Reconnaissance et synthèse d'expressions faciales par modèle d'apparence

Facial expression recognition and synthesis based on an appearance model

Bouchra Abboud Franck Davoine Mô Dang

Laboratoire Heudiasyc, Université de Technologie de Compiègne

BP 20529, 60205 COMPIEGNE Cedex.
Bouchra.Abboud@hds.utc.fr

Résumé

Cet article traite d'une application nouvelle des modèles actifs d'apparence pour l'analyse et la synthèse de visages expressifs, ainsi que pour la reconnaissance d'expressions faciales. Nous considérons les six expressions universelles que sont les expressions de joie, colère, peur, dégoût, tristesse et surprise. Après une description de ce type de modèle (calculé à l'aide de trois ACP ou d'une seule), nous proposons une méthode originale d'analyse et de synthèse permettant, à partir d'une seule photo, d'annuler l'expression d'un visage quelconque, puis de synthétiser une expression faciale artificielle sur ce même visage. Nous proposons pour cela deux approches permettant la modélisation des expressions faciales. Enfin, nous montrons l'intérêt du modèle actif d'apparence pour la reconnaissance automatique d'expressions faciales.

Mots Clef

Représentation, reconnaissance et synthèse de visages, de gestes faciaux et d'expressions faciales. Modélisation des images.

Abstract

This article addresses the issue of expressive face analysis and synthesis as well as facial expression recognition using active appearance models. We consider the six universal emotional categories namely joy, anger, fear, disgust, sadness and surprise. After a description of the active appearance model (computed with 3 PCA or with one PCA), we propose a new method for analysis and synthesis allowing, from a single photo, to cancel the facial expression on a given face and to artificially synthesize novel expressions on this same face. In this framework, we propose two facial expression modelling approaches. Finally we address the active appearance model contribution to automatic facial expression recognition.

Keywords

Face and facial expression representation, recognition and synthesis. Image modelling.

1 Introduction

Les travaux de recherche en psychologie ont démontré que les expressions faciales jouent un rôle prépondérant dans la coordination de la conversation humaine [3], et ont un impact plus important sur l'auditeur que le contenu textuel du message exprimé. En effet, Mehrabian [26] remarque que la contribution du contenu textuel d'un message verbal à son impact global se limite à 7% alors que l'intonation de la voix du locuteur et son expression faciale contribuent respectivement à 38% et 55% de l'impact global du message exprimé. Par conséquent, les expressions faciales constituent une modalité essentielle de la communication humaine.

L'analyse automatique des expressions faciales constitue un outil important pour la recherche dans les domaines des sciences de l'étude du comportement et de la psychologie, ainsi que dans les domaines de la compression d'images et de l'animation de visages synthétiques [21]. De plus, des travaux précédents ont démontré que les expressions faciales de six catégories émotionnelles de base sont universellement reconnues, à savoir : la colère, le dégoût, la peur, la joie, la surprise et la tristesse [16].

Dans le passé, les travaux de recherche sur l'analyse des expressions faciales se situaient principalement dans le cadre de la psychologie. Les progrès effectués dans des domaines connexes tels que la détection, le suivi et la reconnaissance de visages [32] ainsi que dans le traitement d'images et la reconnaissance de formes, ont apporté une contribution significative à la recherche dans le domaine de l'analyse, de la synthèse et de la reconnaissance d'expressions faciales.

Avant de procéder à l'analyse d'une expression faciale sur une image ou sur une séquence vidéo, il convient de détecter ou de suivre le visage observé [31] puis d'en extraire les informations pertinentes pour l'analyse de l'expression

faciale. Il existe différentes méthodes pour l'extraction de ces informations. Les méthodes globales modélisent le visage dans son intégralité alors que les méthodes reposant sur des traits caractéristiques permettent une modélisation locale du visage dans les régions susceptibles de se modifier selon les expressions faciales affichées.

a- Extraction des informations. Black et Yacoob [2] utilisent des modèles locaux paramétriques pour représenter le mouvement des visages. Ils estiment le mouvement relatif des traits caractéristiques dans le repère du visage. Les paramètres de ce mouvement servent par la suite à représenter l'expression faciale. De manière identique, Cohn *et al.* [8] utilise un algorithme hiérarchique pour effectuer le suivi des traits caractéristiques par estimation du flux optique. Les vecteurs de déplacement représentent l'information sur les changements d'expression faciale. De même, Padgett et Cottrell [27] utilisent des gabarits d'œil et de bouche, calculés par Analyse en Composantes Principales d'un ensemble d'apprentissage, en conjonction avec des réseaux de neurones. D'autre part, Hong *et al.* [22] utilise un modèle global basé sur des graphes étiquetés construits à partir de points de repère distribués sur le visage. Les noeuds de ces graphes sont formés par des vecteurs dont chaque élément est la réponse à un filtrage de Gabor extraite en un point donné de l'image. Finalement, Cootes *et al.* [9] utilise une représentation par modèle actif d'apparence (AAM) pour extraire automatiquement des paramètres caractérisant le visage.

b- Reconnaissance d'expressions. Après avoir détecté le visage et extrait les informations pertinentes, l'étape suivante consiste à identifier l'expression faciale affichée. Pour classer l'expression faciale dans l'une des six catégories de base citées dans [16] en plus de la catégorie neutre, Hong *et al.* [22] part du principe que deux personnes qui se ressemblent affichent la même expression faciale de manière similaire. Un graphe étiqueté est attribué à l'image de test puis la personne connue la plus proche est déterminée à l'aide d'une méthode de mise en correspondance de graphes élastiques. La galerie personnalisée de cette personne est alors utilisée pour reconnaitre l'expression faciale de l'image de test. Un graphe étiqueté par des réponses de filtres de Gabor est par ailleurs utilisé par Lyons *et al.* [25] et Bartlett et *al.* [1]. L'ensemble des graphes construits sur un ensemble d'apprentissage est ensuite soumis à une ACP puis analysé à l'aide d'une analyse discriminante linéaire (ADL) afin de séparer les vecteurs dans des classes ayant des attributs faciaux différents. Le graphe étiqueté de l'image testée sera alors projeté sur les vecteurs discriminants de chaque classe afin de déterminer son éventuelle appartenance à cette classe. Dans une finalité identique, Essa et Pentland [18] extraient des gabarits spatio-temporels de l'énergie du mouvement du visage pour chaque expression faciale. Le critère de similarité repose sur la distance Euclidienne entre ces gabarits et l'énergie du mouvement de l'image observée. Heisele, Ho et Poggio [20] utilisent des machines à vecteurs de support (SVM) dans le cadre de la reconnaissance de visages par des méthodes globales ainsi que par des méthodes reposant sur des traits caractéristiques. De manière identique, l'algorithme de reconnaissance de visages FaceIt est basé sur une technique d'analyse locale des traits caractéristiques (LFA) développée par Penev et Atick [28]. Draper *et al.* [12] compare les performances de l'analyse en composantes principales et de l'analyse en composantes indépendantes (ICA) pour la reconnaissance de visages et d'expressions faciales en se basant sur le système de codage FACS [17]. Par contre, Yang [30] utilise une représentation par noyaux (KPCA) pour la reconnaissance de visages. Finalement, Edwards, Cootes et Taylor [15] utilisent le modèle actif d'apparence pour reconnaître l'identité d'un individu observé de manière robuste par rapport à l'expression faciale ainsi que l'illumination et la pose. Pour ceci, le critère de similarité utilisé repose sur la distance de Mahalanobis, et une ADL est appliquée afin de maximiser la séparation des classes.

