

ECOLE POLYTECHNIQUE DE L'UNIVERSITÉ FRANÇOIS RABELAIS DE TOURS
Département Informatique
64 avenue Jean Portalis
37200 Tours, France
Tél. +33 (0)2 47 36 14 14
www.polytech.univ-tours.fr

**Projet Recherche & Développement
2015-2016**

Application d'aide à l'interaction homme/machine pour les personnes handicapées

Tuteurs académiques
Mohamed SLIMANE
Donatello CONTE

Étudiants
Florian TISSIER (DI5)

Liste des intervenants

Nom	Mail	Qualité
Florian TISSIER	florian.tissier@etu.univ-tours.fr	Étudiant DI5
Mohamed SLIMANE	mohamed.slimane@univ-tours.fr	Tuteur académique, Département infomatique
Donatello CONTE	donetello.conte@univ-tours.fr	Tuteur académique, Département infomatique

Avertissement

Ce document a été rédigé par Florian Tissier susnommé les auteurs.

L'école polytechnique de l'université François Rabelais de Tours est représentée par Mohamed Slimane et Donatello Conte susnommé les tuteurs académiques.

Par l'utilisation de ce modèle de document, l'ensemble des intervenants du projet acceptent les conditions définies ci-après.

Les auteurs reconnaissent assumer l'entière responsabilité du contenu du document ainsi que toutes suites judiciaires qui pourraient en découler du fait du non respects des lois ou des droits d'auteur.

Les auteurs attestent que les propos du document sont sincères et assument l'entière responsabilité de la véracité des propos.

Les auteurs attestent ne pas s'appropriier le travail d'autrui et que le document ne contient aucun plagiat.

Les auteurs attestent que le document ne contient aucun propos diffamatoire ou condamnable devant la loi.

Les auteurs reconnaissent qu'ils ne peuvent diffuser ce document en partie ou en intégralité sous quelque forme que ce soit sans l'accord préalable des tuteurs académiques.

Les auteurs autorisent l'école polytechnique de l'université François Rabelais de Tours à diffuser tout ou partie de ce document, sous quelque forme que ce soit, y compris après transformation en citant la source. Cette diffusion devra se faire gracieusement et être accompagnée du présent avertissement.

Pour citer ce document :

Florian Tissier, *Application d'aide à l'interaction homme/machine pour les personnes handicapées*, Projet Recherche & Développement, Ecole Polytechnique de l'Université François Rabelais de Tours, Tours, France, 2015-2016.

```
@mastersthesis{
  author={Tissier, Florian},
  title={Application d'aide à l'interaction homme/machine pour les personnes handicapées},
  type={Projet Recherche & Développement},
  school={Ecole Polytechnique de l'Université François Rabelais de Tours},
  address={Tours, France},
  year={2015-2016}
}
```

Table des matières

Introduction	1
I Recherche	2
1 Cahier des charges du projet	3
1 MOA.....	3
2 MOE.....	3
3 Contexte et présentation	3
4 Problématique et objectif	3
5 Périmètre.....	4
6 Description fonctionnelle.....	4
6.1 Repérer le visage.....	4
6.2 Reconnaître l'expression.....	4
7 Budget.....	5
8 Délai.....	5
2 Émotions universelles et mouvements du visage	6
1 Les 7 émotions universelles.....	7
2 Le FACS.....	7
3 La norme MPEG-4.....	8
3 État de l'art	10
1 3D.....	11
1.1 Techniques d'acquisition	11
1.1.1 Reconstruction à partir d'une image.....	11
1.1.2 Lumière structurée	11
1.1.3 Stéréo photométrique	11
1.1.4 Stéréo multi-vue	12

1.2	Dispositifs d'acquisition d'images.....	13
1.2.1	Kinect.....	13
1.2.2	Minolta Vivid 910.....	13
1.3	Base de données de visages 3D.....	14
1.3.1	BU-3DFE.....	14
1.3.2	BU-4DFE.....	15
1.3.3	Bosphorus 3D Face Database	15
1.3.4	Comparatif des différentes bases de données 3D	16
2	2D.....	17
2.1	Dispositifs d'acquisition d'images 2D	17
2.2	Bases de données de visages 2D	18
2.2.1	Cohn-Kanade (CK) [7].....	18
2.2.2	CK+ [10].....	19
2.2.3	Man-Machine Interaction (MMI) [18][13].....	19
2.2.4	MHI Mimicry [1] [9]	19
2.2.5	HCI Tagging [16] [8]	20
2.2.6	Comparatif des bases de données 2D.....	22
4	Architecture d'un système de reconnaissance d'émotion	23
1	Détection du visage	24
1.1	Absolu	24
1.2	Différentiel.....	24
1.3	Comparatif.....	25
2	Extraction des <i>features</i>	25
2.1	Filtres de Gabor	25
2.2	Composantes pseudo-Haar.....	25
2.3	Local Binary Pattern (LBP)	25
3	Classification	25
3.1	Classificateurs binaires.....	25
3.2	Classificateurs multi-classes.....	26
5	Critiques de l'existant et nouvelle approche	28
6	Spécifications de l'application	29
7	Planning et outils utilisés	30
II	Développement	31
	Conclusion	32
	Annexes	33

Table des figures

2 Émotions universelles et mouvements du visage

1	Exemples d'AUs	7
2	Quelques Facial Features Points utilisés par la norme MPEG-4	9
3	Facial Animation Parameter Units utilisés par la norme MPEG-4.....	9

3 État de l'art

1	Reconstruction d'un visage 2D (1) en 3D (2) grâce à la méthode 3DMM.....	11
2	Exemple de lumière structurée	12
3	Exemple de reconstruction d'un objet grâce à la stéréo photométrie.....	12
4	Version 2 de la Kinect de Microsoft	13
5	Minolta Vivid 910.....	14
6	Comparatif de la qualité entre Kinect et Minolta.....	14
7	Exemple de données contenues dans BU-3DFE	15
8	Exemple de données contenues dans la base de données Bosphorus	15
9	Exemple de caméra PTZ.....	17
10	Exemple de données de CK (de haut en bas et de gauche à droite : neutre, surprise, joie, colère, dégoût)	18
11	Exemple de données de MMI	19
12	Exemple de données de MHI Mimicry.....	20
13	Même image avec 2 tags différents : le premier erroné (Kiss), le deuxième correct (Handshake)	21
14	Données contenues dans HCI Tagging.....	22

4 Architecture d'un système de reconnaissance d'émotion

1	Exemple de projection dans un espace plus grand réalisé par SVM.....	26
2	Fonctionnement du kPPV avec $k=3$, 3 classes et 4 échantillons dont un indécis.....	27



Liste des tableaux

2 Émotions universelles et mouvements du visage	
1 Exemple de Facial Action Parameters	8
3 État de l'art	
1 Comparaison de bases de données 3D	16
2 Comparaison de bases de données 2D	22
4 Architecture d'un système de reconnaissance d'émotion	
1 Avantages/inconvénients des 2 types de détecteurs.....	25



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. Nous pouvons par exemple citer *Siri* d'Apple qui permet de communiquer avec son smartphone simplement en parlant.

C'est dans cet optique de facilitation du quotidien grâce à l'interaction avec une machine que ce projet prend place.

Le Projet de Recherche et Développement (anciennement Projet de Fin d'Études) est un projet se déroulant durant toute la 5ème année de master ingénieur au sein de l'école Polytech Tours. Ce rapport va présenter les travaux que j'ai effectué durant toute la durée de ce projet.

Ce rapport va donc se diviser en deux grandes parties : tout d'abord la partie Recherche qui va contenir le cahier des charges du projet, l'état de l'art en matière de reconnaissance faciale d'émotions ainsi que toutes les bases théoriques dont j'aurais besoin par la suite. La deuxième partie est la partie Développement qui va se concentrer sur les différentes étapes du développement de cette application d'aide aux personnes handicapées.

Dans la partie Recherche de ce rapport, après avoir défini le cahier des charges, je vais vous présenter les sept émotions universelles ainsi que deux normes majeures permettant de définir les mouvements du visages qui composent une expression.

Je vous présenterai ensuite un état de l'art en matière de reconnaissance faciale d'expression, notamment les différentes techniques permettant de capturer un visage et de reconnaître une émotion, tout d'abord en 3D puis en 2D, ainsi que les avantages et inconvénients de chacune de ces techniques. Dans chacune de ces étapes, je présenterai l'état de l'art actuelle en terme de matériel, de méthodes et également de base de données disponibles.

Je continuerai ensuite en présentant les différentes étapes nécessaires à la construction d'un système de reconnaissance faciale d'émotions performant et efficace.

Je définirai ensuite les spécifications de l'application ainsi que les choix qui ont permit cette définition. Je conclurai par un planning prévisionnel ainsi que par les méthodologies et les outils de suivi utilisés durant ce projet.

Dans la partie Développement, [rédaction en deuxième partie de l'année]

Pour ce projet de recherche et développement, j'ai été encadré par Donatello Conte et Mohamed Slimane.

Première partie

Recherche

1

Cahier des charges du projet

1 MOA

- Donatello CONTE : Maître de conférence et enseignant chercheur en informatique à l'école Polytech Tours
- Mohamed SLIMANE : Professeur des Universités et enseignant chercheur en informatique à l'école Polytech Tours

2 MOE

Florian TISSIER : élève en dernière année de master ingénieur en informatique à Polytech Tours

3 Contexte et présentation

Ce projet de recherche et développement prend place dans le cursus de dernière année de master ingénieur dispensé à l'école Polytech Tours.

Monsieur Slimane étant membre d'une association s'occupant de personnes handicapées, il souhaitait pouvoir aider et faciliter la vie de ces derniers via un projet réalisé au sein de l'école.

