



HAL
open science

Etude de la dynamique du visage en situation d'interaction naturelle

Benjamin Allaert, José Mennesson, Ioan Marius Bilasco, Chaabane Djeraba

► To cite this version:

Benjamin Allaert, José Mennesson, Ioan Marius Bilasco, Chaabane Djeraba. Etude de la dynamique du visage en situation d'interaction naturelle. COMpression et REpresentation des Signaux Audiovisuels (CORESA) 2016, May 2016, Nancy, France. hal-01306045

HAL Id: hal-01306045

<https://hal.science/hal-01306045>

Submitted on 29 Sep 2018

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Etude de la dynamique du visage en situation d'interaction naturelle

Benjamin ALLAERT¹ José MENNESSON¹ Ioan Marius BILASCO¹ Chabane DJERABA¹

¹ Univ. Lille, CNRS, Centrale Lille, UMR 9189 - CRISTAL -
Centre de Recherche en Informatique Signal et Automatique de Lille, F-59000 Lille, France

{benjamin.allaert}@ed.univ-lille1.fr
{jose.mennesson, marius.bilasco, chabane.djeraba}@univ-lille1.fr

Résumé

L'importance de la dynamique faciale pour la reconnaissance d'expressions spontanées a été prouvée dans l'identification subtile des déformations physiques du visage. Les approches courantes de reconnaissance d'expressions faciales sont performantes sur des datasets où l'environnement est contrôlé et les expressions sont actées. Cependant, ces datasets ne reflètent pas les conditions rencontrées dans une situation d'interaction naturelle où la personne est libre de ses mouvements. Dans ces contextes, la présence de variations de pose et de larges déplacements rend l'analyse difficile. Cet article présente une synthèse originale des solutions de normalisation et d'extraction de mouvements du visage en présence d'expressions spontanées et propose d'étudier l'impact de ces méthodes de normalisation sur la dynamique faciale.

Mots clefs

Expressions faciales, Etude du mouvement, Normalisation

1 Introduction

La reconnaissance d'émotions à partir de flux vidéo se base généralement sur l'analyse des expressions faciales reposant sur le système *Facial Action Coding System (FACS)* de Paul Ekman [1]. Les six expressions faciales universelles (i.e. Joie, Peur, Tristesse, Surprise, Colère, Dégoût) ($6+N$) sont définies par une combinaison linéaire d'unités d'actions caractérisant les mouvements sur le visage et par un terme aléatoire capturant l'incertitude.

La majorité des systèmes proposés mesure leur performance sur des collections d'images où des acteurs reproduisent des expressions faciales face à une caméra. Au vu des dernières études sur la reconnaissance des expressions faciales [2, 3, 4], les méthodes obtiennent de très bonnes performances sur les datasets où les expressions sont actées. Cependant ces *datasets* ne font pas l'unanimité car en présence d'expressions spontanées, des problèmes de variation de pose et de large déplacement apparaissent. Des *datasets* permettant d'analyser des expressions faciales spontanées mettent en évidence des problèmes de robustesse des précédentes approches et une baisse des performances sur la reconnaissance des expressions spontanées.

Les expressions spontanées sont plus difficiles à caractériser que les expressions actées car les déformations physiques du visage sont moins grandes entre deux images successives. Dans ce contexte, les systèmes s'appuyant sur l'analyse du mouvement obtiennent de meilleures performances [5, 6] car ils permettent de détecter de subtils changements entre deux images, tout en prenant en compte les contraintes physiques du visage.

Plusieurs approches passent par une étape de pré-traitement avant d'extraire le mouvement du visage afin de s'abstraire du mouvement global dans la scène. La première catégorie d'approche conserve uniquement les visages frontaux afin de revenir dans un contexte maîtrisé comme dans les *datasets* contraints [7]. Cette méthode est principalement utilisée pour vérifier les performances des systèmes. La seconde catégorie consiste à analyser le visage dans sa position courante, avec une hypothèse forte de symétrie de l'expression faciale. Enfin, la troisième catégorie utilise des algorithmes de recalage pour amener le visage dans une configuration frontale [8, 9].

