Caractérisation locale des changements de texture pour la reconnaissance d'expressions faciales spontanées

Walid Adaidi, Adel Lablack, Ioan Marius Bilasco

Laboratoire d'Informatique Fondamentale de Lille, Université de Lille 1, France

Résumé

Malgré les avancées récentes, la reconnaissance des émotions et des expressions faciales reste un challenge intéressant. Dans cet article, une approche permettant la reconnaissance d'expressions faciales spontanées grâce à une représentation appropriée des traits du visage sur des flux vidéo et des images statiques est proposée. Une mesure sensible aux changements dans les traits du visage est utilisée dans des régions d'intérêt identifiées pour détecter la présence de chaque expression de base. L'expérimentation a été réalisée sur un ensemble de données standard composées de vidéos et d'images statiques et a montré des résultats prometteurs.

Recognizing human facial expression and emotion by computer is an interesting and challenging problem. In this paper, a method for recognizing spontaneous facial expressions through an appropriate representation of facial features from relevant face regions displayed in video streams and still images is proposed. A measure that is sensitive to facial movements is used in predefined regions of interest to detect the basic emotions. The experimentation has been performed on a standard dataset composed of video streams and static images and has showed promising results.

Mots clé : Reconnaissance d'expressions faciales spontanées, approche locale, régions d'intérêt

1. Introduction

La reconnaissance automatique des expressions faciales est un sujet de recherche actif. Les techniques proposées actuellement pour la détection et le suivi du visage, l'extraction des caractéristiques du visage et les méthodes utilisées pour la classification des expressions sont plus robustes qu'auparavant. L'expression faciale peut être utilisée comme un module important dans les interactions homme-machine ou pour étudier le comportement des personnes. L'interprétation des expressions faciales diffère d'un domaine d'application à un autre. C'est une tâche difficile qui permet au système d'être réactif et d'améliorer l'expérience utilisateur.

Dans un magasin par exemple, un retour positif (un sourire) peut être interprété comme un signe d'intérêt, alors qu'une grimace pourrait être interprétée comme un signe de répulsion. Selon la réactivité souhaitée, le système pourrait présenter plus de détails sur un produit ou changer le produit affiché à l'utilisateur. Dans les applications de e-learning, il faut considérer une échelle temporelle plus large car l'interaction de l'utilisateur avec le système ou bien avec l'enseignant est censée se dérouler en continue. Dans ce cas, des états tels que l'intérêt de l'utilisateur, l'incompréhension, ou la frustration peuvent être considérer comme des actions. Ces deux exemples montrent différentes façon de représenter l'état émotionnel de l'utilisateur.

Dans la littérature deux approches principales pour représenter les expressions sont utilisés : une représentation discrète en catégories introduite par Ekman [EF78] qui utilise six expressions faciales universelles que sont la colère, le dégoût, la peur, la joie, la tristesse et la surprise. La représentation dimensionnelle est une alternative et permet de caractériser un état affectif en termes de dimensions latentes [CD10] telles que l'évaluation, l'activation, le contrôle, la puissance, etc.

En présence d'expressions spontanées et non actées dans les domaines applicatifs visés, nous nous sommes intéressés dans cet article à la reconnaissance d'expressions faciales spontanées dans des environnements non contrôlés. En effet, l'identification des unités d'actions (AUs) à l'aide d'un modèle et leurs suivis est difficile sur le visage à cause du bruit. Ce bruit est souvent confondu avec des mouvements de très faible amplitude. Par ailleurs, dans une situation d'interactions spontanées les règles FACS proposées par Ekman [EF78] ne couvrent pas l'intégralité des situations.

A l'image de Mavadati et al. [MMB*13] qui s'intéressent à la caractérisation des expressions faciales spontanées, nous commençons par étudier le lien entre les changements de texture dans le visage et les expressions de base. En effet, nous adoptons une analyse des changements intervenus sur

des régions spécifiques de visage afin de s'affranchir du délicat problème d'extraction directe des AUs dans un environnement non contrôlé, nous adoptons une analyse des changements intervenus sur des régions spécifiques de visage. Une étude préalable entre les changements observés et les expressions a été conduite sur la base DISFA [MMB*13]. Cette base est constituée de vidéos de 27 personnes avec une annotation manuelle de l'intensité de 12 AUs et de la position de 66 points de contrôle sur le visage.