c- Synthèse d'expressions. La synthèse d'expressions faciales est une tâche difficile compte tenu de la complexité de la forme et de la texture des visages. De plus le visage présente un grand nombre de rides et de plis ainsi que des variations subtiles de forme et de texture qui ont une importance cruciale dans la compréhension et la représentation des expressions faciales. Dans cette perspective, les techniques d'interpolation offrent une approche intuitive pour l'animation de visages. Pighin *et al.* [29] utilise des techniques de morphing 2D combinées avec des transformations d'un modèle géométrique 3D, pour créer des modèles faciaux réalistes tridimensionnels à partir de photographies, et pour construire des transitions lisses entre les différentes expressions faciales. Dans la même optique, Chen *et al.* [5] applique le morphing au cas 3D. En outre, dans le cadre du "Video-Rewrite", Bregler *et al.* [4] utilise des techniques de suivi de points 2D sur la bouche d'un orateur dans une séquence d'apprentissage et des techniques de morphing pour combiner ces mouvements dans une vidéo finale montrant une personne différente prononçant les même paroles. Dans une finalité analogue, Ezzat *et al.* [19] utilise une représentation par modèle déformable multidimensonnel et une technique de synthèse de trajectoire pour modifier les mouvements de la bouche d'un visage parlant. Cette représentation est capable de synthétiser des configurations inconnues de lèvres parlantes "vidéo-réalistes" à partir d'une séquence initiale, en utilisant des techniques de morphing. Dans la même optique, Chuang *et al.* [6] utilise une ACP combinée à un modèle de factorisation bilinéaire pour synthétiser une nouvelle expression sur un visage parlant. Finalement, Kang *et al.* [24] utilise le modèle actif d'apparence en conjonction avec des techniques de régression linéaire pour annuler l'expression faciale d'un visage dans le but d'améliorer les performances de la technique de reconnaissance de visages par AAM décrite dans [15].

Dans cet article, nous proposons une variante au modèle AAM décrit dans [9] pour la représentation de visages en

n'utilisant qu'une seule ACP. Nous montrons que cette représentation plus directe donne des résultats sensiblement comparables au modèle AAM standard pour la représentation et la reconnaissance d'expressions faciales. Nous proposons également une nouvelle approche de la reconnaissance d'expressions faciales basée sur le modèle AAM ainsi que sur sa variante. Finalement, nous proposons une extension de la méthode décrite dans [24] pour l'annulation de l'expression faciale, à une application de synthèse de nouvelles expressions faciales. De plus nous introduisons une nouvelle méthode d'interpolation, basée sur la représentation par le modèle actif d'apparence AAM, pour la synthèse et l'annulation d'expressions faciales.

2 Le modèle actif d'apparence

Cette section donne une description rapide du modèle actif d'apparence, sa construction est d'abord expliquée et quelques résultats expérimentaux montrant l'adaptation du modèle sur des visages inconnus sont ensuite donnés. La construction d'un modèle actif d'apparence avec une méthode alternative plus directe est également proposée et des résultats d'adaptation sur un visage inconnu sont montrés pour illustrer les performances de cette méthode.

2.1 Description

Il a déjà été démontré [9] que le modèle actif d'apparence est un outil puissant permettant de représenter des visages de façon réaliste et naturelle. Il se base sur une technique d'ACP permettant de représenter conjointement les variations de forme et de texture présentes dans un ensemble d'apprentissage. En effet, après avoir aligné toutes les formes de l'ensemble d'apprentissage par rapport à la forme moyenne à l'aide d'une transformation Procrustéenne [13], le modèle statistique de forme est donné par :

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

où $\mathbf{s}_i$ est la forme synthétisée, Φ_s est une matrice contenant les principaux modes de variation de forme, et $\mathbf{b}_{si}$ est un vecteur contrôlant la forme synthétisée. Il est alors possible de déformer les textures de l'ensemble d'apprentissage sur la forme moyenne $\bar{\mathbf{s}}$ pour séparer la texture de la forme. De manière similaire, après avoir calculé la texture sans forme moyenne $\bar{\mathbf{g}}$ et normalisé toutes les textures de l'ensemble d'apprentissage par rapport à la texture moyenne à l'aide d'une translation et d'une remise à l'échelle des niveaux de gris, le modèle statistique de texture est donné par :

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

où $\mathbf{g}_i$ est la texture sans forme synthétisée, Φ_t est une matrice contenant les modes principaux de variation de texture dans l'ensemble d'apprentissage, et $\mathbf{b}_{ti}$ est un vecteur contrôlant la texture synthétisée, sans forme.
Après pondération des valeurs des composantes de $\mathbf{b}_{si}$ avec des poids convenablement choisis pour rendre leur ordre de grandeur comparable à $\mathbf{b}_{ti}$ [9], ces vecteurs sont concaténés. En pratique les composantes de $\mathbf{b}_{si}$ sont pondérées par un poids égal à la racine carrée du rapport de la somme des valeurs propres associées à Φ_t sur la somme des valeurs propres associées à Φ_s. Une ACP supplémentaire sur les nouveaux vecteurs obtenus retourne le modèle statistique d'apparence donné par :

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

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

où Q_s et Q_t sont des matrices contenant les principaux modes de variation d'apparence dans l'ensemble d'apprentissage, et $\mathbf{c}_i$ est un vecteur de paramètres d'apparence permettant de contrôler simultanément la forme et la texture de l'objet.
Une autre manière de représenter le modèle combiné consiste à concaténer directement, pour chaque image i de l'ensemble d'apprentissage, les vecteurs de forme $\mathbf{s}_i$ et de texture $\mathbf{g}_i$ pour former le vecteur $\mathbf{b}_{sti}$. Les vecteurs $\mathbf{s}_i$ seront multipliés par une matrice de poids W_s convenablement choisie afin de rendre leur ordre de grandeur comparable à $\mathbf{g}_i$. En pratique, les vecteurs $\mathbf{s}_i$ sont pondérées par un poids égal au rapport de l'écart-type des intensités normalisées par l'écart-type des formes normalisées de l'ensemble d'apprentissage.
En appliquant une ACP aux vecteurs $\mathbf{b}_{sti}$ obtenus le modèle d'apparence modifié est donné par :

$$\begin{pmatrix} \mathbf{W}_s \mathbf{s}_i \\ \mathbf{g}_i \end{pmatrix} = \begin{pmatrix} \bar{\mathbf{s}} \\ \bar{\mathbf{g}} \end{pmatrix} + \mathbf{Q}_{st} \mathbf{c}_i, \tag{5}$$

où $\mathbf{Q}_{st}$ est une matrice représentant les principaux modes de variation des vecteurs combinés de forme et de texture, et $\mathbf{c}_i$ est le vecteur de paramètres d'apparence contrôlant conjointement la forme et la texture synthétisées par le modèle.

