

N° d'ordre :

N° de série :

REPUBLIQUE ALGERIENNE DEMOCRATIQUE ET POPULAIRE
Ministère de l'Enseignement Supérieur et de la Recherche Scientifique



UNIVERSITÉ ECHAHID HAMMA LAKHDAR EL OUED

FACULTE DES SCIENCES EXACTES
Département de l'informatique



Mémoire de fin d'étude

Présenté pour l'obtention du diplôme de

MASTER ACADEMIQUE

Domaine: Mathématiques et Informatique

Filière: Informatique

Spécialité: Systèmes Distribués et Intelligence Artificielle

Thème

***Un miroir intelligent interactif
basé sur Raspberry Pi***

Présenté par:

Daha Boubaker

Medileh Mounir

Soutenue le : 24-09-2018, Devant le jury :

- Mlle. Guettas Chourouk
- M. Othmani Samir
- M. Nedioui Abdelhamid

Rapporteur
Président
Examineur

Année universitaire 2017 - 2018

“The whole purpose of education is to turn mirrors into windows.”

—*Sydney J. Harris*

Remercîment

*En premier lieu, nous remercions « Allah »
de nous avoir donné la patience, la santé et le courage
pour arriver jusqu'à là.*

*Nous remercions
Miss Guettas Chourouk pour son encadrement, sa disponibilité,
son suivi, ses conseils et ses critiques constructives.*

*Nous remercions aussi
Tous les professeurs dans le département de l'informatique
pour leur soutien durant les années des études universitaires*

Dédicaces

Je dédie ce travail à l'esprit de mon père pur, à la créature la plus précieuse de ma mère, aucune dédicace ne saurait exprimer l'amour, l'estime, le dévouement et le respect que j'ai toujours eu

pour vous, à ma chère femme Nora,

à mes filles Maroua, Ikram, Salma

et à mes enfants Med el Bachir et Ahmed,

A mon binôme Mounir (YAGA)

une spéciale dédicace à mon encadrante

M^{lle} Guettas Chourouk, à Fassemi Othman,

professeur de la langue française

qui nous aider à notre travail,

à mes collègues Aissa Trad, Abdel Fettah Mansour,

et mes collègue de travail

Tayeb, Zeghedi, Hamid, Bachir, Kamal, Messaoud et Laid.

A tous les professeurs dans le département de l'informatique.

YAGA B

Dedications

To my Father and M♡T♡H♡E♡R

To my Brothers and Sisters

To my beloved Wife and daughters AYA & ISRA

To All my friends

Specially

*Johan, Kaizo, Kensei, Sniper, Jacob, Dok, Zaki
HES, YouniRex, Aissa T., Doha B, Saihi Bachir*

Tarek H. and Amel B.

Sergio, Stacy & Carina

To my coworkers for their support

Especially Mr. Mehda & Mr. Beggar

Yaga MNR

Résumé

Le miroir intelligent a devenu un aspect très important dans notre vie actuelle soit comme une interface centralisée de différents aménagements tels que les restaurants, les hôtels, les bâtiments intelligents, les salons de coiffure...etc, ou bien pour l'utilisation personnelle. Notre travail est une tentative de réaliser un miroir basé sur le Raspberry Pi pour afficher des informations générales (l'heure, météo, ..) et qui est capable de détecter l'humeur de son utilisateur et d'interagir selon son expression faciale. Le miroir utilise un classificateur d'expression faciale basé sur l'apprentissage au profond (Deep learning) et qui applique un algorithme de réseaux de neurones convolutif (LeNet).

Mots clés:

Expression faciale, émotion, Deep Learning, Réseaux de neurone convolutif, Raspberry Pi.

Abstract

The smart mirror has become a very important aspect in our present life either as a centralized interface of different facilities such as restaurants, hotels, smart buildings, hairdressing salons... etc, or for personal use. Our work is an attempt to make a mirror based on the Raspberry Pi to display general information (time, weather, ...) and able to detect the mood of its user and interact according to his facial expression. The mirror uses a facial expression classifier based on deep learning and uses a convolutional neural network algorithm (LeNet).

Keywords:

Facial Expression, Emotion, Deep Learning, Convolutional Neural Networks, Raspberry Pi.

ملخص

أصبحت للمرأة الذكية جانب مهم جدا في حياتنا الحالية، إما كواجهة مركزية لمرافق مختلفة مثل المطاعم والفنادق والمباني الذكية وصالونات تصفيف الشعر ... إلخ أو للاستخدام الشخصي. قمنا بمحاولة لإنجاز مرآة ذكية تشتغل على جهاز الراز بيري باي (Raspberry Pi) تقوم بعرض معلومات عامة (الوقت، الطقس، ...) وفي نفس الوقت تكون قادرة على معرفة مزاج مستخدميها والتفاعل معهم وفقا لخاصية التعرف على تعابير الوجه. تستخدم هذه المرآة الذكية مصنفا لتعابير الوجه، والذي يعتمد على التعلم العميق ويستخدم خوارزمية شبكة عصبية إلتفافية (LeNet).