L'article est organisé comme suit. Dans la section 2, nous présentons brièvement l'état de l'art. La méthodologie proposée organisée en deux étapes est décrite dans la section 3. Nous présentons ensuite les résultats obtenus en utilisant les bases DISFA et KDEF [LFO98] dans la section 4. Enfin, nous concluons par discuter les pistes ouvertes par notre présente contribution dans la section 5.

2. État de l'art

Le domaine de la reconnaissance d'émotions a été étudié activement au cours de ces dernières années. En général, le but des systèmes proposés est de reconnaître des classes d'expressions ou bien l'état d'une dimension affective. Dans la littérature, les techniques utilisées pour détecter des expressions dans des flux vidéo peuvent être classés en deux ensembles : (i) les approches géométrique pour détecter et suivre des points caractéristiques du visage qui sont ensuite utilisés en entrée dans la classification, (ii) les approches d'apparence qui utilisent le mouvement et le changement de texture [SRS*11].

Les approches géométriques s'appuient essentiellement sur le système Facial Action Coding System (FACS) qui permet de mesurer les changements subtils dans l'apparence du visage causées par des contractions des muscles faciaux en associant une action unitaire à chaque mouvement musculaire d'une partie du visage [EF78]. Gonzalez et al. [GSEV11] ont appliqué adaboost à un ensemble de caractéristiques extrait dans les régions du visage permettant de reconnaitre les AUs. Popa et al. [PRW*11] ont mené une étude sur la reconnaissance des AUs. Ils se sont appuyés sur le principe que les yeux, le nez et la bouche contiennent beaucoup d'informations qu'il faut ensuite affiner en se basant sur un AAM [CET01]. Ils utilisent le flux optique pour extraire les informations contenues dans les régions du visage mais cette approche est sensible aux mouvements. Lablack et al. [LDBD14] se sont intéressés uniquement aux émotions négatives et ont proposé une approche locale autour de la région englobant l'AU4.

Des méthodes globales telles que celle présentée par Danisman et al. [DBMD13] qui utilisent un perceptron multicouche pour reconnaitre la joie semblent plus robustes aux bruits mais ne sont pas adaptées à toutes les expressions.

Nous proposons une approche locale pour détecter les expressions faciales tout en gardant les bénéfices des méthodes globales. Ainsi, nous étudions les changements de texture consécutifs à l'apparition de certaines AUs sur le visage. Nous proposons une analyse spécifique des régions à proximité ou englobant ces AUs sur le visage afin de détecter les changements qui apparaissent en présence d'une expression.

3. Méthodologie

Nous proposons une méthodologie qui permet de caractériser les expressions faciales spontanées de base en s'appuyant sur une analyse des parties locales du visage. Notre proposition se divise en deux étapes importantes : (i) l'étude préalable des régions qui permet d'identifier les régions d'intérêt du visage sur lesquels un changement apparait lors de l'expression d'une émotion par l'utilisateur (ii) extraire les informations caractéristiques des régions obtenues en associant une métrique à chaque expression faciale. La Figure 1 illustre le schéma proposé en deux étapes.

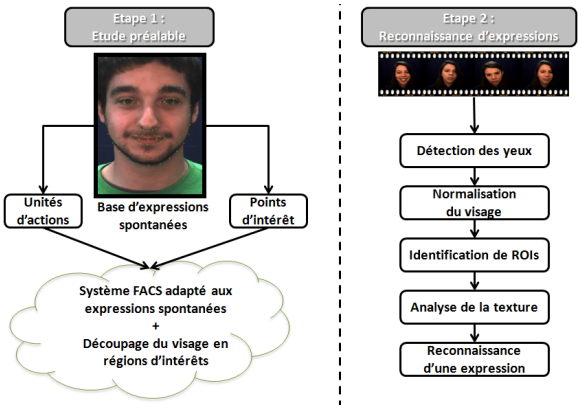
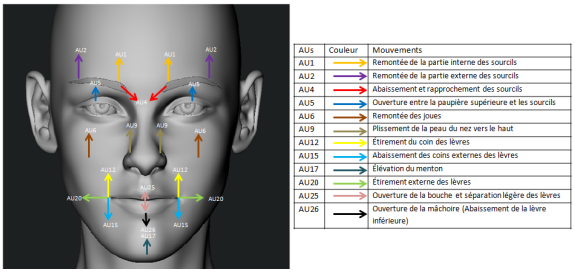


