Expressive face recognition and synthesis

Bouchra Abboud, Franck Davoine, Mô Dang
Heudiasyc Laboratory.
CNRS, University of Technology of Compiègne.
BP 20529, 60205 COMPIEGNE Cedex, FRANCE.

Abstract

Facial expression interpretation, recognition and analysis is a key issue in visual communication and man to machine interaction. In this paper, we present a technique for extracting appearance parameters from a natural image or video sequence, which allow reproduction of natural looking expressive synthetic faces. This technique was used to perform face synthesis and tracking in video sequences as well as facial expression recognition and control.

1 Introduction

Natural human-machine interaction is becoming an active and important research area. Adequate feedback like speech, facial expression and body gestures is an essential component of such interaction since these communicative events satisfy certain communication expectations in human-human interaction. Furthermore, there has been an increased interest in the last years in trying to introduce human to human communication modalities into human-computer interaction. This includes a number of approaches like speech recognition, eye tracking, data fusion and emotion understanding through facial expression recognition. Consistent machine interaction with the received and processed multi-modal stimuli will be conveyed through a photo-realistic virtual face, voice, body and gestures in order to simulate a humane machine. Possible applications are tourism, cultural heritage, eCommerce, technology enhanced learning, multimedia production, handicap, cognitive robots, vehicle on-board safety systems, etc.

The human face comprises major information about identity and emotion. It constitutes a source of many informative social signs [1] and improves the response to communication expectations. For instance, lip-reading helps to understand speech, gaze patterns as well as head gestures like nodding or frowning direct the attention and regulate the conversation, whereas facial expressions communicate information about the feelings or mental states of other dialogue participants.

Regarding emotions, previous works showed that in nearly all cultures, facial expressions of six basic emotional categories are universally recognized, namely: joy, sadness, anger, disgust, fear and surprise [2]. Several other emotions and many combinations of emotions have been studied but remain unconfirmed as universally distinguishable. Thus, most of the research up to now has been oriented towards detecting these six universal expressions [3, 4, 5, 6].

In this paper we will address the issue of face tracking, video realistic face generation and facial expression synthesis, control and recognition using statistical active facial appearance models for face analysis [7].

2 Active facial appearance models

2.1 Model description

It has been shown that the active appearance model [7] is a powerful tool for face synthesis and tracking. It uses Principal Component Analysis to model both shape and texture variations seen in a training set of visual objects. After computing the mean shape $\bar{\mathbf{s}}$ and aligning all shapes from the training set by means of a Procrustes analysis, the statistical shape model is given by:

$$\mathbf{s}_i = \bar{\mathbf{s}} + \Phi_s \mathbf{b}_{si} \tag{1}$$

where $\mathbf{s}_i$ is the synthesized shape, Φ_s is a truncated matrix describing the principal modes of shape variations in the training set and $\mathbf{b}_{si}$ is a vector that controls the synthesized shape.

It is then possible to warp textures from the training set of faces onto the mean shape $\bar{\mathbf{s}}$ in order to obtain shape-free textures. Similarly, after computing the mean shape-free texture $\bar{\mathbf{g}}$ and normalizing all textures from the training set relatively to $\bar{\mathbf{g}}$ by scaling and offset of the luminance values, the statistical texture model is given by:

$$\mathbf{g}_i = \bar{\mathbf{g}} + \Phi_t \mathbf{b}_{ti} \tag{2}$$

where $\mathbf{g}_i$ is the synthesized shape-free texture, Φ_t is a truncated matrix describing the principal modes of texture variations in the training set and $\mathbf{b}_{ti}$ is a vector that controls the synthesized shape-free texture.

By combining the training shape and texture vectors $\mathbf{b}_{si}$ and $\mathbf{b}_{ti}$ and applying further PCA the statistical appearance model is given by:

$$\mathbf{s}_i = \bar{\mathbf{s}} + Q_s \mathbf{c}_i \tag{3}$$

$$\mathbf{g}_i = \bar{\mathbf{g}} + Q_t \mathbf{c}_i \tag{4}$$

where Q_s and Q_t are truncated matrices describing the principal modes of combined appearance variations in the training set, and $\mathbf{c}_i$ is a vector of appearance parameters simultaneously controling both shape and texture.