الكلمات المفتاحية:

التعرف على تعابير الوجه ، العاطفة ، التعلم العميق ، الشبكات العصبية الإلتفافية ، الراز بيري باي

Table des matières

<i>Remercîment</i>	<i>I</i>
<i>Dédicaces</i>	<i>II</i>
<i>Résumé</i>	<i>IV</i>
<i>Liste des figures</i>	<i>IX</i>
<i>Liste des Tableaux</i>	<i>X</i>
<i>Liste des abréviations</i>	<i>XI</i>
<i>Introduction générale</i>	<i>2</i>
<i>Chapitre I : Fondements</i>	
1. INTRODUCTION	5
2. L'INTERACTION HOMME-MACHINE	5
3. RECONNAISSANCE DES VISAGES	5
3.1. PRINCIPALES DIFFICULTES DE LA RECONNAISSANCE DE VISAGE	6
3.1.1. Facteur de changement d'illumination	6
3.1.2. Facteur de variation de pose	7
3.1.3. Facteur d'expressions faciales.....	7
3.1.4. Facteur de présence ou absence des composants structurels.....	8
3.1.5. Facteur d'occultations partielles	8
4. L'EMOTION ET L'EXPRESSION FACIALE:	8
4.1. QU'EST QU'UNE EXPRESSION FACIALE?	9
4.2. POURQUOI LES EXPRESSIONS FACIALES EXISTENT- ELLES?	9
5. LES RESEAUX DE NEURONES ARTIFICIELS	10
5.1. CONCEPT	11
5.2. LES RESEAUX DE NEURONES CONVOLUTIONNELS	11
5.3. APPRENTISSAGE.....	13
5.4. PERCEPTRON MULTICOUCHE.....	14
6. LE DEEP LEARNING	16
6.1. CONCEPTS ET TECHNIQUES:	16
6.2. LES CHAMPS D'APPLICATION	18
7. CONCLUSION	20
<i>Chapitre II : Travaux Connexes</i>	
1. INTRODUCTION	22
2. EXPRESSIONS FACIALES: ETAT DE L'ART.....	22
2.1. LA DETECTION DE VISAGE.....	22
2.1.1. Approche basée sur les connaissances acquises	23
2.1.2. Approche basée sur "le Template matching " ou le modèle de correspondance	23
2.1.3. Approche basé sur l'apparence	24
2.1.4. Approche basée sur des caractéristiques invariantes.....	24
2.1.5. Détection des Objets par la méthode de Viola et Jones.....	24
2.1.6. L'algorithme de Viola et Michael Jones (Classificateur Haar-cascade)	24
2.2. LES METHODES DE RECONNAISSANCE FACIALES	26
2.2.1. Méthodes globales	26
2.2.2. Méthodes locales	27
2.2.3. Méthodes hybrides.....	28
2.3. LE DEVELOPPEMENT EMOTIONNEL.....	29
2.3.1. Méthode subjective	29
2.3.2. Méthode objective :	29
2.3.3. Neuraux imagerie fonctionnelle	30
3. SMART MIRRORS: ETAT DE L'ART	30
3.1. PANASONIC SMART MIRRORS.....	30