a b c

FIG. 1 – *a : visage inconnu. b : initialization du modèle. c : adaptation du modèle classique.*

Ayant le vecteur de paramètres $\mathbf{c}_i$, la forme et la texture correspondantes peuvent être retrouvées respectivement à partir des équations (3) et (4) pour la représentation standard et (5) pour la représentation modifiée. La texture sans

forme reconstruire est alors déformée sur la forme reconstruite afin d'obtenir l'apparence globale du visage reconstruit. De plus, il est nécessaire de rajouter au vecteur de paramètres d'apparence $\mathbf{c}_i$, un vecteur de paramètres de pose $\mathbf{p}_i$ permettant de contrôler l'échelle, l'orientation et la position du visage reconstruit.
En partant du principe que les vecteurs d'apparence $\mathbf{c}$ et de pose $\mathbf{p}$ constituent une représentation d'un visage, le modèle actif d'apparence AAM est capable d'ajuster automatiquement ces paramètres à un visage cible inconnu [10] en minimisant une image résiduelle $\mathbf{r}(\mathbf{c},\mathbf{p})$ qui n'est autre que la différence de texture entre le visage synthétisé et le masque de l'image cible correspondant. Ces résidus sont utilisés pour calculer les matrices $\mathbf{R_a}$ et $\mathbf{R_t}$ permettant de relier linéairement les paramètres d'apparence et de pose aux résidus correspondants $\delta(\mathbf{c}) = -\mathbf{R_a}\ \mathbf{r}(\mathbf{c},\mathbf{p})$ et $\delta(\mathbf{p}) = -\mathbf{R_t}\ \mathbf{r}(\mathbf{c},\mathbf{p})$, de telle sorte à minimiser $|\mathbf{r}((\mathbf{c},\mathbf{p})+\delta(\mathbf{c},\mathbf{p}))|^2$. Un développement de Taylor du premier ordre donne la solution suivante:

$$\mathbf{R_a} = (\frac{\partial \mathbf{r^T}}{\partial \mathbf{c}} \frac{\partial \mathbf{r}}{\partial \mathbf{c}})^{-1} \frac{\partial \mathbf{r^T}}{\partial \mathbf{c}} \text{ et } \mathbf{R_t} = (\frac{\partial \mathbf{r^T}}{\partial \mathbf{p}} \frac{\partial \mathbf{r}}{\partial \mathbf{p}})^{-1} \frac{\partial \mathbf{r^T}}{\partial \mathbf{p}} \quad (6)$$

avec $\frac{\partial \mathbf{r}}{\partial \mathbf{p}}$ constant estimé à partir de l'ensemble d'apprentissage en déplaçant les paramètres d'apparence et de pose de leurs valeurs optimales. Les paramètres d'apparence et de pose sont alors adaptés au visage cible à l'aide d'une procédure itérative d'optimisation. Dans la suite, les paramètres d'apparence et de pose optimaux seront dénommés respectivement $\mathbf{c}_{op}$ et $\mathbf{p}_{op}$.

2.2 Résultats expérimentaux

Le modèle est construit à partir de 375 visages essentiellement extraits de la base CMU [23] représentant les différentes expressions de base colère, dégoût, peur, joie surprise et tristesse avec des intensités fortes et modérées ainsi que des expressions neutres. 53 points de repère sont manuellement positionnés sur ces images et des vecteur de texture de 5871 pixels de taille en sont extraits. Le modèle standard est construit en conservant 50 modes de forme, 170 modes de texture et 120 modes d'apparence au final. Par contre le modèle modifié est construit en gardant 170 modes d'apparence conservant ainsi pour les deux repésentations 98% de la variabilité totale des vecteurs combinés.

FIG. 2 – *Adaptation itérative du modèle, en partant d'une pose et d'une apparence proches du visage cible situé en bas à droite.*

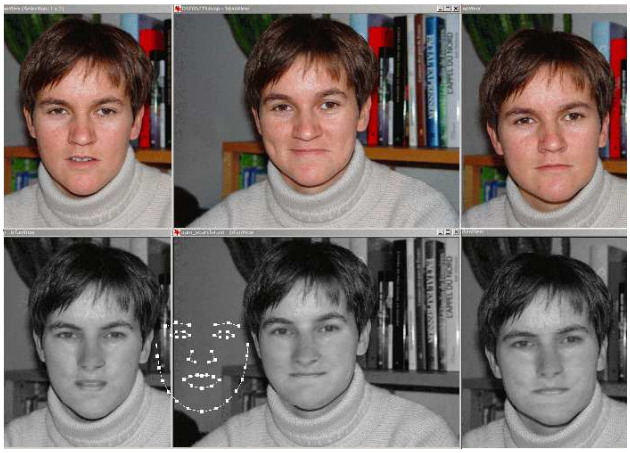


FIG. 3 – *En haut : visages originaux, inconnus du modèle. En bas : adaptation automatique d'un modèle actif d'apparence, calculé sur une base de 375 visages neutres et expressifs, en l'initialisant sur une position proche du visage original. Le visage synthétique est superposé au visage original. La forme moyenne $\bar{\mathbf{s}}$ du modèle est illustrée par les points blancs.*

L'adaptation du modèle standard sur un visage inconnu en partant de la forme et de la texture moyennes est illustrée sur la figure 1.

a b c

FIG. 4 – *a : visage inconnu. b : initialization du modèle. c : adaptation du modèle modifié.*

Les figures 3 et 2 illustrent l'adaptation d'un modèle standard de plus haute résolution composé, dans ce cas, de 13085 pixels.
Nous vérifions expérimentalement que la qualité de l'adaptation du modèle modifié est comparable à celle du modèle classique comme montré dans la figure 4.

3 Reconnaissance d'expressions faciales

Nous proposons dans cette section une solution permettant de reconnaitre l'expression faciale d'un visage cible, par discrimination paramétrique linéaire, en partant du vecteur de paramètres d'apparence $\mathbf{c}_{op}$ qui lui est associé. Une analyse discriminante linéaire [14] est utilisée afin d'évaluer

la qualité de la séparation des classes. L'apprentissage prédictif est effectué sur un ensemble constitué par 7 classes correpondant aux 6 expressions de base en plus de l'expression neutre. Chaque classe contient 26 vecteurs $\mathbf{c}_i$ de taille 120 chacun pour la représentation standard et 170 chacun pour la représentation par AAM modifié. Le critère de similarité utilisé pour la classification d'un vecteur de paramètres $\mathbf{c}_{op}$ repose sur la distance Euclidienne comme montré dans l'équation (7), où $\mathbf{c}_{op}^{lda}$ est la projection du vecteur $\mathbf{c}_{op}$ dans l'espace discriminant et $\bar{\mathbf{c}}_j^{lda}$ est le vecteur moyenne de la classe j dans l'espace discriminant. Finalement $\mathbf{c}_{op}^{lda}$ sera assigné à la classe ayant la moyenne la plus proche.

