View-Based Active Appearance Models

T.F.Cootes [a] * G.V.Wheeler[a] K.N.Walker[b] C.J.Taylor[a]

[a]Division Imaging Science and Biomedical Engineering, University of Manchester M13 9PT U.K.

[b]Image-Metrics PLC, Regent House, Heaton Lane, Stockport SK4 3LE, UK

We demonstrate that a small number of 2D linear statistical models are sufficient to capture the shape and appearance of a face from a wide range of viewpoints. Such models can be used to estimate head orientation and track faces through large angles. Given multiple images of the same face we can learn a coupled model describing the relationship between the frontal appearance and the profile of a face. This relationship can be used to predict new views of a face seen from one view and to constrain search algorithms which seek to locate a face in multiple views simultaneously.

1. Introduction

The appearance of an object in a 2D image can change dramatically as the viewing angle changes. To deal effectively with real world scenes models must be developed which can represent this variability. For instance, the majority of work on face tracking and recognition assumes near fronto-parallel views, and tends to break down when presented with large rotations or profile views. Three general approaches have been used to deal with this; a) use a full 3D model (22; 4; 16), b) introduce non-linearities into a 2D model (6; 17; 19) and c) use a set of models to represent appearance from different view-points (14; 8; 23). In this paper we explore the last approach, using statistical models of shape and appearance to represent the variations in appearance from a particular viewpoint and the correlations between models of different view-points.

The appearance models are trained on example images labelled with sets of landmarks to define the correspondences between images (1). Lanitis *et.al.* (11) showed that a linear model was sufficient to simulate considerable changes in viewpoint, as long as all the modelled features (the landmarks) remained visible. A model trained on near fronto-parallel face images can cope with pose variations of up to 45^o either side. For larger angle displacements, some features become occluded, and the assumptions of the model break down.

We demonstrate that to deal with full 180^o rotation (from left profile to right profile), one needs only 5 models, roughly centred on viewpoints at -90^o,-45^o,0^o,45^o,90^o (where 0^o corresponds to fronto-parallel). The pairs of models at $\pm 90^o$ (full profile) and $\pm 45^o$(half profile) are simply reflections of each other, so only 3 distinct models are required. These models can be used for estimating head pose, for tracking faces through wide changes in orientation and for synthesizing new views of a subject given a single view.

Each model is trained on labelled images of a variety of people with a range of head orientations, chosen so none of the features for that model become occluded. The different models use different sets of features (see Figure 2). Each example view can then be approximated using the appropriate appearance model with a vector of parameters, $\mathbf{c}$. We assume that as the orientation changes, the parameters trace out an approximately elliptical path. We can learn the relationship between $\mathbf{c}$ and head orientation, allowing us to both estimate the orientation of any head and to be able to synthesize a face at any orientation.

By using the Active Appearance Model algorithm (3; 1) we can match any of the individual

*Whilst doing this work, Dr Cootes was funded by an EPSRC Advanced Fellowship.

models to a new image rapidly. If we know in advance the approximate pose, we can easily select the most suitable model. If we do not know, we can search with each of the five models and choose the one which achieves the best match. Once a model is selected and matched, we can estimate the head pose, and thus track the face, switching to a new model if the head pose varies significantly.

There are clearly correlations between the parameters of one view model and those of a different view model. In order to learn these, we need images taken from two views simultaneously. For our experiments we achieved this using a judiciously placed mirror, giving a frontal and a profile view (Figure 1).

Figure 1. Using a mirror we capture frontal and profile appearance simultaneously

By annotating such images and matching frontal and profile models, we obtain corresponding sets of parameters. These can be analyzed to produce a joint model which controls both frontal and profile appearance. Such a joint model can be used to synthesize new views given a single view. Though this can perhaps be done most effectively with a full 3D model (22), we demonstrate that good results can be achieved just with a set of 2D models. The joint model can also be used to constrain an Active Appearance Model search (3; 1), allowing simultaneous matching of frontal and profile models to pairs of images.

In the following we describe the techniques in more detail and give examples of the models, their ability to estimate pose, to track faces and to synthesize new views.

2. Background

Statistical models of shape and texture have been widely used for recognition, tracking and synthesis (7; 11; 3; 20), but have tended to only be used with near fronto-parallel images.