Le but de ce projet de recherche et développement est de construire un système permettant, à partir d'un flux vidéo acquis grâce à une caméra, de détecter l'expression du visage actuelle d'une personne handicapée dans le but de réaliser certaines actions pouvant améliorer son bien être.

4 Problématique et objectif

Comment faciliter la vie quotidienne des personnes handicapées grâce à leurs émotions ?

L'objectif est de réaliser une application qui pourra détecter en temps réel les émotions d'une personne handicapée se trouvant dans une pièce à l'aide d'une caméra fixée à un mur de cette même pièce. La caméra devra être capable de bouger et de zoomer pour suivre le visage de la personne.

Une fois l'émotion détectée, des actions spécifiques devront être réalisées (ex : changement de la couleur de la lumière, lecture de musique douce...)

5 Périmètre

Ce projet se concentre principalement sur les personnes handicapées mais l'application qui résultera de ce projet pourra également être utilisée pour des personnes non handicapées. L'application devra être fonctionnel dans n'importe quelle pièce d'une maison ou d'une structure spécialisée dans l'accueil de personnes handicapées.

6 Description fonctionnelle

Le projet se découpe en 2 fonctions principales :

- Repérer le visage
- Reconnaître l'expression

Chacune de ces fonctions se décomposent en plusieurs sous-fonctions.

6.1 Repérer le visage

Cette première fonction principale va permettre de trouver le visage d'une personne, de le suivre et de zoomer dessus.

Les sous-fonctions suivantes seront donc nécessaires :

- Un algorithme de détection de visage dans une image
- Un algorithme de suivi de visage
- Un algorithme de zoom

Vous trouverez ci-après un descriptif plus précis de chacune de ces sous-fonctions.

Fonction : Repérer le visage/Détection du visage	
Objectif	Détecter un visage dans un environnement quelconque.
Description	En analysant les frames d'un flux vidéo, cet algorithme nous renverra la position d'un cadre entourant le visage trouvé.
Contraintes	Cet algorithme doit être rapide.
Niveau de priorité	Haute
Fonction : Repérer le visage/Suivi du visage	
Objectif	Suivre le visage entre plusieurs frames d'un flux vidéo dans le but de ne pas le perdre.
Description	En comparant 2 frames consécutives d'un flux vidéo, l'algorithme devra nous dire le déplacement du visage pour pouvoir le suivre avec la caméra. Il devra également être capable de faire le suivi si jamais le visage se retrouve occulté pendant quelques secondes.
Contraintes	Éviter le plus possible l'accumulation d'erreur de précision pouvant amener à la perte du visage.
Niveau de priorité	Haute
Fonction : Repérer le visage/Zoom	
Objectif	Zoomer sur une zone de l'image.
Description	L'algorithme devra pouvoir zoomer sur le cadre contenant le visage renvoyé par la sous-fonction de détection du visage.
Contraintes	Disposer d'une caméra ayant la possibilité de zoomer.
Niveau de priorité	Moyenne

6.2 Reconnaître l'expression

Cette deuxième fonction principale va permettre d'identifier l'expression faciale de la personne.

Les sous-fonctions suivantes seront donc nécessaires :

- Un algorithme d'apprentissage
- Un algorithme d'extraction des points clés (*features*) du visage
- Un algorithme de classification

Vous trouverez ci-après un descriptif plus précis de chacune de ces sous-fonctions.

Fonction : Reconnaître l'expression/Apprentissage	
Objectif	Apprendre au système à classifier les expressions en fonction d'une base d'apprentissage.
Description	Pour chaque élément dans la base d'apprentissage, une émotion lui sera associé. Cela va permettre au système d'apprendre à quel expression du visage appartient une émotion.
Contraintes	La base d'apprentissage doit être assez fourni et pertinente pour permettre un apprentissage performant.
Niveau de priorité	Haute

Fonction : Reconnaître l'expression/Extraction des <i>features</i> du visage.	
Objectif	Extraire les <i>features</i> du visage pour pouvoir ensuite réaliser une classification.
Description	Cet algorithme devra extraire les <i>features</i> du visage d'une personne se trouvant sur une frame d'un flux vidéo. Les <i>features</i> d'un visage sont par exemple le coin des yeux, la position de la pupille, les coins de la bouche, le nez, les joues... Les positions des <i>features</i> retournées permettront la classification.
Contraintes	
Niveau de priorité	Haute

Fonction : Reconnaître l'expression/Classification	
Objectif	Classifier l'expression du visage et retourner l'émotion associée.
Description	L'algorithme devra classifier l'expression du visage de la personne récupéré depuis un flux vidéo, grâce aux positions des <i>features</i> , en fonction de l'apprentissage qui aura été effectué précédemment.
Contraintes	Cet algorithme doit être rapide et fiable (au moins 90% de reconnaissance).
Niveau de priorité	Haute

7 Budget

Ce projet ne dispose pas d'un budget précis.

Néanmoins en réalisant un état de l'art des matériels disponibles à notre projet, nous avons décidé de choisir une caméra ne dépassant pas les 500€.

8 Délai

La partie recherche devra être fini pour le 15 janvier 2016 et la partie développement (et donc le projet complet) devra être fini pour la fin du second semestre, c'est-à-dire fin mars/début avril.

2

Émotions universelles et mouvements du visage

Ce chapitre va me permettre d'introduire les notions nécessaires à la décomposition et donc à la reconnaissance d'une expression faciale et de son émotion associée.

Pour un simple sourire, nous utilisons une vingtaine de muscles (les muscles zygomatiques), il ne peut donc pas être décrit par un seul mouvement du visage mais plusieurs. C'est exactement la même chose pour une expression, on ne la reconnaît que grâce à l'ensemble des mouvements faciaux qui la composent. C'est sur ce principe qu'on a créé les deux normes de description des mouvements du visage que je vais vous présenter : le FACS et MPEG-4.

L'utilisation de l'une ou l'autre de ces normes permet de définir entièrement le spectre des mouvements rentrant en jeu dans n'importe quelle émotion.

Mais tout d'abord, je vais vous introduire les sept émotions universelles qui seront utilisées tout au long de mon projet et donc de ce rapport.

1 Les 7 émotions universelles

A ce jour, il a été démontré qu'il existait 7 émotions qui partagent une expression universellement compréhensible.

On considère qu'une émotion possède une expression universelle si tout individu est capable d'exprimer cette émotion et est également capable de la reconnaître et de l'interpréter chez autrui.

Les sept émotions universelles sont donc les suivantes :

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

C'est Charles Darwin qui, en 1872 dans son livre [4], a introduit cette idée d'émotions universelles entre les hommes mais également entre différentes espèces. Il a observé que les hommes et les animaux partagent des émotions comprises par tous et qui sont nécessaires à leur survie.

Mais ce n'est qu'en 1971 que le psychologue Paul Ekman, après un voyage en Papouasie-Nouvelle-Guinée, a confirmé les théories de Darwin. Dans son article [6] écrit avec la participation de Wallace Friesen, il définit les 7 émotions universelles citées plus haut.

2 Le FACS

En 1978, Ekman et Friesen publie [5] et apporte une nouvelle pierre à l'édifice en définissant un système de codification manuelle des expressions du visage : le **Facial Action Coding System** (FACS).

Ce système décompose tous les mouvements du visage en 46 **Action Units** (AU), chacune décrivant la contraction ou la décontraction d'un ou plusieurs muscles du visage. La Figure 1 représente certaines de ces AUs.















AU1  Inner brow raiser	AU2  Outer brow raiser	AU4  Brow Lowerer	AU5  Upper lid raiser	AU6  Cheek raiser
AU7  Lid tighten	AU9  Nose wrinkle	AU12  Lip corner puller	AU15  Lip corner depressor	AU17  Chin raiser
AU23  Lip tighten	AU24  Lip presser	AU25  Lips part	AU27  Mouth stretch	

Figure 1 – Exemples d'AUs

La composition de plusieurs AU permet donc de décrire une expression et donc de reconnaître une émotion. Par exemple, un sourire et donc l'émotion de la joie est composé des AUs 6 (remontée des joues) et 12 (étirement du coin des lèvres).

La tristesse quand à elle va être composée des AUs 1, 4 et 15 et la colère des AUs 4, 5, 7 et 23.

N'importe quelle expression du visage peut donc être représentée par une combinaison d'AU, ce qui fait de FACS le système le plus utilisé par les psychologues ainsi que par les personnes travaillant sur la

reconnaissance faciale d'émotions.

Le système **FACS** possède également un degré d'intensité allant de A à E et permettant de spécifier l'intensité d'une AU :

- A : Très Faible
- B : Minimale
- C : Moyen
- D : Sévère
- E : Maximum

La surprise peut donc être défini comme la combinaison des AUs 1, 2, 5B et 26.

Enfin des ajouts ont été apportés à cette norme. 13 nouveaux AUs ont été ajoutées pour décrire le mouvement de la tête et 7 autres pour le mouvement des yeux.

Nous arrivons donc à un total de 66 AUs, chacune possédant 5 intensités, permettant de décrire les mouvements faciaux.

Néanmoins, un autre système de codification fait concurrence au FACS et est également bien implanté dans le milieu de la reconnaissance d'émotions.

3 La norme MPEG-4

La norme MPEG-4, qui est une norme de codage vidéo, dispose de son propre système permettant de normaliser les mouvements du visage et de reconnaître des expressions.

Pour cela, ce système définit des points clés du visage (**Figure 2**) appelés **Facial Features Points** (FFP) auxquels seront appliqués des mesures pour créer des distances entre ces FFP (**Figure 3**) appelées **Facial Animation Parameter Units** (FAPU).

