Appearance factorization for facial expression analysis

Bouchra Abboud, Franck Davoine
Heudiasyc Laboratory
CNRS, University of Technology of Compiègne
{bouchra.abboud, fdavoine}@hds.utc.fr

Abstract

This paper addresses the issue of face representations for facial expression analysis and synthesis. In this context, a global appearance model is used and two bilinear factorization models are subsequently proposed to separate expression and identity factors from the global appearance parameters. A feature extraction technique inspired from the above representations is then proposed which consists in automatically computing the optimal identity and expression components that best adapt to an unknown target face. The proposed representation can be seen as an alternative to the costly AAM gradient matrix construction and iterative search and is exploited in the context of facial expression control. Results are compared with the ones obtained using bilinear factorization and linear regression in the space of AAM parameters.

1 Introduction

Humans are able to communicate in a variety of ways, besides the use of words, including face gestures and facial expressions. As a matter of fact the idiom "poker face" evokes an attitude of blank expression to prevent detection of intent which suggests that facial expressions constitute an essential modality in human communication. Furthermore, Ekman and Friesen postulated that six basic emotional categories are universally recognized namely: joy, sadness, anger, disgust, fear and surprise [4]. Several other emotions and many combinations of emotions have been studied but remain unconfirmed as universally distinguishable.

This paper addresses the issue of face representations for facial expression analysis and synthesis. In this context, a global appearance model is used and two bilinear factorization models are subsequently proposed to separate expression and identity factors from the global appearance parameters. A feature extraction technique inspired from the above representations is then proposed which consists in automatically computing the optimal identity and expression components that best adapt to an unknown target face. This representation is based on the fact that SVD and PCA are closely related and will be referred to as Factorized Appearance Model (FAM). It can be seen as an alternative to the costly Active Appearance Model (AAM) gradient matrix construction and iterative search.

Facial expressions control is achieved through replacement of the extracted expression factors. Both bilinear and FAM representations yield very interesting synthesis performances in terms of visual quality of the synthetic faces. Indeed, synthetic open mouths

BMVC 2004 doi:10.5244/C.18.53

reconstruction either with or without teeth apparition is of better quality with bilinear and FAM based synthesis than with linear regression based synthesis [1], in the space of AAM parameters.

2 Face appearance representation

In the context of face appearance representation, the Active Appearance Model [2] is a powerful tool allowing to extract from any unknown target face, a set of appearance parameters coding a synthetic face similar to the target (in terms of minimum texture error). AAM uses Principal Component Analysis to model each of the shape and texture variations seen in a training set. Faces are aligned using a generalized Procrustes analysis and illumination is normalized via pixel grey level scaling and translation. The global appearance model is built by performing a separate PCA on each of the normalized shapes and textures and one further PCA on the combined principal shapes and textures.

An alternative representation of the combined shape and texture data consists in directly combining the normalized training shapes and textures. A single PCA is then performed on the combined vectors. The training shapes are properly weighed to compensate for the difference between pixel position and intensity values.

$$\mathbf{c}_j = \mathbf{Q}_{st}^T \begin{pmatrix} \mathbf{g}_j - \overline{\mathbf{g}} \\ \mathbf{W}_{st}(\mathbf{s}_j - \overline{\mathbf{s}}) \end{pmatrix} = \mathbf{Q}_{st}^T \tilde{\mathbf{h}}_j \quad \text{and} \quad \tilde{\mathbf{h}}_j \simeq \mathbf{Q}_{st} \mathbf{c}_j \tag{1}$$

$\mathbf{Q}_{st}$ is a truncated matrix of eigenvectors describing the principal modes of combined shape and texture variations, and $\mathbf{c}_j$ is a vector of appearance parameters simultaneously controlling the synthesized shape and texture. In practice, for a training set of 375 different expressive faces extracted from the CMU database [5], the dimension of $\mathbf{c}_j$ varies between 120 and 170 depending on the chosen representation (three-PCA or one-PCA).