$$d_E^{lda}(\mathbf{c}_{op}^{lda},\bar{\mathbf{c}}_j^{lda}) = (\mathbf{c}_{op}^{lda} - \bar{\mathbf{c}}_j^{lda})^t(\mathbf{c}_{op}^{lda} - \bar{\mathbf{c}}_j^{lda}) \quad (7)$$

Les résultats de reconnaissance pour les représentations par AAM standard et AAM modifié sont montrés dans les tables (1) et (2) en considérant des vecteurs d'apparence tronqués de dimension 80. Evidemment, la classification des vecteurs $\mathbf{c}_{op}$ pourrait tout aussi bien se faire avec une performance similaire, dans l'espace des paramètes d'apparence en utilisant une métrique reposant sur la distance de Mahalanobis.

	neut.	ang.	disg.	fea.	joy	surp.	sad.
neut.	40	1	2	0	0	0	3
ang.	6	11	1	0	0	0	3
disg.	0	0	17	2	1	0	1
fear	1	0	1	13	3	1	0
joy	0	0	0	2	26	0	0
surp.	0	0	0	0	0	27	0
sad.	2	2	0	0	0	1	14

TAB. 1 – *Matrice de confusion pour la reconnaissance d'expressions faciales avec le modèle classique sur 181 images de test inconnues. Le pourcentage global de bonne classification est de 81.77%.*

	neut.	col.	dég.	peur.	joie	surp.	tris.
neutre	44	1	0	0	0	0	1
colère	9	10	0	0	0	0	2
dégout	1	2	14	3	1	0	0
peur	1	0	1	14	3	0	0
joie	0	1	0	1	26	0	0
surprise	0	1	0	0	0	25	1
tristesse	3	5	0	1	0	1	9

TAB. 2 – *Matrice de confusion pour la reconnaissance d'expressions faciales avec le modèle à 1 ACP sur 181 images de test inconnues. Le pourcentage global de bonne classification est de 78.45%.*

Les pourcentages de bonne classification pour chaque expression faciale dans les deux approches considérées sont donnés dans la table 3.

	3 ACP	1 ACP
neutre	86.9	95.65
colère	52.38	47.62
dégoût	80.95	66.67
peur	68.42	73.68
joie	92.86	92.86
surprise	100	92.59
tristesse	73.68	47.37
total	81.77	78.45

TAB. 3 – *Pourcentages de bonne classification de toutes les expressions expressions faciales pour les differentes approches de modélisation et avec 181 images de test inconnues.*

L'analyse discriminante linéaire a pour but d'évaluer la séparabilité des classes comme montré dans la figure 5. Une mesure de la séparation des classes dans le nouvel espace est donnée par le critère de Fisher ($trace(\mathbf{S}_w^{-1}\mathbf{S}_b)$), où $\mathbf{S}_w$ et $\mathbf{S}_b$ représentent respectivement les matrices de variance intra et inter-classes dans l'espace de Fisher.

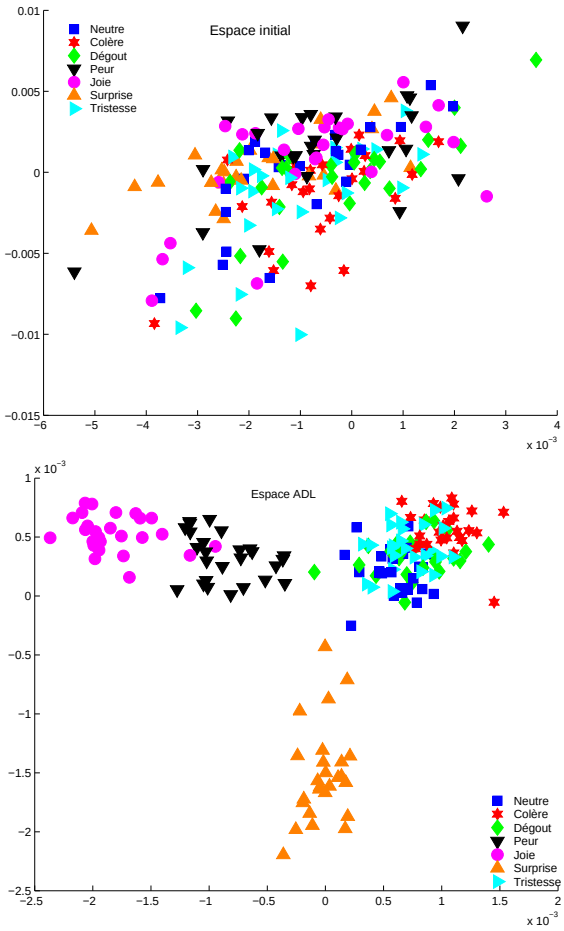


FIG. 5 – *Representation des 2 premiers modes de l'ensemble d'apprentissage construit avec 3 ACP. En haut : dans l'espace original. En bas : projection dans l'espace ADL avec un critère de Fischer* $trace(S_w^{-1}S_b) = 37.19$.

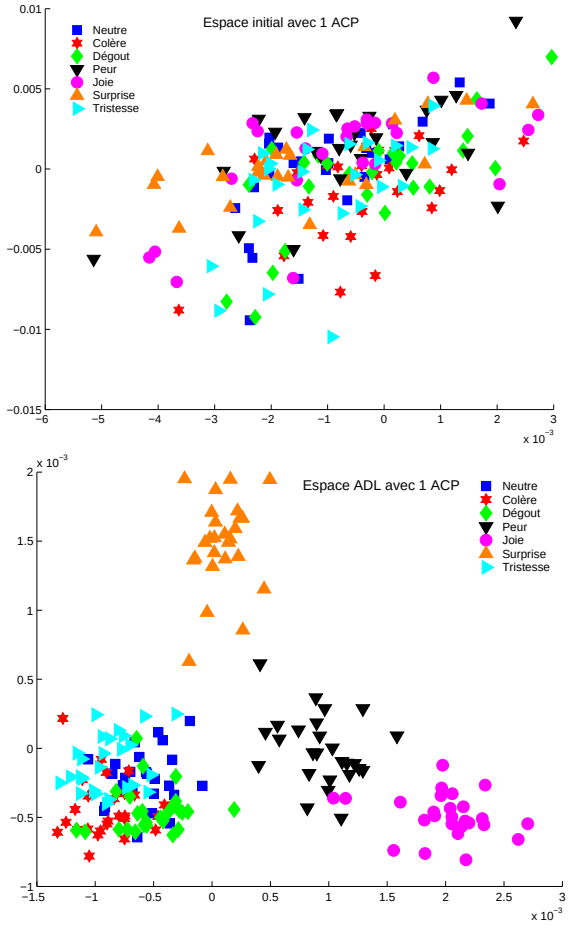


FIG. 6 – *Representation des 2 premiers modes de l'ensemble d'apprentissage construit avec 1 ACP. En haut : dans l'espace original. En bas : projection dans l'espace ADL avec un critère de Fischer* $trace(S_w^{-1}S_b) = 36.3$.