Les études de synthèse [10, 11] permettent d'avoir une vision globale sur les approches de reconnaissance d'expressions faciales en passant en revue les différentes étapes de traitement, de l'identification à la classification. Ils s'accordent majoritairement sur le fait que l'extraction du mouvement dans une situation d'interaction naturelle reste un problème ouvert dû aux grandes difficultés rencontrées dans un contexte d'interaction naturelle.

Deux défis majeurs pour la reconnaissance d'expression en présence d'interactions naturelles sont étudiés dans cet article : les variations de pose, et la présence de mouvement global de la personne dans le cadre de la scène. Dans la section 2, nous discutons de l'impact des variations de pose et de larges déplacements sur la reconnaissance d'expressions faciales. Pour chacune de ces problématiques, nous discutons des solutions existantes dans la littérature et des déformations du visage induites par les approches de normalisation. Les *datasets* les plus utilisés dans la littérature pour analyser les expressions faciales sont présentés en section 3. Dans la section 4, les approches récentes sont comparées sur les différents *datasets* et le bruit généré par les approches de normalisation du visage pour l'extraction du mouvement facial est étudié. Dans la section 5, les

limites des approches et des *datasets* existants sont discutées, et des pistes pour limiter le déphasage entre les situations d'interactions contrôlées et les situations d'interactions naturelles.

2 Grandes Déplacements (LD) et Variations de Poses (VP)

En présence de larges déplacements, une étape d'alignement est souvent nécessaire afin d'enlever le mouvement global du visage et de conserver uniquement le mouvement local. Une solution simple consiste à appliquer une transformation entre les deux images pour aligner les visages [12, 13]. Ces solutions ont tendance à déformer le visage afin de le rendre invariant aux translations, rotations dans le plan et aux changements d'échelle. Souvent utilisés pour identifier les expressions faciales, les visages normalisés peuvent produire des changements morphologiques néfastes [14]. Les travaux de Black *et al.* [15] ont montré qu'une normalisation locale sur plusieurs régions est mieux adaptée pour s'abstraire du mouvement global tout en réduisant les déformations induites par la transformation.

En présence de changements de pose de la tête, les visages ne sont pas directement exploitables sur la plupart des systèmes actuels. Un premier traitement consiste à normaliser toutes les images de visage. L'alignement du visage consiste à modifier la position du visage dans l'espace 3D afin de l'amener dans une configuration 2D idéale où l'apparence du visage est préservée tout en minimisant les artefacts et la perte d'information. La perte d'information est due aux rotations hors plan du visage, comme illustré dans la figure 1, où la partie gauche du visage disparaît progressivement au fur et à mesure que la tête tourne. La complexité de la tâche d'alignement dépend de la correction à appliquer.

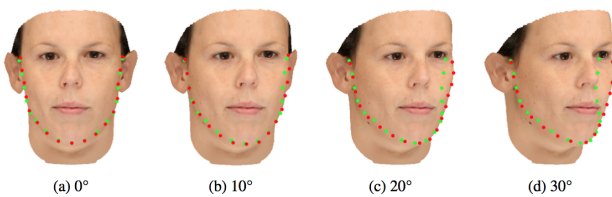


Figure 1 – Erreur de correspondance des points faciaux du visage 2D (rouge) et 3D (vert) sous plusieurs angles [16].

Des systèmes d'alignement sont proposés dans la littérature, cependant, la majorité de ces méthodes ne sont pas entièrement satisfaisantes. La majorité des systèmes aligne le visage en appliquant une transformation affine par rapport au centre des yeux et du nez car ces points restent stables indépendamment des expressions faciales [8, 13]. Cette solution souffre des mauvaises détections qui perturbent l'alignement. Des extensions de ces travaux utilisent un nombre plus important de points d'intérêt afin de

garantir une meilleure stabilité en cas de mauvaise détection. Certaines approches [3] détectent des points d'intérêt uniquement sur la région centrale du visage. D'autres approches [17] prennent aussi en compte les contours du visage afin d'obtenir des informations complémentaires sur la morphologie du visage. Cependant, en présence de certaines expressions faciales, les contours du visage subissent des déformations impactant la précision de l'alignement.