Furthermore, in order to allow pose displacement of the model, it is necessary to add to the appearance vector $\mathbf{c}_j$ a pose vector $\mathbf{p}_j$ controlling scale, orientation and position of the appearance model in the image support.

The active appearance model automatically adjusts vectors $\mathbf{c}$ and $\mathbf{p}$ to a target face by minimizing the squared norm of a residual image $\mathbf{r}(\mathbf{c},\mathbf{p})$ which is the texture difference between the synthesized face and the corresponding region of the image it covers. The optimization scheme used in this paper is based on the first order Taylor expansion described in [2]. In this context, matrices $\mathbf{R}_a$ and $\mathbf{R}_t$ are computed, establishing the linear relationships:

$$\delta(\mathbf{c}) = -\mathbf{R}_a\, \mathbf{r}(\mathbf{c},\mathbf{p}) \;\text{ and }\; \delta(\mathbf{p}) = -\mathbf{R}_t\, \mathbf{r}(\mathbf{c},\mathbf{p}) \tag{2}$$

A first order Taylor development gives the following solutions:

$$\mathbf{R}_a = \left(\frac{\partial \mathbf{r}^T}{\partial \mathbf{c}} \frac{\partial \mathbf{r}}{\partial \mathbf{c}}\right)^{-1} \frac{\partial \mathbf{r}^T}{\partial \mathbf{c}} \text{ and } \mathbf{R}_t = \left(\frac{\partial \mathbf{r}^T}{\partial \mathbf{p}} \frac{\partial \mathbf{r}}{\partial \mathbf{p}}\right)^{-1} \frac{\partial \mathbf{r}^T}{\partial \mathbf{p}} \tag{3}$$

Computing $\frac{\partial \mathbf{r}}{\partial \mathbf{c}}$ and $\frac{\partial \mathbf{r}}{\partial \mathbf{p}}$ is a heavy procedure conducted by numeric differentiation [2]. An iterative model refinement procedure is then used to drive the appearance model towards the actual target face. In the following, the appearance and pose vectors obtained by this optimization procedure are denoted respectively as $\mathbf{c}_{op}$ and $\mathbf{p}_{op}$.

The $\mathbf{c}_{op}$ vector controls simultaneously the face shape and texture including information about the reconstructed face identity and facial expression [3]. However, it is interesting to extract from the global appearance vector the corresponding expression factors. Indeed, adequate classification and control of such factors would allow to perform facial expression recognition and synthesis. In this perspective, we propose to model the mapping from expression and identity parameters to natural faces using a bilinear factorization model.

3 Bilinear modelling

Bilinear models are two-factor models with the property that their outputs are linear in either factors when the other is held constant. They provide rich factor interactions by allowing factors to modulate each other's contributions multiplicatively.

Two types of bilinear models already proposed in [6] are described in this section, namely the symmetric bilinear model and the asymmetric bilinear model. The general symmetric model allows to represent, in the present work, the interaction between expression $\mathbf{a}_{op}$ and identity $\mathbf{b}_{op}$ factors for a given appearance vector $\mathbf{c}_{op}$ coding a face of unknown identity and expression. The simpler asymmetric model is expression specific and requires expression to be known in advance.

3.1 The symmetric bilinear model

A bilinear symmetric model represents the interaction between expression $\mathbf{a}_{op}$ and identity $\mathbf{b}_{op}$ factors for a given observation $\mathbf{c}_{op}$ according to:

$$\mathbf{c}_{op}(k) = \mathbf{a}_{op}^T \mathbf{w}_k \mathbf{b}_{op} \tag{4}$$

where $\mathbf{c}_{op}(k)$ represents the k^{th} component of $\mathbf{c}_{op}$ and $\mathbf{w}_k$ is an expression and identity independent matrix characterizing their interaction.