3.2. THE WIZE MIRROR	31
3.3. GOOGLE ANDROID SMART MIRROR	31
3.4. MICROSOFT IoT MAGIC MIRROR.....	32
3.5. PERSEUS, LE PREMIER MIROIR INTELLIGENT AU MONDE.....	32
4. CONCLUSION	32
<i>Chapitre III : Conception</i>	
1. INTRODUCTION	34
2. METHODOLOGIE DU TRAVAIL	34
3. CLASSIFICATION EN UTILISANT L'APPRENTISSAGE AUTOMATIQUE TRADITIONNEL PAR RAPPORT A L'APPRENTISSAGE EN PROFONDEUR	35
4. CONSTRUIRE UN CLASSIFICATEUR DES EXPRESSIONS FACIALES EN UTILISANT UN RESEAU DE NEURONES CONVOLUTIF	37
4.1. LeNET:.....	37
5. LES PHASES DE L'APPRENTISSAGE ET DE LA PREDICTION.....	37
5.1. LA PHASE DE L'APPRENTISSAGE	37
5.2. LA PHASE DE LA PREDICTION	38
6. ARCHITECTURE DE MIROIR INTELLIGENT	39
7. CONCLUSION	42
<i>Chapitre IV: Implémentation et résultats</i>	
1. INTRODUCTION	44
2. ENVIRONNEMENT DE DEVELOPPEMENT	44
2.1. ENVIRONNEMENT HARDWARE.....	44
2.1.1. Configuration de la machine	44
2.1.2. Le Raspberry Pi 3.....	45
2.2. LES CAPTEURS ET LES MODULES	46
2.2.1. Module de détection aux ultrasons.....	46
2.2.1. Photorésistance	46
2.2.2. Convertisseur analogique-numérique (MCP3008).....	47
2.3. LE CABLAGE	47
2.4. ENVIRONNEMENT SOFTWARE.....	48
2.4.1. Open CV	48
2.4.2. TensorFlow.....	49
2.4.3. Keras	49
2.4.4. Python.....	49
3. L'APPRENTISSAGE DE CLASSIFICATEUR D'IMAGES	50
3.1. PREPARATION DES DONNEES POUR L'APPRENTISSAGE (L'ENTRAINEMENT)	50
3.2. L'APPRENTISSAGE	51
3.2.1. L'apprentissage de classificateur d'images de réseau neuronal convolutionnel avec Keras et Tensorflow	52
3.3. RESULTAT DE L'APPRENTISSAGE	53
4. ÉVALUATION DE NOTRE CLASSIFICATEUR D'IMAGES	55
4.1. LES TESTS	56
4.1.1. Test en direct avec la webcam.....	56
4.1.2. Test à partir des images enregistré sur disque	58
5. LE FONCTIONNEMENT DU MIROIR	60
6. COMMENT SE FAIT L'INTERACTION DU MIROIR?	61
7. CONCLUSION	62
<i>Conclusion générale</i>	64
<i>Perspectives</i>	66
<i>Bibliographie</i>	68

Liste des figures

FIGURE 1 EXEMPLE DE VARIATION D'ECLAIRAGE.....	6
FIGURE 2 EXEMPLES DE VARIATION DE POSES	7
FIGURE 3 - EXEMPLES DE VARIATION D'EXPRESSIONS	7
FIGURE 4 - MOUVEMENTS FACIAUX GLOBAUX	9
FIGURE 5 - RELATION ENTRE EXPRESSION FACIALE ET EMOTION	10
FIGURE 6 ARCHITECTURE STANDARD D'UN RESEAU DE NEURONE CONVOLUTIONNEL.....	11
FIGURE 7 - LA COURBE DE LA TANGENTE HYPERBOLIQUE ET LE SIGMOÏDE ET LE ReLU	15
FIGURE 8 - PERCEPTRON MULTICOUCHE	15
FIGURE 9 - EXEMPLE D'UN SYSTEME DE DEEP LEARNING	18
FIGURE 10 - PRINCIPE DE DEEPFACE (DE FACEBOOK)	19
FIGURE 11 - LE PRINCIPE DE FONCTIONNEMENT DE BASE D'UN SYSTEME DE RECONNAISSANCE FACIALE.....	22
FIGURE 12 - MODELE DE VISAGE COMPOSE DE 16 REGIONS	24
FIGURE 13 - UN EXEMPLE DES PREMIERES CARACTERISTIQUES PSEUDO-HAAR UTILISEES PAR VIOLA ET JONES EN 2001	25
FIGURE 14 - L'EXTENSION DU CARACTERISTIQUE PSEUDO HAAR PAR LIENHART ET MAYDT.....	25
FIGURE 15 - LE RESULTAT DE DETECTION DE VISAGE EN UTILISANT LE HAAR-CASCADE.....	26
FIGURE 16 - CLASSIFICATION DES ALGORITHMES PRINCIPAUX UTILISES EN RECONNAISSANCE FACIALE.....	29
FIGURE 17 - METHODES D'ANALYSE AUTOMATIQUE D'UN SYSTEME DE RECONNAISSANCE D'EXPRESSIONS FACIALES.....	30
FIGURE 18 – LES PHASES DE MACHINE LEARNING	35
FIGURE 19 – LA DIFFERENCE PRINCIPALE ENTRE L'APPRENTISSAGE AUTOMATIQUE TRADITIONNEL ET LES ALGORITHMES D'APPRENTISSAGE PROFOND	36
FIGURE 20 – ARCHITECTURE DE LeNet	37
FIGURE 21 – LES PHASES DE L'APPRENTISSAGE ET DE LA PREDICTION	38
FIGURE 22 – CONCEPTION DU MIROIR.....	39
FIGURE 23 - GRAPHE DE FLUX	41
FIGURE 24 - RASPBERRY Pi 3.....	45
FIGURE 25 - MODULE DE DETECTION AUX ULTRASON (HC-SR04).....	46
FIGURE 26 - PHOTORESISTANCE	46
FIGURE 27 – CONVERTISSEUR ANALOGIQUE-NUMERIQUE (MCP3008)	47
FIGURE 28 – LE CABLAGE ENTRE LES COMPOSANTS ET LE RASPBERRY Pi.....	47
FIGURE 29 – LES DOSSIERS DES IMAGES	50
FIGURE 30 – UN ENSEMBLE DES IMAGES REPRESENTENT L'EXPRESSION HEUREUSE (HAPPY)	51
FIGURE 31 – UN ENSEMBLE DES IMAGES REPRESENTENT L'EXPRESSION TRISTE (SAD)	51
FIGURE 32 – LE PROCESSUS DE L'APPRENTISSAGE	53
FIGURE 33 – LES GRAPHES DU RESULTATS	55
FIGURE 34 – L'ENSEMBLE DES IMAGES ENREGISTRE SUR LE DISQUE.....	58
FIGURE 35 – UNE VUE SUR LE MIROIR	60

