PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

Model-based interpretation of complex and variable images

C. J. Taylor, T. F. Cootes, A. Lanitis, G. Edwards, P. Smyth and A. C. W. Kotcheff

Phil. Trans. R. Soc. Lond. B 1997 **352**, 1267-1274
doi: 10.1098/rstb.1997.0109

Model-based interpretation of complex and variable images

C. J. TAYLOR*, T. F. COOTES, A. LANITIS, G. EDWARDS, P. SMYTH AND A. C. W. KOTCHEFF

Department of Medical Biophysics, Stopford Building, University of Manchester, Oxford Road, Manchester M13 9PT, UK
(ctaylor@man.ac.uk)

SUMMARY

The ultimate goal of machine vision is image understanding—the ability not only to recover image structure but also to know what it represents. By definition, this involves the use of models which describe and label the expected structure of the world. Over the past decade, model-based vision has been applied successfully to images of man-made objects. It has proved much more difficult to develop model-based approaches to the interpretation of images of complex and variable structures such as faces or the internal organs of the human body (as visualized in medical images). In such cases it has been problematic even to recover image structure reliably, without a model to organize the often noisy and incomplete image evidence. The key problem is that of variability. To be useful, a model needs to be specific—that is, to be capable of representing only 'legal' examples of the modelled object(s). It has proved difficult to achieve this whilst allowing for natural variability. Recent developments have overcome this problem; it has been shown that specific patterns of variability in shape and grey-level appearance can be captured by statistical models that can be used directly in image interpretation. The details of the approach are outlined and practical examples from medical image interpretation and face recognition are used to illustrate how previously intractable problems can now be tackled successfully. It is also interesting to ask whether these results provide any possible insights into natural vision; for example, we show that the apparent changes in shape which result from viewing three-dimensional objects from different viewpoints can be modelled quite well in two dimensions; this may lend some support to the 'characteristic views' model of natural vision.

1. INTRODUCTION

The majority of tasks to which machine vision might usefully be applied are 'hard'. The examples we use in this paper are from medical image interpretation and face recognition, though the same considerations apply to many other domains.

The most obvious reason for the degree of difficulty is that most non-trivial applications 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 characterized by the need to deal with complex and variable structures—faces are a good example—and with images that provide noisy and possibly incomplete evidence—medical images are a good example—where 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 pose, 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—for example, faces—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

* Author for correspondence.

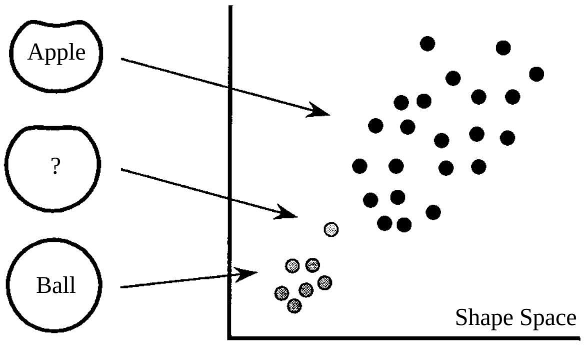

Figure 1. An illustration of why variability must be learnt. The unknown object is more likely to be an apple than a ball, even though it is closer to the mean ball than the mean apple.

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.

Various approaches to modelling variability have been described previously. The most common general approach is to allow a prototype to vary according to a physical model. Kass *et al.* (1987) describe 'snakes' which deform elastically to fit shape contours. Bajcsy & Kovacic (1989) describe a volume model (of the brain) that also deforms elastically to generate new examples. Christensen *et al.* (1995) describe a viscous flow model of deformation which they also apply to the brain. Park *et al.* (1996) and Pentland & Sclaroff (1991) 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 (Scott 1987; Staib & Duncan 1992) and parameterized models, hand-crafted for particular applications (Yuille *et al.* 1992; Lipson *et al.* 1990).