The training set is composed of 70 face images with 10 different persons showing each of the 7 basic facial expressions extracted from the CMU expressive face database [5]. An observation matrix $\mathbf{Y}$ is then built by stacking the corresponding appearance vectors. Each column of $\mathbf{Y}$ contains the AAM appearance vectors of a specific person with different expressions (Neutral, Anger, Disgust, Fear, Joy, Surprise, Sadness (Unhappiness)) whereas each row contains the appearance vectors of all the 10 persons showing a specific expression.

$$\mathbf{Y} = \begin{bmatrix} \mathbf{c}^{\mathbf{n1}} & \dots & \mathbf{c}^{\mathbf{n10}} \\ \mathbf{c}^{\mathbf{a1}} & \dots & \mathbf{c}^{\mathbf{a10}} \\ \mathbf{c}^{\mathbf{d1}} & \dots & \mathbf{c}^{\mathbf{d10}} \\ \mathbf{c}^{\mathbf{f1}} & \dots & \mathbf{c}^{\mathbf{f10}} \\ \mathbf{c}^{\mathbf{j1}} & \dots & \mathbf{c}^{\mathbf{j10}} \\ \mathbf{c}^{\mathbf{s1}} & \dots & \mathbf{c}^{\mathbf{s10}} \\ \mathbf{c}^{\mathbf{u1}} & \dots & \mathbf{c}^{\mathbf{u10}} \end{bmatrix} \tag{5}$$

Training a symmetric bilinear model is achieved with an iterative procedure described in [6] which decomposes the observation matrix into a set of 7 expression factors $\mathbf{a}^{\mathbf{e}}$, 10 identity factors $\mathbf{b}^{\mathbf{i}}$ and weight matrices $\mathbf{w}_k$. In compact matrix form this can be written as:

$$\mathbf{Y} = \left[\mathbf{W}^{VT}\mathbf{A}\right]^{VT}\mathbf{B} \tag{6}$$

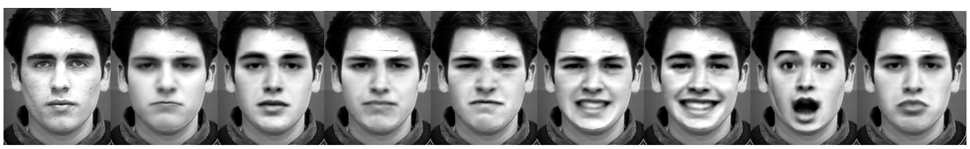
where $\mathbf{A}$ and $\mathbf{B}$ represent the matrices of the stacked expression $\mathbf{a^e}$ and identity $\mathbf{b^i}$ factors. Factors $\mathbf{a^e}$ and $\mathbf{b^i}$ are essentially obtained by repeatedly applying SVD on successively reordered observation matrices to alternate the roles of the expression and identity factors.

We set the dimensionality of expression factors $\mathbf{a^e}$ to be equal to the number of expressions in the training set $I = 7$ to allow maximum expressiveness, and the dimensionality of the identity factors $\mathbf{b^i}$ to be $J = 10$ which corresponds to the maximum number of training identities.

To perform facial expression synthesis on an unknown target face encoded by $\mathbf{c}_{op}$, with an undetermined identity and expression, the bilinear symmetric model is iteratively adapted to the face thus allowing to extract the corresponding expression $\mathbf{a}_{op}$ and identity $\mathbf{b}_{op}$ factors [6].

To synthesize any novel expression "**e'**" while keeping identity intact, an artificial appearance parameter is built by combining the extracted identity factor $\mathbf{b}_{op}$ with the desired expression factor learned from the training set $\mathbf{a^{e'}}$.

Facial expression synthesis using a symmetric bilinear model, on an unknown face with undetermined expression is shown on figure 1.



a b c d e f g h i

Figure 1: Symmetric bilinear expression synthesis. a: Target face b: Symmetric bilinear model fitting. Synthesis of c: Neutral, d: Anger, e: Disgust, f: Fear, g: Joy, h: Surprise, and i: Sadness expression.