Moghaddam and Pentland (15) describe using view-based eigenface models to represent a wide variety of viewpoints. Our work is similar to this, but by including shape variation (rather than the rigid eigen-patches), we require fewer models and can obtain better reconstructions with fewer model parameters. Kruger (8) and Wiskott *et.al.* (23) used flexible graphs of Gabor Jets of frontal, half-profile and full profile views for face recognition and pose estimation. Maurer and von der Malsburg (12) demonstrated tracking heads through wide angles by tracking graphs whose nodes are facial features, located with Gabor jets. The system is effective for tracking, but is not able to synthesize the appearance of the face being tracked. Murase and Nayar (6) showed that the projections of multiple views of a rigid object into an eigenspace fall on a 2D manifold in that space. By modelling this manifold they could recognise objects from arbitrary views. A similar approach has been taken by Gong *et.al.* (18; 9) who use non-linear representations of the projections into an eigen-face space for tracking and pose estimation, and by Graham and Allinson (5) who use it for recognition from unfamiliar viewpoints.

Romdhani *et.al.* (17) have extended the Active Shape Model to deal with full 180^o rotation of a face using a non-linear model. However, the non-linearities mean the method is slow to match to a new image. They have also extended the AAM (19) using a kernel PCA. A non-linear 2D shape model is combined with a non-linear texture model on a 3D texture template. The approach is promising, but considerably more complex than using a small set of linear 2D models.

Vetter (21; 22) has demonstrated how a 3D statistical model of face shape and texture can be used to generate new views given a single view.

The model can be matched to a new image from more or less any viewpoint using a general optimization scheme, though this is slow. Similar work has been described by Fua and Miccio (4) and Pighin *et.al.* (16). By explicitly taking into account the 3D nature of the problem, this approach is likely to yield better reconstructions than the purely 2D method described below. However, the view based models we propose could be used to drive the parameters of the 3D head model, speeding up matching times.

La Cascia *et.al.* (10) describe a related approach to head tracking. They project the face onto a cylinder (or more complex 3D face shape) and use the residual differences between the sampled data and the model texture (generated from the first frame of a sequence) to drive a tracking algorithm, with encouraging results. The model is tuned to the individual by initialization in the first frame, and does not explicitly track expression changes, as the approach here is able to.

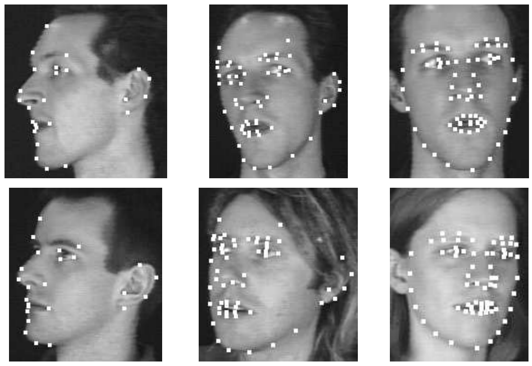
3. View-Based Models of Appearance

An appearance model can represent both the shape and texture variability seen in a training set. The training set consists of labelled images, where key landmark points are marked on each example face. The training sets used in this paper were labelled manually, though automatic methods are being developed (2). For instance, Figure 2 shows examples of labelled images used to train the view-based face models.

Given such a set we can generate a statistical models of shape and texture variation by applying principal component analysis to the training data (see (1; 3) for details). The shape of an object can be represented as a vector $\mathbf{x}$ and the texture (grey-levels or colour values) represented as a vector $\mathbf{g}$. The appearance model has parameters, $\mathbf{c}$, controlling the shape and texture according to

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

where $\bar{\mathbf{x}}$ is the mean shape, $\bar{\mathbf{g}}$ the mean texture and $\mathbf{Q}_s$,$\mathbf{Q}_g$ are matrices describing the modes of variation derived from the training set, and are constructed from the eigenvectors of the covariance matricies of the data (see (1; 3)).



Profile Half Profile Frontal

Figure 2. Examples from the training sets for the models

We trained three distinct models on data similar to that shown in Figure 2. The profile model was trained on 84 landmarked images taken of 15 different individuals from a variety of orientations. The half-profile model was trained on 82 images of 15 individuals, and the frontal model on 288 images of 15 individuals.

An example image can be synthesized 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 models trained on sets of face images, labelled as shown in Figure 2. These parameters change both the shape and the texture component of the synthesized image.

4. Predicting Pose

We assume that the model parameters are related to the viewing angle, θ, approximately as

$$\mathbf{c} = \mathbf{c}_0 + \mathbf{c}_x \cos(\theta) + \mathbf{c}_y \sin(\theta) \quad (2)$$

where $\mathbf{c}_0$, $\mathbf{c}_x$ and $\mathbf{c}_y$ are vectors estimated from training data. Here we consider only rotation about a vertical axis - head turning. Nodding can be dealt with in a similar way.

