Création de l'espace des expressions faciales à partir de modèles bilinéaires asymétriques

Catherine SOLADIÉ[1], Nicolas STOIBER[2], Renaud SÉGUIER[1]

[1]SUPELEC/IETR, équipe SCEE
Avenue de la Boulaie, 35576 Cesson-Sévigné, France

[2]Dynamixyz
80 avenue des Buttes de Coesmes, 35700 Rennes, France
catherine.soladie@supelec.fr, nicolas.stoiber@dynamixyz.com
renaud.seguier@supelec.fr

Résumé – Ce papier étudie l'analyse des expressions non prototypiques et non incluses dans les bases d'apprentissage. La méthode est basée sur un modèle bilinéaire asymétrique appris sur une petite quantité d'expressions. Dans l'espace des expressions ainsi créé, une expression inconnue a une signature qui peut être interprétée comme un mélange des expressions de bases utilisées lors de la construction de l'espace. Trois méthodes sont comparées : une méthode traditionnelle basée sur des vecteurs d'apparence, le modèle bilinéaire asymétrique sur des vecteurs d'apparence indépendants des sujets et le modèle bilinéaire asymétrique sur des vecteurs d'apparence spécifiques aux sujets. Les résultats expérimentaux sur 14 expressions inconnues montrent la pertinence des modèles bilinéaires ainsi que la robustesse des vecteurs d'apparences spécifiques aux sujets.

Abstract – This paper analyzes non prototypic expressions. The method is based on an asymmetric bilinear model learned on a small amount of expressions. In the resulting expression space, a blended unknown expression has a signature, that can be interpreted as a mixture of the basic expressions used in the creation of the space. Three methods are compared. Experimental results on the recognition of 14 blended unknown expressions show the relevance of the bilinear models compared to appearance-based methods and the robustness of the person-specific models according to the types of parameters (shape and/or texture).

1 Introduction

L'analyse des expressions du visage s'est fortement développée ces dernières décennies, notamment dans le but de reconnaitre des émotions. En effet, les travaux d'Ekman [1] ont montré que les 6 émotions de base étaient exprimées de façon universelle par 6 expressions faciales. Jusqu'à récemment, la plupart des systèmes se focalisaient sur la reconnaissance de ces 6 expressions et atteignaient de très bon taux de reconnaissance [2]. Néanmoins, dans la vie de tous les jours, il est rare que ces expressions soient aussi franches et c'est plus généralement des expressions mélangées qui sont réalisées (figure 1). L'analyse des telles expressions et leur interprétation en termes d'émotion restent encore un sujet de préoccupation comme le montrent les résultats du challenge AVEC 2012 [3]. L'espace des expressions étant grand et continu, il n'est pas envisageable d'apprendre l'ensemble des expressions possibles. Notre système possède donc une base d'apprentissage restreinte (8 expressions émotionnelles et visage neutre pour chaque sujet).
Une méthode largement répandue pour décrire les expressions faciales nous vient de la psychologie. Il s'agit du système FACS [4] qui permet de définir une expression comme une somme d'unité d'action. Néanmoins, cette représentation ne permet pas de mettre en évidence le caractère continu de l'espace des expressions et semble mal adaptée aux traitements informatiques. Pour remédier à cela, certains systèmes proposent la représentation des expressions sous forme de variété mathématique [5, 6]. Ces systèmes nécessitent un grand nombre de données dans leur base d'apprentissage. Cheon & Kim [7] ont proposé de s'affranchir de la morphologie des sujets en soustrayant les données du visage neutre aux paramètres d'apparence du visage (Diff-AAM), présumant ainsi que les déformations sont identiques entre les personnes. D'autres systèmes utilisent la décomposition bilinéaire [8, 9, 10]. Les modèles bilinéaires ont pour avantage de nécessiter peu de données. Néanmoins, ils n'ont jamais été testés dans le cas d'expressions non prototypiques. C'est cette étude que nous proposons de mener dans cet article.
La principale contribution est l'application de la décomposition bilinéaire asymétrique pour l'analyse des expressions non contenues dans les bases d'apprentissage. L'originalité de l'approche consiste à déterminer une signature unique pour une expression mélangée via un modèle bilinéaire appris sur un nombre limité d'expressions et à donner un sens aux différentes composantes de la signature obtenue. La pertinence de cette signature est mesurée par un algorithme simple de reconnaissance et le taux de reconnaissance est comparé aux méthodes traditionnelles basées sur l'apparence. Une seconde originalité

