match the model to new face images.[21] The method is slow but can be robust because of the quality of the synthesized images. Vetter[22] uses a 3D variation of this approach, with a general purpose optimization algorithm

to perform the matching. Wang and Staib[23] have incorporated statistical shape information into an image-based elastic matching approach.

Fast matching algorithm for appearance based models have been developed in the tracking community. Gleicher[24] describes a method of tracking objects by allowing a single template to deform under a variety of transformations (affine, projective etc). He chooses the parameters to minimize a sum of squares measure and essentially precomputes derivatives of the difference vector with respect to the parameters of the transformation. Hager and Belhumeur[25] describe a similar approach, but include robust kernels and models of illumination variation. Sclaroff and Isidoro[26] extend the approach to track objects which deform, modeling deformation using the low energy modes of a finite element model of the target. The approach has been used to track heads[27] using a rigid cylindrical model of the head.

3. STATISTICAL MODELS OF APPEARANCE

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).

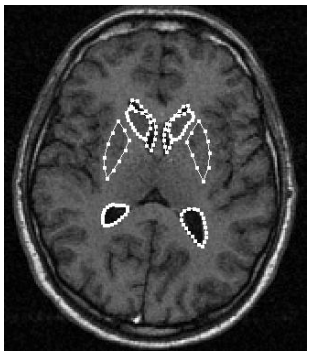


Figure 1. Example of MR brain slice labelled with 123 landmark points around the ventricles, the caudate nucleus and the lentiform nucleus

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[28] 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 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 $\mathbf{c}$

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

A shape in the image frame, $\mathbf{X}$, can be generated by applying a suitable transformation to the points, $\mathbf{x}$: $\mathbf{X} = S_{\mathbf{t}}(\mathbf{x})$. Typically $S_{\mathbf{t}}$ will be a similarity transformation described by a scaling, s, an in-plane rotation, θ, and a translation (t_x, t_y). For linearity we represent the scaling and rotation as (s_x, s_y) where $s_x = (s\cos\theta - 1)$, $s_y = s\sin\theta$. The pose parameter vector $\mathbf{t} = (s_x, s_y, t_x, t_y)^T$ is then zero for the identity transformation and $S_{\mathbf{t}+\delta\mathbf{t}}(\mathbf{x}) \approx S_{\mathbf{t}}(S_{\delta\mathbf{t}}(\mathbf{x}))$.

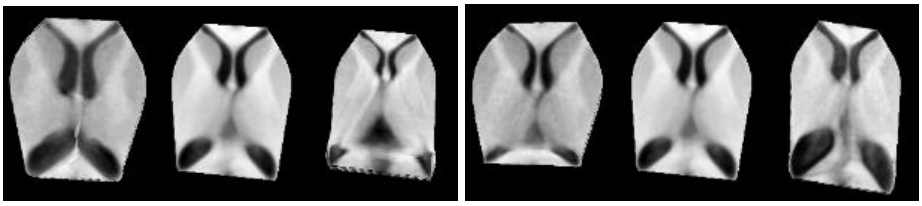
The texture in the image frame is generated by applying a scaling and offset to the intensities, $\mathbf{g}_{im} = T_{\mathbf{u}}(\mathbf{g}) = (u_1 + 1)\mathbf{g}_{im} + u_2\mathbf{1}$, where $\mathbf{u}$ is the vector of transformation parameters, defined so that $\mathbf{u} = \mathbf{0}$ is the identity transformation and $T_{\mathbf{u}+\delta\mathbf{u}}(\mathbf{g}) \approx T_{\mathbf{u}}(T_{\delta\mathbf{u}}(\mathbf{g}))$.

A full reconstruction is given by generating the texture in a mean shaped patch, then warping it so that the model points lie on the image points, $\mathbf{X}$.

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

Figure 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

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

4. ACTIVE SHAPE MODELS

Given a rough starting approximation, an instance of a model can be fit to an image. By choosing a set of shape parameters, $\mathbf{b}$ for the model we define the shape of the object in an object-centred co-ordinate frame. We can create an instance $\mathbf{X}$ of the model in the image frame by defining the position, orientation and scale.

An iterative approach to improving the fit of the instance, $\mathbf{X}$, to an image proceeds as follows:

1. Examine a region of the image around each point $\mathbf{X}_i$ to find the best nearby match for the point $\mathbf{X}'_i$
2. Update the parameters $(\mathbf{t}, \mathbf{b})$ to best fit the new found points $\mathbf{X}$
3. Repeat until convergence.

In practise we look along profiles normal to the model boundary through each model point (Figure 4). If we expect the model boundary to correspond to an edge, we can simply locate the strongest edge (including orientation if known) along the profile. The position of this gives the new suggested location for the model point.