All of these methods can produce models which are general but they do not guarantee specificity. This can only be achieved by learning how the appearance of a class of objects can vary, from a set of examples. Figure 1 provides a simple illustration of why this is the case. Suppose we have two classes of object—apples and balls. Now let us consider a new example (marked with a '?') and ask if it is an apple or a ball. Intuitively we would say that it was an apple. To come to this conclusion we are using learnt knowledge of how the two classes vary. Let us investigate what is happening by working in 'shape space'—a space in which each point represents a particular shape, and metrically similar shapes are represented by nearby points. If we took many examples of our two classes and plotted them in this space we might expect distributions similar to those shown in the figure. The distribution of ball shapes is tighter because the apples are far more variable in shape than the balls. Now consider our new example; it is closer to the average ball than to the average apple—in other words, it could more easily be deformed into a prototype ball than into a prototype apple. It is clear, however, that taking the observed distributions of shapes into account, we should conclude that the object is an example of an apple. In general, we can only acquire such knowledge from a set of examples. This suggests a statistical approach to modelling appearance.

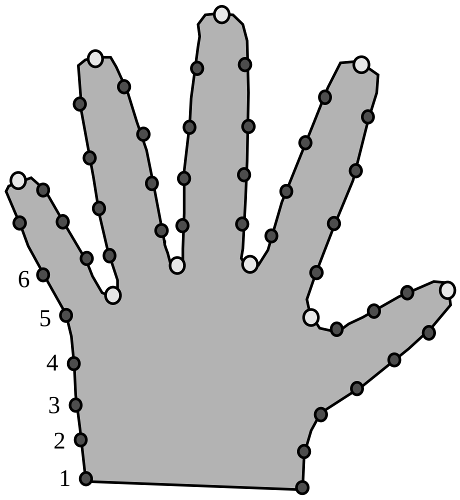

Figure 2. Examples from the training set are represented by points placed on object boundaries at reproducible positions. Here, major landmarks are placed at the tips of the fingers and the cracks between them, secondary landmarks are equally spaced along the boundary.

Prior to our work in this area, Grenander & Miller (1993) and Mardia *et al.* (1991) had described statistical models of shape. These were, however, difficult to use in automated image interpretation. Goodall (1991) and Bookstein (1989) had used statistical techniques for morphometric analysis, but did not address the problem of automated interpretation. Kirby & Sirovich (1990) had described statistical modelling of grey-level appearance (particularly for face images) but did not address shape variability.

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

2. MODELLING SHAPES AND SPATIAL RELATIONSHIPS

Our approach to modelling shapes and spatial relationships has been described previously Cootes *et al.* (1995). The first step is to extract a vector representation of each example shape in a training set. We describe the shapes using a set of landmark points placed at similar positions on each example—this is illustrated in figure 2, using the outline of a hand. In this case, we can achieve a consistent representation

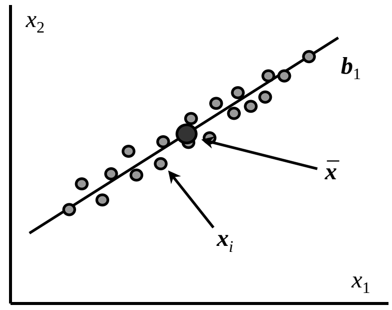

Figure 3. Example shapes plotted in a 'shape space'. Points on the modelled object tend to move in correlated ways, so the 'legal' variation lies in a subspace.

by choosing a set of primary landmark points at the tips of the fingers and the cracks between them, and then adding equally spaced points to represent the rest of the boundary. If there are n points, the result is a vector $\boldsymbol{x}_i = (x_{i1}, y_{i1}, x_{i2}, \ldots x_{in}, y_{in})^{\mathrm{T}}$ containing $2n$ ordinates representing each example i. In order to standardize this representation it is necessary to align the set of examples into a common coordinate frame using, for example, a Procrustes analysis (Goodall 1991).

The next step is to think about the set of training examples in the vector space defined by $\boldsymbol{x}$. In figure 3 we show a two-dimensional space defined by just two of the elements of $\boldsymbol{x}$. Each of the small dots represents an example shape, and the large dot is the mean shape. The important observation is that the values of different components of the vector will tend to be correlated, and that it is these correlations which tell us about the invariant properties of the class of shapes. For example, if we consider x_1 and x_2 as the horizontal positions of two points on the same edge of a finger, they will tend to move together as the hand changes shape. This means that we can define a new coordinate system $\boldsymbol{b}$ in which most of the variation in the training set can be expressed. In general, we find that the subspace in which 'legal' examples of a class of shapes are found has much lower dimensionality than the shape space in which it is embedded.

