Une nouvelle approche à la détection d'émotions pour l'aide à l'interaction homme-machine chez les personnes handicapées

Florian Tissier*, Donatello Conte*, Giuseppe Boccignone† et Mohamed Slimane*
*Université de Francois Rabelais, Laboratoire LI EA 6300, Tours, France
Email : florian.tissier@etu.univ-tours.fr, {donatello.conte, mohamed.slimane}@univ-tours.fr
†Dipartimento di Informatica, Universita degli Studi di Milano, Milano 20135, Italy
Email : giuseppe.boccignone@unimi.it

***Résumé*—Dans cet article une nouvelle approche à la détection d'émotions, par analyse de vidéos, est présentée. Dans le domaine de recherche de l'analyse de vidéos, la reconnaissance d'émotions est une application qui intéresse de plus en plus les chercheurs, pour améliorer l'interaction entre les hommes et les machines. La recherche est encore tout à fait ouverte, car aucune approche présente dans la littérature scientifique a obtenu de bonnes performances, surtout si on considère des vidéos non contrôlées et non construites *ad hoc* pour la détection d'émotion. La problématique est encore plus difficile en présence de personnes handicapées dans la vidéo. Récemment des approches probabilistes ont montrés leur efficacité. L'approche présentée dans cet article s'inspire de ces travaux et proposes de nouvelles modifications afin d'avoir de meilleures performances de détection.**

I. Introduction

L'interaction entre les hommes et les machines a toujours été un enjeu de taille. Arriver à faire communiquer un ordinateur avec un être humain est un défi de tous les jours et est de plus en plus présent dans notre quotidien.
C'est dans cette optique de facilitation du quotidien, et plus précisément de facilitation du quotidien de personnes handicapées, grâce à l'interaction avec une machine, que nos travaux prennent place.
Cet article va décrire les travaux que nous sommes en train de réaliser pour construire un système permettant, à partir d'un flux vidéo acquis grâce à une caméra, de détecter l'expression actuelle du visage d'une personne handicapée. Cette détection d'expressions, et donc d'émotions, se fait dans le but de réaliser des actions, comme par exemple changer la couleur de la lumière de la pièce, jouer de la musique, diffuser un certain parfum relaxant, etc., pouvant améliorer le bien-être de la personne.
Nous allons tout d'abord vous présenter un état de l'art des différentes approches pour un système de reconnaissance faciale d'expressions.
Dans un second temps, sachant que la détection d'émotions chez des personnes handicapées est difficile et que de nouveaux méthodes sont nécessaires, nous vous introduiront une nouvelle approche permettant de reconnaître des émotions via un modèle probabiliste, introduit par Vitale et al. [1]. A partir de ce modèle nous avons introduit des améliorations pour augmenter les taux de reconnaissance de notre application.
Nous finirons par présenter les expérimentations et les résultats obtenues grâce aux modifications que nous aurons apporté.
Ces travaux ont été réalisé dans le cadre d'un projet de recherche et développement au sein de l'école d'ingénieur Polytech Tours, dans le Laboratoire d'Informatique de l'Université de Tours.

II. État de l'art

Dans cette section, nous allons voir un état de l'art des différentes parties qui composent l'architecture d'un système de reconnaissance faciale d'expressions, ainsi que plusieurs bases de données permettant de réaliser l'apprentissage d'un tel système.
Nous ferons ensuite une critique d'un tel système et de ses limitations dans sa forme actuelle par rapport à notre problématique.
Mais tout d'abord, nous allons définir les émotions basiques que nous allons reconnaître via notre système de reconnaissance.

A. Émotions universelles

Définies par Paul Ekman et Wallace Friesen en 1971 dans leur article [2], les 7 émotions universelles sont les suivantes :

- la neutralité
- la joie
- la tristesse
- la colère
- la surprise
- la peur
- le dégoût

Ces émotions sont dites universelles car tout individu est capable de les exprimer mais est également capable de les reconnaître et de les interpréter chez autrui.
Ces 7 émotions sont utilisées dans la plupart des système réalisant de la reconnaissance d'émotions à l'heure actuelle et c'est également celles-ci que nous allons chercher à reconnaître.
À noter également qu'une expression se divise en 3 phases :

l'**onset** (correspondant au début de l'expression), l'**apex** (le pic de l'expression) et l'**offset** (la fin de l'expression).
Ekman et Friesen ont également créé en 1978 une norme de description des mouvements faciaux permettant de décomposer les expressions. Cette norme se nomme FACS, Facial Action Coding System [4]. Elle se compose de 46 Action Units (AU) permettant de représenter la contraction ou décontraction des muscles faciaux. La composition de plusieurs AUs permet donc de décrire une expression.

