Active Shape Models : Evaluation of a Multi-Resolution Method for Improving Image Search

T.F.Cootes, C.J.Taylor, A.Lanitis

Department of Medical Biophysics
University of Manchester
email: bim@wiau.mb.man.ac.uk

Abstract

We describe a multi-resolution technique for locating for variable structures in images. This is an extension of work on Active Shape Models (ASMs) – statistical models which iteratively deform to match image data. An ASM consists of a shape model controlling a set of landmark points, together with a statistical model of the grey–levels expected around each landmark. Both the shape model and the grey–level models are trained on sets of labelled example images. In order to apply a coarse–to–fine search strategy it is necessary to train a *set* of grey–level models for each landmark, one for every level of a multi-resolution image pyramid. During image search the model is started on the coarsest resolution image. As the search progresses it moves to finer and finer resolutions until no further improvement can be made. We describe an automatic technique for deciding when to change resolution and for detecting when the process has converged. We demonstrate the approach on two examples, and give results of quantitative experiments which show a significant increase in both speed and quality of fit compared to previous methods.

1 Introduction

Many applications of computer vision involve locating examples of known objects or structures in new images. Often the structures to be located can vary in shape, either because they are flexible or articulated, or because natural variation is present. Flexible models have proved extremely useful in tackling such problems. Various approaches to modelling shape and shape variation have been described by Yuille *et al.* [1], Kass *et al.* [2] Hinton, Williams and Revow [3] , Staib and Duncan [4], Pentland and Sclaroff [5], Karaolani *et al.* [6], Nastar and Ayache [7], Grenander *et al.* [8] and Mardia *et al.* [9]. Grenander and Miller [10] describe models which represent both the shape and intensity information.

In earlier work (Cootes *et al.* [11]) we described 'Active Shape Models' (ASMs), statistically based flexible models which iteratively move toward structures in images similar to those on which they were trained. An ASM consists of a set of *landmark* points, each representing the position of a particular part of the structure to be located. The model is trained by marking landmark points on each of a set of training images. Statistical analysis of the relative positions of the landmarks in different examples allows both the average shape and shape variation to be modelled. The grey–level landscape in the vicinity of each landmark is also modelled statistically. For image interpretation an example of the model is placed in a previously unseen image. A local search is performed around the current position of each landmark to locate a new position at which there is a better match between the grey–level model for that point and the image. The shape model parameters are iteratively updated to move the landmarks toward these better matched points, with the constraint that the overall shape cannot deform more than the examples seen in the training set. ASMs have been used successfully in a wide range of applications, including lo-

cating organs in medical images [11,12], face recognition and hand–written character recognition [13].

Our recent work has aimed at improving both the speed and the accuracy of the ASM method. In this paper we describe a multi–resolution approach to modelling the grey–levels around each landmark, and a coarse-to-fine strategy for image search. We show that this approach leads to faster and more accurate image interpretation. We describe a method for detecting when the ASM has converged, and show how this can be used to decide when to move to a finer resolution during multi–resolution image search. We give examples of models locating image structures, and results of quantitative experiments showing the improvements gained by using the multi–resolution technique.

2 Active Shape Models

2.1 Shape Models and Grey–Level Models

Cootes *et al.* [11] describe how to build flexible shape models called Point Distribution Models. These are generated from examples of shapes, where each shape is represented by a set of labelled landmark points. The models can be used to generate new shapes using the equation

$$\mathbf{x} = \bar{\mathbf{x}} + \mathbf{Pb} \qquad (1)$$

where $\quad \mathbf{x} = (x_0, \ y_0, \ \dots \ x_{n-1}, \ y_{n-1})^T$

$(x_k, y_k) \quad$ *is the kth model point*
$\bar{\mathbf{x}} \qquad$ *represents the mean shape*
$\mathbf{P} \qquad$ *is a 2n x t matrix of t unit column vectors*
$\mathbf{b} = (b_1...b_t)^T \quad$ *is a set of shape parameters*

The columns of $\mathbf{P}$ are orthogonal and span the space of shape variations observed in the training set. If the shape parameters $\mathbf{b}$ are chosen inside suitable limits (derived from the training set) then the shapes generated by (1) will be similar to those in the training set. Examples of such models are given in [11 – 12].