Ces FAPU vont servir à la description des mouvements musculaires appelés **Facial Action Parameters** (FAP, équivalent des AUs de la norme FACS). 68 FAPs sont recensés à ce jour, j'en ai regroupé quelques uns dans le tableau **Table 1**.

Le descriptif complet de ces FAP se trouve dans le document à cette adresse [[WWW6](#)], dans l'annexe numéro 1.

Table 1 – Exemple de Facial Action Parameters

Numéro	Nom	Description
3	open_jaw	Vertical jaw displacement (does not affect mouth opening)
7	stretch_r_cornerlip	Horizontal displacement of right inner lip corner
10	raise_b_lip_lm	Vertical displacement of midpoint between left corner and middle of bottom inner lip
42	lift_r_cheek	Vertical displacement of right cheek

Cependant, la norme de MPEG-4 est moins réaliste que FACS d'un point de vue musculaire.

Par exemple : l'Au 26 de FACS (« Jaw Drop ») décrit le mouvement d'abaissement du menton, cet abaissement est accompagné d'un abaissement de la lèvre inférieure. Or l'abaissement du menton de MPEG-4 (FAP 3 - open_jaw) ne décrit pas l'abaissement de la lèvre inférieure.

MPEG-4 ne décrit que les mouvements *visibles* du visages, contrairement à FACS qui lui décrit les mouvements *réalistes* du visage.

Nous allons maintenant voir quelles technologies sont disponibles pour réaliser l'acquisition des images qui seront à traiter par la suite.

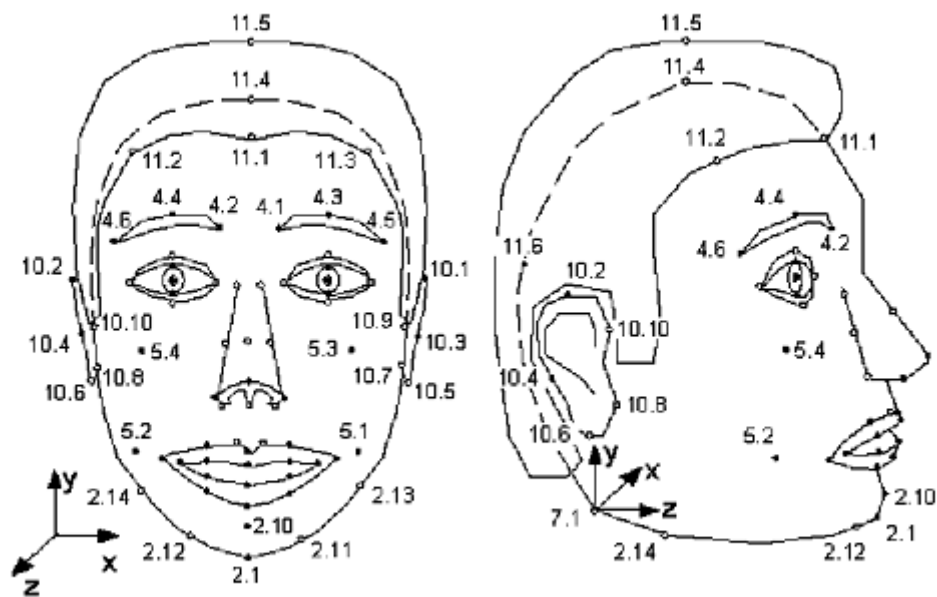


Figure 2 – Quelques Facial Features Points utilisés par la norme MPEG-4

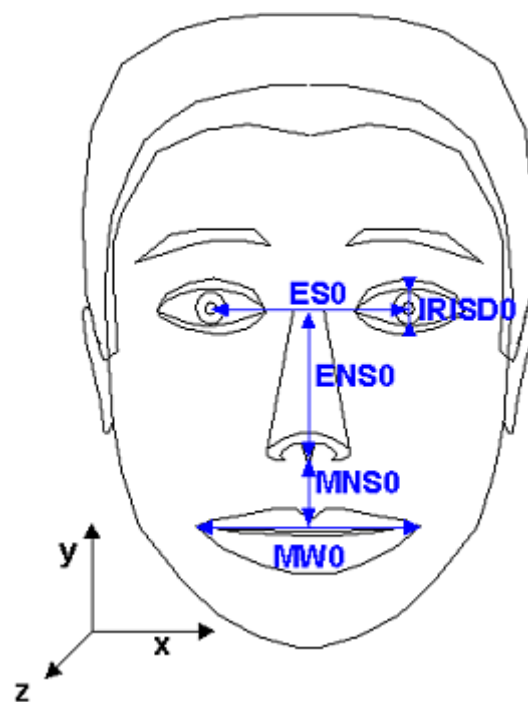


Figure 3 – Facial Animation Parameter Units utilisés par la norme MPEG-4

3

État de l'art

Dans ce chapitre, je vais vous présenter l'état de l'art des systèmes et techniques existants permettant de faire de la reconnaissance facial d'émotion.

Mon projet se basant sur des images et vidéos en 2D, ce chapitre va principalement se concentrer sur les techniques d'acquisition et d'analyse d'images 2D.

Cependant, j'ai tout de même réalisé quelques recherches sur le matériel utilisé pour l'acquisition d'images en 3D ainsi que les bases de données disponibles et c'est donc par ces recherches que je vais entamer ce chapitre.

1 3D

Comme je l'ai expliqué plus haut, mon projet et donc mon système de reconnaissance utilisera des images 2D mais j'ai tout de même mené quelques recherches sur les images en 3D également.

Je vais tout d'abord vous présenter les techniques utilisées pour acquérir des images en 3D. Je vous présenterai ensuite certains matériels déjà existant permettant de faire de la capture d'images 3D et utilisant les techniques que je vous aurais présenter précédemment. Je terminerai enfin cette section sur la 3D en vous présentant les bases de données de visages les plus connues et les plus utilisés par les systèmes de reconnaissance faciale d'émotion en 3D.

Les informations que vous allez trouver dans cette section proviennent en majorité de l'article [14].

1.1 Techniques d'acquisition

1.1.1 Reconstruction à partir d'une image

Il est possible, à partir d'une image en 2D capturée par une caméra basique, d'obtenir une image en 3D. La méthode la plus prometteuse est celle du **3D Morphable Model** (3DMM), qui consiste à apposer un visage en 3D (masque) sur l'image en 2D et que le modifier pour le faire correspondre avec l'image. Sont ensuite extraites les informations correspondant au masque modifié qui vont permettre de créer le visage de l'image en 3D.

La **Figure 1** présente un exemple de visage 3D récupéré depuis une image 2D.

Cette technique est très pratique et répandue car elle ne nécessite pas de matériel au coût exorbitant, une simple caméra est nécessaire.

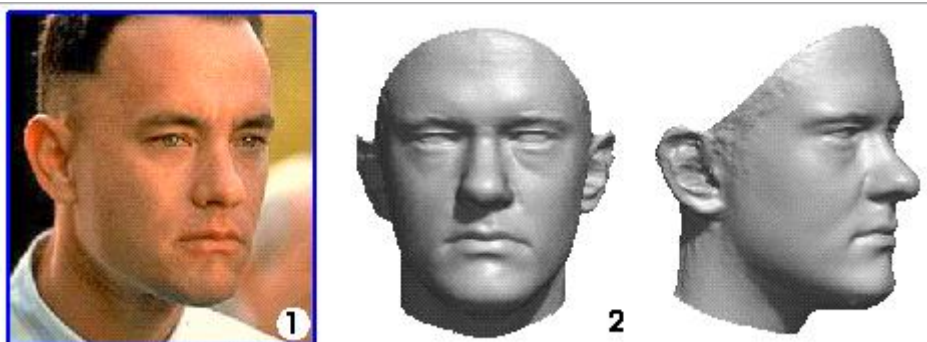


Figure 1 – Reconstruction d'un visage 2D (1) en 3D (2) grâce à la méthode 3DMM

1.1.2 Lumière structurée

Une autre technique, une des plus utilisées, est la technique de la lumière structurée.

Elle consiste à projeter plusieurs rais de lumière (visible ou infra-rouge) de longueur d'onde différente puis, à l'aide d'un capteur, de mesurer la déformation de ces rais de lumière pour construire le visage en 3D.

La **Figure 2** montre un exemple de lumière structurée.

L'utilisation de cette technique requiert d'avoir un matériel spécifique contenant un émetteur et un récepteur, cet outil peut aller de quelques centaines d'euros pour les moins chères à plusieurs dizaines de milliers d'euros pour les plus performantes.

1.1.3 Stéréo photométrie

La technique de la stéréo photométrie consiste à prendre plusieurs photos d'un même objet avec un même appareil sous différentes illuminations (lumière venant de droite, lumière venant de devant ...).

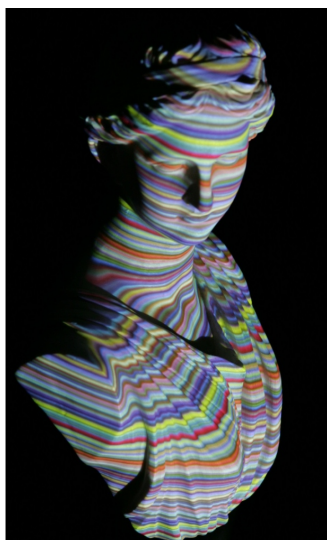


Figure 2 – Exemple de lumière structurée

Pour obtenir un visage en 3D, il ne reste qu'à assembler les photos obtenues.

La **Figure 3** montre un exemple de reconstruction d'un objet grâce à la stéréo photométrie.

Cette technique requiert également un équipement coûteux et n'est donc pas accessible à tout le monde.