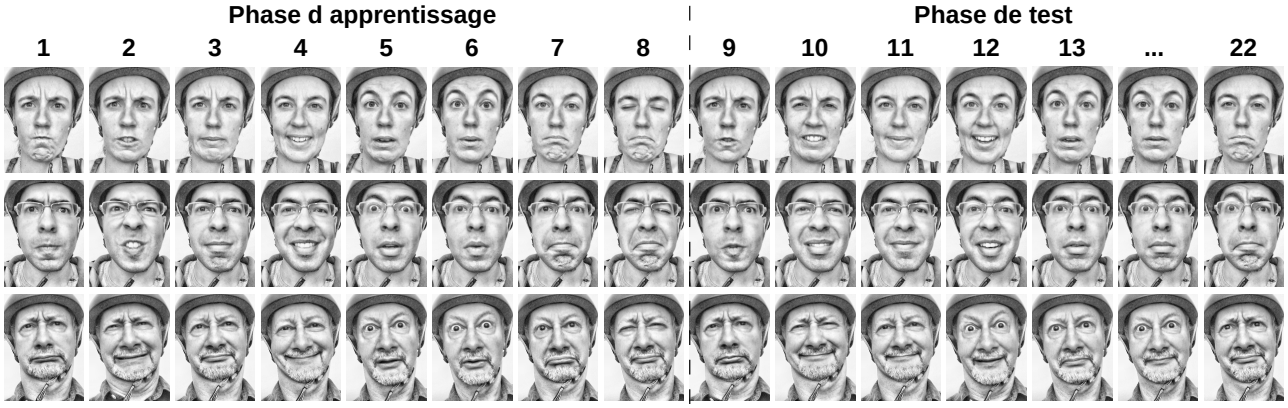

FIGURE 1 – Expressions similaires réalisées par différents sujets.

consiste à appliquer les modèles bilinéaires sur des vecteurs d'apparence spécifiques à la personne (et non sur des vecteurs d'apparence génériques). Nous verrons que cette solution permet de rendre la méthode plus robuste aux types de vecteurs d'apparence (forme et/ou texture des visages).

2 Analyse d'une expression mélangée inconnue par décomposition bilinéaire

2.1 Vecteurs d'apparence

Nous utilisons le principe des modèles actifs d'apparence (AAM) [11] pour définir les vecteurs d'apparence. Chaque image $\mathbf{i}$ est annotée avec plusieurs points caractéristiques (par exemple le coin gauche de la lèvre). Pour chaque image $\mathbf{i}, \mathbf{i} = \mathbf{1}..\mathbf{N}$, ces points caractéristiques sont concaténés dans un vecteur $\mathbf{s_i}$, qui représente la forme du visage. L'intensité des pixels contenus dans la zone du visage forme un vecteur $\mathbf{g_i}$, qui représente la texture. Afin de détecter les déformations de forme et de texture, une analyse en composantes principales (ACP) est réalisée sur chacun des deux vecteurs : $\mathbf{s_i} = \overline{\mathbf{s}} + \mathbf{\Phi_s}.\mathbf{b_i^s}$ et $\mathbf{g_i} = \overline{\mathbf{g}} + \mathbf{\Phi_t}.\mathbf{b_i^t}$. $\mathbf{s_i}$ et $\mathbf{g_i}$ sont appelés les vecteurs d'apparence de forme et de texture. Pour prendre en compte la corrélation entre la forme et la texture du visage, une troisième ACP est réalisée sur un vecteur qui concatène les vecteurs de forme et de texture $\mathbf{b_i} = [w_s.\mathbf{b_i^s}|\mathbf{b_i^t}]$ (w_s est un facteur permettant d'assurer que la forme et la texture ont des variances comparables) : $\mathbf{b_i} = \mathbf{\Phi}.\mathbf{c_i}$. $\mathbf{c_i}$ est le vecteur d'apparence (forme + texture).

2.2 Modèles bilinéaires

Etant donné un corpus de E expressions faciales connues de P personnes, nous souhaitons décomposer les paramètres d'apparence $\mathbf{y}^{pe}$ de dimension K en une signature de l'expression $\mathbf{b}^e$ commune à l'ensemble des sujets de dimension E et un mapping linéaire spécifique au sujet $\mathbf{W}^p$ de dimension $K.E$: $\mathbf{y}^{pe} = \mathbf{W}^p.\mathbf{b}^e$; soit sous forme matricielle : $\mathbf{Y} = \mathbf{W}.\mathbf{B}$.
La décomposition en valeurs singulières (SVD) de la matrice Y permet de répondre à ce problème. Par SVD, $\mathbf{Y} = \mathbf{U}.\mathbf{S}.\mathbf{V}^t$. $\mathbf{W}$ est alors donné par $\mathbf{U}.\mathbf{S}$ et $\mathbf{B}$ par $\mathbf{V}^t$. Chaque vecteur $\mathbf{b}^e, e = 1..E$ de $\mathbf{B}$ représente la signature de l'expression $\mathbf{y}^{pe}$. Elle est identique pour tous les sujets de la base d'apprentissage ($p = 1..P$).