Figure 1: *Le système proposé pour la détection des expressions faciales de base.*

3.1. Étude préalable des régions

Cette étude va nous permettre d'associer à chaque expression de base une région d'intérêt à analyser. Pour cela, nous disposons des coordonnées des points d'intérêts dans la base DISFA [MMB*13]. Les points d'intérêts sont placés tout autour des composants du visage tels que le nez, les sourcils, la bouche et les yeux. Une activation d'une unité d'action (AU) entraine une déformation faciale, donc un déplacement caractérisé par la différence entre les points d'intérêt à l'instant t et à l'instant $t+1$. Nous allons utiliser la combinaison des AUs formant une expression pour définir nos régions d'intérêts. La Figure 2 illustre la position et les mouvements faciaux qui caractérisent les 12 AUs que nous utilisons pour l'identification des régions d'intérêts. Par exemple pour la partie autour des sourcils : l'AU1 correspond à la remontée de la partie interne des sourcils, l'AU2 à la remontée de la partie externe des sourcils, et l'AU4 à l'abaissement et rapprochement des sourcils.

Le système établi par Ekman reste théorique et difficile à appliquer en présence d'expressions faciales spontanées. Nous avons analysés l'activation des AUs sur la base DISFA afin d'identifier et valider le changement des traits du visage en présence d'expressions faciales spontanées. Selon Ekman qui a construit son modèle sur une base d'expressions actées, chaque expression nécessite l'activation de plusieurs AUs. Par exemple, selon Ekman, la surprise est constituée des AU1, AU2, AU5 et AU26 et l'absence d'une seule AU annule le processus de détection. Notre modèle quant à lui se



AUs	Couleur	Mouvements
AU1		Remontée de la partie interne des sourcils
AU2		Remontée de la partie externe des sourcils
AU4		Abaissement et rapprochement des sourcils
AU5		Ouverture entre la paupière supérieure et les sourcils
AU6		Remontée des joues
AU9		Plissement de la peau du nez vers le haut
AU12		Étirement du coin des lèvres
AU15		Abaissement des coins externes des lèvres
AU17		Élévation du menton
AU20		Étirement externe des lèvres
AU25		Ouverture de la bouche et séparation légère des lèvres
AU26		Ouverture de la mâchoire (Abaissement de la lèvre inférieure)

Figure 2: *Localisation des 12 AUs qui caractérisent les expressions de base.*

base sur la probabilité de présence d'une AU ou d'un groupe d'AUs. La surprise peut être détectée avec une combinaison AU1, AU2 et AU5 ou AU1, AU2 et AU26.

Nous avons menée une étude sur la cooccurrence des AUs en présence/absence d'expressions dans la base DISFA. Nous avons constaté par exemple que l'AU26 est souvent présente avec la plupart des autres AUs et se révèle être non pertinente dans la détection des expressions. Le dégout par exemple peut être caractérisé par la présence de l'AU9 avec une grande intensité alors qu'Ekman conditionne sa détection à la présence des AU15 et AU16. Ces observations nous semblent importantes dans l'étude des solutions adaptées pour la détection d'expressions spontanées car au contraire des expressions actées, l'activation de certains AUs requises par le FACS n'est que partielle ou absente. Suite à cette analyse, nous proposons une simplification du modèle classique FACS en réduisant le nombre d'AUs nécessaires à la caractérisation d'une expression faciale. La Figure 3 illustre une représentation en portes logiques des combinaisons des AUs formant une expression faciale spontanée que nous avons retenu en interprétant les données de la base DISFA.

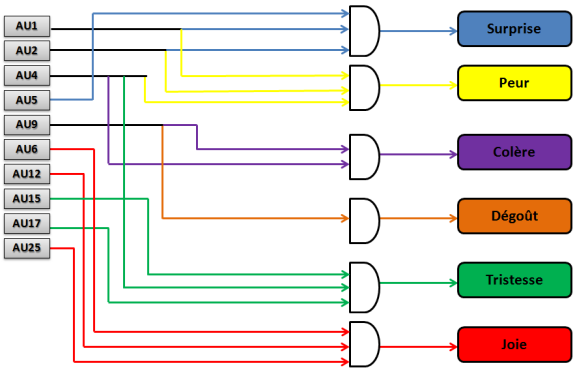


Figure 3: *La représentation en portes logiques des combinaisons des AUs formant un expression faciale.*

Une fois cette association entre les AUs et les expressions établie, nous analysons la position des points d'intérêts en présence/absence de chaque expression. Les zones sont choisies de façon à englober les points d'intérêts lors du déclenchement de l'émotion. Toutefois, nous avons effectuer une normalisation du visage qui permet de compenser les bruits dus à la variation de la position spatiale du visage. En effet, une variation entraîne le déplacement des points d'intérêt et induit donc un bruit dans l'analyse. Nous avons affiner les zones pour obtenir la région minimale qui contiendra tous les points d'intérêts. La Figure 4 illustre un découpage des régions pour les expressions surprise, joie, et dégout.