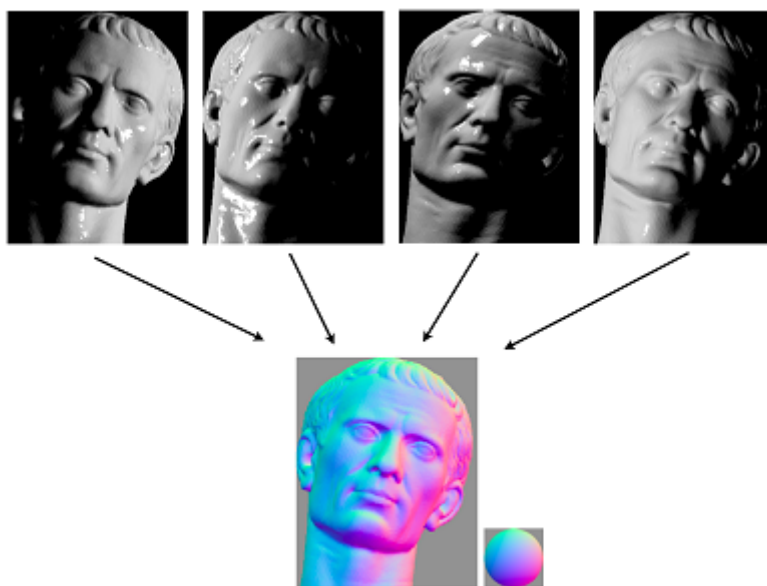


Figure 3 – Exemple de reconstruction d'un objet grâce à la stéréo photométrie

1.1.4 Stéréo multi-vue

C'est une technique très similaire à celle de la stéréo photométrie sauf qu'au lieu de prendre en photo un visage sous différentes illuminations, le visage est pris sous différents angles simultanément avec plusieurs appareils.

Cette technique requiert elle aussi un équipement spécifique coûteux.

1.2 Dispositifs d'acquisition d'images

Plusieurs types de dispositifs différents existent à l'heure actuelle permettant de capturer des images en 3D. La plupart d'entre eux se basent sur les techniques présentées précédemment. Je vais ici vous présenter les deux dispositifs les plus connus.

1.2.1 Kinect

Probablement la caméra 3D la plus connue du grand public : la Kinect de Microsoft. Créée initialement pour créer une immersion plus poussée pour les jeux de la console Xbox 360, elle a depuis été améliorée et ne sert plus exclusivement qu'à jouer aux jeux vidéos. Elle utilise la technique de la lumière structurée (infra-rouge dans ce cas) pour capturer les images en 3D de ce qu'elle filme. La Kinect reste cependant une caméra 3D low-cost, de mauvaise qualité lorsqu'on le compare à d'autres dispositifs du même type, tel que celui que je vais présenter maintenant.



Figure 4 – Version 2 de la Kinect de Microsoft

1.2.2 Minolta Vivid 910

Un autre dispositif très utilisé et utilisant lui aussi la technologie de la lumière structurée est le Minolta Vivid 910.

Un comparatif entre la qualité de Kinect et de Minolta est présenté en **Figure 6**. On s'aperçoit très clairement du gouffre séparant ces deux dispositifs. Bien sûr le prix n'est pas le même, car nous passons de quelques centaines d'euros pour la Kinect à plusieurs dizaines de milliers d'euros pour le Minolta Vivid 910.



Figure 5 – Minolta Vivid 910



Figure 6 – Comparatif de la qualité entre Kinect et Minolta

1.3 Base de données de visages 3D

Les dispositifs tels que le Minolta Vivid 910 présentées précédemment permettent également la création de bases de données de visage en 3D. Leurs grandes qualités permettent d'obtenir des images très précises, facilement exploitable.

Je vais maintenant vous présenter les bases de données les plus connues et pouvant être utilisée par la communauté scientifique.

1.3.1 BU-3DFE

Les premiers efforts pour récolter des données en 3D ont menés à la création de la base de données BU-3DFE (Binghamton University 3D Facial Expression [23]).

Les images contenues dans cette base sont statiques et ont été capturées par le dispositif 3dMD. Elles se composent de 100 sujets âgés de 18 à 70 ans, dont 56% de femmes, et appartenant à différentes ethnicités.

Chaque sujet réalise les 7 expressions basiques (cf [Section 1](#) (Chapitre 2)) et chaque expression, sauf l'expression neutre, est réalisée suivant 4 niveaux d'intensités. Pour chaque sujet, il y a donc 25 images différentes, ce qui nous donne au total 2500 images dans cette base de données.

Un exemple de données contenues dans cette base sont présentés en [Figure 7](#).

Chaque donnée contient également la position de 83 points clés du visage.



Figure 7 – Exemple de données contenues dans BU-3DFE

1.3.2 BU-4DFE

Une extension de la base BU-3DFE a été réalisée dans le but d'obtenir un espace 3D dynamique, c'est-à-dire rajouter la dimension du temps dans les images de la base, en plus des 3 dimensions déjà présentes. En effet, dans la version de base, il n'y avait qu'une seule image présente pour une expression et une intensité, mais dans le but d'obtenir des analyses plus performantes, inclure la notion du temps dans ces images devient indispensable.

La base de données BU-4DFE [22] se compose donc de 101 sujets (58 femmes) de différentes ethnicités, chaque sujet réalisant 6 des expressions basiques (toutes sauf la neutre). Chaque séquence d'émotion contient environ 100 frames, ce qui nous permet d'obtenir environ 60600 différents frames dans la base.

1.3.3 Bosphorus 3D Face Database

La base de données Bosphorus [15] Cette base de données est composé de 105 sujets (45 femmes) dont la plupart sont de type Caucasien et dont un tiers sont des acteurs professionnels.

Chaque sujet réalise environ 35 expressions et toutes les images sont codés en terme de FACS (Section 2 (Chapitre 2)).

Plusieurs illuminations et occlusions du visage (barbe, moustache, lunettes...) sont également présentes pour chaque sujet. 24 points clés du visage sont également définis pour chaque donnée.

Un exemple de donnée est présente en Figure 8

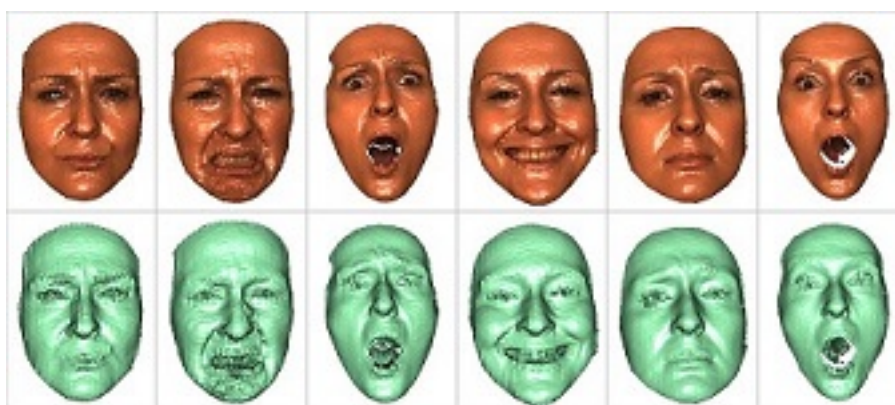


Figure 8 – Exemple de données contenues dans la base de données Bosphorus

1.3.4 Comparatif des différentes bases de données 3D

Précédemment, je ne vous est présenté que les bases de données les plus populaires. Cependant, beaucoup d'autres existent également.

J'ai donc réalisé un tableau (**Table 1**) recensant ces différentes bases de données publiquement disponible, et pouvant représenté un intérêt dans la reconnaissance d'émotions, avec leurs principales caractéristiques (données principalement tirées de [14]).

Table 1 – *Comparaison de bases de données 3D*

Nom	Type de données	Taille	Contenu
BU-3DFE[23]	Statique	100 adultes	6 émotions basiques avec 4 niveaux d'intensités
BU-4DFE[22]	Dynamique	101 adultes	6 émotions basiques
Bosphorus[15]	Statique	105 adultes (dont 25 acteurs)	6 émotions basiques, 24 AUs, occlusions
ICT-3DRFE[17]	Statique	23 adultes	6 émotions basiques, 2 expressions neutres, 4 orientations du regard (haut, bas, gauche, droite) et un visage "grimaçant"
D3DFACS[3]	Dynamique	10 adultes (dont 4 experts en FACS)	Jusqu'à 38 AUs par sujets
Gavabdb[12]	Statique	61 adultes	3 expressions : sourires ouverts/fermés et aléatoire

2 2D

Passons maintenant à ce qui va m'être utile à la réalisation de mon projet : la 2D.

Dans cette section, je présenterai tout d'abord rapidement les dispositifs permettant de récupérer des images en 2D, puis plusieurs bases de données de visages 2D et enfin des méthodes permettant de réaliser de la reconnaissance faciale d'expression.

2.1 Dispositifs d'acquisition d'images 2D

Contrairement à la 3D, il n'est pas nécessaire d'avoir des dispositifs extrêmement coûteux pour acquérir des images en 2D.

En effet, un simple appareil photo ou une simple caméra trouvés dans n'importe quelle commerce suffit amplement.

Plusieurs entreprises se sont spécialisées dans le domaine de reconnaissance d'émotions et permettent, par exemple, aux publicitaires de tester leur publicités et d'avoir un feedback sur ce que ressentent les spectateurs. Pour cela, ces entreprises([WWW2],[WWW1]) utilisent la webcam intégrée dans les ordinateurs pour ensuite traiter les images.

Dans le cadre de ce projet, nous utiliserons une caméra **PTZ** (Pan, Tilt, Zoom). Ces caméras permettent de faire une rotation selon l'axe Z (Pan), une rotation selon l'axe X (Tilt) et de zoomer selon l'axe Y. Souvent utilisé en temps que caméra de surveillance, dans le cadre de ce projet, ce type de caméra va nous permettre de se déplacer pour trouver la personne présente dans la pièce puis de zoomer sur son visage pour pouvoir ensuite l'analyser.