This is an accurate representation of the relationship between the shape, $\mathbf{x}$, and orientation angle under an affine projection (the landmarks trace circles in 3D which are projected to ellipses in 2D), but our experiments suggest it is also an acceptable approximation for the appearance model parameters, $\mathbf{c}$.

We estimate the head orientation in each of our training examples, θ_i, accurate to about $\pm 10^o$. For each such image we find the best fitting model parameters, $\mathbf{c}_i$. We then perform regression between $\{\mathbf{c}_i\}$ and the vectors $\{(1, \cos(\theta_i), \sin(\theta_i))'\}$ to learn $\mathbf{c}_0$,$\mathbf{c}_x$ and $\mathbf{c}_y$.

Figure 4 shows reconstructions in which the orientation, θ, is varied in Equation 2.

Given a new example with parameters $\mathbf{c}$, we can estimate its orientation as follows. Let $\mathbf{R}_c^{-1}$ be the left pseudo-inverse of the matrix $(\mathbf{c}_x|\mathbf{c}_y)$ (thus $\mathbf{R}_c^{-1}(\mathbf{c}_x|\mathbf{c}_y) = \mathbf{I}_2$). Let

$$(x_a, y_a)' = \mathbf{R}_c^{-1}(\mathbf{c} - \mathbf{c}_0) \tag{3}$$

then $(\cos\theta, \sin\theta) \approx (x_a, y_a)$ and the best estimate of the orientation is $\theta = \tan^{-1}(y_a/x_a)$.

Figure 5 shows the predicted orientations vs the actual orientations for the training sets for each of the models. It demonstrates that equation 2 is an acceptable model of parameter variation under rotation, and that the estimate is accurate to about $\pm 5^o$.

5. Tracking through wide angles

We can use the set of models to track faces through wide angle changes (full left profile to full right profile). We use a simple scheme in which we keep an estimate of the current head orientation and use it to choose which model should be used to match to the next image.

We assume we can locate the face in the first frame using suitable initialization or a face detector. We then project the current best model instance into the next frame and run a multi-resolution search with the AAM. We estimate the head orientation from the results of the search, as described above. We then use the orientation to choose the most appropriate model with which to continue. Each model is valid over a particular range of angles, determined from its training set (see Table 1). If the orientation suggests changing to a new model, we estimate the parameters of the new model from those of the current best fit. We then perform an AAM search to match the new model more accurately. This process is repeated for each subsequent frame, switching to new models as the angle estimate dictates.

Table 1
Valid angle ranges for each model

Model	Angle Range
Left Profile	-110^o - -60^o
Left Half-Profile	-60^o - -40^o
Frontal	-40^o - 40^o
Right Half-Profile	40^o - 60^o
Right Profile	60^o - 110^o

When switching to a new model we must estimate the image pose (position, within image orientation and scale) and model parameters of the new example from those of the old. We assume linear relationships which can be determined from the training sets for each model, as long as there are some images (with intermediate head orientations) which belong to the training sets for both models.

Figure 6 shows the results of using the models to track the face in a new test sequence (in this case a previously unseen sequence of a person who is in the training set). The model reconstruction is shown superimposed on frames from the sequence. The method appears to track well, and is able to reconstruct a convincing simulation of the sequence.

We used this system to track 15 new sequences of the people in the training set. Each sequence contained between 20 and 30 frames. Figure 7 shows the estimate of the angle from tracking against the actual angle. In all but one case the tracking succeeded, and a good estimate of the angle is obtained. In one case the models lost track and were unable to recover. This occurred because the local optimisation used from frame to frame failed to converge correctly. A more reliable system would keep multiple hypotheses.

The system currently works off-line, loading sequences from disk. On a 450MHz Pentium III it runs at about 3 frames per second, though so far little work has been done to optimize this.

6. Synthesizing Rotation

Given a single view of a new person, we can find the best model match and determine their head orientation. We can then use the best model to synthesize new views at any orientation that can be represented by the model. If the best matching parameters are $\mathbf{c}$, we use Equation 3 to estimate the angle, θ. Let $\mathbf{r}$ be the residual vector not explained by the rotation model, ie

$$\mathbf{r} = \mathbf{c} - (\mathbf{c}_0 + \mathbf{c}_x \cos(\theta) + \mathbf{c}_y \sin(\theta)) \tag{4}$$

To reconstruct at a new angle, α, we simply use the parameters

$$\mathbf{c}(\alpha) = \mathbf{c}_0 + \mathbf{c}_x \cos(\alpha) + \mathbf{c}_y \sin(\alpha) + \mathbf{r} \tag{5}$$

For instance, Figure 8 shows fitting a model to a roughly frontal image and rotating it. The top example uses a new view of someone in the training set. The lower example is a previously unseen person from the Surrey face database (13).