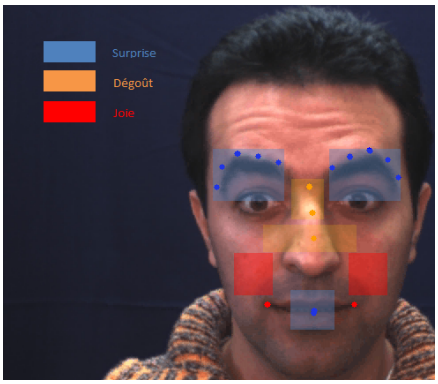


Figure 4: *Découpage du visage en régions d'intérêt.*

3.2. Reconnaissance d'expressions

Après avoir trouvé et affiné les régions correspondant à chaque expression, nous allons analyser ces régions afin d'identifier les changements de texture en présence d'une expression particulière. Cette analyse de la texture est faite en appliquant des filtres tels que le filtre de Gabor ou le filtre LBP. Le filtre LBP possède certaines propriétés qui lui permettent d'être efficace en pratique. Il est robuste aux conditions d'éclairage [AHP04], performant à la sélection de paramètres [OPM02], ne nécessite pas d'initialisation et fonctionne de manière fiable sur des résolutions d'images basses [SGM05].

Nous avons choisi d'appliquer les filtres de Gabor et de LBP avec différents paramètres sur les régions associées à chaque expression. Nous avons ensuite maximisé la valeur de chaque mesure pour obtenir des pics lors de l'activation d'une expression tout en minimisant sa valeur en absence de cette expression.

4. Résultats expérimentaux

Nous avons validé notre approche sur deux types de bases. Une base contenant des vidéos de personnes regardant une webcam et exprimant spontanément différentes expressions. La seconde est composée d'images statiques de personnes exprimant des expressions actées qui nous permet de valider notre étude préliminaire sur la base DISFA.

4.1. La base vidéo DISFA

La Figure 5 présente le résultat obtenu pour la détection de la surprise sur une vidéo de la base DISFA. Le participant a exprimé différentes émotions et les instants où il a exprimé la surprise sont illustrés dans la partie haute de la figure. La partie basse présente les résultats de notre métrique qui n'est pas influencée par les autres expressions puisque les pics surviennent aux moments où la surprise a été annotée sur la vidéo.

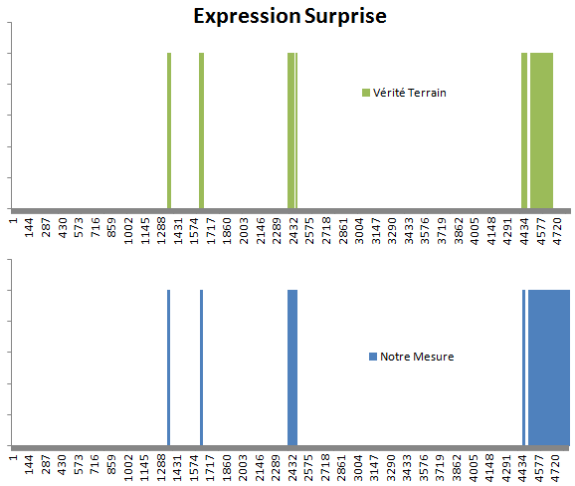


Figure 5: *Détection de la surprise dans une vidéo.*

4.2. La base d'images KDEF

La base Karolinska Directed Emotional Faces (KDEF) [LFO98] contient un ensemble d'expressions faciales actées qui permettent de renforcer la validation de notre étude préalable. La base contient 70 personnes, chacune affichant sept expressions différentes (neutre, joie, colère, peur, dégoût, tristesse, surprise) avec chaque expression photographiée deux fois. Les participants étaient assis à une distance d'environ trois mètres de la caméra. La résolution des images est de 562x762. La Figure 6 présente les résultats obtenus pour les expressions "surprise", "joie" et "dégout" sur la base KDEF. La mesure associée à une expression est considérée détectée correctement lorsque sa valeur est la plus élevée.