B. Architecture d'un système de reconnaissance faciale d'expressions

Un tel système se compose de 3 parties qui sont : détection du visage dans une image, extraction des *features* (points clés) de ce dernier et la classification des *features* pour obtenir l'émotion associée.

1) Détection du visage: Il existe 2 types de détecteurs de visages, les absolus et les différentiels.
Tout d'abord les détecteurs absolus. Ils permettent de déterminer la position d'un visage sur chaque *frame* d'un flux vidéo, indépendamment des *frames* précédentes. Le premier détecteur absolu a été inventé par Paul Viola et Mickaël Jones en 2001 [3] et il est encore à l'heure actuelle la référence de ce type de détecteur d'objets .
Les détecteurs absolus quant à eux vont repérer un visage en utilisant sa position dans la *frame* précédente. Si la position d'un visage est connu à l'instant t, le détecteur différentiel va se servir de cette position pour trouver celle à l'instant $t + 1$. Bien sûr, il faut initialiser la position à l'instant 0 pour faire fonctionner ces détecteurs. Pour cela il est possible d'utiliser un détecteur absolu sur la toute première *frame* du flux vidéo. L'avantage de ces détecteurs est leur grande rapidité et précision. Cependant, l'inconvénient est que l'accumulation de petites erreurs sur les positions peut mener à la "dérive" et donc la perte du visage.

2) Extraction des features: La deuxième partie d'un système de reconnaissance faciale d'émotions est l'extraction des *features* du visage précédemment repéré. Ces features sont les points clés d'un visage et se composent par exemple de la position du coin des yeux, des lèvres, du nez, des joues, de la position de la pupille, etc.
Pour extraire ces points et leurs positions, plusieurs méthodes existent.
Nous pouvons par exemple citer les filtres de Gabor qui appliquent un filtre linéaire à chaque pixel d'une image et dont la réponse est une sinusoïde modulée grâce à une fonction gaussienne. Un filtre de Gabor possède plusieurs paramètres lui permettant de pouvoir s'adapter à tous les types de problèmes de détection d'objets ou de textures dans une image.
Une autre méthode existante est celle des composantes pseudo-Haar. Elle a été introduite par Viola et Jones lors du développement de leur détecteur de visage [3]. Elle consiste à appliquer un masque en forme de "boîte" (composé d'une partie noire et d'une partie blanche) à toutes les positions possibles sur l'image, tout d'abord de petite taille (20*20 pixels par exemple) puis en les agrandissant. A chaque position, on soustrait la somme des pixels contenus dans la partie noire à la somme des pixels contenus dans la partie blanche. Le résultat d'une caractéristique à une certaine position est donc un nombre réel qui va coder les variations du contenu pixellique. Cette méthode est d'une grande rapidité lorsqu'elle est utilisée avec des images intégrales ([3]).
D'autres méthodes existent pour réaliser la détection d'objet dans une image, telle que les motifs binaire locaux ou l'utilisation de flots optiques.

3) Classification: Une fois les features extraites, il ne reste plus qu'à les analyser pour les classifier.
Deux types de classificateurs existent : les systèmes basés sur des règles d'experts (*rule-based expert systems*) et les classificateurs avec apprentissage (*machine learning classifiers*). Les premiers sont de moins en moins répandus car ils sont beaucoup plus compliqués à utiliser que les autres. C'est donc sur les classificateurs avec apprentissage que nous allons nous concentrer. Parmi ces classificateurs, nous pouvons les classer en 2 catégories : les binaires et les multi-classes.
Comme l'indique son nom, un classificateur binaire ne permet de classifier une instance qu'entre 2 classes (ex : sourire VS pas de sourire). Le classificateur binaire le plus connu à l'heure actuelle est le SVM [5] (Support Vector Machine, ou Séparateurs à Vaste Marge en français).
Contrairement aux binaires, les classificateurs multi-classes permettent de classifier une instance selon un nombre de classes supérieur à 2. Un des classificateur multi-classes les plus connu est le k-PPV [6] (k Plus Proche Voisin) qui consiste à chercher les k échantillons (appris lors de la phase d'apprentissage) les plus proches d'un point pour ensuite retourner la classe majoritairement représentée par ces k voisins.