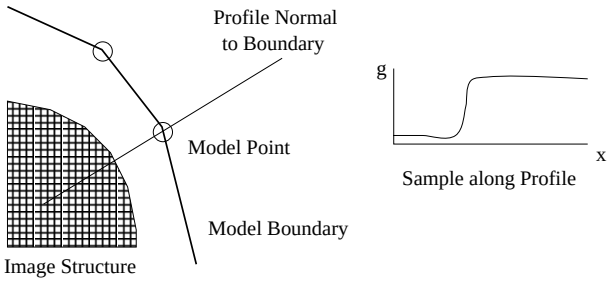


Figure 4. At each model point sample along a profile normal to the boundary

However, model points are not always placed on the strongest edge in the locality - they may represent a weaker secondary edge or some other image structure. The best approach is to learn from the training set what to look for in the target image. This is done by sampling along the profile normal to the boundary in the training set, and building a statistical model of the grey-level structure.

4.1. Modelling Local Structure

Suppose for a given point we sample along a profile k pixels either side of the model point in the i^{th} training image. We have $2k+1$ samples which can be put in a vector $\mathbf{g}_i$. To reduce the effects of global intensity changes we sample the derivative along the profile, rather than the absolute grey-level values. We then normalise the sample by dividing through by the sum of absolute element values,

$$\mathbf{g}_i \rightarrow \frac{1}{\sum_j |g_{ij}|}\mathbf{g}_i \tag{5}$$

We repeat this for each training image, to get a set of normalised samples $\{\mathbf{g}_i\}$ for the given model point. We assume that these are distributed as a multivariate gaussian, and estimate their mean $\hat{\mathbf{g}}$ and covariance $\mathbf{S}_g$. This gives a statistical model for the grey-level profile about the point. This is repeated for every model point, giving one grey-level model for each point.

The quality of fit of a new sample, $\mathbf{g}_s$, to the model is given by

$$f(\mathbf{g}_s) = (\mathbf{g}_s - \hat{\mathbf{g}})^T \mathbf{S}_g^{-1} (\mathbf{g}_s - \hat{\mathbf{g}}) \tag{6}$$

This is the Mahalanobis distance of the sample from the model mean, and is linearly related to the log of the probability that $\mathbf{g}_s$ is drawn from the distribution. Minimising $f(\mathbf{g}_s)$ is equivalent to maximising the probability that $\mathbf{g}_s$ comes from the distribution.

During search we sample a profile m pixels either side of the current point ($m > k$). We then test the quality of fit of the corresponding grey-level model at each of the $2(m-k)+1$ possible positions along the sample (Figure 5) and choose the one which gives the best match (lowest value of $f(\mathbf{g}_s)$).

This is repeated for every model point, giving a suggested new position for each point. We then compute the pose and shape parameters which best match the model to the new points, effectively imposing shape constraints on the allowed point positions.

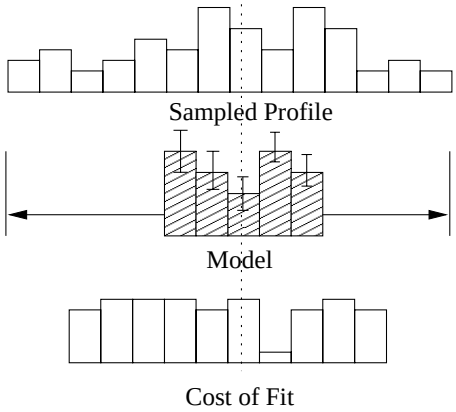


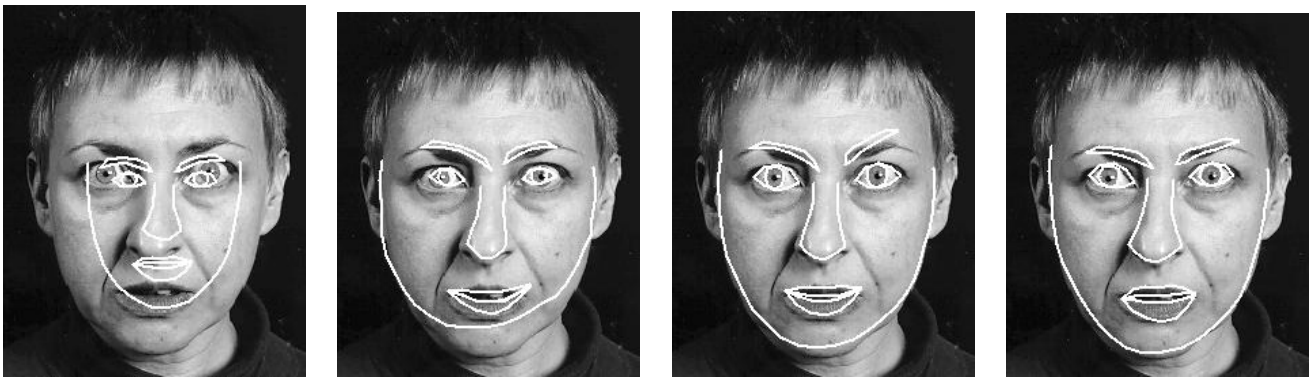
Figure 5. Search along sampled profile to find best fit of grey-level model

4.2. Multi-Resolution Active Shape Models

To improve the efficiency and robustness of the algorithm, it is implement in a multi-resolution framework. This involves first searching for the object in a coarse image, then refining the location in a series of finer resolution images. This leads to a faster algorithm, and one which is less likely to get stuck on the wrong image structure.