In addition the local grey–level environment about each landmark point can be modelled. Statistical information is gathered about the mean and covariances of the values of the pixels in the vicinity of each landmark, typically on profiles normal to the object boundary at that point. This data can be used to assess how well the grey–levels in a particular area of an image match those expected around a given model landmark point.

2.2 Image Search Using An Active Shape Model

Given a rough starting approximation an Active Shape Model can be iteratively fitted to an image [11]. By choosing a set of shape parameters $\mathbf{b}$ for a Point Distribution Model, we define the shape of a model 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:

$$\mathbf{X} = M(s, \theta)[\mathbf{x}] + \mathbf{X}_c \qquad (2)$$

where $\mathbf{X}_c = (X_c, Y_c, \ldots, X_c, Y_c)^T$

$M(s, \theta)[.]$ performs a rotation by θ and a scaling by s.

(X_c, Y_c) is the position of the centre of the model in the image frame.

An iterative approach to improving the fit of the instance, **X**, to an image proceeds as follows :

i) Examine a region of the image around each point to calculate the displacement of the point required to move it to a better location.

ii) From these displacements calculate adjustments to the pose and the shape parameters.

iii) Update the model parameters; by enforcing limits on the shape para- meters, global shape constraints can be applied ensuring the shape of the model instance remains similar to those of the training set.

The procedure is repeated until no significant changes result. Because the models deform to better fit the data, but only in ways which are consistent with the shapes found in the training set they are called 'Active Shape Models' (ASMs).

To find a better location for each model point (step (i) above) we sample a profile perpendicular to the boundary at the point, and run the grey–level model along it to find the best match (Figures 1, 2). The suggested movement is then toward the point on the sampled profile which gave the best match to the model. We denote the set of such adjustments as a vector

$$d\mathbf{X} = (dX_0, \ dY_0, \ \ldots, \ dX_{n-1}, \ dY_{n-1})^T.$$
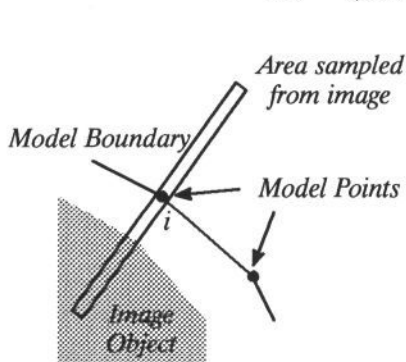

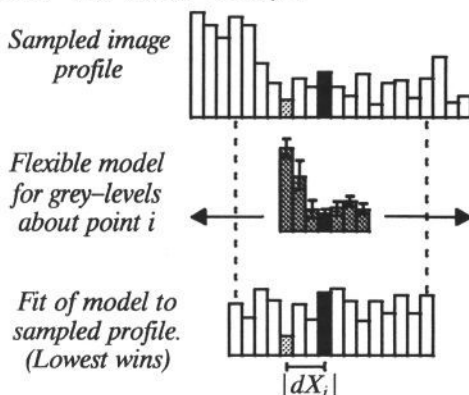*Figure 1. Part of a model bound- ary approximating to the edge of an image object.*

Figure 2. Suggested movement of point dX_i is along normal to boundary, in a direction to- wards the point at which the profile model best fits the profile sampled from the image.

Given such a set of suggested changes we can use a least squares approach to find the best change in pose (dX_c, dY_c, ds, $d\theta$) and changes to the shape parameters ($d\mathbf{b}$) to move the points from their current locations in the image frame, **X**, to be as close as possible to the suggested new locations ($\mathbf{X} + d\mathbf{X}$). We can then update the pose and shape parameters, apply limits to the shape parameters to ensure only 'legal' shapes are generated and repeat. (See [11] for details).