C. Base de données

Pour réaliser un système tel que nous l'avons présenté précédemment, une phase d'apprentissage est nécessaire pour la classification. Cette phase va donc utiliser des images se trouvant dans des bases de données de visages.
Nous avons regroupé plusieurs bases de données ainsi que leurs caractéristiques principales dans le tableau I.
Malheureusement, aucune base de données existante ne concerne les personnes handicapées.

D. Critiques

A l'heure actuelle, tout système de reconnaissance faciale d'émotions tel que présenté précédemment ([10], [11], [12]) a atteint ses limites. Tout d'abord du fait que pour obtenir des résultat convenables, les expressions devant être analysées doivent être posées et donc exagérées, ce qui diffère catégoriquement de la vie réelle. Ensuite la plupart de système permettant la détection d'émotions se basent sur des techniques et base de données, telle que les FACS et Cohn-Kanade [7],

TABLE I
TABLEAU RÉCAPITULATIF DE BASES DE DONNÉES DE VISAGES 2D

Nom	P/S	Taille	Contenu	Label
CK	P	97 sujets (65% femmes, 15% afro-américains et 3% asiatiques ou sud américain) de 18 à 30 ans	486 séquences vidéos des 6 émotions basiques	FACS
CK+	P/S	+ 27%	+ 22%	FACS + émotion
MMI	P/S	75 sujets (48% de femmes, européen, africain et sud américain) de 19 à 62 ans	2900 vidéos des 6 émotions basiques ou d'AUs spécifiques	FACS
MHI Mimicry	S	40 sujets (28 hommes, 12 femmes) entre 18 et 40 ans	54 vidéos (sessions)	Auto-évaluation
HCI Tagging	S	27 sujets (11 hommes, 16 femmes) de 19 à 40 ans	Partie 1 : 20 vidéos ; Partie 2 : 28 images et 14 vidéos	Auto-évaluation
SAL	S	24 sujets	10 heures de vidéo : les sujets parle à une IA ayant différentes personnalités	Feeltrace
JAFFE	P	10 femmes japonaises	213 images des 7 émotions basiques	Noté selon 6 adjectifs d'émotion par 60 sujets Japonais

qui commencent à se faire vieille.
Dans la but d'innover et répondre au mieux à notre problématique de facilitation du quotidien de personnes handicapées, nous avons décidé d'utiliser des images de la base de données HCI-Tagging ([8]), dont les expressions présentées sont spontanées, pour l'apprentissage ainsi qu'un modèle probabiliste qui nous permettra de représenter les émotions sur un espace latent de dimension finie.
C'est maintenant ce modèle que nous allons vous présenter.

III. MÉTHODE PROPOSÉE

Dans un soucis d'innovation, et pour ameliorer les performances, nous avons décidé d'utiliser un modèle probabiliste qui va nous permettre de représenter les émotions sur un espace fini en dimension finie. Ce modèle a été défini par Vitale et al. dans l'article *Affective Facial Expression Processing Via Simulation : A Probabilistic Model* [1] paru en 2014.
Nous allons tout d'abord expliciter ce modèle probabiliste, puis nous introduirons les différentes modifications que nous avons apporté à ce modèle pour l'adapter à notre problématique de détection d'émotions sur les personnes handicapées.

A. Contexte

Pour comprendre les études menées dans l'article [1], poser le problème, et donc le contexte de la simulation de détection d'émotions, est nécessaire.
À un certain moment, un individu, appelé *acteur*, éprouvera un certain état interne, noté X_{act}. Cet état interne X_{act} peut soit être déclenché par un événement extérieur soit induit (remémoration d'un certain souvenir).
X_{act} va déclencher un comportement correspondant, noté Y_{act}, qui peut être par exemple une expression facial, une posture particulière, un changement du rythme cardiaque, etc.
Un second individu, appelé *observateur*, va maintenant utiliser Y_{act} pour trouver son état interne X_{obs} lui permettant d'être dans le même état mental M_{obs} que celui de l'acteur M_{act}.
Si l'observateur est dans le même état mental que l'acteur, cela signifie que l'observateur a réussi à trouver à quoi correspondait Y_{act} et donc l'émotion de l'acteur.
Cette situation de simulation est illustrée en Figure 1 (figure tirée et traduite de [1]).