4.3. Examples of Search

Figure 6 demonstrates using the ASM to locate the features of a face. The model instance is placed near the centre of the image and a coarse to fine search performed. The search starts at level 3 (1/8 the resolution in x and y compared to the original image). Large movements are made in the first few iterations, getting the position and scale roughly correct. As the search progresses to finer resolutions more subtle adjustments are made. The final convergence (after a total of 18 iterations) gives a good match to the target image. In this case at most 5 iterations were allowed at each resolution, and the algorithm converges in much less than a second (on a modern PC).



Initial | After 2 iterations | After 6 iterations | After 18 iterations

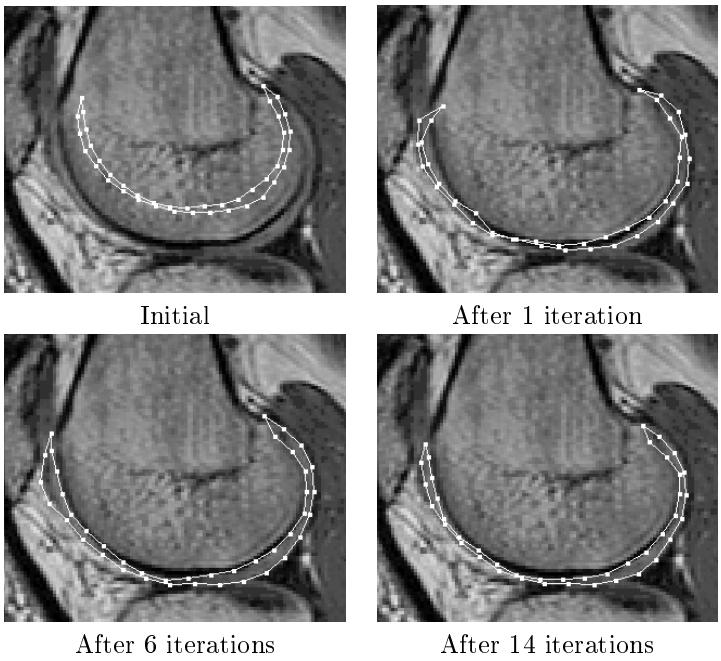
Figure 6. Search using an Active Shape Model of a face

Figure 7 demonstrates how the ASM can fail if the starting position is too far from the target. Since it is only searching along profiles around the current position, it cannot correct for large displacements from the correct position. It will either diverge to infinity, or converge to an incorrect solution, doing its best to match the local image data. In the case shown it has been able to locate half the face, but the other side is too far away.

Figure 8 demonstrates using the ASM of the cartilage to locate the structure in a new image. In this case the search starts at level 2, samples at 2 points either side of the current point and allows at most 5 iterations per level. A detailed description of the application of such a model is given by Solloway *et. al.*.[29]

Initial After 2 iterations After 20 Iterations

Figure 7. Search using Active Shape Model of a face, given a poor starting point. The ASM is a local method, and may fail to locate an acceptable result if initialised too far from the target



Initial After 1 iteration

After 6 iterations After 14 iterations

Figure 8. Search using ASM of cartilage on an MR image of the knee

5. ACTIVE APPEARANCE MODELS

This section outlines the basic AAM matching algorithm. A more comprehensive description is given by Cootes *et. al.*.[30] 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.

5.1. Overview of AAM Search

The appearance model parameters, $\mathbf{c}$, and shape transformation parameters, $\mathbf{t}$, define the position of the model points in the image frame, $\mathbf{X}$, which gives the shape of the image patch to be represented by the model. During matching we sample the pixels in this region of the image, $\mathbf{g}_{im}$, and project into the texture model frame, $\mathbf{g}_s = T^{-1}(\mathbf{g}_{im})$. The current model texture is given by $\mathbf{g}_m = \bar{\mathbf{g}} + \mathbf{P}_g\mathbf{Q}_g\mathbf{c}$. The current difference between model and image (measured in the normalized texture frame) is thus

$$\mathbf{r}(\mathbf{p}) = \mathbf{g}_s - \mathbf{g}_m \tag{7}$$

where $\mathbf{p}$ are the parameters of the model, $\mathbf{p}^T = (\mathbf{c}^T|\mathbf{t}^T|\mathbf{u}^T)$.