Given the parameter vector $\mathbf{c}_i$, the corresponding shape $\mathbf{s}_i$ and shape-free texture $\mathbf{g}_i$ can be computed respectively using equations (3) and (4). The reconstructed shape-free texture is then warped onto the reconstructed shape in order to obtain the full appearance of a face. Furthermore, in order to allow pose displacement of the model, it is necessary to add to the appearance parameter vector $\mathbf{c}_i$ a pose parameter vector $\mathbf{p}_i$ allowing control of scale, orientation and position of the synthesized face.

While a couple of appearance parameter vector $\mathbf{c}$ and pose parameter vector $\mathbf{p}$ represents a face, the active appearance model can automatically adjust those parameters to a target face [8], by minimizing a residual image $\mathbf{r}(\mathbf{c}, \mathbf{p})$ which is the texture difference between the synthesized face and the corresponding mask of the image it covers. For this purpose, a set of training residual images are computed by displacing the appearance and pose parameters whithin allowable limits. These residuals are then used to compute matrices R_a and R_t establishing the linear relationships $\delta(\mathbf{c}) = -R_a\ \mathbf{r}(\mathbf{c}, \mathbf{p})$ and $\delta(\mathbf{p}) = -R_t\ \mathbf{r}(\mathbf{c}, \mathbf{p})$ between the parameter displacements and the corresponding residuals, so as to minimize $|\mathbf{r}((\mathbf{c}, \mathbf{p}) + \delta(\mathbf{c}, \mathbf{p}))|^2$. An iterative model refinement procedure [8] is then used to drive the appearance model towards the actual face in the image.

In the following, the appearance and pose parameters obtained by this optimization procedure will be denoted respectively as $\mathbf{c}_{op}$ and $\mathbf{p}_{op}$.

2.2 Experimental setup

The appearance model is built using the CMU expressive face database [9]. Each sequence of this database contains ten to twenty images, beginning with a neutral expression and ending with a high magnitude expression.

a b c

Figure 1: a: Target (original face). b: Model initialization. c: Iterative model refinement until convergence to the target face.

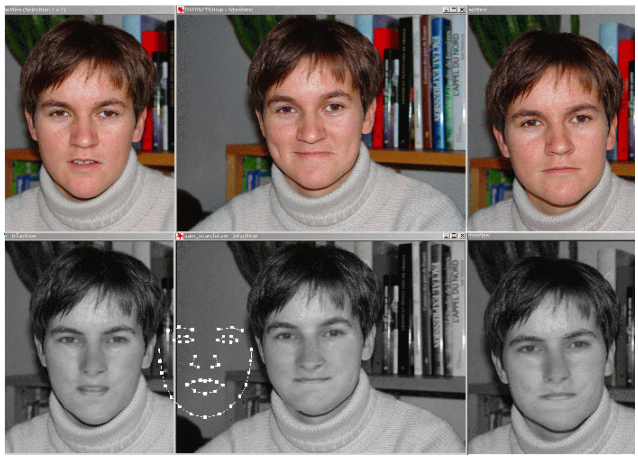

Figure 2: Upper row : original unseen faces. Lower row: reconstructed synthetic faces inserted onto the original ones (the shape of the facial mask represented by the model is illustrated by the set of white points).

We selected 338 frontal still face images composed of 26 neutral expression faces, 26 moderate and 26 high magnitude *anger, disgust, fear, joy, surprise and sadness* expressions. Each moderate expression image has been chosen manually by extracting an intermediate frame from the video sequence.

In order to build the model, a shape $\mathbf{s}$ of 57 landmarks is manually positioned on each of the 338 images, yielding shape-free texture vectors $\mathbf{g}$ of 3493 pixels. The model is built using 50 shape modes, 150 texture modes and 40 appearance modes: the vector $\mathbf{c}$ is composed of 40 components, that retain 98 percent of the combined shape and texture variation. The shape-free texture vector $\mathbf{g}$ is composed of 3493 pixels.

The active appearance search algorithm is illustrated on a previously unseen face (figure 1.a) with the mean shape and texture used as a first approximation (figure 1.b). The model output converges to the target face as shown in figure 1.c.