Un exemple de caméra PTZ est présenté en **Figure 9**.



Figure 9 – Exemple de caméra PTZ

Dans le cas d'images statiques (images, photos), la reconnaissance se fera directement sur l'image.

Par contre dans le cas d'un flux vidéo récupéré via une caméra, c'est les *frames* (images constituant une vidéo) de la vidéo qui vont être analysées.

2.2 Bases de données de visages 2D

Je vais ici vous présenter différentes bases de données de visages 2D, libre d'accès à la communauté scientifique, intéressantes pour la recherche et permettant de réaliser une reconnaissance faciale d'expressions.

2.2.1 Cohn-Kanade (CK) [7]

Probablement la base de données de visage 2D la plus connue et la plus utilisée pour la reconnaissance faciale d'expressions, elle se compose de 97 sujets (65% femmes, 15% afro-américains et 3% asiatiques ou sud américain) âgés de 18 à 30 ans et réalisant les 6 expressions universelles (joie, tristesse, dégoût, peur, surprise et colère).

Cette base contient au total 486 séquences vidéos, du visage neutre au départ jusqu'à l'**apex** (le pic de l'émotion) à la fin. Ces séquences sont en niveaux de gris et digitalisé en tableaux de 640*480 pixels avec donc une précision de 8 bits (dû aux niveaux de gris).

Toutes ces séquences sont entièrement codées en termes de FACS et d'AUs mais ne contiennent pas de label spécifiant l'émotion présentée.

Toutes les expressions présentes dans les séquences vidéos contenues dans cette base sont posées et non spontanées, cela signifie qu'on a demandé à ces personnes de produire telle ou telle émotion, cela ne leur est pas venu d'eux-même spontanément. Les expressions présentées seront donc plus "exagérées" qu'en temps normal.

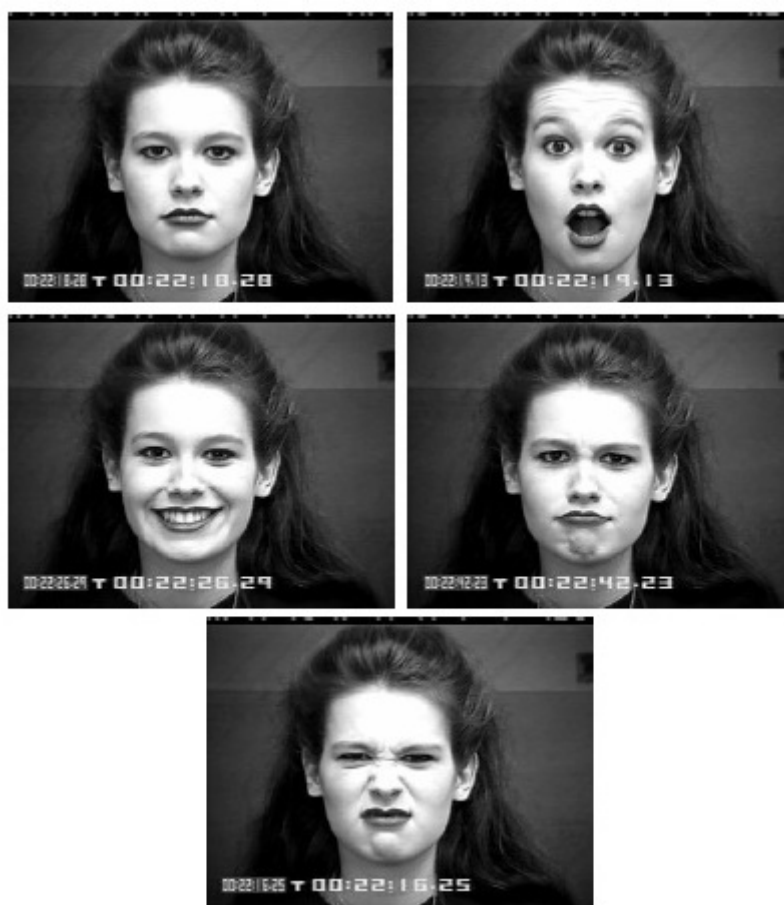


Figure 10 – Exemple de données de CK (de haut en bas et de gauche à droite : neutre, surprise, joie, colère, dégoût)

2.2.2 CK+ [10]

C'est la version améliorée de la base de données CK.

Cette fois-ci, elle contient des expressions posées et spontanées.

Dans le cas des expressions posées, le nombre de sujets a été augmenté 27% et le nombre des séquences de 22%. Les séquences sont toujours codées en termes de FACS et d'AUS avec, cette fois-ci, un label associé à l'expression présentée qui se trouvera dans les metadata de la vidéos.

Cette version propose également des protocoles et des résultats pour le suivi des points clés du visage ainsi que pour la reconnaissance d'émotion.

Actuellement, une troisième version de CK est en préparation, avec comme principale ajout la synchronisation entre une séquence frontale et une séquence orientée de 30 degrés par rapport à la vue frontal.

2.2.3 Man-Machine Interaction (MMI) [18][13]

Créée en 2002, cette base de données, comme CK+, se compose de deux parties : une partie posée et une partie spontanée.

Cette base contient plus de 2900 vidéos ainsi que des images en haute résolution de 75 sujets (48% de femmes, européen, africain et sud américain) âgés de 19 à 62 ans. Chaque séquence vidéos contient soit une expression dans son entièreté, soit un AU spécifique

Les données de cette base sont également codées en terme de FACS et d'AUs.

L'avantage de cette base de données est qu'elle contient toute l'étendue d'une expression : visage neutre - **onset** (début de l'expression) - **apex** (pic) - **offset** (fin de l'expression) - visage neutre .

Cela permet de réaliser une reconnaissance d'émotion dans le temps plus précise.

Certaines des séquences vidéos sont en couleur, les autres sont en niveaux de gris.



Figure 11 – Exemple de données de MMI

2.2.4 MHI Mimicry [1] [9]

Cette base de données est intéressante car elle diffère totalement des deux précédentes au niveau de sa construction.

Elle se compose de 54 vidéos (sessions) de 40 sujets (28 hommes, 12 femmes) entre 18 et 40 ans venant de l'Imperial College de Londres. Sur chaque session, 2 participants interagissent et chaque session est divisée en deux parties.

Tout d'abord une partie débat dans laquelle les deux participants vont discuter de politique. Ils seront donc soit du même avis, soit d'un avis opposé.

La deuxième partie consiste en un "jeu de rôle" : un participant joue le rôle d'un étudiant cherchant un appartement et le deuxième participant joue le rôle d'un propriétaire d'appartement. Le but de l'étudiant est donc de trouver un appartement et celui du propriétaire de louer son appartement, ils ont donc un but qui les relie.

Cela permet d'observer le comportement humain lors d'une discussion où l'on est d'accord ou non et lorsque l'on veut convaincre quelqu'un, comment nous "mimons" ou non le comportement de notre interlocuteur dans le but de montrer nos intentions, de nous faire accepter.

Chaque session est divisée en "épisodes d'intérêt" et chaque épisode est labellisé en fonction de ce qu'on y voit : sourire, hochement de tête, penchement du corps en avant ou en arrière.

Sont aussi labellisés le rôle de chaque participant (s'il parle ou s'il écoute) ainsi que leurs intentions (2 catégories) :

- Social Signal Expression (inconscient) : compréhension, accord, confusion, "liking"
- Desired Goal (conscient) : flatter l'autre, souligner la compréhension, exprimer l'accord, partager l'empathie, augmenter l'acceptation.

Tous les enregistrements ont été réalisés avec 15 caméras (7 par participants et une vue d'ensemble) et 3 micros (un micro de tête par participant et un au milieu de la pièce). Un exemple des images récupérées par les caméras se trouvent en [Figure 12](#)



Figure 12 – Exemple de données de MHI Mimicry

Cependant, sur cette base de données, aucune notion de FACS ni d'AUs n'est présente, les données sont labellisées en fonction du ressenti et de l'auto-évaluation réalisée par les participants..

2.2.5 HCI Tagging [16] [8]

C'est avec cette base de données que j'ai travaillé à la réalisation de mon projet.

Créée par la même équipe que celle qui a créée MHI Mimicry, cette base de données se compose de 27 participants (11 hommes, 16 femmes) de 19 à 40 ans.

Elle est divisée en 2 parties : la première partie contient 20 vidéos et la deuxième 28 vidéos et 14 images. Chaque vidéos contient la vidéo du visage, l'audio et les expressions vocales, la position où regardent les yeux et les signaux physiologiques (température du corps, rythme de la respiration, EEG, rythme cardiaque). Les vidéos ont été filmées via 6 caméras : 5 en niveaux de gris et d'orientations différentes (frontal, bas-gauche, bas-droite, vue d'ensemble et profil gauche) et une en couleur (frontal).

Concernant la partie 1 (Explicit Tagging) : chaque participant regarde plusieurs vidéos et à la fin de chaque vidéo un feedback sur leur ressenti leur ai demandé :

- émotion ressenti (parmi neutre, anxieux, amusé, triste, joyeux, dégoûté, en colère, surpris ou apeuré)
- excitation (sur une échelle de 1 à 9)
- agréabilité (sur une échelle de 1 à 9)
- dominance (sur une échelle de 1 à 9)

— prédictibilité (sur une échelle de 1 à 9)

Les vidéos visionnées par les participants lors de la partie 1 sont séparées par de petits clips courts neutre pour remettre l'expression de la personne à 0, la réinitialiser en quelque sorte.