3.2 The asymmetric bilinear model

The expression specific asymmetric bilinear model decomposes an appearance vector $\mathbf{c}^{\mathbf{e}}_{op}$ coding a face of known expression "**e**" and unknown identity into an expression specific linear mapping $\mathbf{W^e}$ and an identity factor $\mathbf{b}_{op}$.

$$\mathbf{c}^{\mathbf{e}}_{op} = \mathbf{W^e}\mathbf{b}_{op} \tag{7}$$

Training an asymmetric, expression-specific, bilinear model consists in computing the 7 expression specific linear mappings $\mathbf{W^e}$ and 10 identity factors $\mathbf{b^i}$ which minimize the total squared error between the actual and reconstructed observations of the training face set using SVD decomposition [6]. In compact matrix form it can be written as:

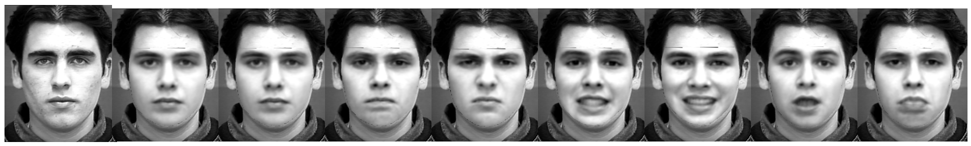
$$\mathbf{Y} = \mathbf{WB} \tag{8}$$

where $\mathbf{W}$ is a matrix containing the stacked expression-specific linear mappings $\mathbf{W^e}$ and $\mathbf{B}$ represents the matrix of the stacked identity factors $\mathbf{b^i}$.

To synthesize a novel expression "**e'**" while keeping identity intact, the extracted identity factor $\mathbf{b}_{op}$ is combined with the desired previously learned weights matrix $\mathbf{W^{e'}}$:

$$\mathbf{c}_{op}^{\mathbf{e'}} = \mathbf{W}^{\mathbf{e'}} \mathbf{b}_{op} \tag{9}$$

Facial expression synthesis, using an asymmetric bilinear model, on an unknown neutral face is shown on figure 2.



a b c d e f g h i

Figure 2: Asymmetric bilinear expression synthesis. a: Target face b: Asymmetric bilinear model fitting. Synthesis of c: Neutral, d: Anger, e: Disgust, f: Fear, g: Joy, h: Surprise, and i: Sadness expression.

It should be noted if we compare figures 1.b and 2.b that asymmetric bilinear model fitting to an unknown face gives a more accurate representation of the target facial expression since this (neutral) expression is supposed to be *a priori* known. Symmetric bilinear fitting shown on figure 1.b for the same target face simulates an expression close to anger, however the extracted identity factors are correct since when a neutral expression is imposed, the similitude of the synthetic output with the target face increases as shown on figure 1.c.

4 A factorized face appearance representation

The bilinear appearance factorization described above offers interesting properties for face and facial expression analysis and control. It allows to extract from any target face with known or unknown expression an identity-specific factor exclusively coding identity, and an expression-specific factor exclusively coding expression. The extraction of such factors offers undeniable advantages for synthesis which becomes immediate.

However, this approach relies on the factorization of the appearance parameters extracted by AAM search, and thus depends strongly on the quality of AAM adaptation. Furthermore AAM search is a relatively heavy procedure which requires offline construction of a gradient matrix estimated by numeric differentiation [2]. Indeed, building matrix $\frac{\partial r}{\partial c}$ in equation (2) implies adding systematic perturbations to every appearance mode (there are 120 or 170 appearance modes depending on the chosen representation) and performing an average over the 375 faces of the training set.

This of course is time consuming and therefore it is of particular interest to bypass the $\mathbf{R}_a$ learning procedure and the iterative AAM adaptation in order to extract from any unknown target face a set of sufficiently representative factorized appearance parameters.

In this context, we propose the Factorized Appearance Model (FAM) which includes the advantages of the bilinear model in terms of relevant information separation, and can be seen as an interesting alternative to AAM in the sense that it doesn't require iterative appearance extraction and gradient matrix learning.