Liste des Tableaux

TABLEAU 1 - RESULTATS DE L'APPRENTISSAGE	54
TABLEAU 2 - TESTS ET RESULTATS AVEC WEBCAM	57
TABLEAU 3 - RESULTATS DES TESTS SUR IMAGES ENREGISTRE SUR DISQUE	59

Liste des abréviations

AAM	Apparence actif model (Les modèles actifs d'apparence)
ADL	Analyse Discriminante Lineaire
AR	The AR Face Database, Aleix Martinez and Robert Benavente
ARM	Advanced RISC Machine
BIC	Binary Interaction Coefficients
CNN	Convlutional Neural Network
EBGM	ElasticBench graph matching
FER	Facial Expression Recognition
FERET	The Facial Recognition Technology
FRVT	Face Recognition Vendor Test
GPIO	General-Purpose Input/Output
GPU	Graphics Processing Units
HMM	Hidden Markov Model
ICA	Independent component analysis (analyse en composantes indépendantes).
K-LDA	Kernel Lineaire Discriminant Analysis
K-PCA	Kernel Principal Component Analysis
LBP	Local binary pattern
LFA	Local Feature analysis .
LG-PCA	l'algorithme Log Gabor PCA
LLE	Locally Linear Embedding Alogorithm.
PCA	Principal Component Analysis (analyse en composantes principales).
PMC	Perceptron MultiCouche
SPI	Serial Peripheral Interface
SVM	Support Vector Machin (les machines à vecteurs de support)

Introduction générale

Introduction général

Grâce à la technologie, l'utilisation de l'intelligence artificielle se répand dans tous les aspects de notre vie, cela donne la naissance de nombreuses disciplines scientifiques, comme l'IoT ou la vision par machine qui combine à la fois l'ordinateur, la caméra et l'internet.

La reconnaissance des visages désigne l'imitation de certaines capacités humaines en utilisant des méthodes et des techniques développées par exemple les miroirs intelligents qui peuvent percevoir l'environnement et commander par la suite.

La reconnaissance de visages est une tâche importante effectuée par les personnes naturellement sans fatigue dans la vie quotidienne, de cette découverte l'entrée d'ordinateurs avec des applications élevées a conduit à l'attention dans le traitement numérique des images et des vidéos dans de nombreux domaines, y compris la reconnaissance des visages est devenue une préoccupation majeure dans la plupart des recherches, résultant en des progrès rapides dans la technologie telles que les appareils photo numériques et les dispositifs mobiles.

On entend souvent ce dicton: « un visage de poker » en anglais « Poker face », qui est utilisé dans les jeux de hasard, où les joueurs essaient de maintenir une expression faciale normale pour empêcher les autres de savoir comment ils vont jouer pour garantir une victoire. Le visage humain peut révéler beaucoup d'informations, par exemple, un médecin peut diagnostiquer un patient juste en regardant son visage, un psychologue peut faire un rapport de diagnostic aussi, et un policier peut juger quelqu'un de l'apparence de son visage. Par conséquent, en regardant simplement la face de quelqu'un, nous pourrions savoir beaucoup de choses sur lui, s'il est heureux ou en colère ou malade, ou s'il est digne de confiance ou non, s'il dit la vérité ou ment ... etc. Il y a tellement de systèmes et d'idées sur l'utilisation de ces connaissances pour faire un système d'intelligence artificielle pour exploiter cette information.

L'objectif de notre mémoire est de réaliser un miroir intelligent en utilisant le Raspberry Pi qui fait l'interaction avec l'utilisateur selon son expression faciale.

Dans le premier chapitre on va expliquer quelques notions de bases liées à notre projet ; l'interaction homme machine, la reconnaissance de visage et ses grandes difficultés. Également on va aborder l'émotion et les expressions faciales en expliquant les principales notions de réseaux neurones artificiels, enfin on présentera le Deep Learning ou l'apprentissage profond, ses procédés, ses techniques, et ses domaines d'application.