Concernant la partie 2 (Implicit Tagging) : on diffuse à chaque participant une image ou une vidéo deux fois. La première fois sans rien puis la deuxième fois avec un *tag* censé décrire ce qu'il se passe dans la vidéo/image. Soit ce tag est correct soit il est erroné, le participant doit donc dire si il est d'accord ou non avec le tag attribué en appuyant respectivement sur un bouton vert ou rouge. Pour illustrer cela, un exemple est présent en **Figure 13**. Cependant, comme les créateurs de cette base n'ont pas les droits pour ces images, les images accessibles que j'ai récupéré ne contiennent que les bords contrairement aux images diffusées aux participants.



Figure 13 – Même image avec 2 tags différents : le premier erroné (Kiss), le deuxième correct (Handshake)

Un exemple des différentes données vidéos contenues dans cette base est en **Figure 14**.

Pareillement à MHI Mimicry, aucune notion de FACS ni d'AUs n'est présente, les données sont labellisées en fonction de l'auto-évaluation réalisée par tous les participants.



Figure 14 – Données contenues dans HCI Tagging

2.2.6 Comparatif des bases de données 2D

Pour récapituler toutes les informations que j'ai pu donner précédemment et pour introduire de nouvelles bases de données accessibles à la communauté scientifique dont je n'ai pas parlé, j'ai donc réalisé le tableau comparatif suivant :

Table 2 – Comparaison de bases de données 2D

Nom	P/S	Taille	Contenu	Label
CK[7]	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+[10]	P/S	Nombre de sujets augmenté 27%	Nombre de séquences augmenté 27%	FACS + émotion
MMI[18] [13]	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 Mimi-cry [1][9]	S	40 sujets (28 hommes, 12 femmes) entre 18 et 40 ans	54 vidéos (sessions)	Auto-évaluation
HCI Tagging [16][8]	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 [WWW5]	S	24 sujets	10 heures de vidéo : les sujets parle à une intelligence artificielle et leurs émotions sont changées en fonction des différentes personnalités de l'IA	Feeltrace
JAFFE[11]	P	10 femmes japonaises	213 images des 7 émotions basiques	Noté selon 6 adjectifs d'émotion par 60 sujets Japonais

4

Architecture d'un système de reconnaissance d'émotion

Dans ce chapitre je vais introduire les différentes parties nécessaires à la construction d'un système permettant de reconnaître des expressions et émotions.

Un tel système se constitue de 4 parties (ou 3 car la première et deuxième peuvent être combiné en une seule) :

- Détection du visage
- Extraction des *features* (nez, bouches, yeux ...)
- Classification

Chacune de ces parties sera décrit plus précisément dans les sections suivantes avec un état de l'art sur les techniques existantes pour chacune d'entre elles.

Une partie des informations se trouvant de ce chapitre sont tirées de [21] écrit par les fondateurs de la société Emotient et qui présente un état de l'art des différentes parties d'un système de reconnaissances d'émotions.

1 Détection du visage

Depuis déjà plusieurs années, la détection de visage dans une image ou vidéo est devenu une réalité grâce aux algorithmes d'apprentissage.

La détection de visage sur une vidéo ou une image se retrouve de plus en plus dans notre quotidien. Tout d'abord dans nos appareils photos ou smartphones mais également sur les réseaux sociaux comme Facebook, qui repère les visages sur une photos lorsque l'on souhaite identifier des personnes, ou encore Snapchat, qui depuis la dernière mise à jour repère notre visage grâce à la caméra frontale d'un smartphone pour ensuite lui appliquer diverses animations et déformations.

Les algorithmes de détection de visage se divisent en 2 catégories : les absolus et les différentiels, chacun ayant leurs avantages et inconvénients.

1.1 Absolu

Les détecteurs absolus, aussi appelés détecteurs *frame-by-frame*, vont déterminer la position d'un visage sur chaque frame d'une vidéo, indépendamment des frames précédentes.

Le principale avantage de ces détecteurs est qu'ils sont très facilement parallélisable sur plusieurs frames d'une vidéo en même temps. Il permet également de réagir très rapidement si jamais le nombre de visages dans l'image change subitement et ne "dérive" pas au court du temps. Ici le terme "dériver" fait référence au fait de perdre la position d'un visage.

Cependant, comme dit précédemment, ces détecteurs n'utilisent pas de notions de temps qui pourraient les rendre plus rapide et précis.

L'algorithme de détection utilisé par la plupart de ces détecteurs est l'algorithme de **Viola-Jones** [19] créé en 2001.

Cet algorithme va tout d'abord apprendre un classificateur à différencier des visages d'autres objets grâce à un apprentissage réalisés avec des images de différentes tailles de visages ou de divers autres objets. A noté qu'il est préférable d'avoir un nombre d'objets quelconque très supérieur au nombre de visages si nous voulons que l'algorithme soit efficace.

Une fois l'apprentissage fini et que l'on passe une nouvelle image à l'algorithme, il va analyser cette dernière en extrayant plusieurs *patches*, des bouts de l'image, de différentes tailles qu'il va ensuite normaliser à une taille précise (par exemple 48*48 pixels) puis les donner au classificateur qui va se charger de définir si oui ou non se trouve un visage dans ce patch.

Cette étape d'extraction de patches peut également être paralléliser pour gagner ne rapidité.

1.2 Différentiel

Contrairement aux détecteurs absolus, les détecteurs différentiels, aussi appelés *face trackers*, déterminent la position d'un visage sur une image grâce à sa position 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 petite erreurs sur les positions peut mener à la "dérive" du détecteur puisqu'il ne se remet jamais à 0 une fois lancé.

L'un des plus connus est l'algorithme **Active Appearance Model** (AAM).

Dans AAM, un visage est représenté comme un modèle en forme de maillage triangulaire composé d'environ 70 points. Ce modèle est construit grâce à un apprentissage sur différentes visages réalisant différentes expressions et dans lesquels les *features* sont connus. Il va donc chercher à faire correspondre ce modèle avec tout visage se trouvant sur une image en le déformant pour faire correspondre les features du modèle à celles du visage. Les déformations possibles ont été calculés au préalable.

Sur la première frame de la vidéo, il suffit d'initialiser les positions des features de chaque visage (manuellement ou via un tracker absolu) puis, sur les frames suivantes, la position de ces features est trackée grâce aux différentes déformations possibles.

AAM ne permet pas que de détecter des visages mais permet également d'en extraire les features.

1.3 Comparatif

Voici un comparatif avantages/inconvénients de ces 2 types de détecteurs.

Table 1 – Avantages/inconvénients des 2 types de détecteurs

	Avantages	Inconvénients
Absolu	Facilement parallélisable, très réactif en cas de changement soudain du nombre de visage et ne "dérive" pas	Plus lent et moins précis que les différentiels
Différentiel	Très rapide et d'une grande précision	Peut "dériver" si de petites erreurs s'accumulent au fur et à mesure

2 Extraction des features

2.1 Filtres de Gabor

2.2 Composantes pseudo-Haar

2.3 Local Binary Pattern (LBP)

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 je vais me concentrer.

Parmi ces classificateurs, nous pouvons les classer en 2 catégories : les binaires et les multi-classes.

3.1 Classificateurs binaires

Comme l'indique son nom, ce type de classificateur ne permet de classifier une instance qu'entre 2 classes (ex : sourire VS pas de sourire).

Le classificateur binaire le plus connu est le **Support Vector Machine (SVM)**.

Les SVMs se basent sur 2 notions très importantes.

La première est la notion de *marge maximale*. Cette marge est la distance entre la frontière de séparation des 2 classes et les échantillons les plus proches appelés vecteurs supports. La frontière de séparation est choisie de telle sorte qu'elle maximise le plus la marge. C'est donc grâce à un apprentissage que l'on peut définir cette frontière.

La deuxième notion est celle de la transformation de l'espace de représentation en un espace de dimension

plus grand, voir infini. cette notion est utile lorsque les données ne sont pas linéairement séparable dans l'espace d'origine. En projetant dans un espace de dimension plus grand, il est plus probable de trouver une séparation **Figure 1**.

Cette transformation est réalisé via une fonction noyau qui va permettre de transformer un produit scalaire dans un espace de grande dimension.

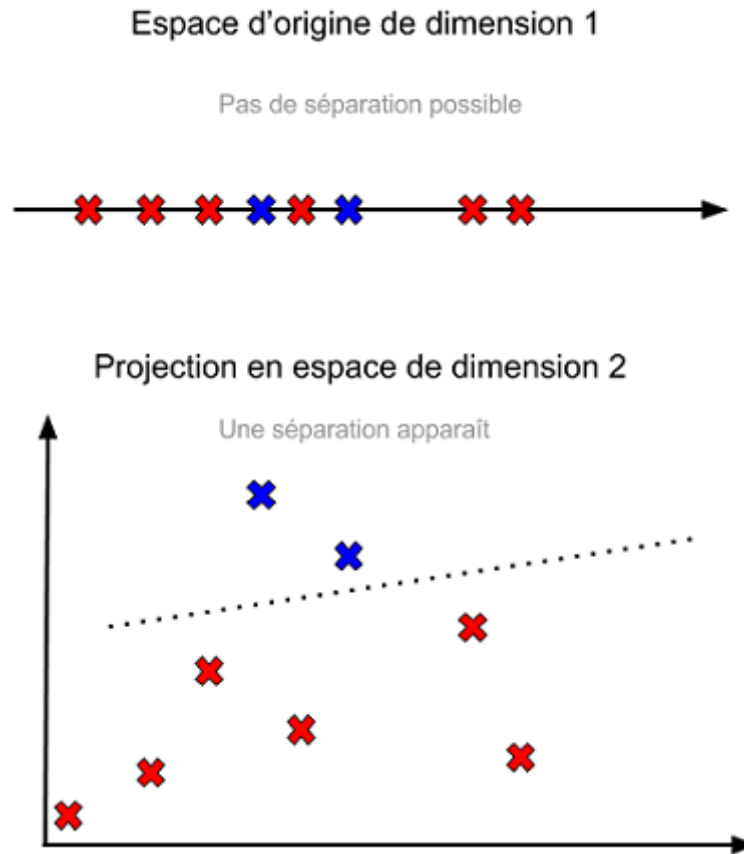


Figure 1 – Exemple de projection dans un espace plus grand réalisé par SVM

3.2 Classificateurs multi-classes

Contrairement aux binaires, ces classificateurs permettent de classer une instance selon un nombre de classes supérieur à 2.