4 Modélisation et synthèse d'expressions faciales

Dans la première partie de cette section, nous proposons une modélisation des expressions faciales par régression linéaire entre les paramètres d'apparence et l'intensité de l'expression faciale (neutre, modérée ou intense). Le modèle linéaire direct ainsi obtenu est utilisé par Kang *et al.* [24] dans le cadre du filtrage des expressions faciales. Nous proposons l'extension de cette méthode à une application en synthèse d'expressions faciales. Quelques résultats expérimentaux illustrant les performances de ce modèle sont donnés. Dans une deuxième partie nous proposons une méthode de modélisation des expressions faciales par régression linéaire entre l'évolution des paramètres d'apparence et l'intensité de l'expression faciale (modérée ou intense). Le modèle linéaire évolutif ainsi obtenu est utilisé dans le cadre de l'annulation et de la synthèse d'expressions faciales artificielles. Finalement, des résultats expérimentaux illustrant la performance du modèle linéaire évolutif sont donnés.

4.1 Modélisation directe de l'expression

En s'inspirant du modèle linéaire proposé par Cootes *et al.* [11] pour représenter les paramètres d'apparence en fonction de l'orientation, un modèle linéaire reliant les paramètres d'apparence à l'intensité de l'expression faciale affichée peut s'écrire comme montré dans l'équation (8) :

$$\mathbf{c_e}(\mathcal{I}) = \mathbf{a_{e0}} + \mathbf{a_{e1}}\mathcal{I}, \tag{8}$$

où $\mathcal{I}$ est un scalaire variant de $\mathcal{I} = 0$ pour indiquer une expression neutre à $\mathcal{I} = 1$ pour indiquer une expression intense, $\mathbf{a_{e0}}$ et $\mathbf{a_{e1}}$ sont des vecteurs de coefficients associés à l'expression faciale $\mathbf{e}$ ($\mathbf{e}$ = colère, dégoût, peur, joie, surprise ou tristesse).
Pour chaque expression faciale e, l'apprentissage des vecteurs $\mathbf{a_{e0}}$ et $\mathbf{a_{e1}}$ se fait par regression linéaire sur un ensemble d'apprentissage en résolvant le système suivant :

$$\mathbf{Y} = \mathbf{X}\begin{pmatrix} \boldsymbol{a}_{\boldsymbol{e0}}^T \\ \boldsymbol{a}_{\boldsymbol{e1}}^T \end{pmatrix}. \tag{9}$$

$\mathbf{Y}$ est une matrice contenant dans les m premières lignes, les vecteurs $\boldsymbol{c_i^T}$ correspondants aux visages de l'ensemble d'apprentissage affichant intensément l'expression faciale considérée et dans les m lignes suivantes les vecteurs $\boldsymbol{c_i^T}$ correspondants aux visages de l'ensemble d'apprentissage affichant la même expression faciale, mais d'intensité moyenne. Les m dernières lignes de $\mathbf{Y}$ contiennent les vecteurs $\boldsymbol{c_i^T}$ correspondants aux visages neutres de l'ensemble d'apprentissage.
La deuxième colonne de $\mathbf{X}$ contient la valeur 1 dans les m premières lignes, 0.5 dans les m lignes suivantes et 0 dans les dernières lignes. La première colonne de $\mathbf{X}$ contient la valeur 1. La solution par régression linéaire est donnée par :

$$\begin{pmatrix} \boldsymbol{a}_{\boldsymbol{e0}}^T \\ \boldsymbol{a}_{\boldsymbol{e1}}^T \end{pmatrix} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{Y}. \tag{10}$$

a- Filtrage de l'expression. Kang *et al.* [24] applique le modèle linéaire ainsi trouvé pour ramener l'expression faciale d'un visage cible inconnu vers une expression neutre. En effet, en connaissant a priori l'expression faciale affichée, les vecteurs $\mathbf{a_{e0}}$ and $\mathbf{a_{e1}}$ sont connus. La première étape consiste à trouver les paramètres $\mathbf{c}_{op}$ permettant d'adapter le modèle au visage cible. Le vecteur $\mathbf{c}_{op}$ nous permet ensuite d'estimer l'intensité de l'expression faciale en inversant le modèle défini dans l'équation (8).

$$\mathcal{I}_{est} = \mathbf{a_{e1}^+}(\mathbf{c}_{op} - \mathbf{a_{e0}}), \tag{11}$$

où $\mathbf{a_{e1}^+}$ est le pseudo-inverse de $\mathbf{a_{e1}}$. Un vecteur résiduel $\mathbf{c}_{res}$ contenant l'information relative à l'identité de la personne cible est obtenu en filtrant l'information contenue dans le modèle d'expression :

$$\mathbf{c}_{res} = \mathbf{c}_{op} - \mathbf{c_e}(\mathcal{I}_{est}). \tag{12}$$

a b c

FIG. 7 – *a : visage joyeux inconnu. b : adaptation du modèle. c : annulation de l'expression par le modèle direct.*

En modifiant la valeur de $\mathcal{I}$ dans le vecteur $\mathbf{c_e}(\mathcal{I}_{est})$, il est possible de modifier l'intensité de l'expression. En particulier en imposant $\mathcal{I} = 0$ l'expression faciale affichée est entièrement filtrée.

$$\mathbf{c_e}(0) = \mathbf{a_{e0}} + \mathbf{a_{e1}} \times 0 = \mathbf{a_{e0}}. \tag{13}$$

Finalement, en conservant intact le vecteur $\mathbf{c}_{res}$, et en filtrant l'expression faciale ($\mathcal{I} = 0$), il est possible d'annuler l'expression faciale affichée sur le visage cible inconnu et de la ramener vers une expression neutre comme montré sur la figure 7, à l'aide de l'équation suivante :

$$\mathbf{c}_{neutre} = \mathbf{c_e}(0) + \mathbf{c}_{res}. \tag{14}$$

b- Synthèse d'expressions. La méthode proposée permet d'annuler l'expression faciale sur un visage expressif inconnu. En étendant cette même méthode à la synthèse, il est possible de générer une expression faciale sur un visage neutre inconnu, notamment sur un visage neutre synthétisé par annulation de son expression.
En effet, soit $\mathbf{e}'$ l'expression désirée et soient $\mathbf{a_{e'0}}$ et $\mathbf{a_{e'1}}$ les paramètres correspondants. Il est alors possible d'estimer l'intensité de l'expression désirée sur le visage neutre de test.

$$\mathcal{I}'_{est} = \mathbf{a_{e'}}_1^{+}(\mathbf{c}_{neutre} - \mathbf{a_{e'0}}). \tag{15}$$