This only allows us to vary the angle in the range defined by the current view model. To generate significantly different views we must learn the relationship between parameters from one view model and those from another.

7. Coupled-View Appearance Models

Given enough pairs of images taken from different view points, we can build a model of the relationship between the model parameters in one view and those in another. Ideally the images should be taken simultaneously, allowing correlations between changes in expression to be learnt. We have achieved this using a single video camera and a mirror (see Figure 1). A looser model can be built from images taken at different times, assuming a similar expression (typically neutral) is adopted in both.

Let $\mathbf{r}_{ij}$ be the residual model parameters for the object in the i^{th} image in view j, formed from the best fitting parameters by removing the contribution from the angle model (Equation 4). We form the combined parameter vectors $\mathbf{j}_i = (\mathbf{r}_{i1}^T, \mathbf{r}_{i2}^T)^T$. We can then perform a principal component analysis on the set of $\{\mathbf{j}_i\}$ to obtain the main modes of variation of a combined model,

$$\mathbf{j} = \hat{\mathbf{j}} + \mathbf{Pb} \tag{6}$$

Where $\mathbf{P}$ is s matrix of modes and $\mathbf{b}$ is a vector of parameters.

Figure 9 shows the effect of varying the first four of the parameters controlling such a model representing both frontal and profile face appearance. The modes mix changes in identity and changes in expression. For instance mode 3 appears to demonstrate the relationship between frontal and profile views during a smile.

7.1. Predicting New Views

We can use the joint model to generate different views of a subject. We find the joint parameters which generate a frontal view best matching the current target, then use the model to generate the corresponding profile view. Figures 10(a,b) show the actual profile and profile predicted from a new view of someone in the training set. In this case the model is able to estimate the expression (a half smile). Because we only have a limited set of images in which we have non-neutral expressions, the joint model built with them is not good at generalising to new people. To deal with this, we built a second joint model, trained on 100 frontal and profile images taken from the Surrey XM2VTS face database (13). These have neutral expressions, but the image pairs are not taken simultaneously, and the head orientation can vary significantly. However, the rotation effects can be removed using the approach described above, and the model can be used to predict unseen views of neutral expressions. Figures 10(c,d) show the actual profile and profile predicted from a new person (the frontal image is shown in Figure 8). With a large enough training set we would be able to deal with both expression changes and a wide variety of people.

8. Coupled Model Matching

Given two different views of a target, and corresponding models, we can exploit the correlations to improve the robustness of matching algorithms. One approach would be to modify the Active Appearance Model search algorithm to drive the parameters, $\mathbf{b}$, of the joint model, together with the current estimates of pose, texture transformation and 3D orientation parameters. However, the approach we have implemented is to train two independent AAMs (one for the frontal model, one for the profile), and to run the search in parallel, constraining the parameters with the joint model at each step. In particular, each iteration of the matching algorithm proceeds as follows:

- Perform one iteration of the AAM on the frontal model, and one on the profile model, to update the current estimate of $\mathbf{c}_1$, $\mathbf{c}_2$ and the associated pose and texture transformation parameters.
- Estimate the relative head orientation with the frontal and profile models, θ_1, θ_2
- Use Equation 4 to estimate the residuals $\mathbf{r}_1$, $\mathbf{r}_2$
- Form the combined vector $\mathbf{j} = (\mathbf{r}_1^T, \mathbf{r}_2^T)^T$
- Compute the best joint parameters, $\mathbf{b} = \mathbf{P}^T(\mathbf{j} - \hat{\mathbf{j}})$ and apply limits to taste.
- Compute the revised residuals using $(\mathbf{r}_1'^T, \mathbf{r}_2'^T)^T = \hat{\mathbf{j}} + \mathbf{Pb}$
- Use Equation 2 to add the effect of head orientation back in

Note that this approach makes no assumptions about the exact relative viewing angles. If appropriate we can learn the relationship between θ_1 and θ_2 ($\theta_1 = \theta_2 + const$). This could be used as a further constraint. Similarly the relative positions and scales could be learnt.