Three other adaptations of a higher resolution appearance model computed on the same learning set are shown on figure 2.

3 Facial expression analysis/synthesis

3.1 Facial expression modeling

The aim of this section is to study a linear model, as it is proposed in [10, 11]. In the same way as Cootes et al. [10] modeled the facial appearance parameters as a linear function of the orientation, Kang et al. [11] propose a linear model that correlates the appearance parameters to facial expression intensity according to:

$$\mathbf{c} = \mathbf{a}_{\mathbf{e}0} + \mathbf{a}_{\mathbf{e}1}\mathcal{I} + \varepsilon \tag{5}$$

where $\mathcal{I}$ is a scalar varying from $\mathcal{I} = 0$ to indicate neutral expression to $\mathcal{I} = 1$ to indicate a high magnitude expression and ε is the approximation error. $\mathbf{a}_{\mathbf{e}0}$ and $\mathbf{a}_{\mathbf{e}1}$ are coefficient vectors learnt for each facial expression ($\mathbf{e}$ is joy, fear, disgust, surprise, fear, sadness or neutral) by linear regression over the training set. The linear regression is performed using 3 control points for each expression namely neutral expression ($\mathcal{I} = 0$), moderate expression ($\mathcal{I} = 0.5$) and high magnitude expression ($\mathcal{I} = 1$).

3.2 Facial expression filtering

Once the coefficient vectors $\mathbf{a}_{\mathbf{e}0}$ and $\mathbf{a}_{\mathbf{e}1}$ have been learnt for a given expression $\mathbf{e}$, the linear model can be used to predict an artificial vector of appearance parameters $\mathbf{c}_{\mathbf{e}}(\mathcal{I})$ for a given intensity $\mathcal{I}$ of the expression $\mathbf{e}$:

$$\mathbf{c}_{\mathbf{e}}(\mathcal{I}) = \mathbf{a}_{\mathbf{e}0} + \mathbf{a}_{\mathbf{e}1}\mathcal{I} \tag{6}$$

Note that a given intensity $\mathcal{I}$ of the expression $\mathbf{e}$ generates an unique value of the vector $\mathbf{c}_{\mathbf{e}}(\mathcal{I})$: the latter encodes thus an average appearance of this expression intensity, independently of any particular person. This means that the information about *identity* is contained in the residual ε in equation 5. Therefore, in order to synthesize a new expression intensity $\mathcal{I}'$ for a given person, the residual for this person has to be added to the average appearance $\mathbf{c}_{\mathbf{e}}(\mathcal{I}')$ of this new intensity. The procedure for doing this is detailed below.

Starting from an unseen face with a given expression (Fig. 3a), an appearance parameter vector $\mathbf{c}_{op}$ is first estimated as described at the end of 2.1, so that $\mathbf{c}_{op}$ synthesizes an artificial face similar to this target face (Fig. 3b).

Having a priori knowledge of the facial expression $\mathbf{e}$ represented on the target face, it is then possible to estimate the intensity of this expression by inverting equation (6):

$$\mathcal{I}_{est} = \mathbf{a}_{\mathbf{e}1}^{+}(\mathbf{c}_{op} - \mathbf{a}_{\mathbf{e}0}) \tag{7}$$

where $\mathbf{a}_{\mathbf{e}1}^{+}$ is the pseudo inverse of $\mathbf{a}_{\mathbf{e}1}$. The information relative to the person's identity will then be retrieved by filtering out the expression information contained in vector $\mathbf{c}_{\mathbf{e}}(\mathcal{I}_{est})$, the latter being evaluated at the estimated expression intensity $\mathcal{I}_{est}$ using equation (6). This gives the identity vector $\mathbf{c}_{res}$:

$$\mathbf{c}_{res} = \mathbf{c}_{op} - \mathbf{c}_{\mathbf{e}}(\mathcal{I}_{est}) \tag{8}$$

Having this, it is possible to modify the facial expression intensity represented in vector $\mathbf{c}_{\mathbf{e}}(\mathcal{I}_{est})$ by modifying the $\mathcal{I}$ value in equation (6). In particular it is possible to filter out the expression by setting $\mathcal{I} = 0$.