Dans le deuxième chapitre on va citer les approches proposées dans la littérature pour la reconnaissance automatique d'expression faciale, l'état de l'art des algorithmes de reconnaissance faciale, et l'état de l'art des travaux liés au domaine de miroir intelligent.

Dans le troisième chapitre on présentera la méthodologie du travail, et la construction du classificateur des expressions faciales en utilisant le Deep learning avec une architecture CNN simple "LeNet", et la conception et l'architecture de miroir intelligent.

Dans le dernier chapitre, on va implémenter le classificateur des expressions faciales et faire les tests et l'analyse des résultats obtenus.

Finalement, une conclusion générale vient pour clôturer ce travail ainsi pour donner une synthèse de tout ce qui a été fait ou acquis, et en ouvrant une parenthèse de perspectives sur ce qu'on projette de faire dans le futur, que ce soit le proche avenir ou à long terme.

Chapitre I

Fondements

1. Introduction

Un miroir intelligent capable de lire vos émotions, le miroir connecté affiche des informations en fonction de votre humeur. Une idée à réaliser afin de faciliter l'intégration de la technologie à notre vie quotidienne.

Dans ce chapitre, on va expliquer quelques notions de bases liées à notre projet, l'interaction homme machine, la reconnaissance de visage et ses grandes difficultés. Également on va aborder l'émotion et les expressions faciales en expliquant les principales notions de réseaux neurones artificiels, enfin on présentera le Deep Learning ou l'apprentissage profond, ses procédés, ses techniques, et ses domaines d'application.

2. L'interaction homme-machine

Plus les recherches basées sur l'usage du triplet: écran-clavier-souris- évoluées, plus l'interaction homme-machine évolue autant au point qu'elle envisage des dispositifs d'entrée/sortie ce qui tend vers l'abolition des frontières séparant le monde physique du monde électronique.

Les nouvelles formes d'interaction dépendent principalement sur la capture du comportement observable de l'utilisateur et de son environnement, dans ce cas-là, on fait appel à la perception artificielle et la vision par ordinateur.

Mais cela ne s'arrête pas là, car il s'agit d'une discipline rapidement évoluée, en effet dans le futur, ces machines deviendront plus naturelle et intuitive possible en y intégrant des nouvelles informations concernant la parole, l'expression faciale et le comportement dynamique.

3. Reconnaissance des visages

Quand un humain regarde à une personne, il peut dire s'il le reconnaît ou non, et il est facile pour lui de détecter un visage humain à partir d'une image ou de téléviseurs ou de tout autre endroit. Cependant, pour une machine ce n'est pas si simple, à cause de la différence de perception entre l'humain et la machine.

De nos jours, les systèmes de reconnaissance faciale sont très utilisés, dans les médias sociaux, les caméras, les systèmes de sécurité ... etc. de manière à identifier les individus, en

utilisant des applications informatiques qui utilisent des images provenant de cadres vidéo ou des images d'individus pour détecter le visage et le comparer à un ensemble d'images dans des bases de données.

3.1. Principales difficultés de la reconnaissance de visage

Selon [OUA12] Le cerveau humain a besoin des facteurs nécessaires pour reconnaître le visage, ce qui est une tâche à haute. Bien que les humains puissent détecter et identifier les visages dans une scène sans trop d'effort, construire un système automatique qui accomplit de telles tâches représente un sérieux défi. Ce défi est d'autant plus grand lorsque les conditions d'acquisition des images sont très variables. Il existe deux types de variations associées aux images de visages: inter et intra sujet. La variation inter sujet est limitée à cause de la ressemblance physique entre les individus. Par contre la variation intra sujet est plus vaste. Elle peut être attribuée à plusieurs facteurs que nous analysons ci-dessous.

3.1.1. Facteur de changement d'illumination

L'apparence d'un visage dans une image varie énormément en fonction de l'illumination de la scène lors de la prise de vue (voir figure 1). Les variations d'éclairage rendent la tâche de reconnaissance de visage très difficile. En effet, le changement d'apparence d'un visage du à l'illumination, se révèle parfois plus critique que la différence physique entre les individus, et peut entraîner une mauvaise classification des images d'entrée. Ceci a été expérimentalement observé dans Adini et Al [ADI97] où les auteurs ont utilisé une base de données de 25 individus. L'identification de visage dans un environnement non contrôlé reste donc un domaine de recherche ouvert. Les évaluations **FRVT** (Face Recognition Vendor Test) ont révélé que le problème de variation d'illumination constitue un défi majeur pour la reconnaissance faciale.