To explore whether these constraints actually improve robustness, we performed the following experiment. We manually labelled 50 images (not in the original training set), then performed multi-resolution search, starting with the mean model parameters in the correct pose. We ran the experiment twice, once using the joint model constraints described above, once without any constraints (treating the two models as completely independent).

Table 2 summarizes the results. After each search we measure the RMS distance between found points and hand labelled points, and the RMS error per pixel between the model reconstruction and the image (the intensity values are in the range [0,255]). The results demonstrate that in this case the use of the constraints between images improved the performance, but not by a great deal. We would expect that adding stronger constraints, such as that between the angles θ_1,θ_2, and the relative scales and positions, would lead to further improvements.

9. Discussion and Conclusions

We have demonstrated that a small number of view-based statistical models of appearance can represent the face from a wide range of viewing angles. Although we have concentrated on rotation about a vertical axis, rotation about a horizontal axis (nodding) could easily be included (and probably wouldn't require any extra models for rotations of less than $\pm 20^o$). The models can be used to estimate head orientation and to track faces through wide angle changes. We have shown that a linear model can represent the correlations between appearance in two views and that such a model can be used to predict appearance from new viewpoints given a single image of a person.

Such models can be used to constrain search when matching models to two views of an object taken simultaneously. We have treated the parameters of each model as equally important, which in the case above is a reasonable approximation. There may be cases where the views should be weighted differently to allow for their relative utility.

One potential application is in face recognition systems. If a system relies on a single image, it can easily be fooled with a photograph. This can be prevented either by using sequences, or by using two cameras with a sufficiently different viewpoint. The methods described above can be used to determine the expected relationship between the different views, and should be able to distinguish a photograph from a true 3D structure. The joint model has implicitly captured the 3D structure of the object.

Table 2
Comparison between coupled search and independent search

Measure	Frontal Model		Profile Model	
	Coupled	Independent	Coupled	Independent
RMS Point Error (pixels)	4.8 ± 0.5	5.1 ± 0.5	3.3 ± 0.15	3.8 ± 0.3
RMS Texture Error (grey-levels)	7.9 ± 0.25	7.9 ± 0.25	8.3 ± 0.25	8.8 ± 0.4

We anticipate the approach will be useful in many applications, including driving animated avatars, calculating head pose and making face recognition systems more invariant to viewing angle.

References

[1] T. F. 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, Berlin, 1998.

[2] R. Davies, T. Cootes, and C. Taylor. A minimum description length approach to statistical shape modelling. In 17^{th} *Conference on Information Processing in Medical Imaging*, pages 50–63, 2001.

[3] 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.

[4] P. Fua and C. Miccio. From regular images to animated heads: A least squares approach. In H.Burkhardt and B. Neumann, editors, 5^{th} *European Conference on Computer Vision*, volume 1, pages 188–202. Springer, Berlin, 1998.

[5] D. Graham and N. Allinson. Face recognition from unfamiliar views: Subspace methods and pose dependency. In 3^{rd} *International Conference on Automatic Face and Gesture Recognition 1998*, pages 348–353, Japan, 1998.

[6] H.Murase and S. Nayar. Learning and recognition of 3d objects from appearance. *International Journal of Computer Vision*, pages 5–25, Jan. 1995.

[7] M. J. Jones and T. Poggio. Multidimensional morphable models : A framework for representing and matching object classes. *International Journal of Computer Vision*, 2(29):107–131, 1998.

[8] N. Kruger. An algorithm for the learning of weights in discrimination functions using a priori constraints. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19(7):764–768, 1997.

[9] J. Kwong and S. Gong. Learning support vector machines for a multi-view face model. In T. Pridmore and D. Elliman, editors, 10^{th} *British Machine Vison Conference*, volume 2, pages 503–512, Nottingham, UK, Sept. 1999. BMVA Press.

[10] M. La Cascia, S. Sclaroff, and V. Athitsos. Fast, reliable head tracking under varying illumination: An approach based on registration of texture mapped 3d models. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(4):322–336, 2000.

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

[12] T. Maurer and C. von der Malsburg. Tracking and learning graphs and pose on image sequences of faces. In 2^{nd} *International Conference on Automatic Face and Gesture Recognition 1997*, pages 176–181, Los Alamitos, Cal-

ifornia, Oct. 1996. IEEE Computer Society Press.