2.3 Calcul de la signature d'une expression mélangée inconnue d'une personne connue

Les paramètres d'apparence des expressions des différents sujets sont issus d'un modèle actif d'apparence (AAM) générique, appris sur E expressions connues et similaires plus le visage neutre des P sujets.
La **phase d'apprentissage** est réalisée par la décomposition bilinéaire asymétrique présentée précédemment sur ces mêmes E expressions faciales connues des P personnes.
Nous souhaitons **analyser une expression inconnue mélangée** $\mathbf{y}^{P_i e}$ d'une personne P_i de la base d'apprentissage (expression autre que l'une des E expressions de la phase d'apprentissage). La signature $\mathbf{b}^e$ de l'expression $\mathbf{y}^{P_i e}$ est estimée pour les paramètres $\mathbf{W}^{P_i}$ du modèle d'apprentissage par : $\mathbf{y}^{p_i e} = \mathbf{W}^{p_i}.\mathbf{b}^e$.
La matrice $\mathbf{B}$ est par construction une matrice de vecteurs orthonormaux et peut donc être **interprétée** comme une base orthonormale. Chaque signature peut alors être caractérisée dans cette base, sous forme d'une signature relative $\mathbf{b}'$ par rapport aux E expressions de base utilisées dans la phase d'entrainement. Par rotation, nous avons $\mathbf{b}' = \mathbf{B}^t.\mathbf{b}$. Ainsi, les paramètres peuvent être interprétés comme une expression mélangée entre plusieurs expressions connues.

Afin de montrer la pertinence de la signature obtenue, **plusieurs méthodes** sont comparées. Tout d'abord, deux méthodes traditionnelles basées apparence : ABM (Méthode Basée sur des paramètres d'Apparence), pour laquelle les paramètres d'apparence sont directement utilisés pour caractériser l'expression ; et DABM (Méthode Basée sur des paramètres d'Apparence Différentiels), pour laquelle les données du visage neutre sont soustraites des paramètres d'apparence du visage, afin d'atténuer les caractéristiques morphologiques des sujets. Ensuite, deux méthodes bilinéaires : BDM (Méthode de Décomposi-

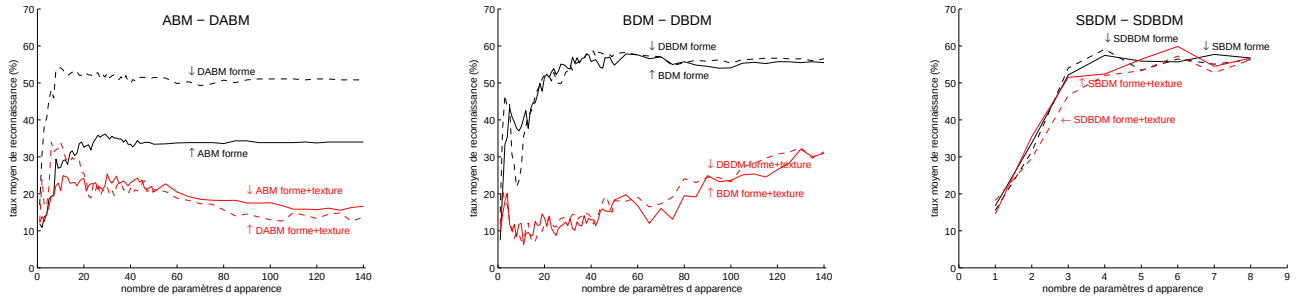

FIGURE 2 – Comparaison des méthodes ABM, DABM, BDM, DBDM, SBDM and SDBDM avec ou sans l'information de texture. Taux de reconnaissance de 14 expressions faciales inconnues sur des sujets connus, selon la taille des vecteurs d'apparence. En pointillés, les méthodes sur les paramètres différentiels ; en noir, sur les vecteurs de forme uniquement ; en rouge, sur les vecteurs de forme et de texture.

tion Bilinéaire), lorsque les calculs présentés dans ce papier sont réalisés sur les paramètres d'apparence ; DBDM (Méthode de Décomposition Bilinéaire sur des paramètres Différentiels), lorsqu'ils sont réalisés sur les paramètres d'apparence différentiels. Comme nous n'utilisons que la propriété de linéarité sur les expressions, nous pouvons aussi appliquer la décomposition bilinéaire asymétrique sur des paramètres d'apparence issus de modèles spécifiques au sujet. Pour chaque sujet, nous calculons alors un modèle d'apparence spécifique (AAM) appris sur quelques expressions du sujet (les mêmes que pour les autres méthodes). Cela permet de comparer finalement deux autres méthodes bilinéaires : SBDM (Méthode de Décomposition Bilinéaire sur des paramètres d'apparence Spécifiques au sujet) et SDBDM (Méthode de Décomposition Bilinéaire sur des paramètres d'apparence Différentiels Spécifiques au sujet).