$$\mathbf{c}_{\mathbf{e}}(0) = \mathbf{a}_{\mathbf{e}0} + \mathbf{a}_{\mathbf{e}1} \times 0 = \mathbf{a}_{\mathbf{e}0} \tag{9}$$

Then by adding the identity vector $\mathbf{c}_{res}$ to the corrected expression it is possible to modify the expression intensity shown on the target face from high magnitude to neutral as shown in figure 3c:

$$\mathbf{c}_{neutral} = \mathbf{c}_{\mathbf{e}}(0) + \mathbf{c}_{res}. \tag{10}$$

a b c

Figure 3: a: Target (original face). b: Reconstructed face using $\mathbf{c}_{\mathbf{op}}$ obtained by iterative model adjustment to target face. c: Neutral expression obtained by canceling joy intensity.

3.3 Facial expression synthesis

Starting from the artificially generated neutral expression of the target face, it is possible to artificially generate any desired expression $\mathbf{e}'$ by applying the same method described in 3.2. It is assumed that the linear model (5) for the new expression $\mathbf{e}'$ has been learnt on the training set, giving the corresponding $\mathbf{a}_{\mathbf{e}'0}$ and $\mathbf{a}_{\mathbf{e}'1}$ parameters. In the procedure described above, the appearance parameters describing the target face $\mathbf{c}_{op}$ will now be replaced by $\mathbf{c}_{neutral}$. It is then possible to estimate the intensity of the desired expression on the artificial neutral face. This value should be close to zero.

$$\mathcal{I}'_{est} = \mathbf{a}_{\mathbf{e}'1}^{+}(\mathbf{c}_{neutral} - \mathbf{a}_{\mathbf{e}'0}). \tag{11}$$

The estimated expression information vector at the $\mathcal{I}'_{est}$ intensity of the desired expression is given by:

$$\mathbf{c}_{\mathbf{e}'}(\mathcal{I}'_{est}) = \mathbf{a}_{\mathbf{e}'0} + \mathbf{a}_{\mathbf{e}'1}\mathcal{I}'_{est}. \quad (12)$$

The new residual $\mathbf{c}_{res}$ is then given by:

$$\mathbf{c}_{res} = \mathbf{c}_{neutral} - \mathbf{c}_{\mathbf{e}'}(\mathcal{I}'_{est}). \quad (13)$$

The facial expression intensity represented in vector $\mathbf{c}_{\mathbf{e}'}(\mathcal{I}'_{est})$ can be controlled through the parameter $\mathcal{I}$ in equation (6). In particular it is possible to generate a high magnitude expression parameter estimation by setting $\mathcal{I} = 1$.

Then by adding the identity vector $\mathbf{c}_{res}$ to the corrected expression estimation vector $\mathbf{c}_{\mathbf{e}'}(\mathcal{I}'_{est})$ it is possible to modify the expression intensity shown on the target face from neutral to high magnitude as shown in figure 4.

$$\mathbf{c}_{intense} = \mathbf{c}_{\mathbf{e}'}(1) + \mathbf{c}_{res}. \quad (14)$$

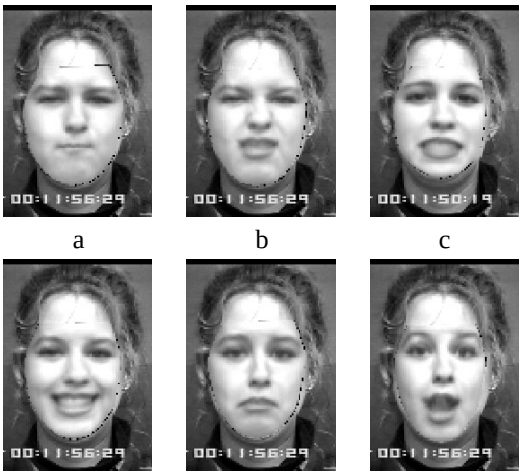

Figure 4: Generation of six synthetic expressions starting from the expression filtered face of figure 3.c. a: Anger. b: Disgust. c: Fear. d: Joy. e: Surprise. f: Sadness

3.4 Expression evolution over a video sequence

In order to analyze the temporal behaviour of the linear model obtained in section 3.1, a series of experiments has been performed on a set of 15 videos representing different persons showing a facial expression gradually evolving from neutral to high magnitude. The active appearance model is fit on each image of a video sequence , using the previous output of the model as a first approximation of appearance and pose parameters as shown on figure 5 (middle row). At each step of the video sequence, the obtained