Figure 1 Exemple de variation d'éclairage.
[OUA12]

3.1.2. Facteur de variation de pose

Le taux de reconnaissance de visage baisse considérablement quand des variations de pose sont présentes dans les images. Cette difficulté a été démontrée par des tests d'évaluation élaborés sur les bases FERET (Face Recognition Technology) et FRVT [PHI03]. La variation de pose est considérée comme un problème majeur pour les systèmes de reconnaissance faciale. Quand le visage est de profil dans le plan image (orientation $< 30^\circ$), il peut être normalisé en détectant au moins deux traits faciaux (passant par les yeux). Cependant, lorsque la rotation est supérieure à 30° , la normalisation géométrique n'est plus possible (voir figure 2).



Figure 2 Exemples de variation de poses
[OUA12]

3.1.3. Facteur d'expressions faciales

Un autre facteur qui affecte l'apparence du visage est l'expression faciale (voir figure 3). La déformation du visage qui est due aux expressions faciales est localisée principalement sur la partie inférieure du visage. L'information faciale se situant dans la partie supérieure du visage reste quasi invariable. Elle est généralement suffisante pour effectuer une identification. Toutefois, étant donné que l'expression faciale modifie l'aspect du visage, elle entraîne forcément une diminution du taux de reconnaissance. L'identification de visage avec expression faciale est un problème difficile qui est toujours d'actualité et qui reste non résolu. L'information temporelle fournit une connaissance additionnelle significative qui peut être utilisée pour résoudre ce problème [GRE01].



Figure 3 - Exemples de variation d'expressions
[OUA12]

3.1.4. Facteur de présence ou absence des composants structurels

La présence des composants structurels telle que la barbe, la moustache, ou bien les lunettes peut modifier énormément les caractéristiques faciales telles que la forme, la couleur, ou la taille du visage. De plus, ces composants peuvent cacher les caractéristiques faciales de base causant ainsi une défaillance du système de reconnaissance. Par exemple, des lunettes opaques ne permettent pas de bien distinguer la forme et la couleur des yeux, et une moustache ou une barbe modifie la forme du visage.

3.1.5. Facteur d'occultations partielles

Selon [OUA12], le visage peut être partiellement masqué par des objets dans la scène, ou par le port d'accessoire tels que lunettes, écharpe... Dans le contexte de la biométrie, les systèmes proposés doivent être non intrusifs c'est-à-dire qu'on ne doit pas compter sur une coopération active du sujet. Par conséquent, il est important de savoir reconnaître des visages partiellement occultés. Gross et Al [GRE01] ont étudié l'impact du port de lunettes de soleil, et du cache-nez occultant la partie inférieure du visage sur la reconnaissance faciale. Ils ont utilisé la base de données **AR** (The AR Face Database, Aleix Martinez and Robert Benavente) [MAR98]. Leurs résultats expérimentaux semblent indiquer que, dans ces conditions, les performances des algorithmes de reconnaissance restent faibles.

4. L'émotion et l'expression faciale:

Les expressions et les émotions sont très liées, en effet l'émotion s'éprouve via la position du corps, la voix et les expressions faciales ce qui fait qu'une telle émotion nécessite une expression faciale correspondante mais le contraire n'est pas juste.

L'expression faciale est une mimique faciale chargée de sens qui peut être l'expression d'une émotion, un indice sémantique ou une intonation dans la langue des signes.

L'interprétation de cette mimique en expression dépend du contexte d'application. Dans le cas d'une application en Interaction Homme – Machine (IHM), on tend à classer les mesures en termes d'émotions en prenant en considération qu'un visage : est l'élément primordial dans les recherches actuelles concernant le perfectionnement de l'IHM.

Les expressions faciales participent à la coordination de la conversation humaine, chose prouvée par des recherches en psychologie, Mehrabian [MEH96] montre que dans un message

verbal en face à face, le contenu textuel contribue en 07% à l'aboutissement de ce message à l'auditeur, les signaux conversationnels en 38% et 55 % pour l'expression faciale.

4.1. Qu'est qu'une expression faciale?

Les expressions faciales ont un rôle significatif dans la conversation humaine ce qui donne la naissance de la reconnaissance des expressions faciales, depuis une vingtaine d'années et qui se développe et développera au futur au point qu'elle on se contente pas de la souris, du clavier, mais on aura besoin d'interpréter les expressions faciales, pour étudier la reconnaissance des émotions faciales, plusieurs paradigmes expérimentaux ont été développés. Le plus commun est celui créé par Ekman en 1976 [EKM76] en utilisant un modèle catégorique d'émotions divisé en six émotions principales: la tristesse, le dégoût, la joie, la peur, la surprise et la colère (voir figure 4) à laquelle est ajouté le visage neutre. Ce modèle est le plus utilisé dans la littérature pour évaluer la reconnaissance des émotions faciales grâce à des batteries d'images [EKM76] [MAT88].