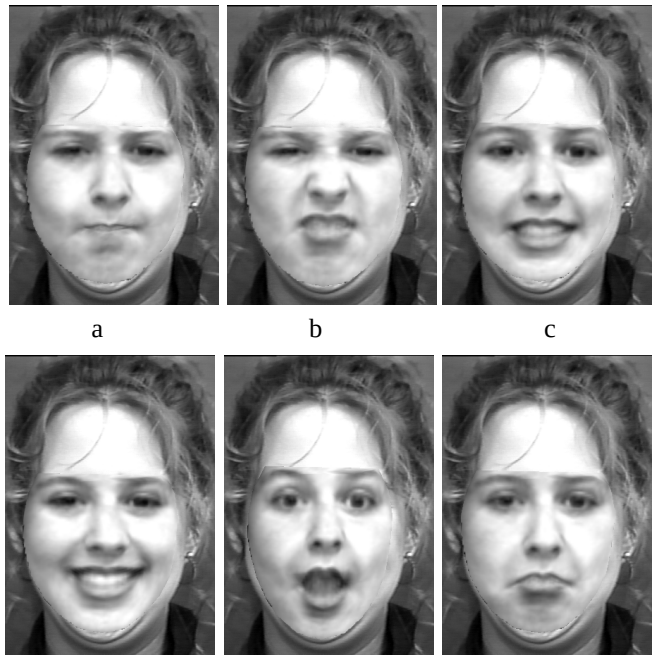
Le vecteur expressif moyen estimé à cette intensité est donné par :

$$\mathbf{c_{e'}}(\mathcal{I}'_{est}) = \mathbf{a_{e'0}} + \mathbf{a_{e'1}}\mathcal{I}'_{est}. \tag{16}$$

Le nouveau résidu $\mathbf{c}_{res}$ est donné par :

$$\mathbf{c}_{res} = \mathbf{c}_{neutre} - \mathbf{c_{e'}}(\mathcal{I}'_{est}). \tag{17}$$

L'intensité de l'expression faciale représentée par le vecteur $\mathbf{c_{e'}}(\mathcal{I}'_{est})$ peut être contrôlée à l'aide du paramètre $\mathcal{I}$ dans l'équation (8). En particulier, il est possible de générer une expression intense en posant $\mathcal{I} = 1$.
Il est alors possible de synthétiser une expression intense comme montré sur la figure 8.



a b c

d e f

FIG. 8 – *Génération de six expressions faciales synthétiques par le modèle linéaire direct en partant du visage neutre synthétique de la figure 7.c. a : colère. b : dégoût. c : peur. d : joie. e : surprise. f : tristesse.*

4.2 Un modèle linéaire évolutif

Le modèle linéaire évolutif proposé consiste à déterminer pour chaque expression faciale $\mathbf{e}$ les paramètres de l'équation :

$$\boldsymbol{c_{op}} - \boldsymbol{c_{op_n}} = \boldsymbol{a_{e1}}\mathcal{I} + \boldsymbol{\varepsilon}, \tag{18}$$

$\boldsymbol{c_{op}} - \boldsymbol{c_{op_n}}$ étant la différence entre les paramètres d'apparence optimaux $\boldsymbol{c_{op}}$ permettant de synthétiser un visage ressemblant au visage cible et les paramètres $\boldsymbol{c_{op_n}}$ permettant de synthétiser ce même visage avec une expression neutre. Comme dans l'équation (8), $\mathcal{I}$ est un scalaire variant de 0 à 1. Le vecteur de coefficients $\mathbf{a_{e1}}$ dépend de l'expression faciale affichée et est appris par régression linéaire sur un ensemble d'apprentissage initial en résolvant le système suivant :

$$\mathbf{Z} = \mathbf{X}\boldsymbol{a}_{\boldsymbol{e1}}^{T} + \boldsymbol{\varepsilon}, \tag{19}$$

où $\mathbf{Z}$ est constituée par les $2m$ premières lignes de la matrice $\mathbf{Y}$ décrite en 4.1 dont on a soustrait les paramètres codant les mêmes visages avec une expression neutre.
La deuxième colonne de $\mathbf{X}$ contient la valeur 1 dans les m premières lignes et 0.5 dans les m lignes suivantes. La première colonne de $\mathbf{X}$ est composée de valeurs 1.
Pour un visage expressif inconnu on peut écrire l'équation :

$$\boldsymbol{c_{op}} - \boldsymbol{c_{op_n}} = \boldsymbol{a_{e1}}\mathcal{I}, \tag{20}$$

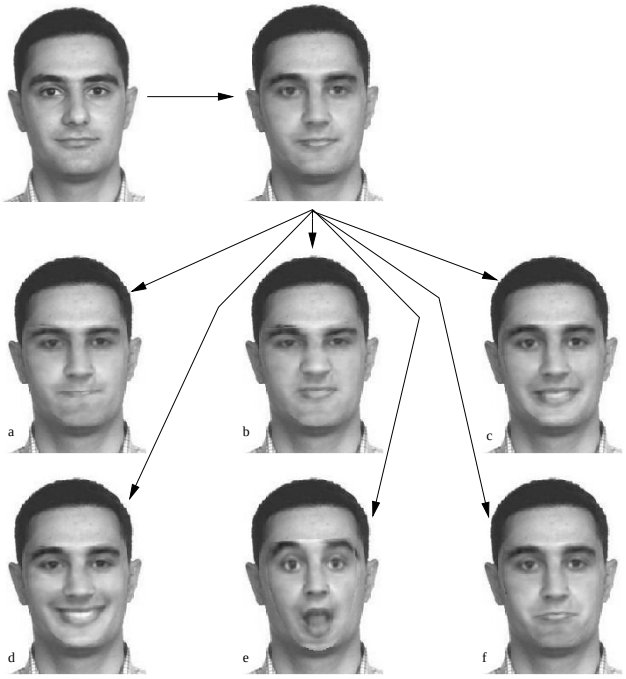


FIG. 9 – *Génération de six expressions faciales synthétiques par le modèle linéaire évolutif en partant d'uu visage neutre. a : colère. b : dégoût. c : peur. d : joie. e : surprise. f : tristesse.*

$\boldsymbol{c_{op_n}}$ étant le vecteur codant le visage neutre correspondant au visage expressif codé par $\boldsymbol{c_{op}}$. Il est possible d'annuler l'expression faciale en supposant que l'expression affichée ainsi que son intensité sont connues (a priori $\mathcal{I} = 1$ pour une expression intense).

De la même manière, il est possible de synthétiser une nouvelle expression à partir d'un visage neutre, à l'aide de l'équation :

$$\boldsymbol{c_{op_e}} = \boldsymbol{c_{op_n}} + \boldsymbol{a_{e1}}\mathcal{I}, \tag{21}$$

$\boldsymbol{c_{op_n}}$ étant le vecteur de paramètres d'apparence codant le visage neutre inconnu.

Le modèle linéaire évolutif proposé en 4.2 constitue une approche plus directe pour la synthèse d'expressions faciales. Dans cette approche, les hyperplans permettant de modéliser chacune des six expressions faciales de base passent tous par la même origine; alors que dans la méthode directe décrite en 4.1 chaque hyperplan caractérisant une expression faciale possède une origine propre dépendant de cette expression. En ce qui concerne la qualité visuelle des images synthétisées, les performances du modèle évolutif (Fig. 9) sont comparables, voire légèrement supérieures, à celles du modèle classique (Fig. 8).

4.3 Evolution de l'expression sur une séquence vidéo

Afin d'analyser le comportement temporel des modèles linéaires obtenus en 4.1 et 4.2, une série d'expériences sera effectuée sur un ensemble de 8 vidéos montrant une évolution graduelle de l'expression faciale allant d'une expression neutre à une expression intense pour différentes personnes.

