# Active Shape Model Search using Local Grey-Level Models : A Quantitative Evaluation

T.F.Cootes, C.J.Taylor

Department of Medical Biophysics
University of Manchester
Oxford Road
Manchester M13 9PT
email: bim@wiau.mb.man.ac.uk

## Abstract

We describe methods for locating known structures in images. We have previously described statistical models of shape and shape variability which can be used for this purpose (Active Shape Models). In this paper we show how statistical models of grey-level appearance can be incorporated, leading to improved reliability and accuracy. We describe experiments designed to (i) test how well an ASM can locate an object in a new image, (ii) to assess the effects on performance of varying the model parameters, and (iii) to compare the results using grey-level models with those using a search for strongest edges. The results demonstrate that the addition of grey-level models leads to considerable improvement over earlier schemes.

## 1 Introduction

Recently there has been increasing interest in image interpretation using flexible models or deformable templates. Different approaches have been described by Yuille *et al* [2], Kass *et al* [1] 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].

In previous papers [10,11] we have shown how one can build models of the shape of deformable objects; we have also described a local search technique which allows initial estimates of the pose and shape parameters to be iteratively refined. The shape models rely on representing objects by sets of labelled points; each point is placed on a particular part of the object. By examining the statistics of the positions of the labelled points a 'Point Distribution Model' is derived. The model gives the average positions of the points, and has a number of parameters which control the main modes of variation found in the training set. Such models give compact and specific descriptions of both object shape and the spatial relationships between objects.

Given a shape model and an image containing an example of the object modelled, interpretation involves choosing values for each of the model parameters so as to best fit the model to the image. We previously described a technique which allows an initial guess for the best shape, orientation, scale and position to be refined by comparing the hypothesised model example with image data and using differences between model and image to deform the shape (Active Shape Models [11]).

This paper describes how it is possible to model the grey levels expected at each point of the shape model, and how such grey-level models can be used in Active Shape Model search [12,13]. We also present the results of systematic experiments to assess how well an ASM can locate the model points on objects in unseen images.

BMVC 1993 doi:10.5244/C.7.64

Experiments have been performed using different types of grey-level model and using different weighting schemes in the search algorithms. We compare the performances of the ASMs under these different regimes; we also show that the results are significantly better than those obtained using the techniques described in our earlier work [11], using only strong edges during search rather than explicit models of the grey-level about each point.

## 2 Background

In [10] we 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 points. A given point corresponds to a particular location on each shape or object to be modelled. The example shapes are all aligned into a standard co-ordinate frame, and a principal component analysis is applied to the co-ordinates of the points. This produces the mean position for each of the points and a description of the main ways in which the points tend to move together. The model can be used to generate new shapes using the equation

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

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

$(x_k, y_k)$ is the $k^{th}$ model point

$\bar{\mathbf{x}}$ represents the mean shape

$\mathbf{P}$ is a $2n \times t$ matrix of $t$ unit eigenvectors

$\mathbf{b} = (b_1 \ldots b_t)^T$ – a set of shape parameters

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 given in the original training set.

## 3 Modelling the Local Grey-Level Environment

We can also model the appearance of an object by examining the statistics of the grey levels in regions around each of the labelled model points in the training images. Since a given point corresponds to a particular part of the object, the grey-level patterns about corresponding points in images of different examples will often be similar. As with the shape, the grey-level environment can be modelled by a mean and a number of modes of allowed variation.

During image search we wish to locate the best position for each model point. This can be done by finding the area near to the current position where the image best matches the grey-level environment model for the point. In order to achieve this we need to associate an orientation with each point of our shape model in order to align the region correctly. A convenient way to do this is to define an orientation with respect to nearby model points. For instance, if the points lie along a boundary, we can choose to align the region with the normal to the boundary. Although in general we can consider a region of any shape around each point, we concentrate here on one-dimensional profiles normal to the arcs passing through each point.