Le classificateur multi-classes le plus connu est le **k-Plus Proche Voisin** (kPPV).

Le kPPV fonctionne d'une manière très simple.

Tout d'abord un apprentissage grâce à une base d'apprentissage composé de couples "entrée-sortie".

Puis la phase de classification se fait en cherchant la distance minimum entre un nouvel échantillon d'entrée dont la sortie doit être déterminée et les k échantillons d'apprentissage dont l'entrée est la plus proche **Figure 2**.

Une fois ces k plus proche voisins trouvés, il suffit de trouver quelle classe est majoritairement présente pour trouver la sortie associée et donc à quelle classe appartient le nouvel échantillon. Si jamais il y a un nombre égale d'échantillon de plusieurs classes (par exemple avec $k=3$ on obtient 1 échantillons de 3 différentes classes), on choisi alors aléatoirement la classe d'appartenance.

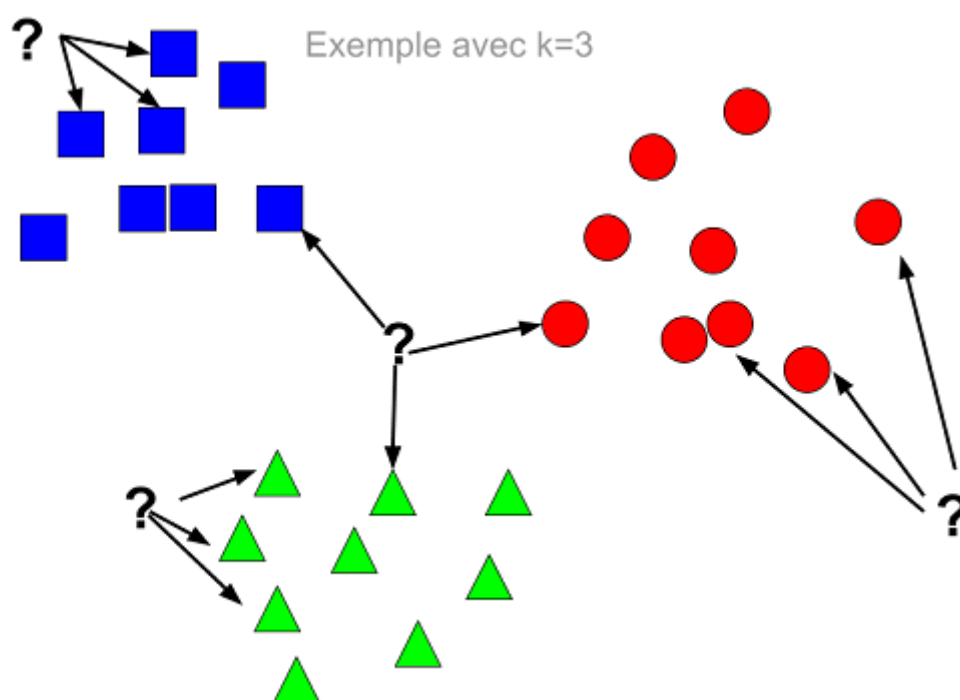


Figure 2 – Fonctionnement du kPPV avec $k=3$, 3 classes et 4 échantillons dont un indécis

5

Critiques de l'existant et nouvelle approche

6

Spécifications de l'application

7

Planning et outils utilisés

Deuxième partie

Développement



Conclusion

Annexes



Comptes rendus hebdomadaires

Compte rendu n°1 du 17/09/2015

Découverte de 2 grandes méthodes de description des mouvements du visage existent : le FACS (Facial Action Coding System) mis en place par P. Ekman et W. Friesen en 1978 et le FAPU (Facial Animation Parameter Units) introduit par la norme de codage vidéo MPEG-4.
Recherche sur les types d'acquisitions d'images en 2D ou en 3D avec le matériel nécessaires à chaque fois ainsi que les algorithmes disponibles.

Compte rendu n°2 du 24/09/2015

Recherche sur comment se décompose un bon système de reconnaissance facial d'émotions.
Décomposition en 4 parties (récupération du visage, normalisation, extraction des points clés, classification) et recherche plus poussée sur les 2 premières parties.

Compte rendu n°3 du 01/10/2015

Continuation des recherches sur les 2 dernières parties du système.
Recherche également sur les différentes bases de données 2D publics disponibles à l'utilisation.

Compte rendu n°4 du 08/10/2015

Recherche plus approfondies sur les filtres de Gabor. Présentation de mes recherches à Messieurs Conte et Slimane.
Commencement de l'écriture du rapport.

Compte rendu n°5 du 15/10/2015

Continuation des recherches sur les filtres de Gabor et leur fonctionnement. J'ai essayé de comprendre le fonctionnement des filtres et l'impact des différents paramètres. Grâce à un simulateur que j'ai trouvé en ligne ([[WWW4](#)]) et aux instructions associées ([[WWW3](#)]), j'ai pu constater l'effet qu'ont les différents paramètres sur le résultat final.

Compte rendu n°6 du 22/10/2015

Étude approfondi des bases de données MMI Mimicry et HCI Tagging.

Compte rendu n°7 du 05/11/2015

Documentation sur les caractéristiques pseudo-Haar et leur fonctionnement.
Recherche de techniques permettant de placer les points clés d'un visage sans FACS (ASM, AAM ...)

Compte rendu n°8 du 12/11/2015

Réunion avec Mrs Conte et Slimane : décision de l'arrêt de la phase état de l'art pour commencer le développement ; prise de décision sur les spécifications de notre système.

Compte rendu n°9 du 19/11/2015

Documentation plus poussée sur ASM, récupération d'un programme Matlab d'analyse d'émotions réalisé par des collègues italiens à Mr Conte et tentative de le faire fonctionner sous Octave vu que nous ne possédons pas de licence Matlab.
Commencement de la prise en main de la librairie C++ OpenCV mais suite à un entretien avec Mr Conte, la décision a été prise de changer les spécifications de notre programme pour continuer le travail qui a déjà été réalisés par ses collègues italiens.

Compte rendu n°10 du 26/11/2015

Étude approfondi de l'article écrit par Vitale et al. ([20]) et rendez vous avec Mr Conte pour faire fonctionner le programme Matlab, presque fonctionnel au final.

Compte rendu n°11 du 03/12/2015

Fin de l'étude approfondi de l'article de Vitale et al.
Travail sur le programme Matlab pour le faire fonctionner à 100%.

Compte rendu n°12 du 10/12/2015

Rédaction du rapport.
Téléchargement de Matlab pour faire fonctionner le programme car il est impossible de le faire fonctionner avec Octave.
RDV avec Mr Conte pour vérifier la compréhension de l'article, des zones d'ombre persistent.

Compte rendu n°13 du 17/12/2015

Étude du programme Matlab fonctionnel et comparaison de l'article au programme pour trouver à quelle partie du programme correspond chaque partie de l'article dans le but de mieux le comprendre.

Compte rendu n°14 du 07/01/2016**Compte rendu n°15 du 14/01/2016****Compte rendu n°16 du 21/01/2016**

Webographie

- [WWW1] AFFECTIVA. *Demonstration de la reconnaissance d’emotion via la webcam par la societe Affectiva*. URL : <https://labs.affectiva.com/superbowl/affdexweb.html> (visité le 03/12/2015).

ANNOTATION: Exemple de reconnaissance d’emotion via la webcam de notre ordinateur realise par la societe Affectiva

- [WWW2] EMOTIENT. *Demonstration de la reconnaissance d’emotion via la webcam par la societe Emotient*. URL : <http://emotient.com/livecam-demo/> (visité le 03/12/2015).

ANNOTATION: Exemple de reconnaissance d’emotion via la webcam de notre ordinateur realise par la societe Affectiva

- [WWW3] N.PETKOV. *Instructions pour le simulateur de filtres de Gabor*. Sous la dir. d’University of GRONINGEN. URL : http://matlabserver.cs.rug.nl/edgedetectionweb/web/edgedetection_params.html.

ANNOTATION: Decrit chaque parametre du simulateur de filtres de Gabor et leurs impacts et explique comment l’utiliser au mieux

- [WWW4] N.PETKOV. *Simulateur de filtres de Gabor*. Sous la dir. d’University of GRONINGEN. URL : <http://matlabserver.cs.rug.nl/edgedetectionweb/web/index.html>.

ANNOTATION: Site permettant de simuler le comportement des filtres de Gabor sur une image de notre choix. Tous les parametres des filtres peuvent etre modifies.

- [WWW5] *Site officiel de SAL*. URL : <http://emotion-research.net/toolbox/toolboxdatabase> 2006-09-26.5667892524 (visité le 16/12/2015).

ANNOTATION: Site decrivant la base de donnees de visages SAL

- [WWW6] Visage TECHNOLOGIES. *MPEG-4 Face and Body Animation (MPEG-4 FBA) : An overview*. URL : <http://www.visagetechologies.com/uploads/2012/08/MPEG-4FBA0verview.pdf> (visité le 01/11/2015).