4.1 The asymmetric FAM

The asymmetric FAM proposed here is inspired from the asymmetric bilinear model which requires one factor to be known in advance, and from the one-PCA AAM (equation (1)). In this paragraph, we will first address asymmetric FAM building and adaptation using an iterative procedure similar to AAM.

Immediate FAM fitting to an unknown target face is then addressed and the results are compared with iterative FAM fitting and with AAM fitting for an AAM model built using the same training set as FAM.

On the one hand, in the one-PCA AAM, each centered training observation $\tilde{\mathbf{h}}_j$ is obtained by direct concatenation of properly extracted, weighted and centered shape $\mathbf{W}_{st}(\mathbf{s}_j - \overline{\mathbf{s}})$ and texture $\mathbf{g}_j - \overline{\mathbf{g}}$.

$$\tilde{\mathbf{h}}_j = \begin{pmatrix} \mathbf{g}_j - \overline{\mathbf{g}} \\ \mathbf{W}_{st}(\mathbf{s}_j - \overline{\mathbf{s}}) \end{pmatrix} \quad (10)$$

The appearance vector $\mathbf{c}_j$ is subsequently computed through PCA of the training observations (see equation (1)).

On the other hand, the asymmetric bilinear model suggests that an observation vector, containing embedded identity and expression information, can be decomposed into an identity factor and an expression specific linear mapping using SVD decomposition.

Let $\tilde{\mathbf{H}}$ be the matrix of all the centered training observations stacked in such a way that each column corresponds to the 7 facial expressions of a given identity, whereas each row corresponds to all the identities showing a given expression.

$$\tilde{\mathbf{H}} = \begin{bmatrix} \tilde{\mathbf{h}}^{\mathbf{n1}} & \dots & \tilde{\mathbf{h}}^{\mathbf{n10}} \\ \tilde{\mathbf{h}}^{\mathbf{a1}} & \dots & \tilde{\mathbf{h}}^{\mathbf{a10}} \\ \tilde{\mathbf{h}}^{\mathbf{d1}} & \dots & \tilde{\mathbf{h}}^{\mathbf{d10}} \\ \tilde{\mathbf{h}}^{\mathbf{f1}} & \dots & \tilde{\mathbf{h}}^{\mathbf{f10}} \\ \tilde{\mathbf{h}}^{\mathbf{j1}} & \dots & \tilde{\mathbf{h}}^{\mathbf{j10}} \\ \tilde{\mathbf{h}}^{\mathbf{s1}} & \dots & \tilde{\mathbf{h}}^{\mathbf{s10}} \\ \tilde{\mathbf{h}}^{\mathbf{u1}} & \dots & \tilde{\mathbf{h}}^{\mathbf{u10}} \end{bmatrix} \quad (11)$$

SVD decomposition gives:

$$\tilde{\mathbf{H}} = \tilde{\mathbf{Q}}\tilde{\mathbf{B}} \quad (12)$$

where rows of $\tilde{\mathbf{B}}$ are the eigenvectors of the covariance matrix $\tilde{\mathbf{H}}^T\tilde{\mathbf{H}}$ and $\tilde{\mathbf{Q}}$ contains the corresponding principal components. This last equation reminds the matrix form of the asymmetric bilinear model (Equation (8)), and therefore we can write for a given observation with expression "**e**" and identity "**i**":

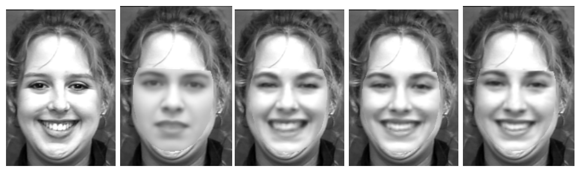
$$\tilde{\mathbf{h}}^{\mathbf{ei}} = \tilde{\mathbf{Q}}^{\mathbf{e}}\tilde{\mathbf{b}}^{\mathbf{i}} \quad (13)$$

Equation (13) reminds the one-PCA AAM (equation (1)), if we write $\tilde{\mathbf{Q}}^{\mathbf{e}}$ as $\tilde{\mathbf{Q}}_{st}^T$ and $\tilde{\mathbf{b}}^{\mathbf{i}}$ as $\mathbf{c}_j$.