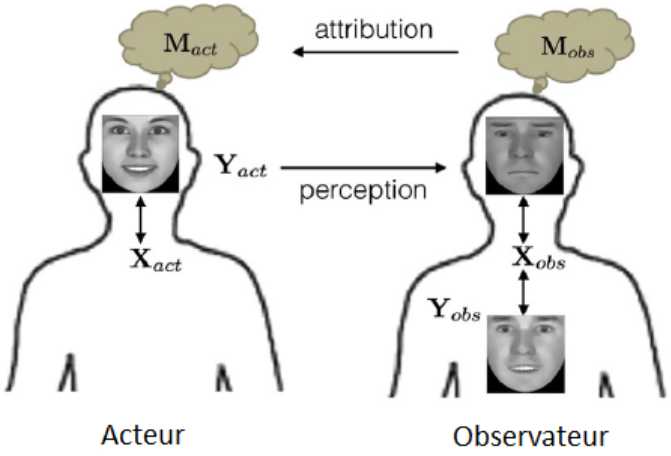

FIGURE 1. Contexte de la simulation du modèle probabiliste

B. Modèle

Pour la réalisation du modèle probabiliste, plusieurs probabilités ont été définies :

- $P(Y_{obs}/Y_{act})$: représente la probabilité pour l'observateur de montrer l'état externe (expression faciale) Y_{obs} quand l'acteur affiche Y_{act}. Cette probabilité est appelée *transcodage*.
- $P(X_{obs}/Y_{obs})$: représente la probabilité d'être dans un état interne X_{obs} sachant Y_{obs}. Cette probabilité est appelée *correspondance inverse*.
- $P(Y_{obs}/X_{obs})$: représente la probabilité que l'observateur génère un état externe Y_{obs} sachant l'état interne X_{obs}. Cette probabilité est appelée *correspondance vers l'avant*.

Une fonction de décision $\mathcal{D}(\bullet)$ est également nécessaire et permettra de comparer l'état externe Y_{act} de l'acteur avec un état simulé $\check{Y}_{obs}$.
Ces probabilités nous permettent d'obtenir les variables suivantes :

$$y_{obs} \sim P(Y_{obs}/Y_{act} = y_{act}) \quad (1)$$

qui permet de définir le processus de transcodage transformant une instance de l'expression faciale de l'acteur y_{act} en une instance de représentation interne de l'observateur y_{obs},

$$x_{obs} \sim P(P(X_{obs}/Y_{obs} = y_{obs}) \quad (2)$$

qui décrit le processus de correspondance inverse donc la détection de l'état x_{obs} à partir de l'expression faciale (interne) y_{obs}, et

$$\check{y}_{obs} \sim P(Y_{obs}/X_{obs} = x_{obs}) \tag{3}$$

qui décrit le processus de correspondance vers l'avant et donc la correspondance de l'expression simulé par l'observateur $\check{y}_{obs}$ quand il se trouve dans l'état interne x_{obs}.

La fonction $\mathcal{D}(\bullet)$ compare le y_{obs} transcodé de l'équation (1) à l'expression générée intérieurement $\check{y}_{obs}$ de l'équation (3), et est également utilisé via l'équation (2) pour contrôler quand le processus de comparaison aura convergé vers la solution la plus probable.

Les processus de transcodage, de correspondance inverse et de correspondance vers l'avant se produisent dans l'espace latent d'auto-projection, noté Z, défini de la façon suivante :

$$P(Z) = \mathcal{N}(0, I_L) \tag{4}$$

qui permet de donner la priorité des points dans l'espace Z de dimension L,

$$P(y_{act}/z) = \mathcal{N}(W_{act}z + \mu_{act}, \sigma^2_{act}I_D) \tag{5}$$

qui permet d'avoir la probabilité d'obtenir l'expression faciale de l'acteur y_{act}, avec D$\gg$ L, sachant z$\in$ Z, et

$$P(y_{obs}/z) = \mathcal{N}(W_{obs}z + \mu_{act}, \sigma^2_{obs}I_D) \tag{6}$$

qui permet d'avoir la probabilité d'obtenir l'expression faciale de l'observateur y_{act}, avec D$\gg$ L, sachant z$\in$ Z, avec :
$\mathcal{N}(\bullet)$: distribution Gaussienne ; W_{act} et W_{obs} : paramètres de correspondance de l'acteur et de l'observateur ; μ : moyenne ; σ : variance ; I_L et I_D : matrices identités de dimensions L et D.