L'inconvénient majeur des systèmes précédents réside dans les informations disparues et cela met en cause la robustesse des systèmes d'alignement. En effet, un système d'acquisition d'images ne fournit que la projection des scènes observées sur un plan en deux dimensions. L'image ne permet donc d'exploiter que le mouvement résultant de la projection sur le plan 2D, d'où la perte d'information.

Des travaux récents proposent d'utiliser l'information 3D pour aligner le visage en minimisant les pertes. Zhu *et al.* [18] proposent une approche robuste d'alignement du visage en appliquant la texture 2D sur un modèle 3D. La transformation géométrique est appliquée sur le modèle 3D afin d'aligner correctement le visage dans l'espace 2D. Le modèle 3D garantit la conservation de la morphologie du visage permettant d'identifier sa forme globale et ses contours. Comme le montre la Figure 2, l'utilisation du modèle 3D permet de trouver un alignement convenable du visage mais le visage n'est pas entièrement reconstitué. En effet, les parties cachées ne sont pas texturées car il est impossible d'obtenir ces informations avec une seule camera. Pour combler la perte d'information, des méthodes de remplissage de texture [19] sont utilisées afin de reconstruire le visage. Le mouvement est ensuite extrait sur ce visage aligné afin de reconnaître les expressions faciales. Plus ces régions sont importantes et texturées, plus la reconstruction est difficile et la perte d'expression faciale est importante. Ainsi, la prise en compte des zones reconstruites dans le calcul de la dynamique faciale est problématique.

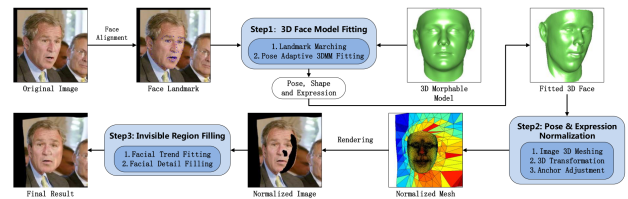


Figure 2 – Alignement du visage en plaquant la texture sur un modèle 3D, extrait des travaux de Zhu *et al.* [18].

Il est important de souligner que les approches d'alignement du visage sont employées principalement dans les systèmes de reconnaissance faciale où la présence de certains artefacts ou la perte de certaines informations est peu importante par rapport aux systèmes de reconnaissance d'expressions faciales. Or, lorsque l'on s'intéresse au mouvement, la présence des déformations liées aux méthodes d'alignement provoque l'apparition de faux mouvements

qui ne peuvent être supprimés sans impacter les mouvements pertinents [14]. Il faut considérer les limitations de l’alignement lorsque l’on extrait le mouvement, sur un visage, dans une situation d’interaction naturelle.

Dans la section suivante, nous analysons comment les critères définis précédemment sont reflétés dans les différents corpus utilisés pour l’analyse des expressions faciales.

3 Les datasets

La majorité des systèmes de reconnaissance d’expressions mesurent leur performance dans des situations actées [20, 21], où le visage reste statique, ce qui permet d’analyser la dynamique faciale dans d’excellentes conditions. Dans ces *datasets*, l’expression est exagérée, jouée par des acteurs afin d’obtenir des déformations anormalement amplifiées par rapport au naturel. A l’inverse, il existe des *datasets* où le contexte d’interaction est naturel et les expressions faciales sont spontanées [22, 22, 23].

La caractérisation des émotions permet d’identifier la manière dont les états affectifs sont annotés. Dans la littérature, certaines émotions de base sont universellement reconnues (i.e. Joie, Peur, Tristesse, Surprise, Colère, Dégoût) ($6+N$). Une alternative à cette représentation est l’utilisation de dimensions émotionnelles : "agréable ou non agréable" (Valence), "réveil ou soumission" (Arousal). La représentation dimensionnelle a un certain nombre d’avantages par rapport à la représentation catégorique. En effet, elle permet d’analyser un large panel d’émotions avec différentes nuances d’intensité qui semble plus adapté à caractériser une situation d’interaction naturelle.

Pour décrire les changements du visage, le système d’action faciale (FACS) proposé par Ekman [1] est largement utilisé. Certains *datasets* [24, 25, 26] contiennent une annotation continue sur l’enclenchement des muscles faciaux. Ces annotations sont utilisées pour caractériser les expressions faciales et permettent de mesurer l’extraction de mouvements sur le visage.