3 Extension to Multi–Resolution Images

The above formulation works well for many different examples. However, there is a problem with choosing the length of profile along which to search with the grey–level model. If the search profile is too short, the model landmark points must be close to their targets in the image before they can 'latch on' and pull the shape model into place. If they are too long the search becomes computationally expensive and the grey–level models are more likely to latch on to distracting structures in the image away from the target object, preventing the ASM from converging to the correct shape.

Ideally the search should first look far from its current location and make large jumps, then as it homes in on a target structure, it should restrict its examination of the image to the immediate locality. This suggests a multi–resolution approach, with the models applied first to a coarse, low resolution version of an image, then refined on higher resolution versions. We can generate such images from an original using gaussian smoothing and sub–sampling to produce a multi–resolution pyramid [14]. Level 0 in such a pyramid is the original image. Level 1 is an image with half the number of pixels along each axis. In our implementation each pixel in a given level is generated by smoothing the image at the level below with a 5 x 5 gaussian mask (which is linearly decomposed into two 1–5–8–5–1 convolutions) and then sub–sampling every other pixel.

In order to run an ASM at a particular level in the pyramid grey–level models trained on the data at that level are required because gaussian smoothing and sub–sampling significantly modify image structure. Instead of a single grey–level model for each landmark point, we thus generate a set of models, one for each level of the pyramid we wish to use. We let the models use the same number of pixels at each level, so those at level $l + 1$ cover twice as much the image area as those at level l (Figure 3). Thus each grey–level model at the coarser levels may cover large parts of the structure of interest, while at fine levels each model is much more localised. On a given layer, we need only search along a short profile sampled from the image. At the coarse resolutions this will allow large movements, at the fine resolutions only small movements are allowed, and the grey–level models are less likely to be distracted by image features away from the current target structure.

In order to perform local search, we start at the highest level of the pyramid and run a number of iterations of the ASM using the models trained at that level. During each iteration we look in the region around each point for a better position. We need only examine the image data at a small number of nearby locations, for instance at the current point and n_l pixels either side (Figure 4). Having run a number of iterations at one resolution, we move to the next level down in the pyramid, and search at a finer resolution. Again we only need look n_l pixels either side of each current point. The process is repeated until a suitable number of iterations are run on the original image at level 0 in the pyramid.

3.1 When to Change Levels

For maximum efficiency it is necessary to devise a method of deciding automatically when to change to a finer resolution. The simplest approach is to run a fixed number of iterations at each level. However, using this approach the model may not have moved sufficiently close to the target image structure after the chosen number of
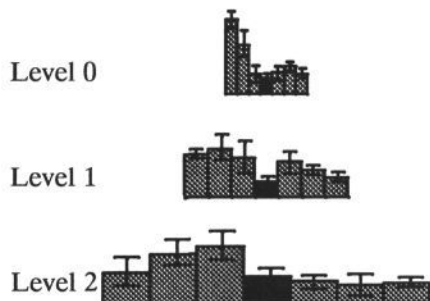
Level 0

Level 1

Level 2

Figure 3 : Grey–level models for a land-mark point at different levels of the gaussian pyramid.

Profile sampled normal to model boundary in image at given level.

Search with grey–level model

Locate best fit (lowest value) of model to sampled profile.
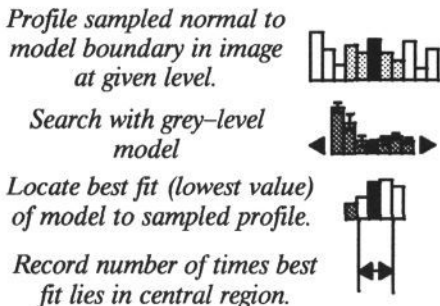
Record number of times best fit lies in central region.