On peut également définir les variables suivantes :
W=$\begin{pmatrix} W_{act} \\ W_{obs} \end{pmatrix}$, $\mu = \begin{pmatrix} \mu_{act} \\ \mu_{obs} \end{pmatrix}$ et $\Phi = \begin{pmatrix} \sigma^2_{act}I_D & 0 \\ 0 & \sigma^2_{obs}I_D \end{pmatrix}$.
Vu que le modèle est gaussien, y_{act} et y_{obs} suivent une distribution gaussienne de la forme $\mathcal{N}(\mu, \Sigma)$ avec $\Sigma = \Phi + WW^T$ correspondant à la matrice de covariance.
On obtient donc :

$$P(y_{obs}|y_{act}) \sim \mathcal{N}(\hat{\mu}_{obs}, \hat{\Sigma}_{obs}) \tag{7}$$

où $\hat{\mu}_{obs} = \mu_{obs} + \Sigma_c^T\Sigma_a^{-1}$ et $\hat{\Sigma}_{obs} = \Sigma_b - \Sigma_c^T\Sigma_a^{-1}\Sigma_c$ est le complément de Schur de Σ réécrit sous la forme du bloc $\Sigma = \begin{pmatrix} \Sigma_a & \Sigma_c \\ \Sigma_c^T & \Sigma_b \end{pmatrix}$.
L'équation (7) nous retourne la probabilité de correspondance de l'expression faciale de l'observateur sachant une expression faciale similaire de l'acteur. Pour faire cela, elle utilise W_{act} et W_{obs} en distribution normale multivariée.

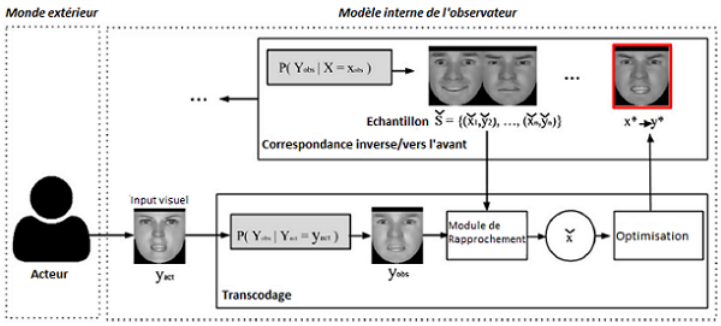

FIGURE 2. Fonctionnement du modèle probabiliste

Pour résoudre $\check{y}_{obs}$, on passe par l'utilisation du GPLVM (Gaussian Process Latent Variable Model) et l'on obtient

$$P(Y_{obs}|X_{obs}) = \frac{1}{\sqrt{(2\pi)^{ND}|K^D|}} \exp(-\frac{1}{2}tr(K^{-1}Y_{obs}Y_{obs}^T)) \tag{8}$$

avec Y_{obs} correspondant à l'ensemble d'apprentissage et K à la matrice de noyau dont les éléments sont définis par $(K)_{i,j} = \Phi(x_{i,obs}, x_{j,obs})$.

Une fois que $P(Y_{obs}/X_{obs})$ est appris, il est très facile d'obtenir $P(X_{obs}/Y_{obs})$ par GPLVM.

1) Fonctionnement: Le fonctionnement de ce modèle probabiliste, utilisant les notions précédentes, est décrit dans la Figure 2 ([1]). y_{act} est transcodé via l'espace latent d'auto-projection Z en y_{obs}.
Au même moment, un ensemble $\check{S}$ d'échantillons d'expressions de l'observateur est généré via l'équation (8). Avec $\mathcal{D}$, une mesure de similarité est utilisé pour évaluer la vraisemblance entre les échantillons de $\check{S}$ et de y_{obs}. L'état initial $\check{x}$ est sélectionné de cette manière :

$$\check{x} \in \check{S}|\check{x} \mapsto \check{y}_{obs} \wedge \check{y}_{obs} = arg_y \max \mathcal{D}(\check{y}_{obs}, y)$$

2) Comparaison: Pour fonctionner, le module de rapprochement (cf Figure 2) utilise la mesure **SSIM** (Structural SIMilarity).
Cette mesure est effectuée sur plusieurs "parties" de l'image, appelées fenêtres, et ensuite une moyenne est calculée pour obtenir le résultat final. Le mesure entre 2 fenêtres x et y se calcule comme suit :

$$SSIM(x, y) = \frac{(2\mu_x\mu_y + c_1)(2\sigma_{xy} + c_2)}{(\mu_x^2 + \mu_y^2 + c_1)(\sigma_x^2 + \sigma_y^2 + c_2)}$$

avec μ_x et μ_y : moyennes de x et y ; σ_x^2 et σ_y^2 : variances de x et y ; σ_{xy} : covariance de x et y ; c_1 et c_2 : variables de stabilisation de la division quand celle-ci à un dénominateur faible.