[13] K. Messer, J. Matas, J. Kittler, J. Luettin, and G. Maitre. Xm2vtsdb: The extended m2vts database. In *Proc. 2nd Conf. on Audio and Video-based Biometric Personal Verification*. Springer Verlag, 1999.

[14] B. Moghaddam and A. Pentland. Face recognition using view-based and modular eigenspaces. In *SPIE*, volume 2277, pages 12–21, 1994.

[15] B. Moghaddam and A. Pentland. Probabilistic visual learning for object recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19(7):696–710, 1997.

[16] F. Pighin, R. Szeliski, and D. Salesin. Resynthesizing facial animation through 3d model-based tracking. In 7^{th} *International Conference on Computer Vision*, pages 143–150, 1999.

[17] S. Romdhani, S. Gong, and A. Psarrou. A multi-view non-linear active shape model using kernel pca. In T. Pridmore and D. Elliman, editors, 10^{th} *British Machine Vison Conference*, volume 2, pages 483–492, Nottingham, UK, Sept. 1999. BMVA Press.

[18] J. Sherrah, S. Gong, and E. Ong. Understanding pose discrimination in similarity space. In T. Pridmore and D. Elliman, editors, 10^{th} *British Machine Vison Conference*, volume 2, pages 523–532, Nottingham, UK, Sept. 1999. BMVA Press.

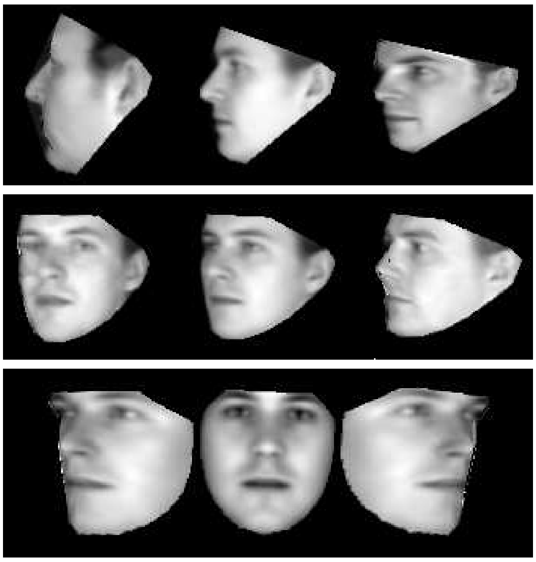
[19] S.Romdhani, A.Psarrou, and S. Gong. On utilising template and feature-based correspondence in multi-view appearance models. In 6^{th} *European Conference on Computer Vision*, volume 1, pages 799–813. Springer, 2000.

[20] M. Turk and A. Pentland. Eigenfaces for recognition. *Journal of Cognitive Neuroscience*, 3(1):71–86, 1991.

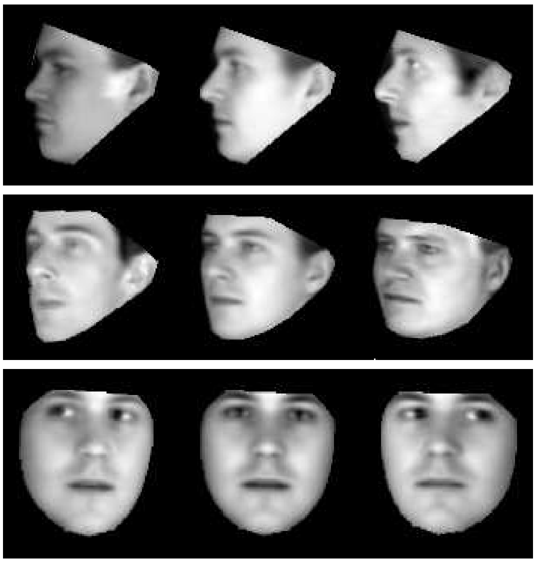
[21] T. Vetter. Learning novel views to a single face image. In 2^{nd} *International Conference on Automatic Face and Gesture Recognition 1997*, pages 22–27, Los Alamitos, California, Oct. 1996. IEEE Computer Society Press.

[22] T. Vetter and V. Blanz. Estimating coloured 3d face models from single images: an example based approach. In H.Burkhardt and B. Neumann, editors, 5^{th} *European Conference on Computer Vision*, volume 2, pages 499–513. Springer, Berlin, 1998.