Figure 4 : At each level and each land-mark the grey–level model fit at a small number of positions about the current point is calculated, and number of best fit's lying in a central region is recorded.

iterations or it may converge at a particular level after only a small number of iterations, so any further processing at that level is wasted. An alternative approach is therefore to assess whether the model has converged at a given level, and if so, move to the next level.

A simple method of testing for convergence is as follows. At every point we assess the fit of its grey–level model at n_l locations either side of the point, along a profile perpendicular to the model boundary (Figure 4). We then calculate the proportion of the points for which the best fit lies within the central 50% of the locations. When this proportion rises above a limit, say 95%, we declare the model to have converged, and move to the next level in the pyramid. The most appropriate choice for the proportion limit should be ascertained by experiment. If it is too low, the model may change resolution too soon, before it has settled at the current level. The model may then be too far away from the target to converge correctly at the next level. If the limit is too high, it may never be reached. To prevent getting stuck an upper limit is applied to the number of iterations, after which the search is forced to the next level. It is assumed that it has got as good as it is going to get at the current level. In our experiments $n_l = 2$, we assess the fit at 5 locations, and we count the 'hits' within the central 3 pixels.

4 Examples of Multi–Resolution Search

4.1 Face Model Example

We trained a face model using 169 landmark points planted by hand on 11 images of a single person's face. We trained sets of grey–level models for each landmark, one at each pyramid level. The grey–level models were 7 pixels long. Figure 5 shows an example of one of the training images, and the mean model shape. Figure 6 shows the face model iterating to fit to a new image (taken with different lighting conditions and background). The search runs at most 10 iterations on each level of a gaussian pyramid generated from the original image, starting with a level 5 image.

During the first iterations the ASM works on a coarse image (1/32 the size of the original in each dimension). It thus makes large changes, when viewed at the original scale. As the search progresses the ASM works on increasingly fine resolution images, so makes only small adjustments. The search is stopped when it has run at
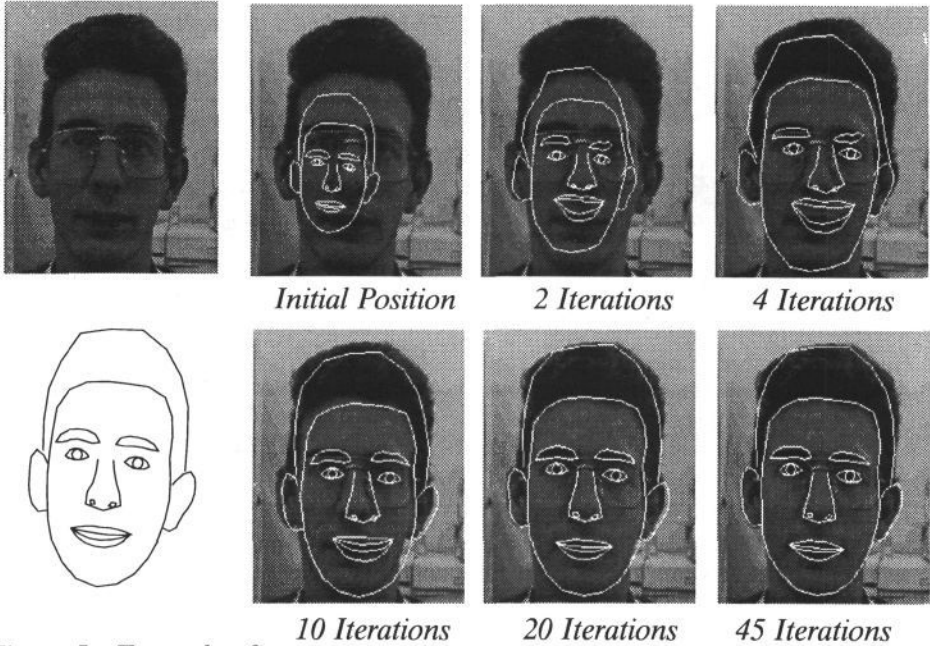
Initial Position   *2 Iterations*   *4 Iterations*

10 Iterations   *20 Iterations*   *45 Iterations*

Figure 5 : Example of face training image and mean model.