The coordinate system, $\boldsymbol{b}$, can be found by principal component analysis (PCA) (or nonlinear equivalents; Sozou *et al.* 1995*a*, *b*) and results in a linear model which is able to reconstruct any of the examples in the training set:

$$\boldsymbol{x}_i = \bar{\boldsymbol{x}} + \boldsymbol{P}\boldsymbol{b}_i. \quad (1)$$

A given example $\boldsymbol{x}_i$, can be reconstructed from a weighted sum of the mean shape $\bar{\boldsymbol{x}}$ and a set of linearly independent modes of variation, $\boldsymbol{P}$. These modes are the eigenvectors of the covariance matrix of the set of training shapes; the weight vector $\boldsymbol{b}_i$ is a set of shape parameters which, given the model, provide a unique description of the example shape. This is illustrated in figure 4, which shows the shapes generated by a 'hand' model as elements of $\boldsymbol{b}_i$ are varied. The training set contained hand outlines in various 'open' poses. The important point to note is that, despite the high degree of variability, the model is specific—it only generates

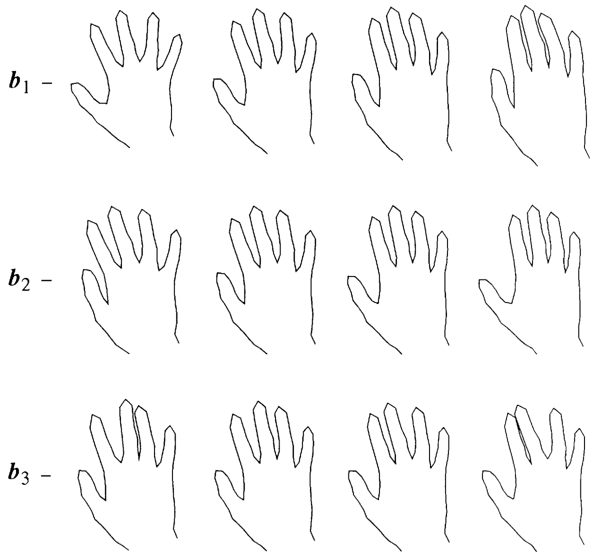

Figure 4. The effect of varying the first three shape parameters of a 'hand' model.

'legal' examples of hand outlines. The shape parameters are kept within limits determined from the training set. Since the eigenvectors which form $\boldsymbol{P}$ are orthogonal, equation (1) can be solved easily for the shape parameters:

$$\boldsymbol{b}_i = \boldsymbol{P}^T(\boldsymbol{x}_i - \bar{\boldsymbol{x}}). \quad (2)$$

Although the illustrative example we have used so far is a simple closed shape, it is important to note that, since the basis for our representation is simply a set of points, complex, multi-part objects can also be modelled, allowing shapes and spatial relationships to be treated in a unified manner. Examples are given later in the paper.

3. MODELLING LOCAL GREY-LEVEL APPEARANCE

In parallel with building a model of shape and spatial relationships, we build a statistical model of the grey-level pattern in the vicinity of each model point. These local grey-level models are typically chosen to represent the grey-level appearance along linear profiles sampled at the model points, perpendicular to the model boundary, and take the form of factor models. These models are important in image search and allow a matching score to be defined between any image patch and the expected grey-level pattern at a given model point. Details have been described previously (Cootes *et al.* 1994).

4. INTERPRETATION: ACTIVE SHAPE MODELS

So far we have seen how we can build statistical models of shapes, spatial relationships and local grey-level appearance. In this section we show how these models can be used in automatic image interpretation. We have investigated various methods (Hill & Taylor

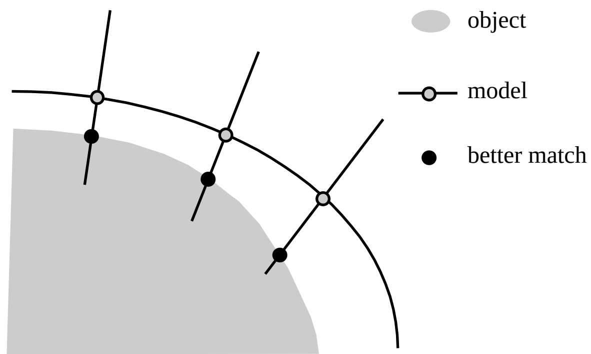

Figure 5. Active shape model (ASM) search. A search is made along the normal at the current position of each model point. The best match to the local grey-level model for the model point is selected.