Certains *datasets* mettent en évidence plusieurs problèmes présentés dans la section 2. Il existe des *datasets* actés [20, 21, 27, 24] qui incluent partiellement des problématiques illustrées dans la section 2. Des *datasets* spontanés [28, 29, 22, 23, 26, 25] répondent à ces conditions en proposant d’étudier les problèmes de variations de pose (VP) et de larges déplacements (LD).

Les *datasets* couramment utilisés dans la littérature sont recensés dans le tableau 1. A ce jour, peu de *datasets* permettent d’analyser les émotions dans une situation d’interaction naturelle où les expressions sont spontanées. Certains *datasets* actés [20, 21] contiennent des extensions permettant d’étudier des expressions spontanées, cependant la majorité des travaux exploitent uniquement les données actés sur ces *datasets*. En présence d’interactions spontanées, il est difficile d’obtenir une annotation précise sur ces *datasets* car la durée des séquences est souvent longue et que les conditions de capture rendent l’annotation manuelle difficile.

Dataset	Type	Caract.	Longueur Seq.	AUs	VP	LD
CK+ [20]	Acté	6+N	10-60 img/seq	-	+	-
MMI [21]	Acté	6+N	40-520 img/seq	-	+	-
JAFFE [27]	Acté	6+N	2-3 img/seq	-	-	-
GEMEP [24]	Acté	6+N	100 img/seq	✓	+	+
SEMAINE [22]	Spont.	A/V	2K-25K img/seq	partielle	++	++
DISFA [25]	Spont.	-	4844 img/seq	✓	-	-

Tableau 1 – Recensement des datasets couramment utilisés pour la reconnaissance des expressions faciales.

Dans la section suivante, nous discutons de la manière dont les méthodes de l’état de l’art répondent aux défis posés par ces *datasets*.

4 Reconnaissance des émotions

Dans la littérature, nous distinguons deux catégories d’approche pour analyser les déformations du visage : les approches basées sur l’étude des caractéristiques du visage et les approches basées sur l’étude du mouvement. L’extraction du mouvement consiste à analyser les déformations apparentes du visage, causées par le mouvement relatif entre la caméra et la scène. Ambadar *et al.* [5] montrent l’importance de l’analyse de la dynamique faciale pour reconnaître les expressions car cela permet d’identifier plus subtilement les déformations physiques du visage.

Les récentes approches de la littérature qui extraient les expressions faciales pour analyser les états affectifs sont référencées dans le tableau 2. Ces approches ont été appliquées sur les différents *datasets* du tableau 1. Au vu des performances sur les *datasets* CK+, MMI, JAFFE, on constate que l’extraction des expressions faciales, où l’environnement est contrôlé et les expressions sont actées, est maîtrisée. En revanche, ce n’est pas le cas sur les *datasets* SEMAINE et DISFA.

Des méthodes de normalisation sont appliquées dans la majorité des approches quel que soit le *dataset* utilisé. En effet, la normalisation est importante dans une situation d’interaction naturelle afin de s’assurer de la correspondance des régions. Cependant l’impact de la normalisation n’aura pas les mêmes conséquences sur les visages contraints car les problèmes de variation de pose et de larges déplacements sont très faibles. C’est pourquoi, une normalisation sur deux/trois points, comme le centre des yeux et le nez, est souvent employé dans un contexte contraint [8, 9, 4, 12, 30]. En revanche, dans une situation d’interaction naturelle, ces méthodes ne sont plus adaptées et nécessitent des techniques plus complexes [31, 32, 33].

La reconnaissance d’émotions via l’analyse des expressions faciales peut être réalisée en mode interactif ou en définissant une séquence type consistant en un ordre de mouvements faciaux. Dans un mode interactif, chaque image est analysée puis associée à une classe puis un taux de reconnaissance (ar) est calculé en sortie. La décomposition en séquence consiste à étudier l’activation et la durée des expressions faciales pour conserver l’information temporelle. La durée de ces séquences permet de distin-

guer des nuances d'intensité dans les expressions et varie en fonction des personnes. Les performances de ces approches sont calculées avec des méthodes de corrélation croisée (cc).