C. Modifications

Pour aller de pair avec l'article [1], les auteurs ont développé un programme qui permet de visualiser les émotions sur l'espace latent Z. Cet espace est créé grâce

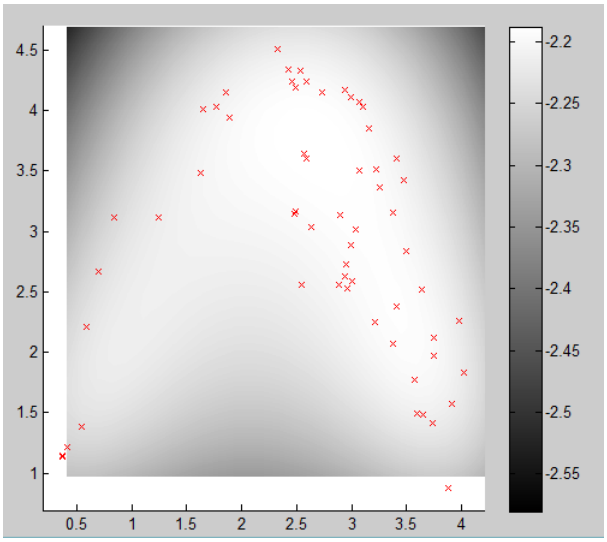

FIGURE 3. Espace latent Z représentant les émotions

à une base d'apprentissage de 60 images de taille 29*19 pixels. Cet espace est présenté en Figure 3. La principale fonctionnalité de ce programme est de pouvoir se déplacer sur l'espace Z et de voir en temps réel l'émotion associée.

Le modèle présentée est très intéressant et efficace, mais il présente encore quelque limitation. Dans cet article nous présentons les modifications apportées afin de l'améliorer.

Tout d'abord le modèle de représentation des données. Le modèle original traite les pixels d'une image et nous souhaiterions plutôt utiliser une représentation vectorielle qui représentera les mouvements du visage. Cela permettra d'être moins dépendante des pixels mais également du sujet présent sur l'image. Cela permettra, en plus, de ne pas utiliser l'algorithme PCA qui était utilisé dans le modèle original pour réduire le nombre de variables et qui est extrêmement chronophage.
Une deuxième amélioration consiste à analyser un ensemble d'images pour détecter une émotion, plutôt qu'une seule. En effet une émotion est composé, généralement, de plusieurs expressions faciales, et analyser plusieurs images d'une vidéo à la place d'une seule, nous permettrait d'être plus précis dans la reconnaissance. Dans notre modèle, cela se traduit à représenter une émotion, non plus par un point dans l'espace latent Z, mais plutôt à travers des trajectoires dans cet espace. Cela nous permettra aussi d'utiliser des modèles tels que les chaines de Markov cachés (HMM [9]) pour la phase de reconnaissance.

IV. EXPÉRIMENTATION

A l'heure où nous écrivons la méthode n'as pas encore été testée. On a, cependant, défini le protocole de notre expérimentation et en particulier les données à utiliser pour l'apprentissage et pour les tests. Pour cela nous avons choisi 3 sujets de la base de données HCI Tagging [8] : les sujets 1, 3 et 6. Pour chacun de ces sujets, nous avons choisi plusieurs vidéos correspondantes à différentes émotions : la joie, le dégoût, la surprise, la tristesse et la colère. Pour chacune de ces vidéos, nous en avons extrait les *frames* (1 toutes les 75) et en avons sélectionné 15 consécutives permettant de voir l'évolution d'une émotion au cours du temps et donc de construire la trajectoire. Les 15 *frames* consécutives ont été choisi de telle sorte que l'on puisse observer l'onset, l'apex et l'offset de l'émotion .
Le tableau II permet de présenter quels sujets nous avons choisi pour une émotion donnée, avec le numéro de la session correspondante dans la base de données HCI Tagging.