Figure 6 : Example of a face ASM iterating to fit a new image.

most 10 iterations on the level 0 image. Each iteration (at any level) takes less than 150ms on a Sun Sparc10 Workstation – the search is completed in under 7 seconds.

4.2 Vertebra Model Example

A flexible model of a single lumbar vertebra was trained using examples taken from 30 lateral radiographs of the spine from different patients. An example is shown in Figure 8. The model consisted of 204 landmark points, and a set of 7 pixel grey–level models trained for each point. 10 modes of shape variation were sufficient to explain 95% of the variation in the training set. Figure 7 shows the first two modes of the model. Figure 8 shows the ASM iterating to fit to a new radiograph. Again, the search started on level 5 of the image pyramid, and no more than 10 iterations were applied at each level. Each iteration takes less than 200ms, convergence is reached within 8 seconds.

5 Quantitative Experiments

We wished to compare the performance of the new multi–resolution technique with earlier methods, and determine the effects of varying various parameters used in the search. We did this by running the ASM on a set of images, and comparing the ASM landmark points with sets of points labelled manually on each image.

5.1 Method

The procedure for testing was as follows. Given an image containing an example of the modelled object and the positions of the model points, X_{known}, annotated manually :

- Calculate the pose which maps the mean shape model points onto the known image points – the base pose.

$b_1 = -3\sigma_1$ $b_1 = -1.5\sigma_1$ *Mean Shape* $b_1 = 1.5\sigma_1$ $b_1 = 3\sigma_1$

$b_2 = -3\sigma_2$ $b_2 = -1.5\sigma_2$ *Mean Shape* $b_2 = 1.5\sigma_2$ $b_2 = 3\sigma_2$
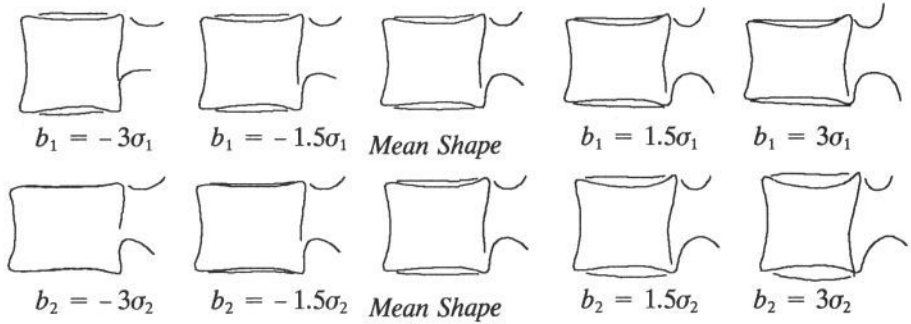
Figure 7 : The effects of varying the first two shape model parameters on the model of a vertebra (setting the others to zero). σ_i is the standard deviation of b_i over the training set.
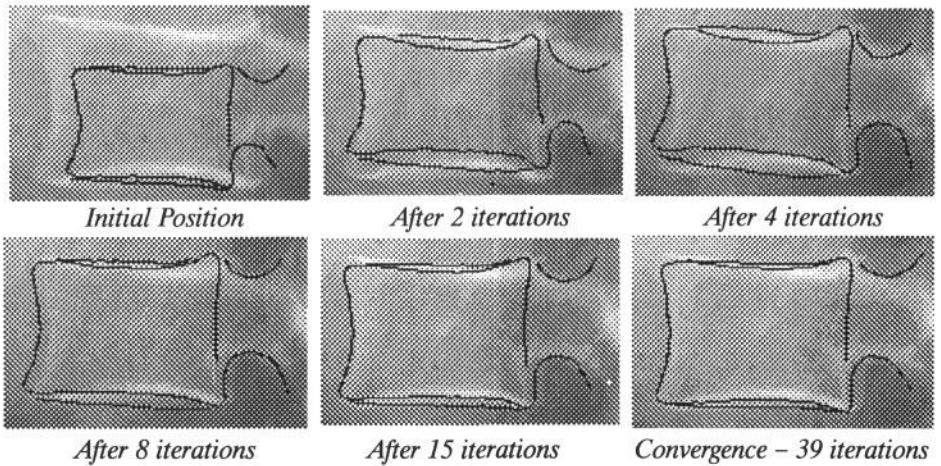