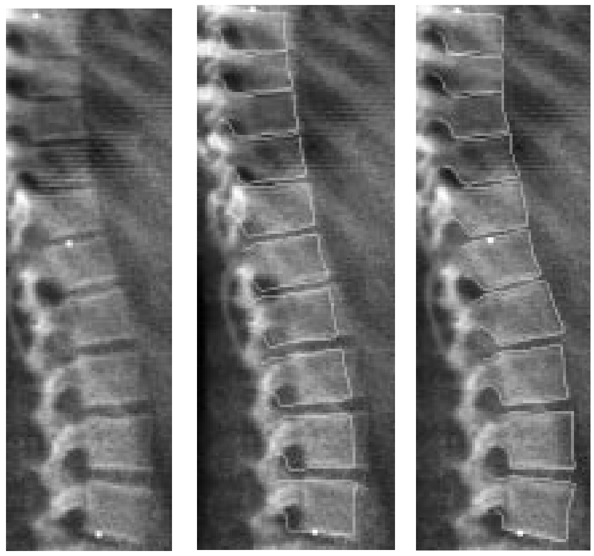

Figure 6. ASM search applied to DEXA images of the spine: (*a*) original image, (*b*) initial model position, and (*c*) solution after convergence of ASM search.

1992), but the 'active shape model' (ASM) approach described here is generally the most successful (Cootes *et al*. 1995). It has much in common with the 'snakes' or 'active contours' of Kass *et al*. (1987), but with the crucial difference that we use the model to apply global constraints to shapes and spatial relationships. The idea is to place an initial model instance into the image and to refine it iteratively. Each model point tries to move towards the appropriate image feature by finding a point close to its current position at which there is a better match to its local grey-level model.

This is illustrated diagrammatically in figure 5, which shows the current model contour and a set of search profiles set up normal to it. At some position along each profile—hopefully at the true boundary contour of the object—we find a better match to the local grey-level model. The key step is that we try to move towards these better matches by updating the parameters of the model—its position, scale, orientation and shape—not the positions of the points directly. This involves two steps: first, we cast the proposed shape into the model frame by finding the translation, orientation and scale which align it as closely as possible to the current model; second, we compute new shape parameters using equation (2), impose limits on their values to ensure a plausible shape, and project back into the image using equation (1). This ensures that our new estimate is always a 'legal' solution, because the model is only capable of generating legal solutions.

In practice, the speed and robustness of ASM search can be improved significantly by using a multi-resolution approach (Cootes *et al*. 1994*a*). During training, a Gaussian pyramid is constructed from each image, and local grey-level models are trained for each level of the pyramid. For new images, ASM search starts at the coarsest level of a similar image pyramid, using the corresponding grey-level model; the search profile length is chosen to allow model points to move some distance to their targets. When convergence is detected at the current scale, the next finer scale is selected, and model refinement continues from the existing solution using a shorter search profile. This is repeated until a solution has been found at the finest scale. Using the multi-scale approach, ASMs typically converge to the correct solution, even given a very poor initialization.

5. PRACTICAL APPLICATIONS OF ACTIVE SHAPE MODELS

In this section we show results from several practical applications of object modelling and ASM search. In each case a quantitative evaluation (beyond the scope of this paper) has shown that accurate and robust interpretation can be achieved. Analysis times are quoted for a SunSparc 20 which operates at 44 mflops. Other examples of practical applications include analysis of: industrial inspection images (Hunter *et al*. 1994), hand gestures (Ahmad *et al*. 1995), echocardiograms (Cootes *et al*. 1995), mammograms (Ellis 1997), surveillance images (Baumberg & Hogg 1994). The method extends to 3D and has been used to interpret 3D MR images of the brain (Hill *et al*. 1993) and knee cartilage (Solloway *et al*. 1996).

(*a*) *DEXA images of the spine*

Dual energy X-ray absorbtiometry (DEXA) images of the spine can be obtained quickly and at low patient dose. An example image is shown in figure 6*a*. There is considerable interest in using these images to monitor the progress of osteoporosis (Steiger *et al*. 1994) by measuring changes in the shapes of vertebrae over time. We have shown that the shapes of the vertebrae can be recovered automatically, using a statistical shape model of the important structures and ASM search, with sufficient accuracy to be of potential clinical value (Smyth *et al*. 1996). Figures 6*b*, *c* show the model at initialization and at convergence respectively. It can be seen that the structures of interest are often poorly defined, but the anatomical knowledge captured in the model of the whole spine prevents spurious solutions. The analysis takes approximately 30 s on a SunSparc 2c.