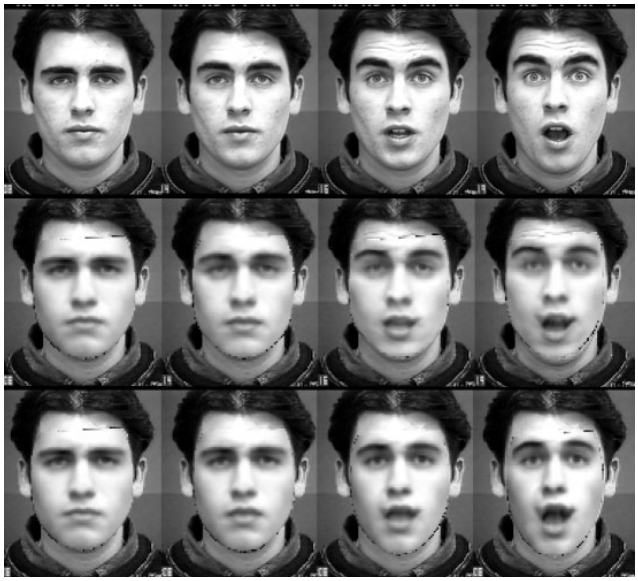

Figure 5: Upper row : original surprised face. Middle row: synthetic expressive face inserted onto the original one. The appearance model may be used to track the face expression in a video. Lower row: synthetic expressive face linearly predicted using eq. 6

$\mathbf{c}_{op}$ parameter vector allows to estimate the facial expression intensity $\mathcal{I}_{est}$ using equation (7), as well as the linearly predicted vector of appearance parameters $\mathbf{c}_{\mathbf{e}}(\mathcal{I}_{est})$ at this intensity.

For each facial expression, the $\mathbf{c}_{\mathbf{e}}(\mathcal{I}_{est})$ parameters tend to follow a well defined trajectory whose behavior can be linearly approximated by the linear model computed in section 3.1. This is illustrated in figure 6 showing the evolution of the first variation mode (first coefficient of $\mathbf{c}_{\mathbf{e}}(\mathcal{I}_{est})$) over 15 video sequences on each sequence, the face displays an expression going from neutral to high magnitude, respectively for anger (left plot) and sadness (right plot).

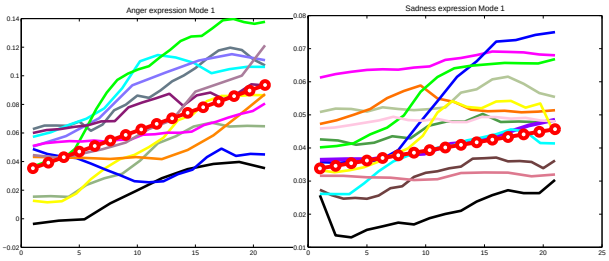

Figure 6: For anger and sadness, evolution of 1st mode of $\mathbf{c}_{\mathbf{e}}(\mathcal{I}_{est})$ over each video sequence. Straight line: 1st mode of $\mathbf{c}_{\mathbf{e}}(\mathcal{I})$, with $\mathcal{I}$ linearly varying from 0 to 1.

4 Facial expression recognition

Automatic recognition of facial expressions has been studied in many previous works. Local representations have been used, such as graphs of 2D Gabor wavelet responses in Lyons et al. [4], or direction of motions computed from the optical flow as in Yacoob and Davies [12]. Global representations have also been much investigated: Donato et al. [3] review and compare holistic representations including eigenfaces, Fisherfaces, independent component analysis, applied on difference texture images or optical flow. Lien et al. [13] propose to recognize facial actions by modeling the temporal evolution of local features as a hidden Markov model.