For every shape model point in each training image, $j$ ($j = 1..N_s$), we can extract a profile, $\mathbf{g}_j$, of length $n_p$ pixels, centred at the point. If the profile runs from $\mathbf{p}_{start}$ to $\mathbf{p}_{end}$ and is of length $n_p$ pixels, the $k^{th}$ element of the profile is

$$g_{jk} = I_j(\mathbf{y}_k) \tag{2}$$

where $\mathbf{y_k}$ is the $k^{th}$ point along the profile :

$$\mathbf{y}_k = \mathbf{p}_{start} + \frac{k-1}{n_p - 1}(\mathbf{p}_{end} - \mathbf{p}_{start}) \tag{3}$$

and $I_j(y_k)$ is the grey level in image j at that point.

For each model point we can calculate a mean profile, $\bar{\mathbf{g}}$, and an $n_p \times n_p$ covariance matrix, $\mathbf{S_g}$, giving a second order statistical description of the expected profiles at the point. Principal Component Analysis can be used to produce a compact model of the allowable variation in the grey level profiles as it was for flexible shape models [10]. The variation about the mean is described by $\mathbf{P_g}$, the eigenvectors of $\mathbf{S_g}$, corresponding to the $t_g$ ($\leq n_p$) largest eigenvalues. By analogy with (1) we can write

$$\mathbf{g}_{new} = \bar{\mathbf{g}} + \mathbf{P_g}\mathbf{b}_{g new} \tag{4}$$

Where $\mathbf{b}_{g new}$ is a set of $t_g$ parameters describing the profile model.

Suppose we wish to measure how well a new profile, $\mathbf{g}$, extracted from an image fits a model for a point, represented by $\bar{\mathbf{g}}$, $\mathbf{P_g}$ and the $t_g$ largest eigenvalues $\lambda_j$, $j = 1..t_g$. The parameters required to best fit the model to $\mathbf{g}$ are given by

$$\mathbf{b}_g = \mathbf{P}_g^T(\mathbf{g} - \bar{\mathbf{g}}) \tag{5}$$

The best fit of the model to $\mathbf{g}$ is then given by (4),

$$\mathbf{g}_{best\_fit} = \bar{\mathbf{g}} + \mathbf{P_g}\mathbf{b}_g \tag{6}$$

If all the eigenvectors were used in the model ($t_g = n_p$) then this would be exactly equal to $\mathbf{g}$. However, since typically the expansion is truncated ($t_g < n_p$), there will be a difference between the two. The sum of squares of differences, $R^2$, is given by

$$R^2 = (\mathbf{g} - \mathbf{g}_{best\_fit})^T(\mathbf{g} - \mathbf{g}_{best\_fit})$$

It can be shown that

$$R^2 = (\mathbf{g} - \bar{\mathbf{g}})^T(\mathbf{g} - \bar{\mathbf{g}}) - \mathbf{b}_g^T\mathbf{b}_g \tag{7}$$

Figure 1 demonstrates this for a simple model with a single mode of variation.

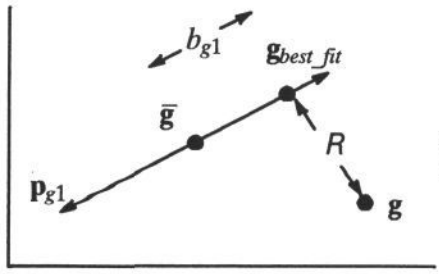


*Figure 1. A simple model consists of the mean $\bar{\mathbf{g}}$ and a single mode of variation $\mathbf{p}_{g1}$. $\mathbf{g}_{best\_fit}$ is then the projection of $\mathbf{g}$ onto the line of $\mathbf{p}_{g1}$ through the mean, and the residual, R, is the difference between $\mathbf{g}$ and $\mathbf{g}_{best\_fit}$.*

If all the eigenvectors were used in the model ($t_g = n_p$) then a measure of how well the model fits the profile would be given by the Mahalanobis distance,

$$M = \sum_{j=1}^{n_p} \frac{b_{gj}^2}{\lambda_j} \tag{8}$$

where $\lambda_i$ is the eigenvalue corresponding to the the $i^{th}$ eigenvector and $\lambda_i \geq \lambda_{i+1}$.