Initial Position   *After 2 iterations*   *After 4 iterations*

After 8 iterations   *After 15 iterations*   *Convergence – 39 iterations*

Figure 8 : An ASM for a vertebra iterating to fit to a radiograph.

- Run the Active Shape Model search on the image, starting it with the pose parameters *perturbed from the base pose values* and the shape parameters set to zero (ie the mean model shape).

- After every iteration calculate the mean distance of the model points from the known image points.

- For each image repeat with a variety of different perturbations of the initial pose parameters and consolidate the results.

- Repeat for a number of images.

The resulting graphs of mean distance against iteration number allow comparison between different techniques and parameter settings.

The experiments were conducted using the flexible model of a face described earlier. The model was trained on image pyramids with five levels derived from 15 512^2 images of one person (Figure 5), with 169 points being used to represent the shape. Normalised, derivative grey–level profile models of length 7 pixels were used for both the multi–resolution and single resolution experiments.

5.2 Results

Results were obtained by attempting to locate the labelled points in the original training images using an ASM, starting with a pose systematically displaced from the correct pose. Miss–one–out experiments were performed, training a model on 14 images and testing it on the 15th. 8 runs were performed for each image, displacing the centre by $(\pm 30, \pm 30)$ pixels, the orientation by $\pm 6°$ and the scale by -30%. Each face was about 200 pixels across in the original image.

Figure 9 shows the effect of using different numbers of iterations at each level. The more iterations used the better the final fit, but the longer it will take to achieve. The graph demonstrates that good results can be achieved most quickly with 5 or 7 iterations. For the experiments in which a fixed number of iterations were run at each level, the points at which the search changes level show clearly as steps in the curves. These steps suggest that a more rapid improvement in fit would occur if the 'bottoming out' of the curves is used as a trigger for a move to a finer level. The automatic approach gives a smoother, more rapid improvement in quality of fit, avoiding these 'steps'. Figure 10 shows examples of the automatic approach and the effect of varying the threshold for the proportion of 'hits' required for a change to a finer resolution during the search. In this case a maximum of 7 iterations were run at each iteration, with fewer running if the threshold was reached. The results suggest that for this case a threshold of 90–95% gives the best convergence rate combined with a good final fit.
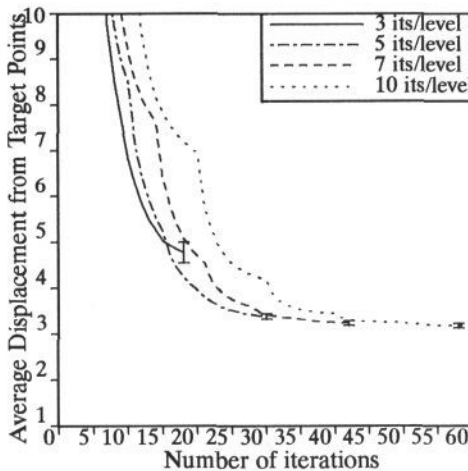