L'extraction de caractéristiques visuelles est utilisée par la majorité des approches. La prise en compte de l'information temporelle dans plusieurs travaux [8, 2, 34] apporte une hausse des performances sur la reconnaissance des émotions. Ceci met en évidence l'importance de l'évolution du contexte pour reconnaître un état émotionnel. Les approches basées sur les caractéristiques visuelles et/ou sur le mouvement sont appliquées localement ou globalement afin d'analyser les expressions faciales. Sanchez *et al.* [35] comparent deux approches de mouvement (dense, 15 points d'intérêt) au cours d'une même expérience sur le *dataset* MMI [21]. Ils constatent une amélioration de 3% avec l'approche dense. Les approches d'extraction des mouvements denses [12, 36, 4, 37] montrent de bonnes performances sur l'analyse des expressions actées. Comme souligné dans la section 2, la normalisation est un processus difficilement maîtrisé dans une situation d'interaction naturelle, ainsi, il est difficile de récupérer tous les avantages des approches denses dans ces conditions. Dans la suite de cette section nous nous intéressons aux approches denses et aux solutions pour adapter ces approches aux situations d'interactions naturelles.

Reference	Methode	Norm.	Type	Performances
Zhao <i>et al.</i> [8]	LBP-TOP	2pts	6+N	CK+ ar :95.2%
Sikka <i>et al.</i> [9]	Dense BoW	2pts	6+N	CK+ ar :95.9%
Happy <i>et al.</i> [30]	Salient Patches	2pts	6+N	CK+ ar :94.14%
Su <i>et al.</i> [37]	Optical flow	Edges	6+N	CK+ ar :93.89%
Koelstra <i>et al.</i> [4]	FFDs	2pts	AUs	MMI cr :94.3%
Jiang <i>et al.</i> [2]	LPQ-TOP	4pts	AUs	MMI cr :94.7%
Jiang <i>et al.</i> [38]	LPQ	N/A	AUs	GEMEP cr :66%
Yang <i>et al.</i> [33]	LBPLPQ	3D	6+N	GEMEP ar :84%
Liao <i>et al.</i> [12]	Optical Flow	2pts	6+N	JAFFE ar :92.5%
Zhang <i>et al.</i> [36]	patch based Gabor	N/A	6+N	JAFFE ar :92.93%
Cruz <i>et al.</i> [31]	LPQ	3D	A/V	SEMAINE ar :55%
Nicolle <i>et al.</i> [32]	App&Geom Feat.	CLM	A/V	SEMAINE cc :0.46
Sandbach <i>et al.</i> [13]	LBP	3pts	AUs	DISFA cc :0.342

Tableau 2 – Les récentes approches de la littérature. (cc : *person's cross correlation*, ar : *average recognition rate*, cr : *classification rate*)

Contrairement aux approches basées caractéristiques, les approches denses ne nécessitent pas d'étape de segmentation d'image et permettent ainsi d'exploiter toute l'information contenue dans l'image en restant au plus près de la réalité. Parmi ces approches, les méthodes de flux optiques ont été largement utilisées et ont montrées leur efficacité pour analyser les déformations physiques du visage [15, 39]. Cependant, ces approches sont très sensibles à l'environnement et il est difficile de dissocier le mouvement réel du mouvement apparent dans les images. En effet, un système d'acquisition d'images ne fournit que la

projection des scènes observées sur un plan en deux dimensions. L'image ne permet donc d'exploiter que le mouvement résultant de la projection sur le plan 2D du mouvement réel 3D, couplé avec le mouvement de la caméra. Liao *et al.* [12] utilisent les flux optiques pour analyser les régions les plus distinctives afin de reconnaître les expressions faciales. Chaque modèle de mouvement observé est associé à l'une des 6 expressions universelles

L'analyse du mouvement à partir de séquences d'images décrivant la dynamique faciale dans une situation d'interaction naturelle est particulièrement délicat. La difficulté réside dans la spécificité des mouvements associés au visage. Ces mouvements se caractérisent entre autres par des fortes variations de pose et de larges déplacements. Dans ce contexte, les techniques standards issues de la littérature, qui s'appuient sur des caractéristiques stables de la fonction de luminance et sur l'hypothèse d'un mouvement rigide, s'avèrent en général mal adaptées.