In this work, we study the discriminating power of the combined appearance parameter $\mathbf{c}$. To classify a new face represented by the parameter vector $\mathbf{c}_i$, and assuming that the expression classes have a common covariance matrix, we measure the squared Mahalanobis distance d_M as shown in equation 15 from $\mathbf{c}_i$ to each of the j estimated mean vector $\bar{\mathbf{c}}_j$ in the original space, and assign $\mathbf{c}_i$ to the class of the nearest mean. $j \in$ [joy, fear, disgust, surprise, fear, sadness, neutral] and Σ is the estimated common covariance matrix of the training set. We repeat the same procedure within a Linear Discriminant Analysis subspace [14], and we measure the squared Mahalanobis distance d_M^{lda} as shown in equation 16, where $\mathbf{c}_i^{lda}$ is the projection of the $\mathbf{c}_i$ vector on the LDA space, $\bar{\mathbf{c}}_j^{lda}$ is the j estimated mean vector in the LDA space and Σ_{lda} is the estimated common covariance matrix of the projected training set. Finally we assign $\mathbf{c}_i^{lda}$ to the class of the nearest mean.

$$d_M(\mathbf{c}_i, \bar{\mathbf{c}}_j) = (\mathbf{c}_i - \bar{\mathbf{c}}_j)^t \Sigma^{-1} (\mathbf{c}_i - \bar{\mathbf{c}}_j) \quad (15)$$

$$d_M^{lda}(\mathbf{c}_i^{lda}, \bar{\mathbf{c}}_j^{lda}) = (\mathbf{c}_i^{lda} - \bar{\mathbf{c}}_j^{lda})^t \Sigma_{lda}^{-1} (\mathbf{c}_i^{lda} - \bar{\mathbf{c}}_j^{lda}) \quad (16)$$

We also use a nearest neighbor classification scheme in the original and LDA spaces using the Mahalanobis distances d_n and d_n^{lda}. Results are shown in tables 1 to 4. Percentages of good classifications of each facial expression for the different schemes are shown in table 5 for comparison.

	neut.	ang.	disg.	fea.	joy	surp.	sad.
neutral	38	1	1	4	0	0	8
anger	1	10	0	0	0	0	1
disgust	0	0	12	0	0	0	0
fear	0	0	1	8	0	0	0
joy	0	0	0	2	11	0	0
surprise	1	0	0	0	0	13	1
sadness	3	0	0	0	0	0	14

Table 1: Confusion matrix for the Mahalanobis distance expression classifier in the original space, using 130 unknown test images.

	neut.	ang.	disg.	fea.	joy	surp.	sad.
neutral	40	1	0	4	0	0	7
anger	2	10	0	0	0	0	1
disgust	0	1	11	0	0	0	0
fear	2	0	0	7	0	0	0
joy	0	0	0	0	13	0	0
surprise	2	0	0	0	0	13	2
sadness	9	2	0	0	0	0	6

Table 2: Confusion matrix for the Mahalanobis distance expression classifier in the LDA space, using 130 unknown test images.

	neut.	ang.	disg.	fea.	joy	surp.	sad.
neutral	35	4	2	4	0	0	7
anger	3	7	2	0	0	0	0
disgust	1	1	10	0	0	0	0
fear	3	0	0	4	1	0	1
joy	0	0	1	2	10	0	0
surprise	0	0	0	1	0	14	0
sadness	3	2	0	1	0	0	11

Table 3: Confusion matrix, for a nearest neighbor classifier in the original space, using 130 unknown test images.

	neut.	ang.	disg.	fea.	joy	surp.	sad.
neutral	29	1	1	5	1	0	15
anger	1	10	0	0	0	0	1
disgust	0	1	10	1	0	0	0
fear	0	0	1	7	0	0	1
joy	0	0	0	1	12	0	0
surprise	0	0	0	1	0	13	1
sadness	2	3	0	0	0	0	12

Table 4: Confusion matrix, for a nearest neighbor classifier in the LDA space, using 130 unknown test images.

	d_M	d_M^{lda}	d_n	d_n^{lda}
neutral	73.1	76.9	67.3	55.8
anger	83.3	83.3	58.3	83.3
disgust	100	91.7	83.3	83.3
fear	88.9	77.8	44.4	77.8
joy	84.6	100	76.9	92.3
surprise	86.7	86.7	93.3	86.7
sadness	82.3	35.3	64.7	70.6
total	81.5	76.9	70.0	71.5

Table 5: Percentages of good classification of all facial expressions for different classification schemes in the original and LDA spaces using 130 test images.