[23] L. Wiskott, J. Fellous, N. Kruger, and C. der Malsburg. Face recognition by elastic bunch graph matching. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19(7):775–779, 1997.



c_1 varies ± 2 s.d.s



c_2 varies ± 2 s.d.s

Figure 3. First two modes of the face models (top to bottom: profile, half-profile and frontal), changing each parameter by ± 2 standard deviations (s.d.s)

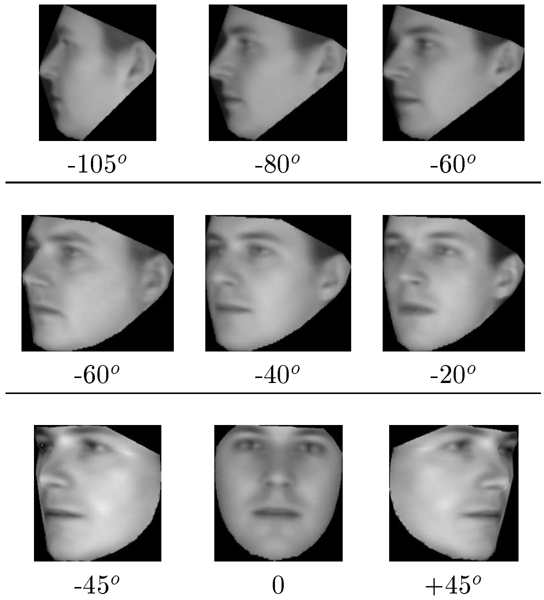


Figure 4. Rotation modes of three face models

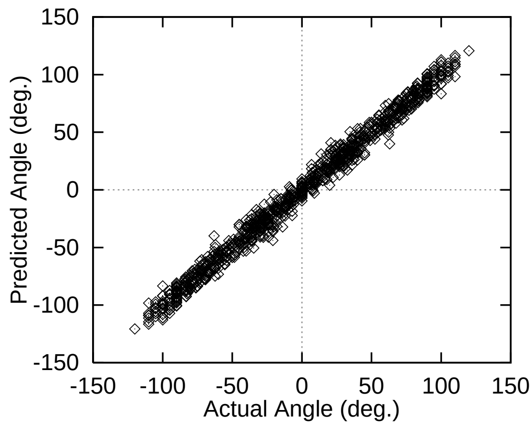


Figure 5. Prediction vs actual angle across training set

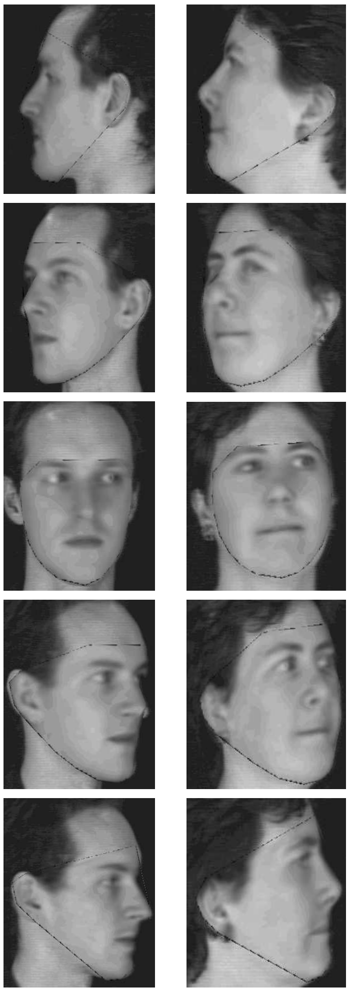


Figure 6. Reconstruction of tracked faces superimposed on sequences

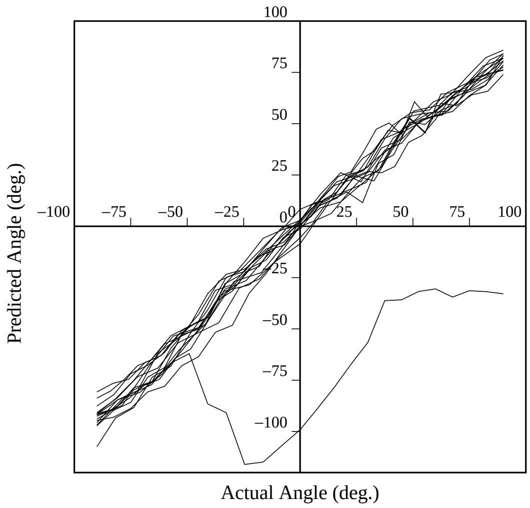


Figure 7. Comparison of angle derived from AAM tracking with actual angle (15 sequences)