La méthode de Farneback [40] permet de calculer le flux optique dense rapidement en utilisant des polynômes, où le mouvement local est diffusé sur l'ensemble de ces voisins pour calculer le mouvement global. Cependant, en présence de discontinuité de mouvement, l'erreur produite localement est propagée sur l'ensemble des voisins avec une intensité moindre, ce qui réduit la qualité du mouvement global. C'est pourquoi, il est important de prendre en compte les variations de pose et de large déplacement lorsqu'on utilise les méthodes de flux optique sur un visage.

La conception de méthodes alternatives dédiées aux mouvements d'objet non rigide constitue un vaste champ d'investigation relativement peu abordé. Les régions occultées de l'image courante sont définies par un ensemble de pixels qui disparaissent dans l'image suivante en présence de large déplacement et de variation de pose comme illustré sur la figure 1. Ces pixels n'ont pas de correspondance dans l'image suivante et donc le mouvement associé n'est pas observable dans ces régions. Pour résoudre ces problèmes, des méthodes sont proposées pour détecter les contours assujettis aux risques d'occultations en s'appuyant sur les flux optiques [41]. La majorité des méthodes calcule le flux optique sur l'ensemble du visage et supprime l'information de mouvement aux alentours de ces contours pour réduire le bruit provoqué par ces discontinuités. Cependant, les fortes variations de pose impliquent une perte importante des informations du visage et de ce fait, réduisent la précision des algorithmes de reconnaissance d'expressions faciales. D'autres méthodes utilisent ces contours pour calculer un flux optique localement afin de réduire ces discontinuités de mouvement. Pour cela, ils utilisent des méthodes de remplissage de mouvement en se basant sur les flux optiques du voisinage et sur les contraintes physiques [42].

La prise en compte de toutes ces contraintes augmente considérablement le temps de calcul. Toutefois, ces solutions permettent d'adapter les méthodes de flux optiques pour analyser subtilement les expressions faciales dans une situations d'interaction naturelle.

5 Discussion

Les récentes approches de reconnaissance d'expressions faciales sont validées sur des *datasets* où l'interaction est naturelle car les expressions observées sont spontanées et reflètent mieux la réalité. Dans ce contexte, l'analyse des expressions est soumise à certaines contraintes liées à la liberté de mouvement du sujet (notamment de sa tête) et au système d'acquisition. De fait que la plupart des approches requièrent des images de visage quasi-frontale, une étape de sélection des visages à traiter ou une étape de normalisation du visage sont nécessaires pour garantir, en entrée du processus d'analyse, uniquement des visages frontaux. Toutefois, il est important d'adapter les méthodes de normalisation en fonction du contexte et des caractéristiques des approches sous-jacentes. La normalisation permet d'améliorer les performances des systèmes caractérisant statiquement la texture du visage. En revanche, elle induit des pertes d'information ou des modifications morphologiques du visage qui provoquent des incohérences sur les données extraites lorsque la dynamique de la texture est caractérisée, par exemple, par l'intermédiaire des flux optiques.

La prise en compte de l'information temporelle, de l'étude du mouvement et des méthodes denses nous semblent mieux adaptées pour répondre aux problématiques soulevées par les nouveaux *datasets* qui s'approchent de plus en plus d'un environnement de captation naturel où les sujets s'expriment de manière non-actée et non-posée. Cette observation est également confortée par le renouveau des approches caractérisant la dynamique du visage par rapport aux nombreuses approches traitant uniquement de l'analyse statiques du visage. Il en est de même pour les méthodes de classifications où l'analyse continue est mieux adaptée pour caractériser un état affectif qu'une analyse discrète.