TABLE II
LES SUJETS ET LES ÉMOTIONS CONSIDÉRÉES DANS NOTRE EXPÉRIMENTATION (EXTRAITS DE LA BASE HCI-TAGGING).

	N^o sujet	N^o session	Émotion
1	1	6	joie
2	3	264	dégoût
3	1	28	tristesse
4	6	662	colère
5	3	290	surprise

Une autre partie des mêmes vidéos utilisées pour l'apprentissage, sera employée pour les tests de détection. A l'heure actuelle nous sommes en train de terminer les expérimentations pour analyser ensuite les résultats.

V. CONCLUSION

Les systèmes de reconnaissance faciale d'expressions majoritairement utilisés à l'heure actuelle commencent à atteindre leurs limites, notamment lorsqu'ils doivent être utilisés avec des personnes handicapées.
De nouvelles méthodes et approches ont donc dû être mises en place et notamment la méthode probabiliste introduite par Vitale et al. [1]. C'est en se basant sur cette méthode et sur les travaux réalisés par ces même auteurs que nous avons commencé à construire notre application de reconnaissance d'émotions chez les personnes handicapées.
Plusieurs modifications ont été réalisées dans le but d'obtenir une reconnaissance des plus fidèles. La construction de trajectoires sur l'espace latent va nous permettre de voir l'évolution d'une émotion au cours du temps et donc de comparer plus fidèlement différentes émotions. De plus l'utilisation d'images de la base de données HCI Tagging nous permet de travailler avec des expressions spontanées et donc d'être au plus proche de la réalité, même si il aurait été préférable d'utiliser une base contenant directement des images de personnes handicapées, ce qui n'existe pas à l'heure actuelle malheureusement.
Dans le futur nous envisageons de créer des bases de données de vidéos avec des personnes handicapés pour pouvoir tester le système et apporter éventuellement d'ultérieures améliorations.

RÉFÉRENCES

[1] J. Vitale, M-A. Williams, B. Johnston et G. Boccignone, *Affective Facial Expression Processing Via Simulation : A Probabilistic Model*, Biologically Inspired Cognitive Architectures, 2014.

[2] P. Ekman et W. V. Friesen, *Constants across culture in the face and emotion*, Journal of Personality and Social Psychology, 1971.

[3] P. Viola et M. Jones, *Robust Real-Time Face Detection*, 2001.

[4] P. Ekman, *Facial Action Coding System (FACS) : Manual*, Palo Alto : Consulting Psychologists Press, 1978.

[5] C. Cortes, V. Vapnik, *Support-vector networks*, Machine Learning, Vol. 20-3, pp. 273-297, 1995.

[6] N. S. Altman, *An introduction to kernel and nearest-neighbor nonparametric regression*. The American Statistician, Vol. 46-3, pp. 175–185, 1992.

[7] P. Lucey, J. F. Cohn, T. Kanade, J. Saragih, Z. Ambadar et I. Matthews, *The Extended Cohn-Kanade Dataset (CK+) : A complete expression dataset for action unit and emotion-specified expression*. Proceedings of the Third International Workshop on CVPR for Human Communicative Behavior Analysis (CVPR4HB 2010), San Francisco, USA, pp. 94-101. 2010.

[8] M. Soleymani, J. Lichtenauer, T. Pun et M. Pantic., *A multimodal database for affect recognition and implicit tagging*, IEEE Transactions on Affective Computing, Vol. 3-1, pp. 42 - 55, 2012.

[9] L. E. Baum et T. Petrie, *Statistical Inference for Probabilistic Functions of Finite State Markov Chains*. The Annals of Mathematical Statistics, Vol. 37-6, pp. 1554–1563. 1966.

[10] M. Soleymani, S. Asghari-esfeden, M. Pantic et Y. Fu, *Continuous Emotion Detection using EEG Signals and Facial Expressions*, Proceedings of IEEE Int'l Conf. Multimedia and Expo (ICME'14), Juillet 2014.

[11] Tingfan Wu, M.S Bartlett et J.R. Movellan, *Facial expression recognition using Gabor motion energy filters*, Computer Vision and Pattern Recognition Workshops (CVPRW), 2010 IEEE Computer Society Conference, pp. 42-47, 13-18 Juin 2010

[12] J. Whitehill, M.S. Bartlett et J.R. Movellan, *Automatic Facial Expression Recognition*, Social Emotions in Nature and Artifact, 2013.