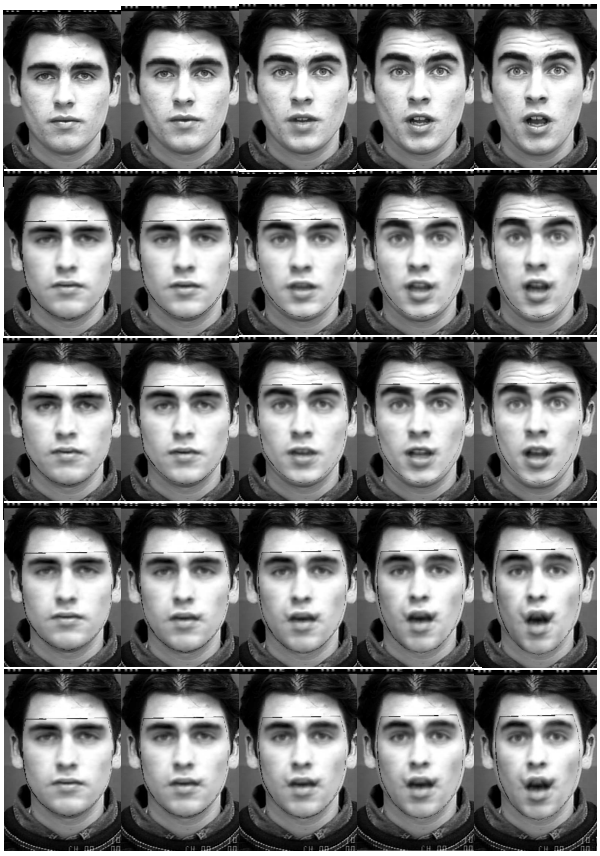


FIG. 10 – *Première ligne : visage initial surpris. Deuxième ligne : adaptation du modèle d'apparence standard. Troisème ligne : adaptation du modèle d'apparence modifié. Le modèle actif d'apparence peut être utilisé pour effectuer un suivi de visages expressifs dans une séquence vidéo. Quatrième ligne : visage surpris synthétique linéairement prédit à l'aide de l'équation (8). Dernière ligne : visage surpris synthétique linéairement prédit à l'aide de l'équation (21).*

Le modèle actif d'apparence classique est adapté sur chaque image de la séquence vidéo, en initialisant les paramètres d'apparence et de pose d'après l'adaptation de l'image précédente comme montré sur la figure 10 (deuxième ligne). Le suivi de l'expression faciale est également effectué à l'aide du modèle AAM modifié et les résultats sont montrés dans la troisième ligne de la figure 10 pour comparaison.
A chaque étape de la séquence vidéo, le vecteur de paramètres $\mathbf{c}_{op}$ obtenu permet d'estimer l'intensité de l'ex-

pression affichée $\mathcal{I}_{est}$ à l'aide de l'équation (11) ainsi que le vecteur de paramètres d'apparence linéairement prédit à cette intensité $\mathbf{c_e}(\mathcal{I}_{est})$ à l'aide de l'équation (8). De même, en supposant que la première image de la séquence correspond à un visage neutre, il est possible de calculer pour chaque image de la séquence l'évolution $\mathbf{c}_{op} - \mathbf{c}_{op_n}$ par rapport au visage neutre $\mathbf{c}_{op_n}$ initial.

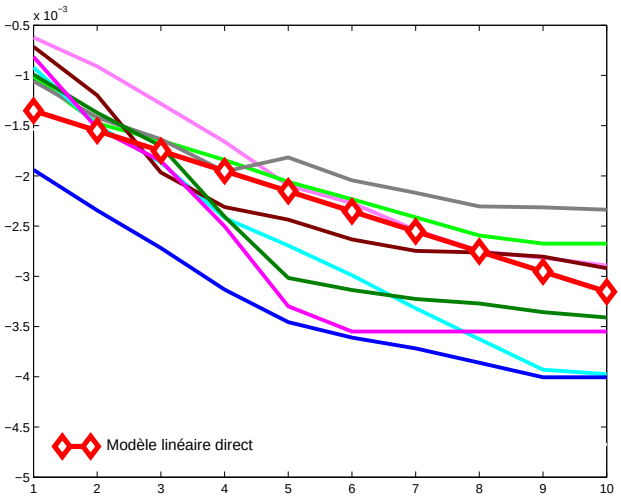


FIG. 11 – *Pour une expression de surprise, évolution du premier mode de* $\mathbf{c_e}(\mathcal{I}_{est})$ *pour chaque séquence vidéo. Ligne droite : Premier mode du modèle linéaire* ***direct****.*

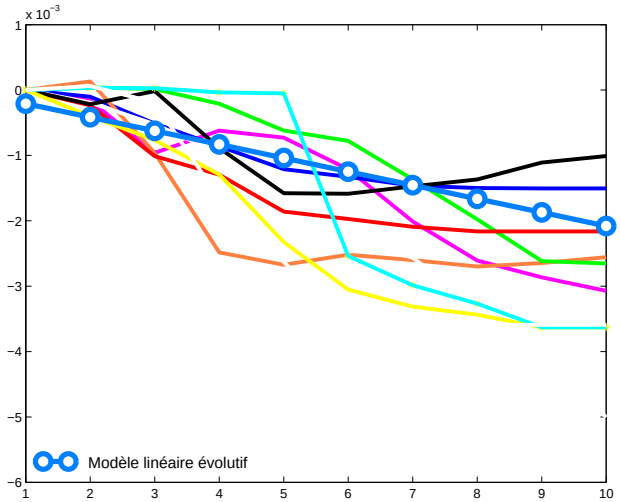


FIG. 12 – *Pour une expression de surprise, évolution du premier mode de* $\mathbf{c}_{op} - \mathbf{c}_{op_n}$ *pour chaque séquence vidéo. Ligne droite : Premier mode du modèle linéaire* ***évolutif****.*

Nous remarquons que pour chaque expression faciale, les paramètres $\mathbf{c_e}(\mathcal{I}_{est})$ tendent à suivre une trajectoire bien définie pouvant être approximée par le modèle linéaire calculé en 4.1. Ceci est illustré dans la figure 11 montrant l'évolution du premier coefficient de $\mathbf{c_e}(\mathcal{I}_{est})$ sur 8 séquences vidéo. Les séquences montrent des visages évoluant d'une expression neutre vers une expression intense de surprise. La ligne droite montre l'évolution de la droite $\mathbf{c_e}(\mathcal{I})$, avec $\mathcal{I}$ variant linéairement de 0 à 1.

De même, les paramètres $\mathbf{c}_{op} - \mathbf{c}_{op_n}$ tendent à suivre une trajectoire bien définie pouvant être approximée par le modèle linéaire calculé en 4.2. Ceci est illustré dans la figure 12 montrant l'évolution du premier coefficient de $\mathbf{c}_{op} - \mathbf{c}_{op_n}$. La ligne droite montre l'évolution de la droite $\boldsymbol{a}_{\boldsymbol{e1}}\mathcal{I}$, avec $\mathcal{I}$ variant linéairement de 0 à 1.