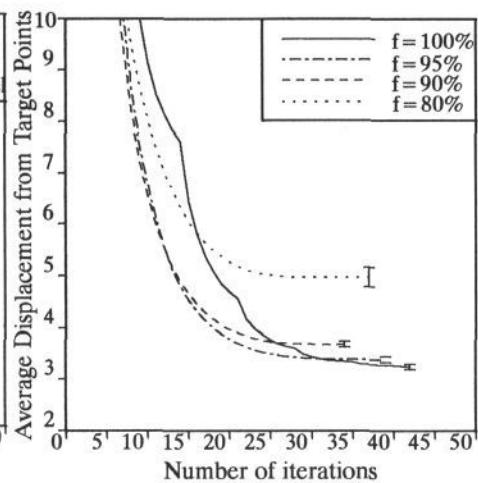Figure 9: Effects of varying number of iterations run at each level.

Figure 10: Effects of varying the threshold (f) for changing to a finer resolution. Max. iterations per level = 7

Figure 11 compares the ASM run only on the original image resolution with results from the new multi–resolution approach. In practice the results for the single resolution method were obtained by only displacing the pose by less than the amount used for the multi–resolution experiments (ie centre $(\pm 10, \pm 10)$ pixels, $\pm 3°$ and using the correct scale). This was because the single resolution experiment takes considerably longer and becomes too prone to getting stuck in local minima when started with large displacements. The results show that despite starting considerably further

away, the multi–resolution approach converges on average more rapidly, and to a better solution.

Figure 12 shows the effect of using different grey–model lengths. In each case the automatic search process was used, running a maximum of 7 iterations at each level and changing levels using a threshold of 95% of central hits. Longer grey–models give improved convergence and final results. Using models longer than 7 pixels gave no further gains.
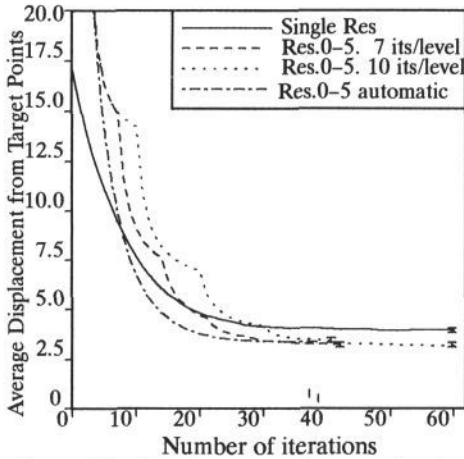

Figure 11 : Results comparing running the ASM at a single resolution with the Multi– Res. approach. Error bars are 1.0 standard error
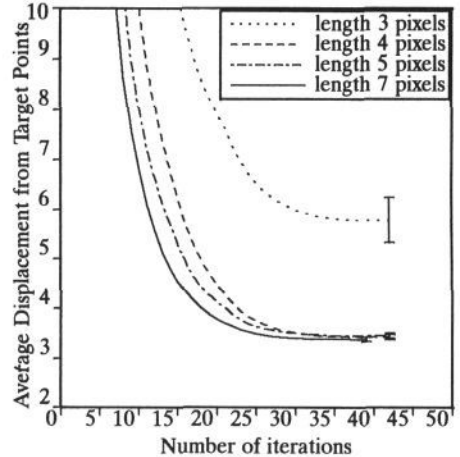
Figure 12 : Experiment comparing different grey–model lengths. Error bars are 1.0 standard error

6 Discussion and Conclusions

We have described a simple method of extending Active Shape Model search to deal with multi–resolution image pyramids. The new method can successfully locate structures from further away, giving better results in fewer iterations than an ASM run at a single resolution. In addition we have developed an automatic method of detecting convergence, giving a way of terminating the iterative search.

The multi–resolution methods are less likely to be caught in nearby local minima. This is partly because the search at coarser resolution avoids the model getting caught on fine–grain clutter. In addition the use of shorter search profiles means that the grey–level models at each point are less likely to be distracted from their target structure by nearby image features. Each iteration can be considerably quicker, since during search only a small number of neighbouring points about each model point are considered. When run on a single resolution image it was necessary to search long profiles with the grey–level models, if the model instance was thought to be a long way from the image structure. Because the new methods also converge after fewer iterations, this leads to multi–resolution search taking typically 10–20% of the CPU time required by the original method.