Figure 4 - Mouvements faciaux globaux
[OUA12]

4.2. Pourquoi les expressions faciales existent- elles?

Il y'a trois réponses à cette question :

1. Les expressions faciales sont au centre de l'explication des émotions.
2. Les émotions faciales sont des signaux de conservation qui réclament des intentions de l'émetteur, du comportement de récepteur et du contexte de l'interaction.
3. Les expressions faciales sont des régulateurs d'émotions dues à la plasticité du cerveau, une émotion peut causer une expression faciale ou non et une expression faciale peut être causée par une émotion ou non (voir figure 5).

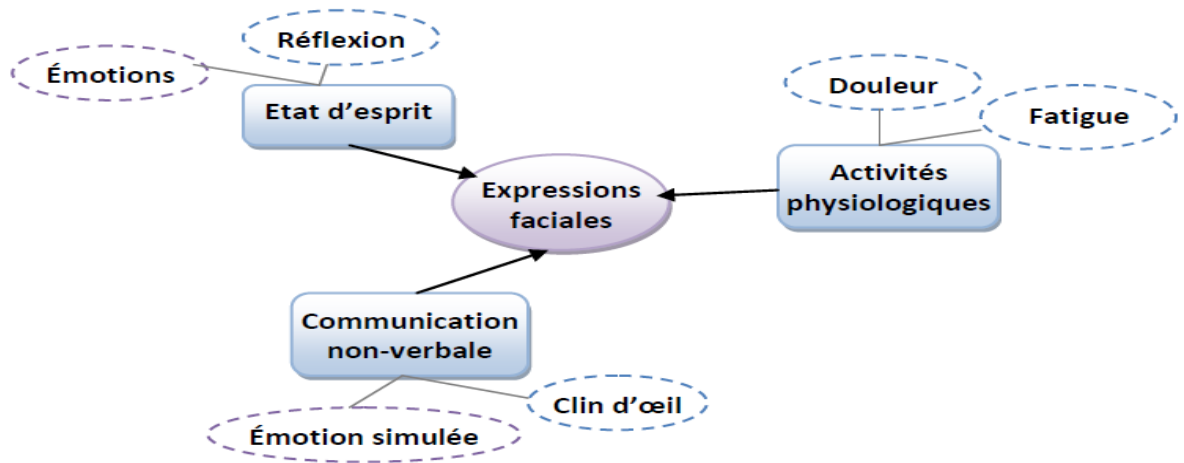


Figure 5 - Relation entre expression faciale et émotion
[OUA12]

5. Les Réseaux de neurones artificiels

Les ordinateurs ont des unités de traitement et de mémoire. Ils permettent à l'ordinateur d'effectuer le calcul numérique le plus complexe durant un temps très court.

Si on compare l'ordinateur et le cerveau, on notera que, théoriquement, l'ordinateur devrait être plus puissant que notre cerveau: Il comprend 10^9 transistors avec un temps de commutation de 10^{-9} secondes. Le cerveau contient 10^{11} neurones, mais ceux-ci ont seulement un temps de commutation d'environ 10^{-3} secondes.

La plus grande partie du cerveau travaille en permanence, tandis que la plus grande partie de l'ordinateur est réservée seulement pour le stockage de données passives. En revanche Un ordinateur est statique, le cerveau comme un réseau neuronal biologique peut se réorganiser au cours de sa durée de vie, et est donc en mesure d'apprendre, pour compenser les erreurs et ainsi de suite [KRI07].

En fait la caractéristique que le cerveau humain peut se réorganiser, est la clé derrière sa capacité d'apprentissage qu'on ne trouve pas dans les ordinateurs. Le but de l'utilisation des réseaux de neurones est de s'inspirer de la biologie pour établir un système artificiel capable d'apprendre et de compenser ses erreurs.

5.1. Concept

Un réseau de neurones artificiels, où réseau neuronal artificiel, est un modèle de calcul dont la conception est très schématiquement inspirée du fonctionnement des neurones biologiques [WIK18a]. Ces réseaux ont été développés pour résoudre des problèmes:

- de contrôle.
- de reconnaissance de formes.
- de décision.
- de mémorisation.