3 Résultats expérimentaux

Les expérimentations ont été réalisées sur une base [1] contenant 22 expressions mélangées similaires plus neutre de 17 sujets. Les modèles ont été entrainés sur 8 expressions plus neutre des 17 sujets. Les 14 autres expressions mélangées restantes ont été utilisées pour tester les modèles.
Afin de mesurer la pertinence des espaces des expressions créés par chaque méthode, le même **algorithme de comparaison** est mis en œuvre. L'objectif est d'avoir la même signature pour chaque expression similaire de sujets différents. Pour chaque expression $\mathbf{i}$ d'un sujet $\mathbf{p}$, nous recherchons donc l'expression $\mathbf{j}$ la plus proche de chaque autre sujet $\mathbf{s} \neq \mathbf{p}$ (Nearest Neighbor Search). Le label de l'expression $\mathbf{i}$ du sujet $\mathbf{p}$ est celui trouvé le plus fréquemment parmi les autres sujets :
$\mathbf{max_s}|\mathbf{min_j}[\mathbf{dist}(\mathbf{b_{i,p}}, \mathbf{b_{j,s}})|$.
Le tableau 1 montre les meilleurs taux de reconnaissance obtenus pour chacune des méthodes. Comme attendu, la méthode ABM donne les moins bons résultats. En effet, les paramètres d'apparence contiennent à la fois des informations d'expression mais aussi de morphologie du sujet. Les méthodes de décomposition bilinéaire donnent de meilleurs résultats que la méthode DABM. Cela peut s'interpréter par le fait que les paramètres d'apparence différentiels contiennent encore une quantité non négligeable d'information spécifique au sujet. En revanche, les modèles bilinéaires apprennent, sur les expressions utilisées dans la phase d'apprentissage, la façon qu'à chaque sujet de réaliser une expression. Ce type de déformation est alors spécifique au sujet et se retrouve dans les expressions mélangées de la phase de reconnaissance. Nous constatons que le modèle bilinéaire appliqué sur les paramètres d'apparence différentiels donne légèrement de meilleurs résultats mais cette différence n'est pas suffisamment significative pour être analysée.

1. Base de données disponible à l'adresse `http://www.rennes.supelec.fr/immemo/`

TABLE 1 – Taux de reconnaissance moyen de 14 expressions mélangées inconnues de sujets connus.

Méthode	ABM	DABM	BDM	DBDM	SBDM	SDBDM
Forme	36.2	54.1	**57.9**	**58.7**	**57.7**	**59.1**
Forme+Text.	25.4	34.0	32.1	32.5	**59.9**	**57.3**

La figure 2 montre pour chaque méthode le taux de reconnaissance selon la taille des vecteurs d'apparence. Nous constatons que, pour toutes les méthodes utilisant des modèles d'apparence génériques, l'ajout de la texture diminue les taux de reconnaissance, et que cela n'est pas le cas pour les méthodes utilisant des modèles d'apparence spécifiques aux sujets. Cela peut s'interpréter par le fait que les variations de texture sont principalement dues aux variations de morphologie dans les modèles d'apparence génériques, alors qu'elles sont dues aux variations d'expressions dans les modèles d'apparence spécifiques aux sujets.
L'analyse, sur des exemples, des composantes des signatures obtenues permet de définir une expression mélangée inconnue comme étant le mélange de plusieurs expressions connues et valide ainsi l'interprétation de la signature d'une expression (voir figure 3).