Projecting a face of know expression "**e**" on the subspace spanned by $\tilde{\mathbf{Q}}^{\mathbf{e}}$ gives the appearance vector $\tilde{\mathbf{b}}^{\mathbf{i}}$ which is also the identity factor and thus allows to factorize the appearance of an observation. Hence this model will be addressed to as the Factorized Appearance Model (FAM).

AAM search can automatically adjust the appearance vector $\mathbf{c}_{op}$ to an unknown target face according to the procedure described in [2]. This same search algorithm can be applied to FAM in order to adapt the identity-specific appearance parameters to an unknown target face with a known expression "$\mathbf{e}$". The optimal appearance vector obtained $\tilde{\mathbf{b}}_{op}$ corresponds to the identity factor of the target face. This iterative search procedure requires much less time than standard AAM search due to the lower number of components in $\tilde{\mathbf{b}}_{op}$ when compared to $\mathbf{c}_{op}$ (10 versus 120 to 170).

Expression specific asymmetric FAM iterative adaptation to an unknown joyful face is shown in figure 3.d.



a b c d e

Figure 3: a: Unknown joyful face. b: AAM initializing c: AAM fitting. d: Joy-specific asymmetric FAM iterative fitting. e: Joy-specific asymmetric FAM immediate fitting.

However, it should be noted that AAM search seeks to minimize the difference between the target and reconstructed textures of a given face. The target texture is acquired by sampling the region of the target face under the reconstructed shape at each iteration. As the model adapts, the reconstructed shape tends to the real shape and landmark points take their true positions around the main facial features. Consequently, upon convergence, the target texture is extracted using a relatively correct shape and can be considered as representative of the true shape-free texture.

Nevertheless, face shape varies mainly with facial expressions and very little with identities. In classical AAM search, the model is initialized with a mean appearance (identity and expression) and the shape varies along the iterations to match the real shape of the target face. However, in the case of asymmetric expression-specific FAM search, the model is initialized with the real (*a priori* known) expression of the target face and a mean identity. The search algorithm gives the optimal appearance parameters which are also the identity factors of the target face and the expression is not changed along the iterations.

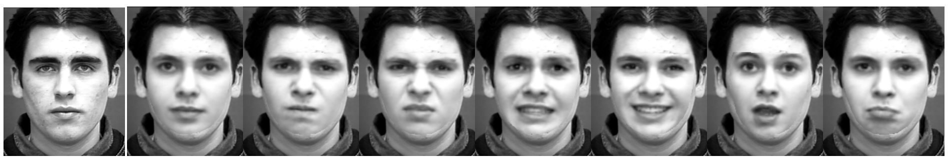
Thus it is reasonable to consider that starting from a correct pose through proper eye detection for example, the initial shape is close enough to the real target shape and hence it is sufficient to sample the texture under this shape to obtain an observation vector $\tilde{\mathbf{h}}^{\mathbf{e}}$ with known expression "$\mathbf{e}$". Projecting this observation vector onto the corresponding expression-specific eigenspace $\mathbf{Q}^{\mathbf{e}}$ allows to immediately extract the corresponding identity-specific appearance vector:

$$\tilde{\mathbf{b}}_{op} = [\tilde{\mathbf{Q}}^{\mathbf{e}}]^{+}\tilde{\mathbf{h}}^{\mathbf{e}} \tag{14}$$

For the unknown joyful face of figure 3.a immediate asymmetric FAM fitting through extraction of the corresponding identity-specific appearance vector is shown on figure 3.e.

Standard AAM adaptation on the same target face and for a model built on the same training set is also shown on figure 3.c for comparison. These images confirm that the visual qualities of the 3 adaptations are comparable. This result is very interesting in the sense that model fitting to an unknown target face becomes immediate.