ANNOTATION: Description du systeme de FAPU introduit par la norme de codage MPEG-4

Bibliographie

- [1] S. BILAKHIA, S. PETRIDIS, A. NIJHOLT et M. PANTIC. « The MAHNOB Mimicry Database - a database of naturalistic human interactions ». In : *Pattern Recognition Letters*, vol. 66, pp. 52-61 (2015).
ANNOTATION: Article decrivant la base de donnees de visage MHI Mimicry
- [2] Timothy F. COOTES, Gareth J. EDWARDS et Christopher J. TAYLOR. « Active Appearance Models ». In : *IEEE transactions on pattern analysis and machine intelligence* (2001).
ANNOTATION: Article decrivant la methode AAM. Creation du modele grace a la base d'apprentissage puis matching du modele avec un autre visage via un processus iteratif puis extraction de la position des features du visage
- [3] D. COSKER, E. KRUMHUBER et A. HILTON. « A FACS valid 3D dynamic action unit database with applications to 3D dynamic morphable facial modeling ». In : *IEEE International Conference on Computer Vision (ICCV)* (2011).
ANNOTATION: Article decrivant la base de donnees de visage D3DFACS
- [4] Charles DARWIN. *L'Expression des émotions chez l'homme et les animaux*. 1872.
ANNOTATION: Ce livre de Charles DARWIN decrit les similitudes entre les hommes et les animaux en matiere d'émotions
- [5] Paul EKMAN. « Facial Action Coding System (FACS) : Manual ». In : *Palo Alto : Consulting Psychologists Press* (1978).
ANNOTATION: Description et manuel d'utilisation du systeme FACS cree par Paul EKMAN
- [6] Paul EKMAN et Wallace V. FRIESEN. « Constants across culture in the face and emotion. » In : *Journal of Personality and Social Psychology* (1971).
ANNOTATION: Article ecrit par Ekman apres son voyage en Papouasie-Nouvelle-Guinee et decrivant pour la premiere fois les 7 emotions universelles (neutre, joie, peur, colere, degout, tristesse et surprise)
- [7] Takeo KANADE, Jeffrey F. COHN et Yingli TIAN. « Comprehensive Database for Facial Expression Analysis ». In : *Proceedings of the Fourth IEEE International Conference on Automatic Face and Gesture Recognition* (2000).

ANNOTATION: Article decrivant la base de donnees de visage Cohn-Kanade (CK)

- [8] Jeroen LICHTENAUER et Mohammad SOLEYMANI. « MAHNOB-HCI-TAGGING DATA-BASE ». In : (2012).

ANNOTATION: Manuel d 'utilisation de la base de donnees de visage HCI Tagging

- [9] Jeroen LICHTENAUER, Michel VALSTAR, Xiaofan SUN, Anton NIJHOLT et Maja PANTIC. « MAHNOB HMI IBUG MIMICRY DATABASE (MHI-MIMICRY) ». In : (2011).

ANNOTATION: Manuel d 'utilisation de la base de donnees de visage MHI Mimicry

- [10] Patrick LUCEY, Jeffrey F. COHN, Takeo KANADE, Jason SARAGIH, Zara AMBADAR et Iain MATTHEWS. « The Extended Cohn-Kanade Dataset (CK+) : A complete dataset for action unit and emotion-specified expression ». In : (2010).

ANNOTATION: Article decrivant la 2eme version de la base de donnees de visage Cohn-Kanade (CK+)

- [11] Michael J. LYONS, Shigeru AKAMATSU, Miyuki KAMACHI et Jiro GYOBA. « Coding Facial Expressions with Gabor Wavelets ». In : *Proceedings, Third IEEE International Conference on Automatic Face and Gesture Recognition* (1998).

ANNOTATION: Article decrivant la base de donnees de visage JAFFE

- [12] A. MORENO et A. SANCHEZ. « Gavabdb : a 3D face database ». In : (2004).

ANNOTATION: Article decrivant la base de donnees de visage Gavabdb

- [13] Maja PANTIC, Michel VALSTAR, Ron RADEMAKER et Ludo MAAT. « Web-based database for facial expression analysis ». In : (2005).

ANNOTATION: Article decrivant la base de donnees de visage MMI

- [14] Georgia SANDBACH, Stefanos ZAFEIRIOU, Maja PANTIC et Lijun YIN. « Static and dynamic 3D facial expression recognition : A comprehensive survey ». In : *IMAGE AND VISION COMPUTING* (2012).

ANNOTATION: Etat de l'art sur les techniques et materiels utilises dans la reconnaissance d'expressions en 3D. Description des differentes techniques utilisees par differents materiels (ex : Kinect) pour recuperer des visages en 3D et les analyser ensuite

- [15] Arman SAVRAN, Nese ALYUZ, Hamdi DIBEKLIOGLU, Oya CELIKTUTAN, Berk GOKBERK, Bulent SANKUR et Lale AKARUN. « Bosphorus database for 3D face analysis ». In : *The First COST 2101 Workshop on Biometrics and Identity Management* (9 mai 2008).

ANNOTATION: Article decrivant la base de donnees de visage Bosphorus

- [16] Mohammad SOLEYMANI, Jeroen LICHTENAUER, Thierry PUN et Maja PANTIC. « A Multi-modal Database for Affect Recognition and Implicit Tagging ». In : *IEEE Transactions on Affective Computing*. 3 : pp. 42 - 55, Issue 1 (2012).

ANNOTATION: Article decrivant la base de donnees de visage HCI Tagging

- [17] Giota STRATOU, Abhijeet GHOSH, Paul DEBEVEC et Louis-Philippe MORENCY. « Effect of Illumination on Automatic Expression Recognition : A Novel 3D Relightable Facial Database ». In : *9th International Conference on Automatic Face and Gesture Recognition* (2011).

ANNOTATION: Article decrivant la base de donnees de visage ICT-3DRFE

- [18] Michel F. VALSTAR et Maja PANTIC. « Induced Disgust, Happiness and Surprise : an Addition to the MMI Facial Expression Database ». In : *Proceedings of Int'l Conf. Language Resources and Evaluation, Workshop on EMOTION* (2010).
 ANNOTATION: Article decrivant la base de donnees de visage MMI
- [19] Paul VIOLA et Mickael JONES. « Robust Real-Time Face Detection ». In : (2001).
 ANNOTATION: Cet article decrit l'algorithme cree par Viola et Jones. Ce fut l'un des premiers algorithmes a detecter efficacement en temps reel des visages dans une image et il est maintenant utilise dans un grand nombre de detecteurs.
- [20] Jonathan VITALE, Mary-Anne WILLIAMS, Benjamin JOHNSTON et Giuseppe BOCCIGNONE. « Affective facial expression processing via simulation : A probabilistic model ». In : *Biologically Inspired Cognitive Architectures* (2014).
 ANNOTATION: Article decrivant une nouvelle approche permettant d'analyser des emotions via une methode probabilistique
- [21] Jacob WHITEHILL, Marian Stewart BARTLETT et Javier R. MOVELLAN. « Automatic Facial Expression Recognition ». In : (2014).
 ANNOTATION: Article ecrit par les fondateurs de l'entreprise Emotient et faisant un etat de l'art des differentes parties d'un systeme de reconnaissances d'emotions ainsi que de plusieurs techniques et bases de donnees
- [22] Lijun YIN, Xiaochen CHEN, Yi SUN, Tony WORM et Michael REALE. « A High-Resolution 3D Dynamic Facial Expression Database ». In : *The 8th International Conference on Automatic Face and Gesture Recognition* (19 sept. 2008).
 ANNOTATION: Article decrivant la base de donnees de visage BU4DFE
- [23] Lijun YIN, Xiaozhou WEI, Yi SUN, Jun WANG et Matthew J. ROSATO. « A 3D Facial Expression Database For Facial Behavior Research ». In : *The 7th International Conference on Automatic Face and Gesture Recognition* (12 avr. 2006).
 ANNOTATION: Article decrivant la base de donnees de visage BU3DFE

Application d'aide à l'interaction homme/machine pour les personnes handicapées

Florian Tissier

Encadrement : Mohamed Slimane et Donatello Conte

Objectif

L'objectif de ce projet est de réaliser une application d'aide aux personnes handicapées.

Cette application, à l'aide d'une caméra, va repérer le visage d'une personne puis zoomer dessus pour ensuite analyser son émotion actuelle.

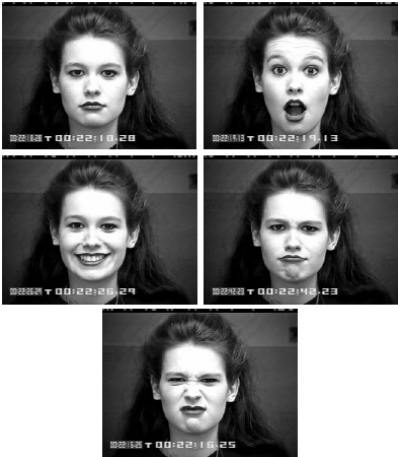
En fonction de l'émotion perçue, une action spécifique sera réalisé (ex: si la personne est triste, lui mettre de la musique joyeuse).



????

AU1 Inner brow raiser	AU2 Outer brow raiser	AU4 Brow Lowerer	AU5 Upper lid raiser	AU6 Cheek raiser
AU7 Lid tighten	AU9 Nose wrinkle	AU12 Lip corner puller	AU15 Lip corner depressor	AU17 Chin raiser
AU23 Lip tighten	AU24 Lip presser	AU25 Lips part	AU27 Mouth stretch	

????



Application d'aide à l'interaction homme/machine pour les personnes handicapées

Résumé

Résumé

Mots-clés

mot, clé, deux mots

Abstract

Abstract

Keywords

word, key, two words, fourth word

Tuteurs académiques

Mohamed SLIMANE

Donatello CONTE

Étudiants

Florian TISSIER (DI5)