If $t_g < n_p$ we cannot calculate $b_{gj}$ for $j > t_g$. We have $\lambda_j \leq \lambda_{t_g}$ for $j > t_g$. If we make the approximation that $\lambda_j = 0.5\lambda_{t_g}$ for $j > t_g$, then

$$M \approx \sum_{j=1}^{t_g} \frac{b_{gj}^2}{\lambda_j} + \frac{2}{\lambda_{t_g}} \sum_{j=t_g+1}^{n_p} b_{gj}^2 \tag{9}$$

It can be shown (using (7)) that

$$\sum_{j=t_g+1}^{n_p} b_{gj}^2 = R^2 \tag{10}$$

Thus a measure of how well the model fits the profile is

$$F = \sum_{j=1}^{t_g} \frac{b_{gj}^2}{\lambda_j} + \frac{2R^2}{\lambda_{t_g}} \tag{11}$$

where $F$ approaches zero as the quality of fit improves.

If the distribution of the original profile data is assumed to be normal, the probability that **g** comes from the same population is approximately

$$p(\mathbf{g}|model) \propto e^{-F} \tag{12}$$

The method above can be applied either to the raw grey levels or the derivative of the grey levels along the profile in the image. Also we can normalise the profiles prior to training and evaluation using

$$\mathbf{g}_{ij}' = \frac{\mathbf{g}_{ij}}{\sum_{k=1}^{n_p} |g_{ijk}'|} \tag{13}$$

The effects on performance of using each of the four types of profile model (grey/derivative, normalised/unnormalised) are examined below.

# 4 Active Shape Models

We have previously described a method of fitting by local search given a starting approximation to the pose and shape parameters required to fit a model to an image [11,12,13]. By choosing a set of shape parameters **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, **X**, of the model in the image frame by defining the position, orientation and scale:

$$\mathbf{X} = M(s,\theta)[\mathbf{x}] + \mathbf{X}_c \quad (14)$$

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

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

and $(X_c, Y_c)$ 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 parameters, 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 we call them 'Active Shape Models' (ASMs).

To find a better location for each model point (step (i) above) we search in its current locality for the region of grey-levels which best matches the grey-level model for that point, using (11) (See [12,13]). We can generate a set of adjustments $(dX_i, dY_i)$ to move each point to a better position. We denote such a set as a vector $d\mathbf{X} = (dX_0, dY_0, \ldots, dX_{n-1}, dY_{n-1})^T$.

## 4.1 Calculating the Adjustments to the Pose and Shape Parameters

We aim to adjust the pose and shape parameters to move the points from their current locations in the image frame, **X**, to be as close to the suggested new locations $(\mathbf{X} + d\mathbf{X})$ as can be arranged *whilst still satisfying the shape constraints of the model.* The required pose adjustment is achieved by finding the translation $(dX_c, dY_c)$, rotation $d\theta$ and scaling factor $(1 + ds)$ which best maps the current set of points, **X**, onto the set of points given by $(\mathbf{X} + d\mathbf{X})$. This can be done by a weighted least squares fit [10]. The choice of possible weights and their effects on performance is discussed below.

Having adjusted the pose variables there remain residual adjustments which can only be achieved by deforming the shape of the model. We wish to calculate the adjustments, $d\mathbf{x}$, to the original model points in the local co-ordinate frame required to cause the scaled, rotated and translated points **X** to move by $d\mathbf{X}$ when combined with the new scale, rotation and translation parameters.

In [11] we show that

$$d\mathbf{x} = M((s(1 + ds))^{-1}, -(\theta + d\theta))[M(s,\theta)[\mathbf{x}] + d\mathbf{X} - d\mathbf{X}_c] - \mathbf{x} \quad (15)$$

The adjustments to the shape parameters, $d\mathbf{b}$, which will best match the model to the suggested new positions are given by solving

$$(\mathbf{P}^T\mathbf{W}_s)d\mathbf{x} = (\mathbf{P}^T\mathbf{W}_s\mathbf{P})d\mathbf{b} \quad (16)$$

This is a set of $t$ linear equations in the $t$ variables of $d\mathbf{b}$, and can be solved using standard matrix algebra. In the special case in which all weights are set to unity, $\mathbf{W}_s = \mathbf{I}$, Eq.(16) simplifies to

$$d\mathbf{b} = \mathbf{P}^T d\mathbf{x} \tag{17}$$

### 4.2 Updating the Pose and Shape Parameters

The equations above allow us to calculate changes to the shape parameters $d\mathbf{b}$ required to improve the match between an object model and image evidence. When these changes are applied we can ensure that the model only deforms into shapes consistent with the training set by placing limits on the values of $b_k$. A new example can be calculated, and new suggested movements derived for each point. See [11,12,13] for examples.