The multi–resolution approach appears to be able to cope with occlusion as well as the single scale ASM – the constraints of the shape model ensure that points in occluded areas are estimated in a 'sensible' manner.

In the current implementation the same model points are used at each level. At the coarse resolutions this is inefficient, since the areas of the image covered by the grey-level models about each landmark point will overlap. We intend examining ways in which the number of model points can be systematically reduced at the coarse resolutions to further increase the speed of the search.

Acknowledgements

References

[1] A.L. Yuille, P. Hallinan and D.S. Cohen Feature extraction from faces using deformable templates, IJCV, **8**, August 1992, pp. 99–112.

[2] M. Kass, A. Witkin and D. Terzopoulos , Snakes: Active Contour Models, in *Proc. First International Conference on Computer Vision*, pp 259–268 IEEE Computer Society Press, 1987.

[3] G.E.Hinton, C.K.I. Williams and M.D. Revow, Adaptive Elastic Models for Hand–Printed Character Recognition. in *Advances in Neural Information Processing Systems 4*, (J.E.Moody, S.J.Hanson, R.P.Lippmann. Ed.s) Morgan Kauffmann, San Mateo, CA, 1992.

[4] L.H. Staib and J.S. Duncan, Parametrically Deformable Contour Models, *IEEE Computer Society Conference on Computer Vision and Pattern Recognition. San Diego, 1989*, pp. 427–430.

[5] A. Pentland and S. Sclaroff, Closed–Form Solutions for Physically Based Modelling and Recognition, *IEEE Trans. on Pattern Analysis and Machine Intelligence.* **13**, 1991, 715–729.

[6] P. Karaolani, G.D. Sullivan, K.D. Baker and M.J. Baines, A Finite Element Method for Deformable Models. *Proceedings of the Fifth Alvey Vision Conference, Reading, 1989*, pp. 73–78.

[7] C. Nastar and N. Ayache, Non–Rigid Motion Analysis in Medical Images : a Physically Based Approach. in *Proceedings of IPMI '93*, (H.H.Barrett, A.F.Gmitro. Ed.s) Springer–Verlag, Berlin 1993, pp.17–32.

[8] U. Grenander, Y. Chow and D.M. Keenan, *Hands. A Pattern Theoretic Study of Biological Shapes.* Springer–Verlag, New York, 1991.

[9] K.V. Mardia, J.T. Kent and A.N. Walder, Statistical Shape Models in Image Analysis, *Proceedings of the 23rd Symposium on the Interface, Seattle 1991*, pp 550–557.

[10] U. Grenander, M.I.Miller, Representations of Knowledge in Complex Systems. *J.R.Statistical Soc.:B* (1994) **56**, No.3.

[11] T.F.Cootes, A.Hill, C.J.Taylor, J.Haslam, The Use of Active Shape Models for Locating Structures in Medical Images. *Image and Vision Computing Vol.12, No.6 July 1994*, 355–366.

[12] A. Hill, A. Thornham and C.J.Taylor, Model–Based Interpretation of 3D Medical Images. in *Proc. British Machine Vision Conference.* Ed. J. Illingworth, BMVA Press, 1993, pp. 339–348.

[13] A. Lanitis, C.J.Taylor, T.F.Cootes, A Generic System for Classifying Variable Objects using Flexible Template Matching. in *Proc. British Machine Vision Conference.* Ed. J. Illingworth, BMVA Press, 1993, pp. 329–338.

[14] P.J.Burt, The Pyramid as a Structure for Efficient Computation. in *Multi–Resolution Image Processing and Analysis.* Ed. Rosenfield, pub. Springer–Verlag. 1984. pp. 6 – 37.