A simple scalar measure of difference is the sum of squares of elements of $\mathbf{r}$, $E(\mathbf{p}) = \mathbf{r}^T\mathbf{r}$.

A first order Taylor expansion of (7) gives

$$\mathbf{r}(\mathbf{p}+\delta\mathbf{p}) = \mathbf{r}(\mathbf{p}) + \frac{\partial \mathbf{r}}{\partial \mathbf{p}}\delta\mathbf{p} \tag{8}$$

Where the ij^{th} element of matrix $\frac{\partial \mathbf{r}}{\partial \mathbf{p}}$ is $\frac{dr_i}{dp_j}$.

Suppose during matching our current residual is $\mathbf{r}$. We wish to choose $\delta\mathbf{p}$ so as to minimize $|\mathbf{r}(\mathbf{p}+\delta\mathbf{p})|^2$. By equating (8) to zero we obtain the RMS solution,

$$\delta\mathbf{p} = -\mathbf{R}\mathbf{r}(\mathbf{p}) \quad where \quad \mathbf{R} = (\frac{\partial \mathbf{r}}{\partial \mathbf{p}}^T \frac{\partial \mathbf{r}}{\partial \mathbf{p}})^{-1} \frac{\partial \mathbf{r}}{\partial \mathbf{p}}^T \tag{9}$$

In a standard optimization scheme it would be necessary to recalculate $\frac{\partial \mathbf{r}}{\partial \mathbf{p}}$ at every step, an expensive operation. However, we assume that since it is being computed in a normalized reference frame, it can be considered approximately fixed. We can thus estimate it once from our training set. We estimate $\frac{\partial \mathbf{r}}{\partial \mathbf{p}}$ by numeric differentiation, systematically displacing each parameter from the known optimal value on typical images and computing an average over the training set. Residuals at displacements of differing magnitudes are measured (typically up to 0.5 standard deviations of each parameter) and combined with a Gaussian kernel to smooth them. We then precompute $\mathbf{R}$ and use it in all subsequent searches with the model.

Images used in the calculation of $\frac{\partial \mathbf{r}}{\partial \mathbf{p}}$ can either be examples from the training set or synthetic images generated using the appearance model itself. Where synthetic images are used, one can either use a suitable (e.g. random) background, or can detect the areas of the model which overlap the background and remove those samples from the model building process. This latter makes the final relationship more independent of the background. Where the background is predictable (e.g. medical images), this is not necessary.

5.2. Iterative Model Refinement

Using equation (9) we can suggest a correction to make in the model parameters based on a measured residual $\mathbf{r}$. This allows us to construct an iterative algorithm for solving our optimization problem. Given a current estimate of the model parameters, $\mathbf{c}$, the pose $\mathbf{t}$, the texture transformation $\mathbf{u}$, and the image sample at the current estimate, $\mathbf{g}_{im}$, one step of the iterative procedure is as follows:

1. Project the texture sample into the texture model frame using $\mathbf{g}_s = T_{\mathbf{u}}^{-1}(\mathbf{g}_{im})$
2. evaluate the error vector, $\mathbf{r} = \mathbf{g}_s - \mathbf{g}_m$, and the current error, $E = |\mathbf{r}|^2$
3. compute the predicted displacements, $\delta\mathbf{p} = -\mathbf{R}\mathbf{r}(\mathbf{p})$
4. update the model parameters $\mathbf{p} \to \mathbf{p} + k\delta\mathbf{p}$, where initially $k = 1$,
5. calculate the new points, $\mathbf{X}'$ and model frame texture $\mathbf{g}'_m$
6. sample the image at the new points to obtain $\mathbf{g}'_{im}$
7. calculate a new error vector, $\mathbf{r}' = T_{\mathbf{u}'}^{-1}(\mathbf{g}'_{im}) - \mathbf{g}'_m$
8. if $|\mathbf{r}'|^2 < E$ then accept the new estimate, otherwise try at $k = 0.5$, $k = 0.25$ etc.

This procedure is repeated until no improvement is made to the error, $|\mathbf{r}|^2$, and convergence is declared. In practice we use a multi-resolution implementation, in which we start at a coarse resolution and 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. The complexity of the algorithm is $O(n_{pixels} n_{modes})$ at a given level. Essentially each iteration involves sampling n_{pixels} points from the image then multiplying by a n_{modes} x n_{pixel} matrix.

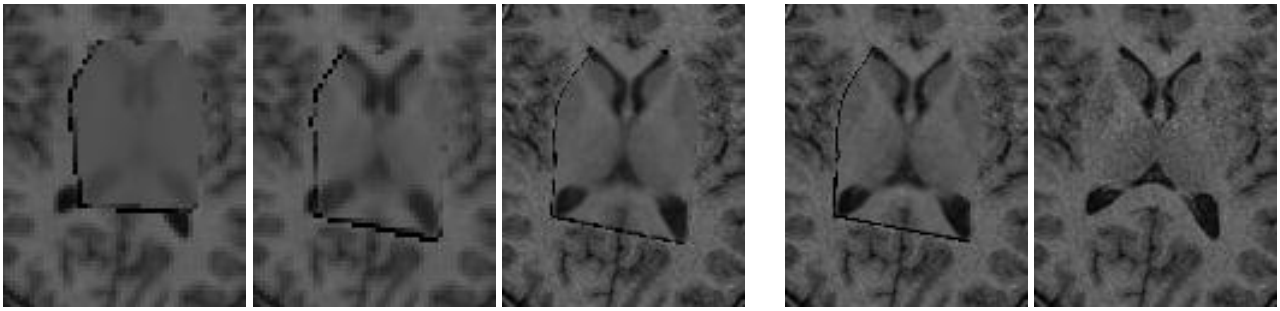


Figure 9. Multi-resolution AAM search from a displaced position

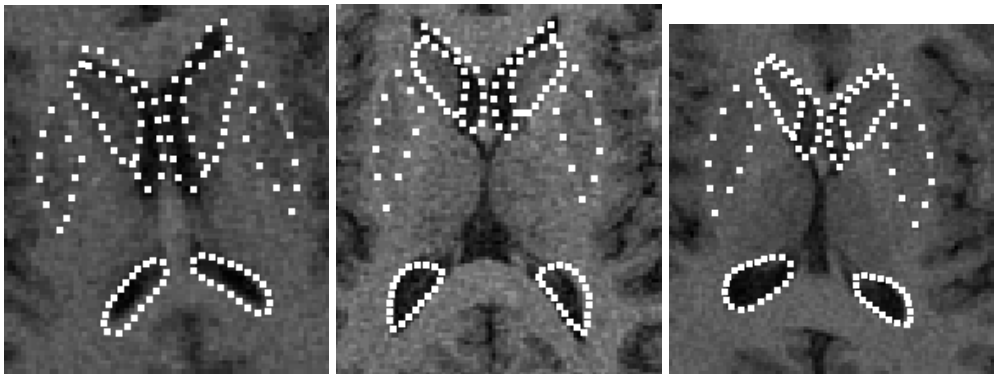


Figure 10. Results of AAM search. Model points superimposed on target image

5.3. Examples of AAM Search

For example, Figure 9 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 **c** parameters. The search took about a second on a modern PC. Figure 10 shows examples of the results of the search, with the found model points superimposed on the target images.

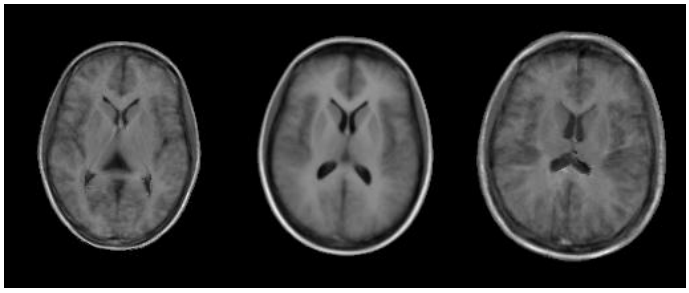
Though we only demonstrated on the central part of the brain, models can be build of the whole cross-section. Figure 11 shows the first two modes of such a model. This was trained from the same 72 example slices as above, but with additional points marked around the outside of the skull. The first modes 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.[31,32]

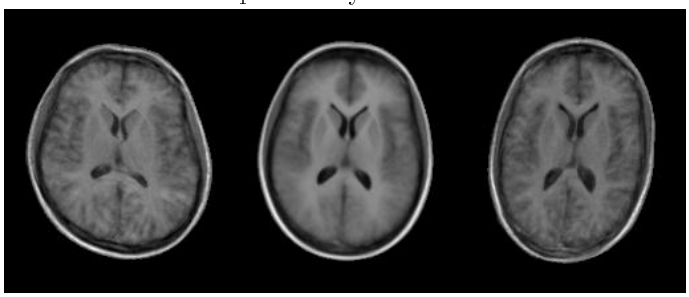
When the AAM converges it will usually be close to the optimal result, but may not achieve the exact position. Fisker[33] has shown that applying a general purpose optimiser can improve the final match.

5.4. Examples of Failure

Figure 12 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.[34] 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).



c_1 varies by ± 2 s.d.s



c_2 varies by ± 2 s.d.s

Figure 11. First two modes of appearance model of full brain cross-section from an MR image

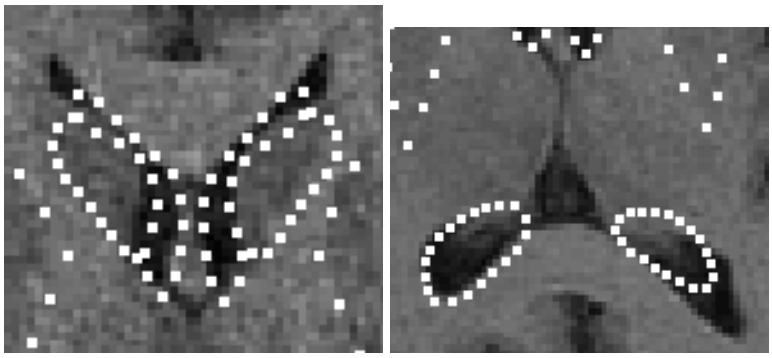


Figure 12. Detail of examples of search failure. The AAM does not always find the correct outer boundaries of the ventricles (see text).

6. DISCUSSION AND CONCLUSIONS

We have demonstrated that image structures can be represented using statistical models of shape and appearance. Both the shape and the appearance of the structures can vary in ways observed in the training set. Arbitrary deformations are not allowed. Matching to a new image can be achieved rapidly using either the Active Shape Model or the Active Appearance Model algorithms.

6.1. Applications

ASMs have been used to locate vertebrae in DEXA Images of the spine,[35] bones and prostheses in radiographs of total hip replacements,[36] structures in MR images of the brain,[37] and the outlines of ventricles in echocardiograms.[37,38] Both ASMs and AAMs have been used in face image interpretation.[39,40] The approaches can be extended to 3D, and have been used for interpretting volume images.[41–45]

6.2. Comparison between ASMs and AAMs

Active Shape Models search around the current location, along profiles, so tend to have a larger capture range than the AAM which only examines the image directly under its current area.[46]

ASMs only use data around the model points, and do not take advantage of all the grey-level information available across an object as the AAM does. Thus they may be less reliable. However, the model points tend to be places of interest (boundaries or corners) where there is the most information. One could train an AAM to only search using information in areas near strong boundaries - this would require less image sampling during search so a potentially quicker algorithm (see for instance work by Fisker[33]). A more formal approach is to learn from the training set which pixels are most useful for search - this was explored in.[47] The resulting search is faster, but tends to be less reliable.

One advantage of the AAM is that one can build a convincing model with a relatively small number of landmarks. Any extra shape variation is expressed in additional modes of the texture model. The ASM needs points around boundaries so as to define suitable directions for search. Because of the considerable work required to get reliable image labelling, the fewer landmarks required, the better.

In general we have found that the ASM is faster and achieves more accurate feature point location than the AAM. However, as it explicitly minimises texture errors the AAM gives a better match to the image texture, and can be more robust.

It is possible to combine the two approaches. For instance, Mitchell *et. al.*[38] used a combination of ASM and AAM to segment cardiac images. At each iteration the two models ran independently to compute new estimates of the pose and shape parameters. These were then combined using a weighted average. They showed that this approach gave better results than the AAM alone.

6.3. Extensions to 3D

The approaches have been demonstrated in 2D, are extensible to 3D.[37] The main complications are the size of the models and the difficulty of obtaining well annotated training data. Obtaining good (dense) correspondences in 3D images is difficult, and is the subject of current research.[41–43,48,44,45]

Extending the ASMs to 3D is relatively straightforward, given a suitable set of annotated images. The profiles modelled and sampled are simply taken along lines through the 3D image orthogonal to the local surface.[37,44] Kelemen *et. al.*[49] describe further modifications, including using a continuous surface representation rather than a set of points.

In theory extending the AAM is straightforward, but in practice the models would be extremely large. Each mode of the appearance model (and corresponding derivative vector) is the size of a full (3D) image, and many modes may be required. A more practical approach is likely to be only sampling in bands around the boundaries of interest.

The approaches can also be extended into the temporal domain, to track objects through sequences, for instance, the heart boundary in echocardiograms.[50]

6.4. Conclusion

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 images. Such models can be matched to new images rapidly and reliably using either the ASM or the AAM algorithms. The methods are applicable to a wide variety of problems and give a useful framework for automatic image interpretation.

ACKNOWLEDGMENTS

Dr Cootes was 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. T. McInerney and D. Terzopoulos, "Deformable models in medical image analysis: a survey," *Medical Image Analysis* **1**(2), pp. 91–108, 1996.
2. J. B. A. Maintz and M. A. Viergever, "A survey of medical image registration," *Medical Image Analysis* **2**(1), pp. 1–36, 1998.

3. M. Kass, A. Witkin, and D. Terzopoulos, "Active contour models," *International Journal of Computer Vision* **1**(4), pp. 321–331, 1987.
4. J. Park, D. Mataxas, A. Young, and L. Axel, "Deformable models with parameter functions for cardiac motion analysis from tagged mri data," *IEEE Transactions on Medical Imaging* **15**, pp. 278–289, 1996.
5. A. P. Pentland and S. Sclaroff, "Closed-form solutions for physically based modelling and recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence* **13**(7), pp. 715–729, 1991.
6. G. L. Scott, "The alternative snake – and other animals," in 3^{rd} *Alvey Vison Conference, Cambridge, England*, pp. 341–347, 1987.
7. L. H. Staib and J. S. Duncan, "Boundary finding with parametrically deformable models," *IEEE Transactions on Pattern Analysis and Machine Intelligence* **14**(11), pp. 1061–1075, 1992.
8. A. L. Yuille, D. S. Cohen, and P. Hallinan, "Feature extraction from faces using deformable templates," *International Journal of Computer Vision* **8**(2), pp. 99–112, 1992.
9. P. Lipson, A. L. Yuille, D. O'Keeffe, J. Cavanaugh, J. Taaffe, and D. Rosenthal, "Deformable templates for feature extraction from medical images," in 1^{st} *European Conference on Computer Vision*, O. Faugeras, ed., pp. 413–417, Springer-Verlag, Berlin/New York, 1990.
10. U. Grenander and M. Miller, "Representations of knowledge in complex systems," *Journal of the Royal Statistical Society B* **56**, pp. 249–603, 1993.
11. I. Dryden and K. V. Mardia, *The Statistical Analysis of Shape*, Wiley, London, 1998.
12. C. Goodall, "Procrustes methods in the statistical analysis of shape," *Journal of the Royal Statistical Society B* **53**(2), pp. 285–339, 1991.
13. F. L. Bookstein, "Principal warps: Thin-plate splines and the decomposition of deformations," *IEEE Transactions on Pattern Analysis and Machine Intelligence* **11**(6), pp. 567–585, 1989.
14. G. Subsol, J. P. Thirion, and N. Ayache, "A general scheme for automatically building 3d morphometric anatomical atlases: application to a skull atlas," *Medical Image Analysis* **2**, pp. 37–60, 1998.
15. Bajcsy and A. Kovacic, "Multiresolution elastic matching," *Computer Graphics and Image Processing* **46**, pp. 1–21, 1989.
16. 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**, pp. 618–625, 1983.
17. G. E. Christensen, R. D. Rabbitt, M. I. Miller, S. C. Joshi, U. Grenander, T. A. Coogan, and D. C. V. Essen, "Topological properties of smooth anatomic maps," in 14^{th} *Conference on Information Processing in Medical Imaging, France*, pp. 101–112, Kluwer Academic Publishers, 1995.
18. G. E. Christensen, S. C. Joshi, and M. Miller, "Volumetric transformation of brain anatomy," *IEEE Trans. Medical Image* **16**, pp. 864–877, 1997.
19. M. Kirby and L. Sirovich, "Appliction of the karhumen-loeve procedure for the characterization of human faces," *IEEE Transactions on Pattern Analysis and Machine Intelligence* **12**(1), pp. 103–108, 1990.
20. C. Nastar, B. Moghaddam, and A. Pentland, "Generalized image matching: Statistical learning of physically-based deformations," *Computer Vision and Image Understanding* **65**(2), pp. 179–191, 1997.
21. M. J. Jones and T. Poggio, "Multidimensional morphable models," in 6^{th} *International Conference on Computer Vision*, pp. 683–688, 1998.
22. T. Vetter, "Learning novel views to a single face image," in 2^{nd} *International Conference on Automatic Face and Gesture Recognition 1997*, pp. 22–27, IEEE Computer Society Press, (Los Alamitos, California), Oct. 1996.
23. Y. Wang and L. H. Staib, "Elastic model based non-rigid registration incorporating statistical shape information," in *MICCAI*, pp. 1162–1173, 1998.
24. M. Gleicher, "Projective registration with difference decomposition," in *IEEE Conference on Computer Vision and Pattern Recognition*, 1997.
25. G. Hager and P. Belhumeur, "Efficient region tracking with parametric models of geometry and illumination," *IEEE Transactions on Pattern Analysis and Machine Intelligence* **20**(10), pp. 1025–39, 1998.
26. S. Sclaroff and J. Isidoro, "Active blobs," in 6^{th} *International Conference on Computer Vision*, pp. 1146–53, 1998.
27. M. La Cascia, S. Sclaroff, and V. Athitsos, "Fast, reliable head tracking under varying illumination: An approach based on registration of texture mapped 3d models," *IEEE Transactions on Pattern Analysis and Machine Intelligence* **22**(4), pp. 322–336, 2000.