## 5 Quantative Assessment of ASM Search Performance

We wished to examine how accurately an ASM can locate examples of the modelled object in images, and how varying different parameters of the model affected the performance. This we did by running the ASM on a set of images, and comparing the ASM points with sets of points labelled manually on each image.

### 5.1 Method

The process for testing the techniques is as follows. Given an image of a modelled object and the positions of the model points, $\mathbf{X}_{known}$, annotated manually :

- Calculate an approximation to the best model pose by calculating the translation, rotation and scaling parameters which map the mean shape model points onto the known image points.
- Run the Active Shape Model search on the image, initialising it with the pose parameters *perturbed from the best 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 vs iterations allow comparison between different techniques, parameter settings and grey models.

Experiments were run to determine the effects on performance of

- Using raw grey-level models compared to derivative models, with and without normalisation.
- Varying the length of the grey-level profile model
- Using different weighting schemes during the calculation of pose and shape parameters.

The methods were also compared with our earlier approach of searching for strong edges in the correct direction [11] to determine what advantages were given by using grey-level models for each point.

### 5.2 Data

The experiments were conducted using a flexible model of a face. The model was trained on 11 $512^2$ images of one person (Figure 2), with 169 points being used to represent the shape.

*Figure 2 : Examples of image and model used.*

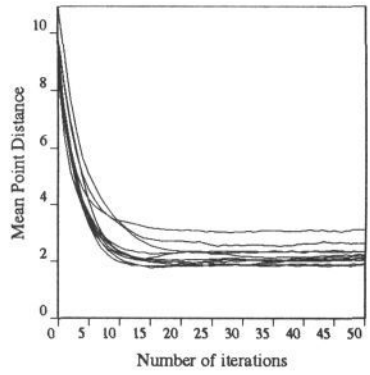


*Figure 3 : Average Performance of an ASM using un-normalised grey level profile models on 11 different images.*

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. 16 runs were done on each image, displacing the centre by $(\pm 10, \pm 10)$ pixels, the orientation by $\pm 0.05$ radians and the scale by $\pm 5\%$. Each face was about 200 pixels across.

# 6 Results

## 6.1 Effects on Performance of Using Different Types of Profile Model

We wished to assess the effects on the rate of convergence and final result of ASM search produced by using different types of profile model : raw grey-level vs. derivative, normalised vs. unnormalised. Figure 3 shows a typical set of curves, one per image, showing the mean distance of the model points from the known labelled points as the iterations progress. These were produced using un-normalised grey level profiles 7 pixels long. Figure 4 compares the average results for the four combinations of profile model, raw grey-level/derivative, un-normalised/normalised. Each profile was 7 pixels long.

This suggests that normalised derivative profile models give the best results, closely followed by un-normalised grey-level profile models. The un-normalised derivative models and the normalised grey-level models both give significantly poorer results. This order agrees with the results of Bailes & Taylor [14] in their work on modelling the grey levels along symmetry chords.

## 6.2 Effects on Performance of Varying the Length of Profile Model and Comparison with Searching for Strongest Edge.

The experiments were repeated using normalised derivative profile models of lengths $n_p = 2,3,5$ and 7 pixels. (The number of modes used in each model, $t_g = n_p - 1$ or $n_p - 2$). Figure 5 shows the results. This clearly indicates that the longer the profile the better the performance. In addition the experiments were performed replacing the grey-level model matching with a search for the strongest edge in the correct direction (the technique used in earlier work [11]). The results are shown in Figure 5. The addition of the grey-level models, particularly the longer ones, gives considerable improvements in final fit.