Figure 7. ASM search applied to radiographs of total hip replacements: (*a*) original image with initial model position superimposed, and (*b*) solution after convergence of ASM search.

(*b*) *Radiographs of total hip replacements*

Standard radiographs are routinely taken pre- and post-operatively and as regular follow-up for patients who receive hip replacements. Studies into the clinical effects of different prosthesis design and surgical technique require detailed measurements of the relative positions of the components of the prosthesis and the remaining bones in each radiograph (Walker *et al.* 1995).

We have shown that the positions of the prosthesis and bones can be recovered automatically using ASM search (Kotcheff *et al.* 1996). Figure 7 shows an example radiograph with the initial and final model positions superimposed. The analysis takes approximately 60 s on a SunSparc 20.

(*c*) *Face images*

The location and recognition of faces in images is of considerable interest for applications such as access control, teleconferencing, human–computer interaction, and surveillance. We can locate faces in images and their individual features accurately and robustly using an ASM. We have used this as the basis for a more sophisticated approach which also includes a global model of grey-level appearance—further details of this aspect are given below. ASM search using a face model is illustrated in figure 8. The analysis takes approximately 10 s on a SunSparc 20.

Figure 8. ASM search applied to face images: (*a*) original image, (*b*) initial model position, and (*c*) solution after convergence of ASM search.

6. MODELLING GLOBAL GREY-LEVEL APPEARANCE: FACES

So far, we have seen how models of shape and spatial relationships can be generated and applied in image interpretation. In this section we show how the idea can be extended to deal with overall grey-level appearance, using the example of face images. We have seen how a face shape model can be fitted to a given face image. Once this has been done we know the positions of a set of landmark points. Using this information we can warp the given face image (Bookstein 1989) to the shape of the mean face to obtain a shape-free image patch describing the grey-level appearance of the face, decoupled from its shape. We can perform PCA on this appearance vector, over the training set, obtaining a shape-free eigenface model of similar form to the shape model. Given a new face image we can now fully describe its appearance as follows. First, we fit the shape model guided by local intensity information. This gives a set of shape parameters $\boldsymbol{b}_{\text{shape}}$. Next, we use this shape information to warp the face image and extract the shape-free patch, which we approximate using the grey-level model to give $\boldsymbol{b}_{\text{grey}}$. Together $\boldsymbol{b}_{\text{shape}}$ and $\boldsymbol{b}_{\text{grey}}$ provide a complete description from which the image can be reconstructed (Lanitis *et al.* 1995). The dimensionality of the model can be further reduced by modelling the combination of $\boldsymbol{b}_{\text{shape}}$ and $\boldsymbol{b}_{\text{grey}}$ in a further PCA model which we have termed a 'combined appearance model' (Edwards *et al.* 1996). Figure 9 shows the effect of varying some of the most important parameters of a combined appearance model. It is interesting to note that, although the model is entirely 2D image-based, it is capable of modelling, convincingly, changes in appearance arising from varying 3D structure and pose.

If we take no precautions, the model we have just described is likely to confound variation due to quite different underlying causes. The appearance will depend on the individual, their direction of gaze, their expression, the lighting and so on. If we wish to interpret the images, it is useful to separate out these different sources of variation. This can be achieved by applying canonical discriminant analysis rather than PCA (McLachlan 1992). The result is a linear model

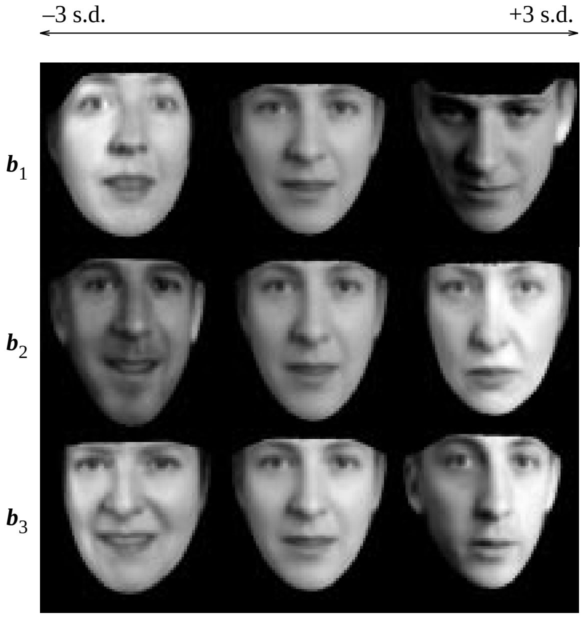

Figure 9. Effect of varying three of the most important parameters of the face combined appearance model.