The LDA subspace is computed so as to maximize class separability, and allows thus an optimal visualization of the

class structure. In Figure 7, the classes appear indeed much better separated in the LDA space than in the original space; that is reflected in the greatly increased value of Fisher criterion $\frac{|S_b|}{|S_w|}$. The "sad" expression appears poorly separated from the other classes, especially from the neutral class, which explains the worse classification rates for this class.

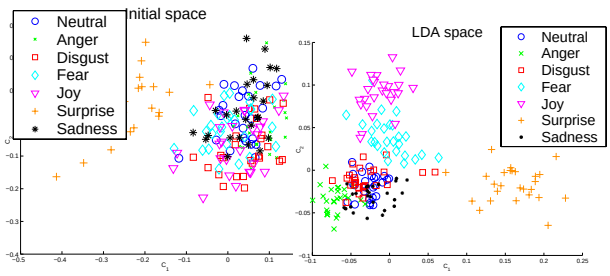

Figure 7: 2D representation of the first 2 modes of the training set. Left: In the original space with Fischer criterion value $\frac{|S_b|}{|S_w|} = 9.48$. Right: Projected in the LDA space with Fischer criterion value $\frac{|S_b|}{|S_w|} = 44.74$.

5 Conclusion and perspectives

In this paper, we have presented different applications of Active Appearance Models for expressive face synthesis and recognition. We are currently investigating other classification approaches, based on linear or non-linear schemes. In addition to the greylevel texture, edges, multi-resolution Gabor responses, or color [15] could also be exploited to perform facial expression recognition.

References

[1] V. Bruce and A. Young, *In the eye of the beholder. The science of face perception*, University Press, Oxford, 1998.

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

[3] G. Donato, M.S. Bartlett, J.C. Hager, P. Ekman, and T.J. Sejnowski, "Classifying facial actions," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 21, no. 10, pp. 974–988, October 1999.

[4] M.J. Lyons, J. Budynek, and S. Akamatsu, "Automatic classification of single facial images," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 21, no. 12, pp. 1357–1362, December 1999.

[5] M. Pantic and L.J.M. Rothkrantz, "Automatic analysis of facial expressions: The state of the art," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 22, no. 12, pp. 1424–1445, December 2000.

[6] B. Fasel and J. Luettin, "Automatic facial expression analysis: a survey," Tech. Rep. RR 99-19, IDIAP, Martigny, Valais, Switzerland, Dec. 2000.

[7] T.F. Cootes, G.J. Edwards, and C.J. Taylor, "Active appearance models," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 23, no. 6, pp. 681–685, June 2001.

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

[9] T. Kanade, J. Cohn, and Y.L. Tian, "Comprehensive database for facial expression analysis," in *International Conference on Automatic Face and Gesture Recognition*, Grenoble, France, March 2000, pp. 46–53.

[10] T.F. Cootes, K. Walker, and C.J. Taylor, "View-based active appearance models," in *International Conference on Automatic Face and Gesture Recognition*, Grenoble, France, March 2000, pp. 227–232.

[11] H. Kang, T.F. Cootes, and C.J. Taylor, "Face expression detection and synthesis using statistical models of appearance," in *Measuring Behavior*, Amsterdam, The Netherlands, August 2002, pp. 126–128.

[12] Y. Yacoob and L. Davis, "Recognizing facial expressions by spatio-temporal analysis," in *Proceedings of the International Conference on Pattern Recognition*, Jerusalem Israel, October 1994, pp. 747–749.

[13] Jenn-Jier James Lien, Takeo Kanade, Jeffrey Cohn, and C. Li, "Detection, tracking, and classification of subtle changes in facial expression," *Journal of Robotics and Autonomous Systems*, vol. 31, pp. 131 – 146, 2000.

[14] S. Dubuisson, F. Davoine, and M. Masson, "A solution for facial expression representation and recognition," *Signal Processing: Image Communication*, vol. 17, no. 9, pp. 657–673, October 2002.

[15] M.B. Stegman and R. Larsen, "Multi-band modelling of appearance," in *First International Workshop on Generative Model-Based Vision GMBV*, Copenhagen Denmark, June 2002, pp. 101–106.