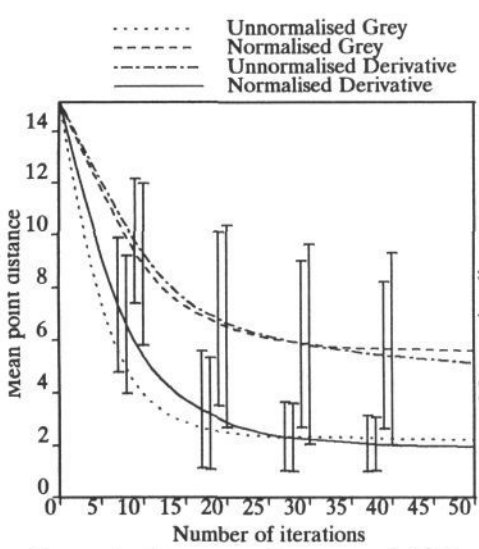


*Figure 4 : Average performance of ASMs using 4 different types of profile model. The error bars are 1.0 s.d.*

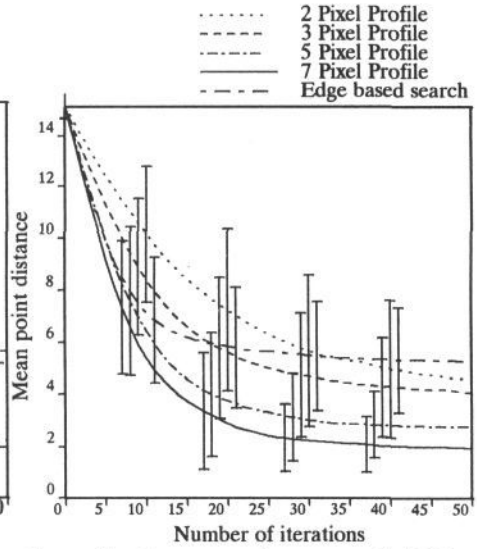


*Figure 5 : Average performance of ASMs using normalised derivative profile models of different lengths, and a strongest edge search. (The error bars are 1.0 s.d.)*

## 6.3 Effects on Performance of Using Different Weighting Schemes

Weights can be introduced at two stages in the ASM calculations (see section 4.1 ). Weights on each point can be used during the calculation of changes to the pose parameters and during calculation of changes to the shape parameters (Eq. 16).

We have experimented with two weighting schemes. In the first the weight on each point, $w_i$, is chosen to penalise against large suggested movements of the point :

$$\text{(a)} \qquad w_i = (1 + (dX_i^2 + dY_i^2))^{-1} \qquad (18)$$

where $(dX_i, dY_i)$ is the suggested movement of the point, the difference between its current position and the position of the area of nearby image which best fits the profile model for the point.

In the second the weight is proportional to the probability that the grey-levels in the area toward which the point is moving are in the class represented by the model:

$$\text{(b)} \qquad w_i = e^{-F_i} \qquad (19)$$

where $-F_i$ is the fit of the model to the best fitting area of image given by Eq. 11.

Figure 6 shows the effects of applying weighting (a) to an ASM searching for strong edges (rather than using a grey-level model). This suggests that although the weighting can slow convergence it can lead to a better final fit. Figure 7 compares the performance of the different schemes on ASMs using normalised derivative profiles of 7 pixels, both using weights only during the pose calculation and using weights during pose and shape calculation. In this case weighting method (a) slows the rate of convergence considerably. Method (b) ( using the probability of fit ) gives no significant difference to the no weights case. However, such a scheme may prove useful in overcoming some occlusion. Model points in occluded regions will give poor fit values, thus low weights, and will be ignored during model fitting; the shape will be dominated by those points found with good fits. Early experiments confirm this.

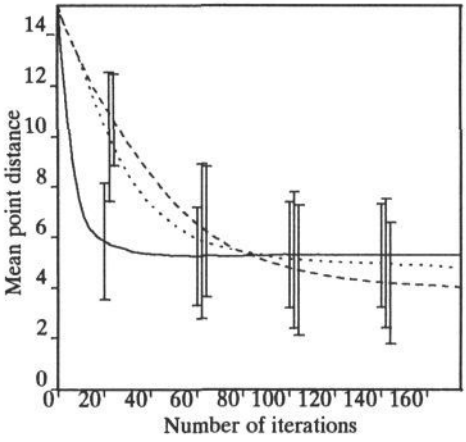


*Figure 6 : Performance of edge based ASMs using different weighting schemes. Error bars are 1.0 s.d.s.*

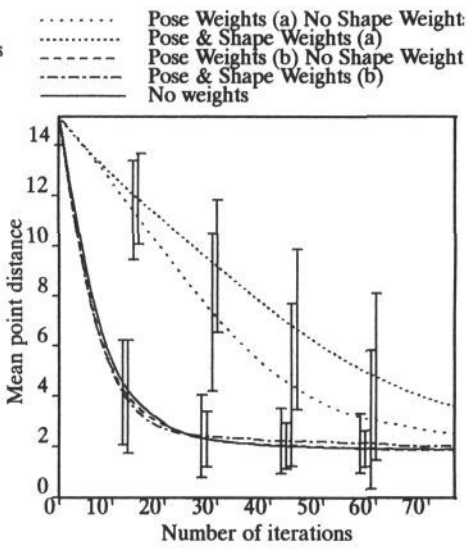


*Figure 7 :Performance of grey-model based ASMs using different weighting schemes. (Profile model length 7 pixels) Error bars are 1.0 s.d.s.*

# 7 Discussion and Conclusions

The method of locating points by finding the best fit of a flexible grey-level model to an area of image is a useful improvement on both edge based and standard correlation techniques. At present we simply find the patch in the search area whose grey-levels best match the model. Better results may be obtained by building a background model for every shape model point, and applying a Baysian classifier to determine the patch most likely to contain the point and least likely to be background. The area modelled need not just be a line, it can be a rectangle or any shaped patch.

We found that using normalised derivative profile models produces the best results. Such models are insensitive to scaling of the grey-levels and addition of a constant (caused by changes in illumination) so a good result would be expected. Why normalised grey models and unnormalised derivative models should perform so relatively poorly is less clear.

The rate of convergence and quality of final result improved as the length of the profile models increased. The longer the model the more likely it is to locate the correct section of the area sampled, and the less it is confused by noise or clutter. The use of the grey-level models gives considerable improvements in final fit over earlier methods of searching for the strongest edge in a given direction.

The experiments above suggest that using the weighting scheme (a) (Eq.18) slows convergence but can give improvements in overall fit when only edge strength is used in the search. This may be because it penalises against points being pulled away from the average position by occasional outlying matches to nearby edges. The weighting gives little or no improvement when the more specific grey-level models are used, perhaps because they produce fewer spurious matches. Weighting scheme (b) appears to have no significant effect compared to using no weights, but may make the system more robust to occlusion.

Including weighting in the pose calculation requires very little extra computation (only the calculation of the weights themselves) so can be done 'for free'. Including weights in the calculation of shape parameters requires solving Eq. 16 rather than using the much simpler Eq. 17, though the increase in computation is small compared to that required to assess the quality of fit of each grey-level model.

## Acknowledgements

This work is funded by SERC. The authors would like to thank the other members of the Wolfson Image Analysis Unit for their help and advice.

## References

[1] 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.

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

[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] T.F.Cootes, C.J.Taylor, D.H.Cooper and J.Graham, Training Models of Shape from Sets of Examples. in *Proc. British Machine Vision Conference*. Springer-Verlag, 1992, pp.9–18.

[11] T.F.Cootes, C.J.Taylor, Active Shape Models - 'Smart Snakes'. in *Proc. British Machine Vision Conference*. Springer-Verlag, 1992, pp.266–275.

[12] T.F.Cootes, C.J.Taylor, A.Lanitis,D.H.Cooper and J.Graham, Building and Using Flexible Models Incorporating Grey-Level Information, *Proc. International Conference on Computer Vision*, May 1993.

[13] T.F.Cootes, A.Hill, C.J.Taylor, J.Haslam, The Use of Active Shape Models for Locating Structures in Medical Images. *Proc. on Image Processing and Medical Imaging*, Arizona, June 1993.

[14] D.R.Bailes, C.J.Taylor, The Use of Symmetry Chords for Expressing Grey Level Constraints. in *Proc. British Machine Vision Conference*. Springer-Verlag, 1992, pp. 296–305.