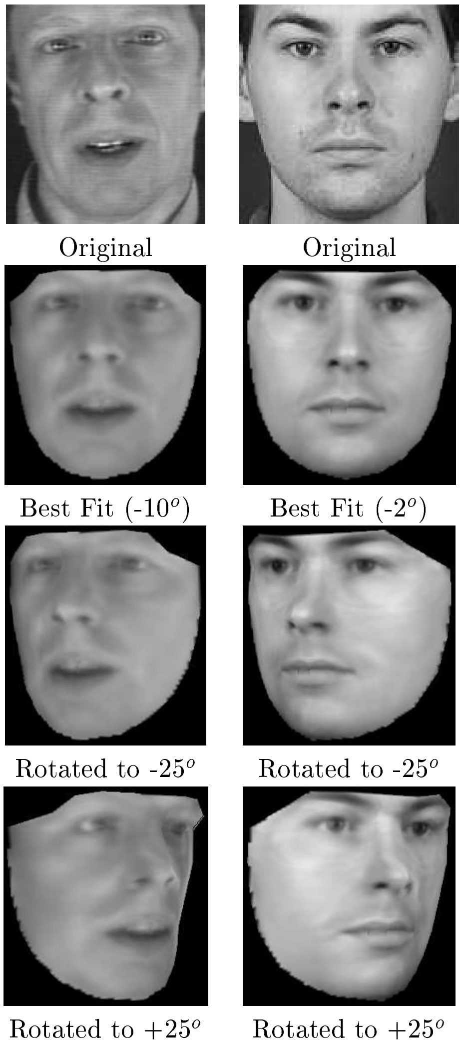


Figure 8. By fitting a model we can estimate the orientation, then synthesize new views

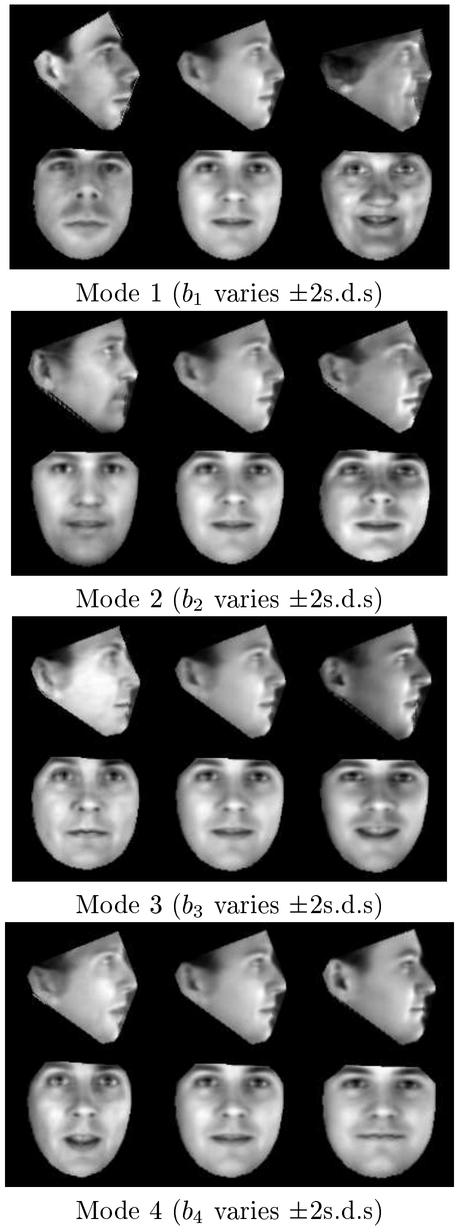
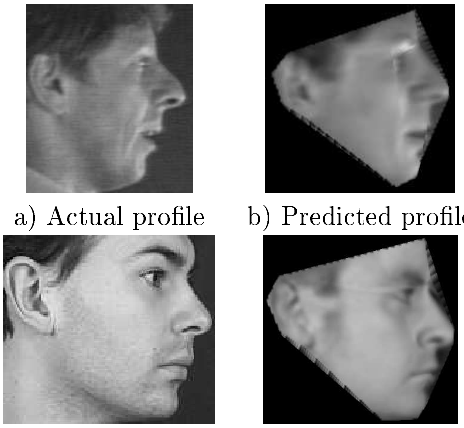


Figure 9. Modes of joint model, controlling frontal and profile appearance



a) Actual profile b) Predicted profile

c) Actual profile d) Predicted profile

Figure 10. The joint model can be used to predict appearance across views (see Fig 8 for frontal view from which the predictions are made)