Furthermore, the immediately extracted appearance vector $\tilde{\mathbf{b}}_{op}$ is also an identity-specific factor having very interesting properties for synthesis. Indeed, combining the extracted identity factor $\mathbf{b}_{op}$ with a given expression mapping $\mathbf{Q}^{\mathbf{e'}}$ allows to construct an artificial face having the same identity and any desired expression "**e'**". This procedure is immediate and no iterations are required. Facial expression synthesis on an unknown neutral face is shown on figure 4.



a b c d e f g h

Figure 4: Asymmetric FAM expression synthesis. a: Target neutral face. Synthesis of b: Neutral, c: Anger, d: Disgust, e: Fear, f: Joy, g: Surprise, and h: Sadness expression.

However, the expression-specific asymmetric FAM requires *a priori* knowledge of the facial expression of a target face in order to allow immediate identity-specific appearance parameters extraction. Since this information is not always available, we propose a symmetric factorized appearance model which bypasses this constraint.

4.2 The symmetric FAM

The symmetric bilinear model represents the interaction between expression $\mathbf{a}^{\mathbf{e}}$ and identity $\mathbf{b}^{\mathbf{i}}$ factors for a given appearance vector $\mathbf{c}^{\mathbf{ei}}$ coding a face with known expression "**e**" and identity "**i**".

$$\mathbf{c}^{\mathbf{ei}}(k) = \mathbf{a}^{\mathbf{e}T}\mathbf{w}_k\mathbf{b}^{\mathbf{i}} \tag{15}$$

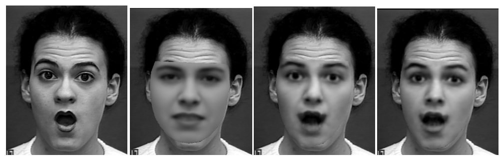
where $\mathbf{c}^{\mathbf{ei}}(k)$ represents the k^{th} component of $\mathbf{c}^{\mathbf{ei}}$ and $\mathbf{w}_k$ is an expression-independent, identity-independent linear mapping characterizing their interaction. Extending this concept to the matrix of observations $\tilde{\mathbf{H}}$ a symmetric FAM can be built iteratively using SVD decomposition. Of course, SVD decomposition is performed using its relationship with PCA which reduces the cost of computation since it allows the decomposition to be performed on a small (10x10) covariance matrix instead of the large (41419x10) observation matrix $\tilde{\mathbf{H}}$.

$$\tilde{\mathbf{H}} = [\tilde{\mathbf{Q}}^{VT}\tilde{\mathbf{A}}]^{VT}\tilde{\mathbf{B}} \tag{16}$$

Symmetric FAM allows to extract from an observation with unknown expression and identity an expression factor $\tilde{\mathbf{a}}_{op}$ and an identity factor $\tilde{\mathbf{b}}_{op}$ according to the iterative procedure described in [6]. As with asymmetric FAM adaptation, we can consider adapting the model by extracting the corresponding expression and identity parameters without using the gradient matrix learning step required by standard AAM search.

At each iteration, the target texture is re-sampled and the residual error is computed. If this error decreases then the identity and expression factors extracted at this iteration are retained, otherwise they are re-estimated.

Symmetric FAM search on an unknown face with unknown expression is illustrated on figure 5.



a b c d

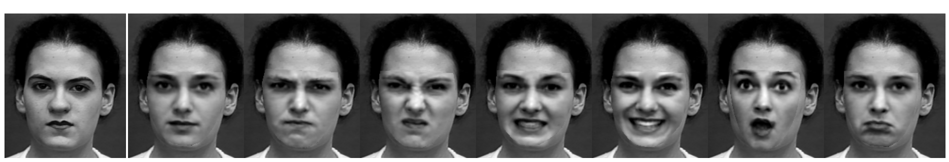
Figure 5: a: Unknown face with undetermined expression. b: Symmetric FAM initialization with mean identity and mean expression. c: Symmetric FAM immediate fitting. d: AAM fitting.