Figure 10. Effect of isolating identity parameters from others. The identity of a real individual is kept constant whilst parameters associated with other forms of variability are modified.

of the same form as equation (1), but where the appearance space defined by the vector of model parameters, $\boldsymbol{b}$, is partitioned into a set of orthogonal subspaces, each coding for a different source of variability. Figure 10 shows the effect of manipulating the parameters of a model in which variation due to personal appearance has been separated from other sources of variation, using discriminant analysis. This allows us to take a face image from an individual, fit the model and use it to reconstruct the appearance, then vary the expression and pose without changing the identity of the individual.

We have used models of the types shown here successfully to perform a range of face image interpretation tasks including person recognition, expression recognition, gaze determination and so on. The approach is in each case the same—to fit the model and then to perform the recognition task using the model parameters (Lanitis *et al*. 1995).

7. DISCUSSION

We are able to produce convincing 2D models of familiar 3D objects, without recourse to knowledge of 3D structure. This provides some support for the 'characteristic views' model of natural vision (Koendrink & Doorn 1976) by suggesting a mechanism that could support it. For a region of the view-sphere within which there is no change in *image* topology it is possible to model the change in appearance with pose purely in 2D. When a 'view catastrophe' is encountered a new 2D model (of different topology) must be selected. Even if this is not the mechanism used in general, it is conceivable that in might be used for important and familiar objects such as faces.

It is interesting to contrast the effects of model complexity in our framework with the well-established data-driven (Marr-like) approach (Marr 1982). The role of the model in this top-down approach is to help organize the image evidence. It is used to select that evidence which is relevant; the selected evidence is then used to update the model parameters. The key point is that, whereas in the data-driven approach there is an unhelpful combinatorial explosion as the number of primitives in the model increases (Grimson 1990), here the combinatorics are on our side. If we have m model points and if p_d is the probability that a model point thrown at random into the image will find a match to its local grey-level model, then the chance of finding an accidental (and erroneous) match to the full model decreases exponentially with m (because $p_d < 1$). The chance of an accidental match is $k(p_d)^m$. This means that the more complex the model, the less likely we are to find an incorrect solution.

We should revisit two aspects of model building which were mentioned without much comment near the beginning. The first is the question of how we define the outlines of the objects in the training set. At present, all the experiments we have performed are with hand annotated training examples. This is clearly not entirely satisfactory—either practically or intellectually. We have started to look at this problem and have published some very preliminary results (Robinson 1996). Since we plan to use the shape loci in image search, we would like to find a way of automatically selecting points and contours which are distinctive and should thus be easy to find in new images. We have looked at an approach where we measure a multi-scale set of differential invariants of the image intensity function, at each pixel, to provide a signature at each point. We then estimate the probability density function for these signatures using a kernel method (Silverman 1986) and choose the loci of low probability as those which should be used for model building. For face images, this results in selecting points which seem intuitively useful, some of which were already included in the hand-annotated model.

Finally, even when the shape loci have been located, we are left with the problem of where to place the landmark points. As we made clear in § 2, it is important to ensure that these are placed consistently on the shape boundaries in the training set. If not, a non-specific model results. We have tried to tackle this by treating landmark placement as an optimization problem. We define a measure of the quality of model generated and try to optimize this with respect to the positions of the landmarks on the set of training examples—a massive optimization problem. The objective function we have used is one which tries to make the model compact, in the sense that the training set fits into as small a volume of shape space as possible. We have produced very good models completely automatically using Genetic Algorithm search to optimize the objective function (Kotcheff & Taylor 1997). This is again work in progress—at present there are problems scaling this to very complex objects and large training sets.