5 Conclusion

Nous avons présenté dans cet article une alternative au modèle actif d'apparence pour la représentation de visages expressifs à l'aide d'une seule ACP. Nous avons également montré l'interêt de ces représentations pour la reconnaissance d'expressions faciales. Le pourcentage global de bonne reconnaissance se situe aux alentours de 80% pour les deux représentations.

Deux approches pour la modélisation d'expressions faciales par régression linéaire sur un ensemble d'apprentissage ont également été abordées. Les modèles linéaires direct et évolutif ainsi obtenus ont été utilisés pour le filtrage et la synthèse d'expressions faciales. Le modèle évolutif proposé constitue une approche plus directe pour la synthèse d'expressions faciales. En effet, dans ce cas, les hyperplans permettant de modéliser chacune des six expressions faciales de base passent tous par la même origine; alors que pour le modèle direct chaque hyperplan caractérisant une expression faciale possède une origine propre dépendant de cette expression. La qualité visuelle des images synthétisées par ces deux modèle est sensiblement comparable.

Nous étudions actuellement l'extension des approches proposées pour la synthèse de mélanges d'expressions. Nous étudions également l'utilisation de mesures de similarité robustes pour rendre l'adaptation du modèle AAM moins dépendante des occlusions et des variations d'illumination.

Références

[1] M. S. Bartlett, G. Littlewort, I. Fasel, J. R. Movellan, Real Time Face Detection and Facial Expression Recognition: Development and Applications to Human Computer Interaction, *IEEE workshop on Computer Vision and Pattern Recognition for Human Computer Interaction (in conjunction with IEEE Intl. Conf. on CVPR)*, Madison, U.S.A., June, 2003

[2] M. J. Black and Y. Yacoob, Recognizing Facial Expressions in Image Sequences Using Local Parametrized Models of Image Motion, *International Journal of Computer Vision*, Vol. 25, Number 1, pp. 23–48, 1997.

[3] E. Boyle, A.H. Anderson, and A. Newlands, The Effects of Visibility on Dialogue in a Cooperative Pro-

blem Solving Task, *Language and Speech*, Vol. 37, pp. 1–20, 1994.

[4] C. Bregler, M. Covel, and M. Slaney, Video Rewrite: Driving Visual Speech with Audio, *Siggraph proceedings*, pp. 353–360, 1997.

[5] D.T. Chen, A. State, and D. Banks, Interactive Shape Metamorphosis, *Siggraph proceedings*, pp. 43–44, 1995.

[6] E. S. Chuang, H. Deshpande and C. Bregler, Facial Expression Space Learning.

[7] J. Cohn, A. Zlochower, J. J. Lien, Y. T. Wu, and T. Kanade, Automated Face Coding: A Computer Vision Based Method of Facial Expression Analysis, *European Conference on Facial Expression Measurement and Meaning*, pp. 329–333, 1997.

[8] J. Cohn, A. Zlochower, and J. J. Lien and T. Kanade, Feature-point Tracking by Optical Flow Discriminates Subtle Differences in Facial Expression, *Proceedings of the 3rd IEEE International Conference on Automatic Face and Gesture Recognition*, pp. 396–401, 1998.

[9] T.F. Cootes and G.J. Edwards and C.J. Taylor, Active Appearance Models, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pp. 681–685, 2001.

[10] T.F. Cootes and P. Kittipanya-ngam, Comparing Variations on the Active Appearance Model Algorithm, *British Machine Vision Conference*, pp. 837–846, 2002.

[11] T.F. Cootes and K. Walker and C.J. Taylor, View-Based Active Appearance Models, *International Conference on Automatic Face and Gesture Recognition*, pp. 227–232, 2000.

[12] B. Draper, K. Baek, M.S. Bartlett and R. Beveridge, Recognizing Faces with PCA and ICA, *Computer Vision and Image Understanding*.

[13] I. L. Dryden, Kanti V. Mardia, *Statistical Shape Analysis*, Wiley, 1998.

[14] S. Dubuisson and F. Davoine and M. Masson, A Solution for Facial Expression Representation and Recognition, *Signal Processing: Image Communication*, Vol. 14, no. 9, pp. 657–673, 2002.

[15] G.J. Edwards, T.F. Cootes, and C.J. Taylor, Face Recognition Using Active Appearance Models, *Proceedings of the European Conference of Computer Vision*, pp. 581–695, 1998.

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

[17] P. Ekman and W. Friesen, Facial Action Coding System: A Technique for the Measurement of Facial Movement.

[18] I. Essa and A. Pentland, Coding, Analysis Interpretation Recognition of Facial Expressions, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 19, Number 7, pp. 757–763, 1997.

[19] T. Ezzat, G. Geiger and T. Poggio, Trainable Videorealistic Speech Animation, *ACM Transactions on Graphics (TOG)*, Vol. 21, Number 3, pp. 388–398, 2002.

[20] B. Heisele, P. Ho and T. Poggio, Face Recognition with Support Vector Machines: Global Versus Component-based Approach, , *International Conference on Computer Vision (ICCV'01)*, Vancouver, Canada, Vol. 2, pp. 688–694, 2001.

[21] M. Hoch, G. Fleischmann, and B. Girod, Modeling and Animation of Facial Expressions based on B-Splines, *The Visual Computer*, pp. 87–95, 1994.

[22] H. Hong, H. Neven and C. von der Malsburg, Online Facial Expression Recognition based on Personalized Gallery, *Intl. Conference on Automatic Face and Gesture Recognition*, IEEE Comp. Soc, pp. 354–359, 1998.

[23] T. Kanade and J. Cohn and Y.L. Tian, Comprehensive Database for Facial Expression Analysis, *International Conference on Automatic Face and Gesture Recognition*, pp. 46–53, 2000.

[24] H. Kang, T.F. Cootes and C.J. Taylor, Face Expression Detection and Synthesis using Statistical Models of Appearance, *Measuring Behavior*, pp. 126–128, 2002.

[25] M.J. Lyons and J. Budynek and S. Akamatsu, Automatic Classification of Single Facial Images, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 21, pp. 1357–1362, 1999.

[26] A. Mehrabian, Communication without Words, *Psychology Today*, Vol. 2, Number 4, pp. 53–56, 1968.

[27] C. Padgett, G. Cottrell and R. Adolphs, Categorical Perception in Facial Emotion Classification, *Siggraph proceedings*, pp. 75–84 1998.

[28] P. Penev and J. Atick, Local Feature Analysis: A general Statistical Theory for Object Representation, 1996.

[29] F. Pighin, J. Hecker, D. Lischinski, R. Szeliski, D.H. Salesin, Synthesizing Realistic Facial Expressions from Photographs, *Proceedings of Cognitive Science Conference*, 1996.

[30] M. H. Yang, Face Recognition Using Kernel Methods, *Advances in Neural Information Processing Systems 14 (NIPS 14)*, pp. 215–220, 2002.

[31] M. H. Yang, D. Kriegman, and N. Ahuja, Detecting Faces in Images: a Survey, *IEEE transactions on pattern analysis and machine intelligence*, January 2002.

[32] W. Zhao, R. Chellappa, A. Rosenfeld and P.J. Phillips. Face recognition: A literature survey. *CVL Technical Report, University of Maryland,*, October 2000.