Références

- [1] P. Ekman et E.L. Rosenberg. *What the face reveals : Basic and applied studies of spontaneous expression using the Facial Action Coding System (FACS)*. Oxford University Press, 1997.
- [2] B. Jiang, M. Valstar, B. Martinez, et M. Pantic. A dynamic appearance descriptor approach to facial actions temporal modeling. *Cybernetics*, 44(2) :161–174, 2014.
- [3] M. Valstar et M. Pantic. Fully automatic recognition of the temporal phases of facial actions. *Systems, Man, and Cybernetics, Part B : Cybernetics*, 42(1) :28–43, 2012.
- [4] S. Koelstra, M. Pantic, et I.Y. Patras. A dynamic texture-based approach to recognition of facial actions and their temporal models. *PAMI*, 32(11) :1940–1954, 2010.
- [5] Z. Ambadar, J.W. Schooler, et J.F. Cohn. Deciphering the enigmatic face the importance of facial dynamics in interpreting subtle facial expressions. *Psychological Science*, 16(5) :403–410, 2005.
- [6] E.G. Krumhuber, A. Kappas, et A.S.R. Manstead. Effects of dynamic aspects of facial expressions : a review. *Emotion Review*, 5(1) :41–46, 2013.
- [7] T. Ahonen, A. Hadid, et M. Pietikainen. Face description with local binary patterns : Application to face recognition. *PAMI*, 28(12) :2037–2041, 2006.
- [8] G. Zhao et M. Pietikainen. Dynamic texture recognition using local binary patterns with an application to facial expressions. *PAMI*, 29(6) :915–928, 2007.
- [9] K. Sikka, T. Wu, J. Susskind, et M. Bartlett. Exploring bag of words architectures in the facial expression domain. Dans *ECCV*, pages 250–259. Springer, 2012.
- [10] E. Sariyanidi, H. Gunes, et A. Cavallaro. Automatic analysis of facial affect : A survey of registration, representation and recognition. 2014.
- [11] Z. Zeng, M. Pantic, G. Roisman, T.S. Huang, et al. A survey of affect recognition methods : Audio, visual, and spontaneous expressions. *PAMI*, 31(1) :39–58, 2009.
- [12] C. Liao, H. Chuang, C. Duan, et S. Lai. Learning spatial weighting via quadratic programming for facial expression analysis. Dans *CVPRW*, pages 86–93. IEEE, 2010.
- [13] G. Sandbach, S. Zafeiriou, et M. Pantic. Markov random field structures for facial action unit intensity estimation. Dans *ICCVW, 2013*, pages 738–745. IEEE, 2013.
- [14] S.W. Chew, P. Lucey, S. Lucey, J. Saragih, J.F. Cohn, I. Matthews, et S. Sridharan. In the pursuit of effective affective computing : The relationship between features and registration. *Systems, Man, and Cybernetics, Part B : Cybernetics, IEEE Transactions on*, 42(4) :1006–1016, 2012.
- [15] M.J. Black et Y. Yacoob. Tracking and recognizing rigid and non-rigid facial motions using local parametric models of image motion. Dans *Computer Vision*, pages 374–381. IEEE, 1995.
- [16] C. Qu12, E. Monari, T. Schuchert, et J. Beyerer21. Adaptive contour fitting for pose-invariant 3d face shape reconstruction. 2015.
- [17] A. Dhalla, A. Asthana, R. Goecke, et T. Gedeon. Emotion recognition using phog and lpq features. Dans *FG*, pages 878–883. IEEE, 2011.
- [18] X. Zhu, Z. Lei, J. Yan, D. Yi, et S.Z. Li. High-fidelity pose and expression normalization for face recognition in the wild. Dans *CVPR*, pages 787–796, 2015.
- [19] S. Li, X. Liu, X. Chai, H. Zhang, S. Lao, et S. Shan. Morphable displacement field based image matching for face recognition across pose. Dans *ECCV*, pages 102–115. Springer, 2012.