Adapting the symmetric FAM to an unknown target face with undetermined expression allows to extract the corresponding identity and expression factors and constitutes a very interesting face representation for facial expression analysis in the sense that it bypasses the costly gradient matrix construction.

Combining the extracted identity factor $\mathbf{b}_{op}$ with any desired expression factor $\mathbf{a}^{\mathbf{e'}}$ allows to construct an artificial face having the same identity and any desired expression "**e'**".

$$\mathbf{h}^{\mathbf{e'}}(k) = \mathbf{a}^{\mathbf{e'}}\mathbf{w}_k\mathbf{b}_{op} \tag{17}$$

Facial expression synthesis on an unknown target face of unknown expression is shown on figure 6.

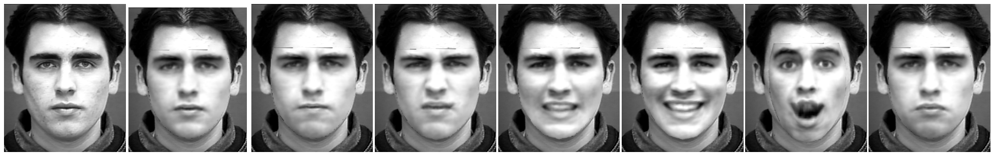


a b c d e f g h

Figure 6: Symmetric FAM expression synthesis. a: Unknown target face with undetermined expression. Synthesis of b: Neutral, c: Anger, d: Disgust, e: Fear, f: Joy, g: Surprise, and h: Sadness expression.

5 Comparison and conclusion

To compare the visual quality of the synthetic faces obtained either with bilinear modelling or with FAM we use a classical linear regression model correlating facial expression intensity to appearance parameters [1]. For the same training set synthesis results for an unknown neutral face are shown on figure 7.



a b c d e f g h

Figure 7: Linear expression synthesis. a: Neutral target face. b: AAM adaptation. Synthesis of c: Anger, d: Disgust, e: Fear, f: Joy, g: Surprise, and h: Sadness expression.

We conclude that the use of bilinear modelling for expression synthesis enhances photorealism since subtle face variations are better represented than with linear expression modelling. In addition, the symmetric bilinear model allows to modify facial expression on an unseen target face without *a priori* knowledge of the shown expression. Facial expression recognition performance is also boosted by bilinear modelling since it allows to extract a set of expression-specific factors. Indeed, a correct recognition rate of 83.33% is achieved with an identity-specific asymmetric bilinear model. Recognition is performed using the maximum number of votes obtained for each expression factor extracted using each of the 10 training identities. Euclidian distance based classification of AAM parameters projected in fisherspace yields only 67.59% correct recognition using the same training and test images, and 30 trained humans achieved a mean recognition rate of 79.36% for the same test faces.

References

[1] B. Abboud, F. Davoine, and M Dang. Expressive face recognition and synthesis. In *IEEE CVPR workshop on Computer Vision and Pattern Recognition for Human Computer Interaction*, Madison, U.S.A., June 2003.

[2] T.F. Cootes and P. Kittipanya-ngam. Comparing variations on the active appearance model algorithm. In *British Machine Vision Conference*, pages 837–846, Cardiff University, September 2002.

[3] G.J. Edwards, T.F. Cootes, and C.J. Taylor. Face recognition using active appearance models. In *IEEE European Conference on Computer Vision*, volume 2, pages 581–695.

[4] P. Ekman. *Facial Expressions*, chapter 16 of Handbook of Cognition and Emotion, T. Dalgleish and M. Power. John Wiley & Sons Ltd., 1999.

[5] T. Kanade, J. Cohn, and Y.L. Tian. Comprehensive database for facial expression analysis. In *Intl. Conference on Automatic Face and Gesture Recognition*, pages 46–53, Grenoble, France, March 2000.

[6] J. B. Tenenbaum and W. T. Freeman. Separating style and content with bilinear models. *Neural Computation*, 12:1247–1283, 2000.