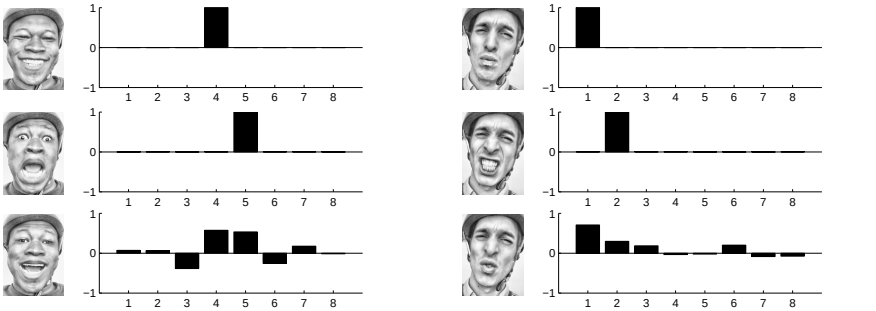

FIGURE 3 – Signification de la signature d'une expression. Sur la gauche, deux expressions connues (sourire - expression 4 et surprise - expression 5) et une expression inconnues (mélange de sourire et de surprise - expression 12). Sur la droite, deux expressions connues (colère de faible intensité - expression 1 et de forte intensité - expression 2) et une expression inconnues (colère d'intensité moyenne - expression 9).

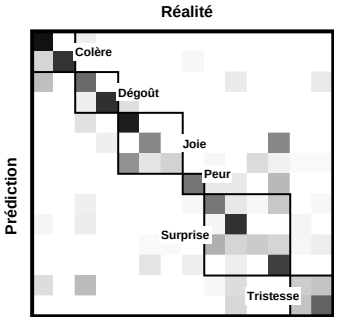

FIGURE 4 – Matrice de confusion des 14 expressions mélangées inconnues sur 17 sujets (DABM, caractéristiques de forme et de texture, 6 paramètres). Plus la case est foncée, plus le taux de reconnaissance est élevé.

Pour finir, la figure 4 montre la matrice de confusion obtenue pour la méthode ayant de meilleurs taux de reconnaissance. Nous constatons que les confusions apparaissent principalement lorsque les expressions correspondent aux mêmes types d'émotions. Pour certaines expressions, nous constatons aussi la confusion classique entre les expressions de colère et de dégoût et entre les expressions de peur et de surprise. A noter que les 8 autres expressions (expressions connues ayant servi à l'apprentissage) sont reconnues par construction avec un taux de reconnaissance de 100%.

4 Conclusion

Cet article présente une méthode de reconnaissance d'expressions mélangées basée sur la décomposition bilinéaire asymétrique. La méthode proposée montre de meilleurs résultats que les méthodes traditionnelles basées sur l'apparence. Elle propose une signature qui peut être interprétée : la valeur des composantes d'une expression mélangée indique les expressions de base de la phase d'apprentissage qui sont mélangées. Dans le futur, nous comptons étendre la méthode aux personnes inconnues du système.

Références

[1] Ekman, P. and Friesen, W. V and Ellsworth, P.. *Emotion in the human face*. Cambridge University Press New York, 1982.

[2] Valstar, M. F and Jiang, B and Mehu, M and Pantic, M and Scherer, K *The first facial expression recognition and analysis challenge*. Automatic Face & Gesture Recognition and Workshops (FG 2011), IEEE International Conference on, 2011

[3] Schuller, B. and Valster, M. and Eyben, F. and Cowie, R. and Pantic, M. *AVEC 2012 : the continuous audio/visual emotion challenge*. Proceedings of the 14th ACM international conference on Multimodal interaction, 2012

[4] Ekman, P. and Friesen, W. V. *Facial action coding system : A technique for the measurement of facial movement*. Consulting Psychologists Press, Palo Alto, CA, 1978.

[5] Stoiber, N. and Seguier, R. and Breton, G. *Automatic design of a control interface for a synthetic face*. International Conference on Intelligent user interfaces, 2009.

[6] Shan, C. and Gong, S. and McOwan, P. W. *Appearance manifold of facial expression*. Computer Vision in Human-Computer Interaction, 2005.

[7] Cheon, Y. and Kim, D.. *Natural facial expression recognition using differential-AAM and manifold learning*. Pattern Recognition, 2009.

[8] Wang, H. and Ahuja, N.. *Facial expression decomposition*. Computer Vision, 2003. Proceedings. Ninth IEEE International Conference on, 2003.

[9] Abboud, B. and Davoine, F.. *Bilinear factorisation for facial expression analysis and synthesis*. Vision, Image and Signal Processing, IEE Proceedings-, 2005.

[10] Mpiperis, I. and Malassiotis, S. and Strintzis, M. G.. *Bilinear elastically deformable models with application to 3d face and facial expression recognition*. Automatic Face & Gesture Recognition, 2008. FG'08.

[11] Cootes, T.F. and Edwards, G.J. and Taylor, C.J.. *Active appearance models*. Pattern Analysis and Machine Intelligence, IEEE Transactions on, 2001.

Cette recherche a été réalisée avec le support du projet REPLICA (ARN Techsan).