- [20] P. Lucey, J.F. Cohn, T. Kanade, J. Saragih, Z. Ambadar, et I. Matthews. The extended cohn-kanade dataset (ck+) : A complete dataset for action unit and emotion-specified expression. Dans *CVPRW*, pages 94–101. IEEE, 2010.
- [21] M. Pantic, M. Valstar, R. Rademaker, et L. Maat. Web-based database for facial expression analysis. Dans *ICME*, pages 5–pp. IEEE, 2005.
- [22] G. McKeown, M. Valstar, R. Cowie, M. Pantic, et M. Schröder. The semaine database : Annotated multimodal records of emotionally colored conversations between a person and a limited agent. *Affective Computing*, 3(1) :5–17, 2012.
- [23] F. Ringeval, A. Sonderegger, J. Sauer, et D. Lalanne. Introducing the recola multimodal corpus of remote collaborative and affective interactions. Dans *FG*, pages 1–8. IEEE, 2013.
- [24] T. Bänziger, M. Mortillaro, et K.R. Scherer. Introducing the geneva multimodal expression corpus for experimental research on emotion perception. *Emotion*, 12(5) :1161, 2012.
- [25] M.S. Bartlett, G.C. Littlewort, M.G. Frank, C. Lainssek, I.R. Fasel, et J.R. Movellan. Automatic recognition of facial actions in spontaneous expressions. *Journal of multimedia*, 1(6) :22–35, 2006.
- [26] X. Zhang, L. Yin, J.F. Cohn, S. Canavan, M. Reale, A. Horowitz, P. Liu, et J.M. Girard. Bp4d-spontaneous : a high-resolution spontaneous 3d dynamic facial expression database. *Image and Vision Computing*, 32(10) :692–706, 2014.
- [27] M. Lyons, S. Akamatsu, M. Kamachi, et J. Gyoba. Coding facial expressions with gabor wavelets. Dans *FG*, pages 200–205. IEEE, 1998.
- [28] L. Yin, X. Chen, Y. Sun, T. Worm, et M. Reale. A high-resolution 3d dynamic facial expression database. Dans *FG*, pages 1–6, 2008.
- [29] I. Sneddon, M. McRorie, G. McKeown, et J. Hanratty. The belfast induced natural emotion database. *Affective Computing*, pages 32–41, 2012.
- [30] S.L. Happy et A. Routray. Automatic facial expression recognition using features of salient facial patches. *Affective Computing, IEEE Transactions on*, 6(1) :1–12, 2015.
- [31] A. Cruz, B. Bhanu, et S. Yang. A psychologically-inspired match-score fusion model for video-based facial expression recognition. Dans *ACII*, pages 341–350. Springer, 2011.
- [32] J. Nicolle, V. Rapp, K. Bailly, L. Prevost, et M. Chehouani. Robust continuous prediction of human emotions using multiscale dynamic cues. Dans *ICMI*, pages 501–508. ACM, 2012.
- [33] S. Yang et B. Bhanu. Facial expression recognition using emotion avatar image. Dans *FG*, pages 866–871. IEEE, 2011.
- [34] C. Weng et S. Lai. Online facial expression recognition based on combining texture and geometric information. Dans *ICIP, 2014*, pages 5976–5980. IEEE, 2014.
- [35] A. Sánchez, J.V. Ruiz, A. Moreno, A.S. Montemayor, J. Hernández, et J.J. Pantrigo. Differential optical flow applied to automatic facial expression recognition. *Neurocomputing*, 74(8) :1272–1282, 2011.
- [36] L. Zhang et D. Tjondronegoro. Facial expression recognition using facial movement features. *Affective Computing*, 2(4) :219–229, 2011.
- [37] M. Su, Y. Hsieh, et D. Huang. A simple approach to facial expression recognition. *Proceeding WSEAS 2007*, 2007.
- [38] B. Jiang, B. Martinez, M. Valstar, et M. Pantic. Decision level fusion of domain specific regions for facial action recognition. Dans *ICPR*, pages 1776–1781. IEEE, 2014.
- [39] J.F. Cohn, A.J. Zlochower, J.J. Lien, et T. Kanade. Feature-point tracking by optical flow discriminates subtle differences in facial expression. Dans *FG*, pages 396–401. IEEE, 1998.
- [40] G. Farnebäck. Two-frame motion estimation based on polynomial expansion. Dans *Image Analysis*, pages 363–370. Springer, 2003.
- [41] P. Sundberg, T. Brox, M. Maire, P. Arbeláez, et J. Malik. Occlusion boundary detection and figure/ground assignment from optical flow. Dans *CVPR*, pages 2233–2240. IEEE, 2011.
- [42] J. Revaud, P. Weinzaepfel, Z. Harchaoui, et C. Schmid. EpicFlow : Edge-Preserving Interpolation of Correspondences for Optical Flow. Dans *Computer Vision and Pattern Recognition*, 2015.