8. CONCLUSIONS

In summary, we have shown that complex and variable objects can be modelled realistically, and that these models can be used effectively in image interpretation. From an engineering point of view, this approach has opened the way to developing practical systems to solve difficult image interpretation problems—particularly in medical image analysis and face recognition, but also in other application domains. Although we were not motivated by a desire to gain insight into natural vision, it is tempting to ask whether there are lessons to be learnt. Although the approach has proved successful there are still a number of significant problems to solve if it is to become completely automatic—at present it is automatic at image-interpretation time, but involves manual intervention at training time. Our viewpoint is that vision is essentially a statistical problem, and there are far more sophisticated methods that can be brought to bear than those which have been deployed to date.

REFERENCES

Ahmad, T., Taylor, C. J., Lanitis, A. & Cootes, T. F. 1995 Tracking and recognizing hand gestures using statistical shape models. In *6th British Machine Vision Conf., September*, (ed. D. Pycock), pp. 403–412. Birmingham, England: BMVA Press.

Bajcsy, R. & Kovacic, A. 1989 Multiresolution elastic matching. *Comput. Graphics Image Process.* **46**, 1–21.

Baumberg, A. M. & Hogg, D. C. 1994 An efficient method for contour tracking using active shape models. Research report series, Division of Artificial Intelligence, School of Computer Studies, University of Leeds.

Bookstein, F. L. 1989 Principal warps: thin-plate splines and the decomposition of deformations. *IEEE Trans. Pattern Analysis Mach. Intell.* **11** (6), 567–585.

Christensen, G. E., Rabbitt, R. D., Miller, M. I., Joshi, S. C., Grenander, U., Coogan, T. A. & Van Essen, D. C. 1995 Topological properties of smooth anatomic maps. In *14th Conf. on Inform. Process. in Med. Imaging*, pp. 101–112. Ile de Berder, France: Kluwer.

Cootes, T. F., Taylor, C. J. & Lanitis, A. 1994 Active shape models: evaluation of a multi-resolution method for improving image search. In *5th British Machine Vision Conf.* (ed. E. Hancock), pp. 327–336. York, England: BMVA Press.

Cootes, T. F., Hill, A., Taylor, C. J. & Haslam, J. 1994 The use of active shape models for locating structures in medical images. *Image Vision Comput.* **12** (6), 276–285.

Cootes, T. F., Taylor, C. J., Cooper, D. H. & Graham, J. 1995 Active shape models—their training and application. *Comput. Vision Image Understanding* **61** (1), 38–59.

Edwards, G. J., Lanitis, A., Taylor, C. J. & Cootes, T. F. 1996 Modelling the variability in face images. In *2nd Int. Conf. on Automatic Face and Gesture Recognition, October*, pp. 328–333. Los Alamitos, California: IEEE Computer Society Press.

Ellis, D. 1997 Trainable methods of mammographic screening. M.Sc. thesis, University of Manchester.

Goodall, C. 1991 Procrustes methods in the statistical analysis of shape. *J. R. Statist. Soc.* B **53** (2), 285–339.

Grenander, U. & Miller, M. I. 1993 Representations of knowledge in complex systems. *J. R. Statist. Soc.* B **56**, 249–603.

Grimson, W. E. L. 1990 *Object recognition by computer*. MIT Press.

Hill, A. & Taylor, C. J. 1992 Model-based image interpretation using genetic algorithms. *Image Vision Comput.* **10** (5), 295–300.

Hill, A., Thornham, A. & Taylor, C. J. 1993 Model-based interpretation of 3D medical images. In *4th British Machine Vision Conf., September* (ed. J. Illingworth), pp. 339–348. York, England: BMVA Press.

Hill, A., Cootes, T. F., Taylor, C. J. & Lindley, K. 1994 Medical image interpretation: a generic approach using deformable templates. *J. Med. Informatics* **19** (1), 47–59.

Hunter, J. J., Graham, J., Cootes, T. F. & Taylor, C. J. 1994 User programmable visual inspection. In *5th British Machine Vision Conf., September* (ed. E. Hancock), pp. 661–670. York, England: BMVA Press.

Kass, M., Witkin, A. & Terzopoulos, D. 1987 Active contour models. *Int. J. Comput. Vision* **1** (4), 321–331.

Kirby, M. & Sirovich, L. 1990 Application of the Karhumen–Loeve procedure for the characterization of human faces. *IEEE Trans. Pattern Analysis Mach. Intell.* **12** (1), 103–108.

Koendrink, J. J. & Van Doorn, A. J. 1979 Internal representation of solid shape with respect to vision. *Biological Cybernetics* **32** (4), 211–216.