28. T. F. Cootes, C. J. Taylor, D. Cooper, and J. Graham, "Active shape models - their training and application," *Computer Vision and Image Understanding* **61**, pp. 38–59, Jan. 1995.
29. S. Solloway, C. Hutchinson, J. Waterton, and C. J. Taylor, "Quantification of articular cartilage from MR images using active shape models," in 4^{th} *European Conference on Computer Vision*, B. Buxton and R. Cipolla, eds., vol. 2, pp. 400–411, Springer-Verlag, (Cambridge, England), April 1996.
30. T. F. Cootes, G. J. Edwards, and C. J. Taylor, "Active appearance models," in 5^{th} *European Conference on Computer Vision*, H.Burkhardt and B. Neumann, eds., vol. 2, pp. 484–498, Springer, Berlin, 1998.
31. A. Caunce and C. J. Taylor, "3d point distribution models of the cortical sulci," in 8^{th} *British Machine Vison Conference*, A. F. Clark, ed., pp. 550–559, BMVA Press, (University of Essex, UK), Sept. 1997.
32. A. Caunce and C. J. Taylor, "Using local geometry to build 3d sulcal models," in 16^{th} *Conference on Information Processing in Medical Imaging*, pp. 196–209, 1999.
33. R. Fisker, *Making Deformable Template Models Operational*. PhD thesis, Informatics and Mathematical Modelling, Technical University of Denmark, 2000.
34. T. F. Cootes, A. Hill, C. J. Taylor, and J. Haslam, "The use of active shape models for locating structures in medical images," *Image and Vision Computing* **12**, pp. 276–285, July 1994.
35. P. P. Smyth, C. J. Taylor, and J. E. Adams, "Automatic measurement of vertebral shape using active shape models," in 7^{th} *British Machine Vison Conference*, pp. 705–714, BMVA Press, (Edinburgh, Scotland), Sept. 1996.
36. A. Kotcheff, A. Redhead, C. J. Taylor, and D. Hukins, "Shape model analysis of thr radiographs," in 13^{th} *International Conference on Pattern Recognition*, vol. 4, pp. 391–395, IEEE Computer Society Press, 1996.
37. A. Hill, T. F. Cootes, C. J. Taylor, and K. Lindley, "Medical image interpretation: A generic approach using deformable templates," *Journal of Medical Informatics* **19**(1), pp. 47–59, 1994.
38. S. Mitchell, B.Lelieveldt, R. van der Geest, J. Schaap, J. Reiber, and M. Sonka, "Segmentation of cardiac mr images: An active appearance model approach," in *SPIE Medical Imaging*, Feb. 2000.
39. A. Lanitis, C. J. Taylor, and T. F. Cootes, "Automatic interpretation and coding of face images using flexible models," *IEEE Transactions on Pattern Analysis and Machine Intelligence* **19**(7), pp. 743–756, 1997.
40. G. Edwards, A. Lanitis, C. Taylor, and T. Cootes, "Statistical models of face images - improving specificity," *Image and Vision Computing* **16**, pp. 203–211, 1998.
41. A. Brett and C. Taylor, "Construction of 3d shape models of femoral articular cartilage using harmonic maps," in *MICCAI*, pp. 1205–1214, 2000.
42. A. D. Brett and C. J. Taylor, "A method of automatic landmark generation for automated 3d pdm construction," in 9^{th} *British Machine Vison Conference*, P. Lewis and M. Nixon, eds., vol. 2, pp. 914–923, BMVA Press, (Southampton, UK), Sept. 1998.
43. A. D. Brett and C. J. Taylor, "A framework for automated landmark generation for automated 3D statistical model construction," in 16^{th} *Conference on Information Processing in Medical Imaging*, pp. 376–381, (Visegrád, Hungary), June 1999.
44. G. Szekely, A. Kelemen, C. Brechbuhler, and G. Gerig, "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**, pp. 19–34, 1996.
45. M. Fleute and S. Lavallee, "Building a complete surface model from sparse data using statistical shape models: Application to computer assisted knee surgery," in *MICCAI*, pp. 878–887, 1998.
46. T. F. Cootes, G. J. Edwards, and C. J. Taylor, "Comparing active shape models with active appearance models," in 10^{th} *British Machine Vison Conference*, T. Pridmore and D. Elliman, eds., vol. 1, pp. 173–182, BMVA Press, (Nottingham, UK), Sept. 1999.
47. T. F. Cootes, G. J. Edwards, and C. J. Taylor, "A comparative evaluation of active appearance model algorithms," in 9^{th} *British Machine Vison Conference*, P. Lewis and M. Nixon, eds., vol. 2, pp. 680–689, BMVA Press, (Southampton, UK), Sept. 1998.
48. A. Kelemen, G. Székely, and G. Guido Gerig, "Three-dimensional Model-based Segmentation," Technical Report 178, Image Science Lab, ETH Zürich, 1997.
49. A. Kelemen, G. Szekely, and G. Gerig, "Elastic model-based segmentation of 3D neurological data sets," *IEEE-TMI* **18(10)**, pp. 828–839, 1999.
50. G. Harmarneh, "Deformable spatio-temporal shape modeling," Master's thesis, Department of Signals and Systems, Chalmers University of Technology, Sweden, 1999.