Expression	Peur	Colère	Dégoût	Joie	Neutre	Tristesse	Surprise
Joie	1,43%	1,43%	0,00%	**92,14%**	0,71%	4,29%	0,00%
Surprise	7,14%	2,41%	1,43%	0,71%	4,29%	6,43%	**77,86%**
Dégoût	2,14%	21,43%	**70,00%**	0,71%	2,14%	1,43%	2,14%

Figure 6: *Les résultats de l'application de la mesure sur les expressions surprise, joie et dégout de la base KDEF.*

5. Conclusion

Dans cet article, nous avons proposé une approche locale pour caractériser les changements de texture pour la reconnaissance d'expressions faciales spontanées. Nous associons à chaque expression une mesure qui est sensible aux changements dans les traits du visage dans des régions d'intérêt localisées. Cette approche ne nécessite pas de procédés spécifiques d'apprentissage préalable ou une reconnaissance d'un ensemble d'unités d'action. Les résultats sont prometteurs en termes de robustesse aussi bien pour la détection d'expressions faciales actées que spontanées. Dans nos futurs travaux, nous allons essayer d'augmenter le taux de reconnaissance en appliquant le filtre LBP sur des images filtrées par Gabor. On veut également modéliser par ce même principe des régions, la présence/absence des mouvements faciaux discriminant les expressions faciales ayant des unités d'actions similaires.

Références

[AHP04] AHONEN T., HADID A., , PIETIKAINENG M. : Face recognition with local binary patterns. In *European Conference on Computer Vision (ECCV)* (2004).

[CD10] CALVO R., D'MELLO S. : Affect detection : An interdisciplinary review of models, methods, and their applications. *IEEE Transactions on Affective Computing (TAC). Vol. 1*, Num. 1 (2010), 18–37.

[CET01] COOTES T., EDWARDS G., TAYLOR C. : Active appearance models. *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI). Vol. 23*, Num. 6 (2001), 681–685.

[DBMD13] DANISMAN T., BILASCO I. M., MARTINET J., DJERABA C. : Intelligent pixels of interest selection with application to facial expression recognition using multilayer perceptron. *Signal Processing. Vol. 93*, Num. 6 (2013), 1547–1556.

[EF78] EKMAN P., FRIESEN W. : *Facial Action Coding System : A technique for the measurement of facial movement*. Consulting Psychologists Press, Palo Alto, 1978.

[GSEV11] GONZALEZ I., SAHLI H., ENESCU V., VERHELST W. : Context-independent facial action unit recognition using shape and gabor phase information. In *4th International Conference on Affective Computing and Intelligent Interaction (ACII)* (2011).

[LDBD14] LABLACK A., DANISMAN T., BILASCO I. M., DJERABA C. : A local approach for negative emotion detection. In *22nd International Conference on Pattern Recognition (ICPR)* (2014).

[LFO98] LUNDQVIST D., FLYKT A., OHMAN A. : *The Karolinska Directed Emotional Faces (KDEF)*. Karolinska Institutet, 1998.

[MMB*13] MAVADATI S. M., MAHOOR M. H., BARTLETT K., TRINH P., COHN J. F. : DISFA : A Spontaneous Facial Action Intensity Database. *IEEE Transactions on Affective Computing. Vol. 4*, Num. 2 (2013), 151–160.

[OPM02] OJALA T., PIETIKÄINEN M., MÄENPÄÄ T. : Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)* (2002), 971–987.

[PRW*11] POPA M., ROTHKRANTZ L., WIGGERS P., BRASPENNING R., SHAN C. : Facial Action Units Recognition - A comparative study. *IEEE Transactions on Multimedia special issue on Multimodal Affective Iinteraction* (2011).

[SGM05] SHAN C., GONG S., MCOWAN P. : Robust facial expression recognition using local binary patterns. In *International Conference on Image Processing (ICIP)* (2005).

[SRS*11] SENECHAL T., RAPP V., SALAM H., SÉGUIER R., BAILLY K., PREVOST L. : Combining AAM coefficients with LGBP histograms in the multi-kernel SVM framework to detect facial action units. In *International Conference on Automatic Face and Gesture Recognition (FG)* (2011).