Kotcheff, A. C. W. & Taylor, C. J. 1997 Automatic construction of eigen-shape models by genetic algorithm. In *15th Conf. on Inform. Process. in Med. Imaging* (ed. J. Duncan & G. Grindi), pp. 1–14. Poultney, VA: Springer.

Kotcheff, A. C. W., Redhead, A., Taylor, C. J. & Hukins, D. 1996 Shape model analysis of THR radiographs. In *13th Int. Conf. on Pattern Recognition*, vol. 4, pp. 391–395. IEEE Computer Society Press.

Lanitis, A., Cootes, A. F. & Taylor, C. J. 1996 A unified approach to coding and interpreting face images. In *5th Int. Conf. on Computer Vision*, pp. 368–373. Cambridge, MA: IEEE Computer Society Press.

Lipson, P., Yuille, A. L., O'Keeffe, D., Cavanaugh, J., Taaffe, J. & Rosenthal, D. 1990 Deformable templates for feature extraction from medical images. In *2nd European Conf. on Computer Vision* (ed. O. Faugeras), pp. 413–417. Berlin/New York: Springer.

Mardia, K. V., Kent, J. T. & Walder, A. N. 1991 Statistical shape models in image analysis. In *23rd Symp. on the Interface, Seattle*, pp. 550–557.

Marr, D. 1982 *Vision*. San Fransisco: W.H. Freeman & Co.

McLachlan, G. J. 1992 *Discriminant analysis and statistical pattern recognition*. Wiley.

Park, J., Mataxas, D., Young, A. A. & Axel, L. 1996 Deformable models with parameter functions for cardiac motion analysis from tagged MRI data. *IEEE Trans. on Medical Imaging* **15**, 278–289.

Pentland, A. P. & Sclaroff, S. 1991 Closed-form solutions for physically based modelling and recognition. *IEEE Trans. Pattern Analysis Mach. Intell.* **13** (7), 715–729.

Robinson, D. 1996 Differential geometry as an approach to automatic landmark point generation. M.Sc. thesis, University of Manchester.

Scott, G. L. 1987 The alternative snake—and other animals. In *3rd Alvey Vision Conf., Cambridge, England*, pp. 341–347.

Silverman, B. W. 1986 *Density estimation for statistics and data analysis*. London: Chapman & Hall.

Smyth, P. P., Taylor, C. J. & Adams, J. E. 1996 Automatic measurement of vertebral shape using active shape models. In *7th British Machine Vision Conf., September*, pp. 705–714. Edinburgh, Scotland: BMVA Press.

Sozou, P. D., Cootes, T. F., Taylor, C. J. & Di Mauro, E. C. 1995*a* A non-linear generalisation of point distribution models using polynomial regression. *Image Vision Comput.* **13**(5), 451–457.

Sozou, P. D., Cootes, T. F., Taylor, C. J. & DiMauro, E. C. 1995*b* Non-linear point distribution modelling using a multi-layer perceptron. In *6th British Machine Vision Conf., September* (ed. D. Pycock), pp. 107–116. Birmingham, England: BMVA Press.

Solloway, S., Hutchinson, C. E., Waterton, J. C. & Taylor, C. J. 1996 Quantification of articular cartilage from MR images using active shape models. In *4th European Conf. on Computer Vision, April* (ed. B. Buxton & R. Cipolla), vol. 2, pp. 400–411. Cambridge, England: Springer.

Staib, L. H. & Duncan, J. S. 1992 Boundary finding with parametrically deformable models. *IEEE Trans. Pattern Analysis Mach. Intell.* **14**(11), 1061–1075.

Steiger, P., Cummings, S. R., Genant, H. K., Weiss, H. & the Study of Osteoporotic Fractures Group 1994 Morphometric X-ray absorptiometry of the spine: correlation *in vivo* with morphometric radiography. *Osteoporosis Int.* **4**, 238–244.

Walker, P. S., Mai, S. F., Cobb, A. G., Bentley, G. & Hua, J. 1995 Prediction of clinical outcome of (THR) from migration measurements on standard radiographs—a study of cemented Charnley and Stanmore femoral stems. *J. Bone Jt Surg. (Br.)* **77**, 705–714.

Yuille, A. L., Cohen, D. S. & Hallinan, P. 1992 Feature extraction from faces using deformable templates. *Int. J. Comput. Vision* **8**(2), 99–112.

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES