Comparing Active Shape Models with Active Appearance Models

T.F. Cootes, G. Edwards and C.J. Taylor
Dept. Medical Biophysics,
Manchester University, UK
email: t.cootes@man.ac.uk

Abstract

Statistical models of the shape and appearance of image structures can be matched to new images using both the Active Shape Model [7] algorithm and the Active Appearance Model algorithm [2]. The former searches along profiles about the current model point positions to update the current estimate of the shape of the object. The latter samples the image data under the current instance and uses the difference between model and sample to update the appearance model parameters. In this paper we compare and contrast the two algorithms, giving the results of experiments testing their performance on two data sets, one of faces, the other of structures in MR brain sections. We find that the ASM is faster and achieves more accurate feature point location than the AAM, but the AAM gives a better match to the texture.

1 Introduction

Interpretting images containing objects whose appearance can vary is difficult. A powerful approach has been to use deformable models, which can represent the variations in shape and/or texture (intensity) of the target objects. In this paper we concentrate on two related algorithms, the Active Shape Model (ASM), which seeks to match a set of model points to an image, constrained by a statistical model of shape, and the Active Appearance Model (AAM), which seeks to match both the position of the model points and a representation of the texture of the object to an image.

Both use the same underlying statistical model of the shape of target objects. This represents shape using a set of landmarks, learning the valid ranges of shape variation from a training set of labelled images [7].

The ASM matches the model points to a new image using an iterative technique which is a variant on the Expectation Maximisation algorithm. A search is made around the current position of each point to find a point nearby which best matches a model of the texture expected at the landmark. The parameters of the shape model controlling the point positions are then updated to move the model points closer to the points found in the image.

The AAM manipulates a full model of appearance, which represents both shape variation and the texture of the region covered by the model. This can be used to generate full synthetic images of modelled objects. The AAM uses the difference between the current synthesised image and the target image to update its parameters.

There are three key differences between the two algorithms:

1. The ASM only uses models of the image texture in small regions about each landmark point, whereas the AAM uses a model of the appearance of the whole of the region (usually inside a convex hull around the points).
2. The ASM searches around the current position, typically along profiles normal to the boundary, whereas the AAM only samples the image under the current position.
3. The ASM essentially seeks to minimise the distance between model points and the corresponding points found in the image, whereas the AAM seeks to minimise the difference between the synthesized model image and the target image.

In the following paper we give a more detailed description of the two algorithms and give results of experiments testing their performance on two data sets, one of faces, the other of structures in MR brain sections. We measure their accuracy in locating landmark points, their capture range and the time required to locate a target structure.

2 Background

There has been a great deal of research into using deformable models to interpret images. Reviews are given in [7, 12]. Active Shape Models were developed by Cootes *et.al.*[7] to match statistical models of object shape to new images. They have been used successfully in many application areas, including face recognition [11], industrial inspection [7] and medical image interpretation [6]. They have been extended to search 3D images [10].

Active Appearance Models were introduced more recently [2]. They have proved very successful for interpreting and tracking images of faces [9], and have been applied to medical image interpretation [1]. They been extended to model and search colour images [8].

3 Appearance Models

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

Given such a set we can generate a statistical models of shape and texture variation (see [2] for details). The shape of an object can be represented as a vector $\mathbf{x}$ and the texture (or grey-levels) represented as a vector $\mathbf{g}$. The appearance model has parameters $\mathbf{c}$ controlling the shape and texture according to

$$\begin{aligned} \mathbf{x} &= \bar{\mathbf{x}} + \mathbf{Q}_s\mathbf{c} \\ \mathbf{g} &= \bar{\mathbf{g}} + \mathbf{Q}_g\mathbf{c} \end{aligned} \tag{1}$$

where $\bar{\mathbf{x}}$ is the mean shape, $\bar{\mathbf{g}}$ the mean texture and $\mathbf{Q}_s$,$\mathbf{Q}_s$ are matrices describing the modes of variation derived from the training set.

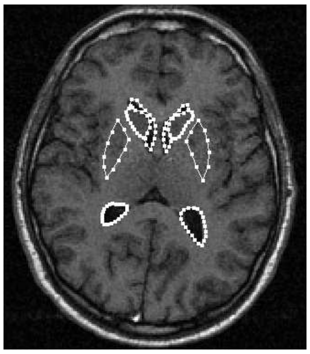


Figure 1: Example of MR brain slice labelled with 123 landmark points around the ventricles, the caudate nucleus and the lentiform nucleus

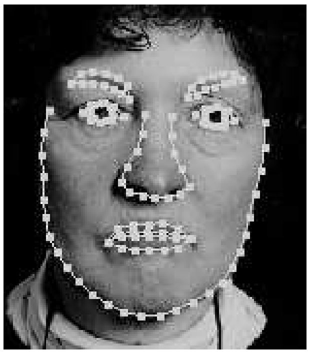


Figure 2: Example of face image labelled with 122 landmark points

An example image can be synthesised for a given $\mathbf{c}$ by generating a texture image from the vector $\mathbf{g}$ and warping it using the control points described by $\mathbf{x}$. For instance, Figure 3 shows the effects of varying the first two appearance model parameters, c_1, c_2, of a model trained on a set of face images, labelled as shown in Figure 2. These change both the shape and the texture component of the synthesised image.

c_1 varies by ± 2 s.d.s $\qquad$ c_2 varies by ± 2 s.d.s

Figure 3: First two modes of an appearance model of a face

4 Active Shape Model Matching

4.1 Active Shape Models

The Active Shape Model algorithm is a fast and robust method of matching a set of points controlled by a shape model to a new image. The shape parameters, $\mathbf{b}$, for the model, along with parameters defining the global pose (the position, orientation and scale), define the position of the model points in an image, $\mathbf{X}$.

Each step in an iterative approach to improving the fit of the points, $\mathbf{X}$, to an image involves first examining the region of the image around each current model point (X_i, Y_i) to find the best nearby match (X_i', Y_i'), then updating the parameters $(t_x, t_y, s, \theta, \mathbf{b})$ to best fit the model to the new found points $\mathbf{X}'$. This is repeated until convergence.

In practice we search for the points along profiles normal to the model boundary through each model point. If we expect the model boundary to correspond to an edge, we can simply locate the strongest edge (including orientation if known) along each profile to give the new suggested location for the model point.

4.2 Local Grey-level Models

Model points are not always placed on the strongest edge in the locality - they may represent a weaker secondary edge or some other image structure. The best approach is to learn from the training set what to look for in the target image. This is achieved by sampling along the profile normal to the boundary in the training set, and building a statistical model of the grey-level structure [5]. The model consists of an estimate of the mean $\bar{\mathbf{g}}$ and covariance $\mathbf{S}_g$ of the grey-levels along the profile.

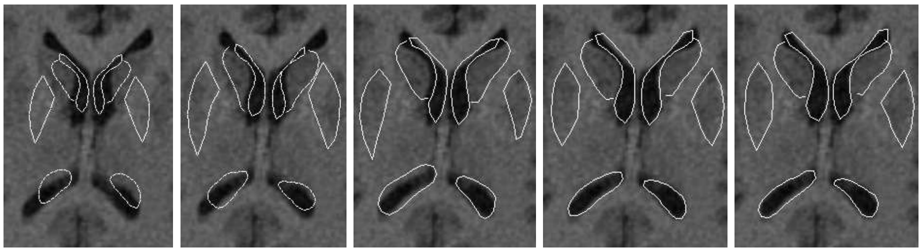
The quality of fit of a new sample, $\mathbf{g}_s$, to the model is given by

$$f(\mathbf{g}_s) = (\mathbf{g}_s - \bar{\mathbf{g}})^T \mathbf{S}_g^{-1} (\mathbf{g}_s - \bar{\mathbf{g}}) \tag{2}$$

During search we sample from a profile m pixels either side of the current point ($m > p$). We then test the quality of fit of the corresponding grey-level model at each of the $2(m - p) + 1$ possible positions along the sample and choose the one which gives the best match (lowest value of $f(\mathbf{g}_s)$). This is repeated for every model point, giving a suggested new position for each point. We then update the current pose and shape parameters to best match the model to the new points [4].

To improve the efficiency and robustness of the ASM algorithm, it is implemented in a multi-resolution framework. This involves first searching for the object in a coarse image, then refining the location in a series of successively finer resolution images.

For example Figure 4 shows the performance of the ASM attempting to match the model of the cortical structures to a new image. The model is trained on 72 example shapes, and uses 25 parameters to represent 98% of the variation. The early stages of the search are at coarse resolutions, allowing large movements. The search runs to convergence in about 220ms on a 450MHz PC.



Initial 1 it. 6 it.s 12 it.s 20 it.s

Figure 4: Multi-resolution ASM search on new image

5 Active Appearance Model Matching

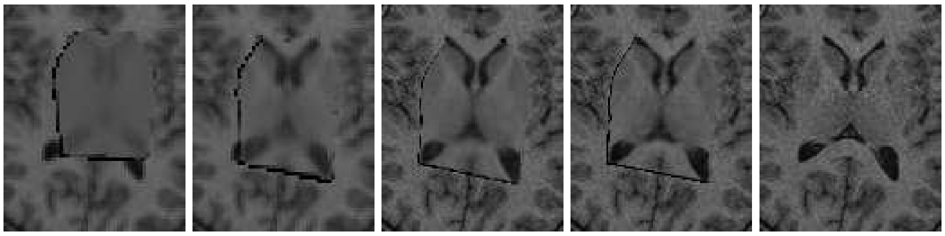
The AAM differs from the ASM in that instead of searching locally about each model point, it seeks to minimise the difference between a new image and one synthesised by the appearance model. Given a set of model parameters, $\mathbf{c}$, we can generate a hypothesis for the shape, $\mathbf{x}$, and texture, $\mathbf{g}_m$, of a model instance. To compare this hypothesis with the image, we use the suggested shape to sample the image texture, $\mathbf{g}_s$, and compute the difference, $\delta\mathbf{g} = \mathbf{g}_s - \mathbf{g}_m$. We seek to minimise the magnitude of $|\delta\mathbf{g}|$.

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

During a training phase, the AAM learns a linear relationship between $\delta\mathbf{g}$ and the parameter perturbation required to correct this, $\delta\mathbf{c} = \mathbf{A}\delta\mathbf{g}$. The matrix $\mathbf{A}$ is obtained by linear regression on random displacements from the true training set positions and the induced image residuals (see [2] for details).

During search we simply iteratively compute $\delta\mathbf{g}$ given the current parameters $\mathbf{c}$ and then update the samples using $\mathbf{c} \rightarrow \mathbf{c} - \delta\mathbf{c}$. This is repeated until no improvement is made to the error, $|\delta\mathbf{g}|^2$, and convergence is declared. Again we use a multi-resolution implementation of search. This is more efficient and can converge to the correct solution from further away than search at a single resolution.

For example, Figure 5 shows an example of an AAM of the central structures of the brain slice converging from a displaced position on a previously unseen image. The model represented about 10000 pixels and had 30 $\mathbf{c}$ parameters. The search took about 330ms on a 450MHz PC. Figure 6 shows examples of the results of the search, with the found model points superimposed on the target images.



Initial 2 its 6 its 16 its original

Figure 5: Multi-resolution AAM search from a displaced position

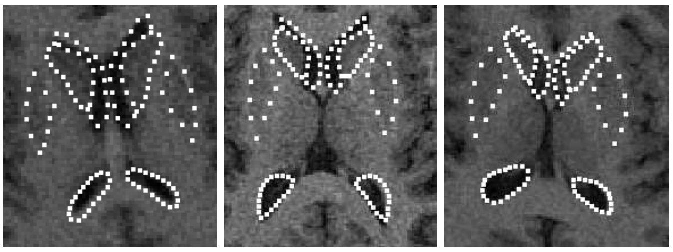


Figure 6: Results of AAM search. Model points superimposed on target image

6 Results of Comparative Experiments

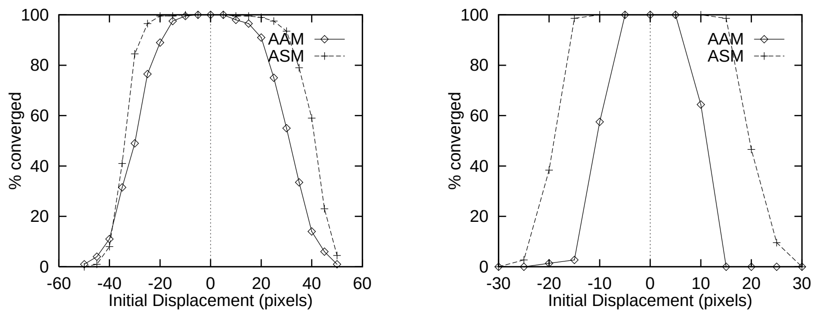
6.1 Data sets

We compared the performance of the ASM and AAM on two data sets. The first contained 400 face images, each marked with 133 points (See Figure 2). The second consisted of 72 slices of MR images of brains, each marked up with 133 points around sub-cortical structures. We performed search experiments on each data set, measuring the accuracy to which points could be re-located, how well texture could be matched and the time required to do so. For the faces, we trained on 200 then tested on the remaining 200. For the brains we performed leave-one-brain-out experiments.

The Appearance model was built to represent 5000 pixels in both cases. Multi-resolution search was used, using 3 levels with resolutions of 25%, 50% and 100% of the original image in each dimension. At most 10 iterations were run at each resolution. The ASM used profile models 11 pixels long (5 either side of the point) at each resolution, and searched 3 pixels either side. The performance of the algorithms can depend on the choice of parameters - we have chosen values which have been found to work well on a variety of applications.

Capture range

We systematically displaced the model instance from the known best position by up to ± 100 pixels in x, then ran the search to attempt to locate the target points. Figure 7 shows the RMS error in the position of the centre of gravity given the different starting positions for both ASMs and AAMs.



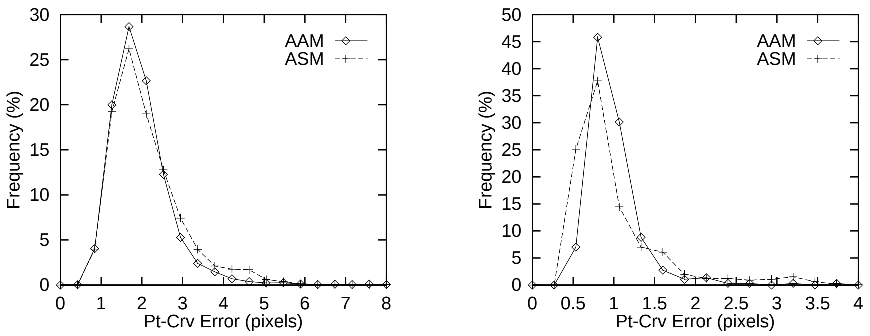
Face data Brain data

Figure 7: Relative capture range of ASM and AAMs

Thus the AAM has a slightly larger capture range for the face, but the ASM has a much larger capture range than the AAM for the brain structures. Of course, the results will depend on the resolutions used, the size of the models used and the search length of the ASM profiles.

Point location accuracy

One each test image we displaced the model instance from the true position by ± 10 in x and y (for the face) and ± 5 in x and y (for the brain), 9 displacements in total, then ran the search starting with the mean shape. On completion the results were compared with hand labelled points. Figure 8 shows frequency histograms for the resulting point-to-boundary errors(the distance from the found points to the associated boundary on the marked images). The ASM gives more accurate results than the AAM for the brain data, and comparable results for the face data.



Face data   Brain data

Figure 8: Histograms of point-boundary errors after search from displaced positions

Table 1 summarises the RMS point-to-point error, the RMS point-to-boundary error, the mean time per search and the proportion of convergence failures. Failure was declared when the RMS Point-Point error is greater than 10 pixels. The searches were performed on a 450MHz PentiumII PC running Linux.

Data	Model	Time/search	Pt-Pt Error	Pt-Crv Error	Failures
Face	ASM	190ms	4.8	2.2	1.0%
	AAM	640ms	4.0	2.1	1.6%
Brain	ASM	220ms	2.2	1.1	0%
	AAM	320ms	2.3	1.1	0%

Table 1: Comparison of performance of ASM and AAM algorithms on face and brain data (See Text)

Thus the ASM runs significantly faster for both models, and locates the points more accurately than the AAM.

6.2 Texture Matching

The AAM explicitly generates a texture image which it seeks to match to the target image. After search we can thus measure the resulting RMS texture error. The ASM only locates points positions. However, given the points found by the ASM we can find the best fit of the texture model to the image, then record the residual. Figure 9 shows frequency histograms for the resulting RMS texture errors per pixel. The images have a contrast range of about [0,255]. The AAM produces a significantly better performance than the ASM on the face data, which is in part to be expected, since it is explicitly attempting to minimise the texture error. However, the ASM produces a better result on the brain data.

This is caused by a combination of experimental set up and the additional constraints imposed by the appearance model.

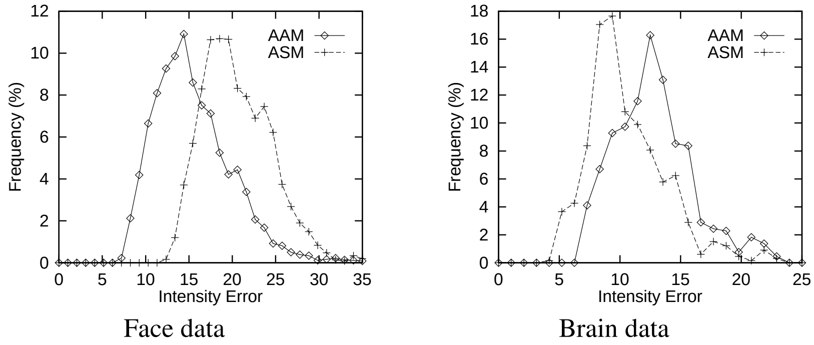


Figure 9: Histograms of RMS texture errors after search from displaced positions

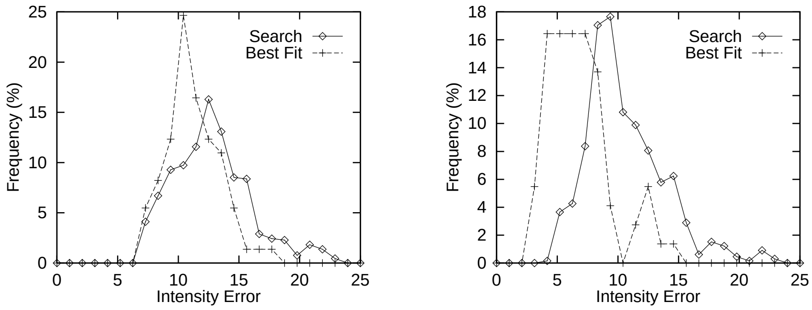
Figure 10 compares the distribution of texture errors found after search with those obtained when the model is fit to the (hand marked) target points in the image (the 'Best Fit' line). This demonstrates that the AAM is able to achieve results much closer to the best fit results than the ASM (because it is more explicitly minimising texture errors). The difference between the best fit lines for the ASM and AAM has two causes;

- For the ASM experiments, though a leave-1-out approach was used for training the shape models and grey profile models, a single texture model trained on all the examples was used for the texture error evaluation. This could fit more accurately to the data than the model used by the AAM, trained in a leave-1-out regime.
- The AAM fits an appearance model which couples shape and texture explicitly - the ASM treats them as independent. For the relatively small training sets used this overconstrained the model, leading to poorer results.

The latter point is demonstrated in Figure 11, which shows the distribution of texture errors when fitting models to the training data. One line shows the errors when fitting a 50 mode texture model to the image (with shape defined by a 50 mode shape model fit to the labelled points). The second shows the best fit of a full 50 mode appearance model to the data. The additional constraints of the latter mean that for a given number of modes it is less able to fit to the data than independent shape and texture models, because the training set is not large enough to properly explore the variations. For a sufficiently large training set we would expect to be able to properly model the correlation between shape and texture, and thus be able to generate an appearance model which performed almost as well as a independent models, each with the same number of modes. Of course, if the total number of modes of the shape and texture model were constrained to that of the appearance model, the latter would perform much better.

7 Discussion and Conclusions

Active Shape Models search around the current location, along profiles, so one would expect them to have a larger capture range than the AAM which only examines the image



AAM ASM

Figure 10: Histograms of RMS texture errors after search from displaced positions

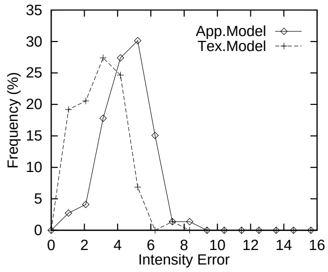


Figure 11: Comparison between texture model best fit and appearance model best fit

directly under its current area. This is clearly demonstrated in the results on the brain data set.

ASMs only use data around the model points, and do not take advantage of all the grey-level information available across an object as the AAM does. Thus they may be less reliable. However, the model points tend to be places of interest (boundaries or corners) where there is the most information. One could train an AAM to only search using information in areas near strong boundaries - this would require less image sampling during search so a potentially quicker algorithm. A more formal approach is to learn from the training set which pixels are most useful for search - this was explored in [3]. The resulting search is faster, but tends to be less reliable.

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

The AAM algorithm relies on finding a linear relationship between model parameter displacements and the induced texture error vector. However, we could augment the error vector with other measurements to give the algorithm more information. In particular one method of combining the ASM and AAM would be to search along profiles at each model point and augment the texture error vector with the distance along each profile of the best match. Like the texture error, this should be driven to zero for a good match. This approach will be the subject of further investigation.

To conclude, we find that the ASM is faster and achieves more accurate feature point location than the AAM. However, as it explicitly minimises texture errors the AAM gives a better match to the image texture.

Acknowledgements

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

References

[1] T. Cootes, C. Beeston, G. J. Edwards, and C. Taylor. A unified framework for atlas matching using active appearance models. In 16^{th} *Conference on Information Processing in Medical Imaging*, page (To appear), 1999.

[2] T. Cootes, G. J. Edwards, and C. J. Taylor. Active appearance models. In H.Burkhardt and B. Neumann, editors, 5^{th} *European Conference on Computer Vision*, volume 2, pages 484–498. Springer, 1998.

[3] T. Cootes, G. J. Edwards, and C. J. Taylor. A comparative evaluation of active appearance model algorithms. In P. Lewis and M. Nixon, editors, 9^{th} *British Machine Vison Conference*, volume 2, pages 680–689, Southampton, UK, Sept. 1998. BMVA Press.

[4] T. Cootes and C. Taylor. A mixture model for representing shape variation. In A. Clarke, editor, 8^{th} *British Machine Vison Conference*, pages 110–119. BMVA Press, Sept. 1997.

[5] T. F. Cootes, A. Hill, and C. J. Taylor. Medical image interpretation using active shape models: Recent advances. In 14^{th} *Conference on Information Processing in Medical Imaging, France*, pages 371–372, June 1995.

[6] T. F. Cootes and C. J. Taylor. Combining point distribution models with shape models based on finite-element analysis. *Image and Vision Computing*, 13(5):403–409, 1995.

[7] T. F. Cootes, C. J. Taylor, D. H. Cooper, and J. Graham. Active shape models - their training and application. *Computer Vision and Image Understanding*, 61(1):38–59, Jan. 1995.

[8] G. Edwards, T. F. Cootes, and C. J. Taylor. Advances in active appearance models. In 7^{th} *International Conference on Computer Vision*, page Submitted, 1999.

[9] G. Edwards, C. J. Taylor, and T. F. Cootes. Interpreting face images using active appearance models. In 3^{rd} *International Conference on Automatic Face and Gesture Recognition 1998*, pages 300–305, Japan, 1998.

[10] A. Hill, T. F. Cootes, C. J. Taylor, and K. Lindley. Medical image interpretation: A generic approach using deformable templates. *Journal of Medical Informatics*, 19(1):47–59, 1994.

[11] A. Lanitis, C. Taylor, and T. Cootes. Automatic interpretation and coding of face images using flexible models. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19(7):743–756, 1997.

[12] T. McInerney and D. Terzopoulos. Deformable models in medical image analysis: a survey. *Medical Image Analysis*, 1(2):91–108, 1996.