5.2. Les réseaux de neurones convolutionnels

Selon [MOK17], les réseaux de neurones convolutionnels sont à ce jour les modèles les plus performants pour classer des images. Désignés par l'acronyme CNN, de l'anglais Convolutional Neural Network, ils comportent deux parties bien distinctes. En entrée, une image est fournie sous la forme d'une matrice de pixels. Elle a 2 dimensions pour une image en niveaux de gris. La couleur est représentée par une troisième dimension, de profondeur 3 pour représenter les couleurs fondamentales [Rouge, Vert, Bleu]. La première partie d'un CNN est la partie convolutive à proprement parler. Elle fonctionne comme un extracteur de caractéristiques des images. Une image est passée à travers une succession de filtres, ou noyaux de convolution, créant de nouvelles images appelées cartes de convolutions. Certains filtres intermédiaires réduisent la résolution de l'image par une opération de maximum local. Au final, les cartes de convolutions sont mises à plat et concaténées en un vecteur de caractéristiques, appelé code CNN (voir figure 6).

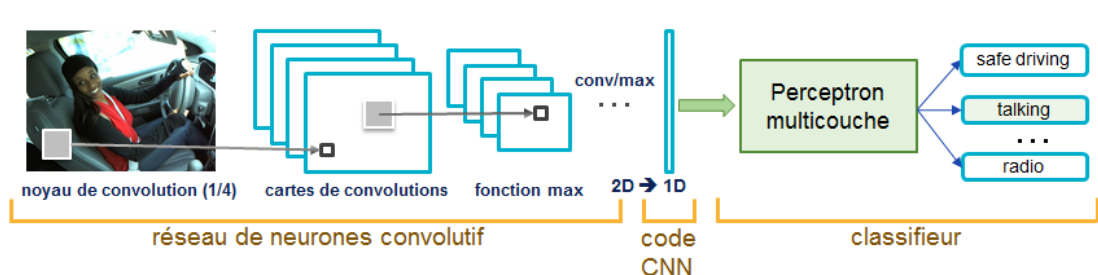


Figure 6 Architecture standard d'un réseau de neurone convolutionnel [MOK17]

Ce code CNN en sortie de la partie convolutive est ensuite branché en entrée d'une deuxième partie, constituée de couches entièrement connectées (perceptron multicouche). Le rôle de cette partie est de combiner les caractéristiques du code CNN pour classer l'image. La sortie est une dernière couche comportant un neurone par catégorie. Les valeurs numériques obtenues sont généralement normalisées entre 0 et 1, de somme 1, pour produire une distribution de probabilité sur les catégories.

Architecture de réseaux de neurone convolutionnel

Selon [MOK17], les réseaux de neurones convolutionnels sont basés sur le perceptron multicouche (MLP), et inspirés du comportement du cortex visuel des vertébrés. Bien qu'efficaces pour le traitement d'images, les MLP ont beaucoup de mal à gérer des images de grande taille, ce qui est dû à la croissance exponentielle du nombre de connexions avec la taille de l'image. Par exemple, si on prend une image de taille $32 \times 32 \times 3$ (32 de large, 32 de haut, 3 canaux de couleur), un seul neurone entièrement connecté dans la première couche cachée du MLP aurait 3072 entrées ($32 \times 32 \times 3$). Une image 200×200 conduirait ainsi à traiter 120 000 entrées par neurone ce qui, multiplié par le nombre de neurones, devient énorme.

Les CNN visent à limiter le nombre d'entrées tout en conservant la forte corrélation « spatialement locale » des images naturelles. Par opposition aux MLP, les CNN ont les traits distinctifs suivants :

1. 'Volumes 3D de neurones': La couche de neurones n'est plus simplement une surface (perceptron), mais devient un volume avec une profondeur. Si on considère un seul champ récepteur du CNN, les n neurones associés (sur la profondeur) forment l'équivalent de la première couche d'un MLP.

2. 'Connectivité locale': Grâce au champ récepteur qui limite le nombre d'entrées du neurone, tout en conservant l'architecture MLP, les CNN assurent ainsi que les «filtres» produisent la réponse la plus forte à un motif d'entrée spatialement localisé, ce qui conduit à une représentation parcimonieuse de l'entrée. Une telle représentation occupe moins d'espace en mémoire. De plus, le nombre de paramètres à estimer étant réduit, leur estimation (statistique) est plus robuste pour un volume de données fixé (comparé à un MLP).

3. 'Poids partagés' : Dans les CNNs, les paramètres de filtrage d'un neurone (pour un champ récepteur donné) sont identiques pour tous les autres neurones d'un même noyau (traitant tous les

autres champs récepteurs de l'image). Ce paramétrage (vecteur de poids et biais) est défini dans une « carte de fonction ». Cela signifie que tous les neurones dans une couche de convolution donnée détectent exactement la même caractéristique. En multipliant les champs récepteurs, il devient possible de détecter des éléments indépendamment de leur position dans le champ visuel, ce qui induit une propriété d'invariance par translation.

Ensemble, ces propriétés permettent aux réseaux de neurones convolutionnels d'obtenir une meilleure généralisation (en termes d'apprentissage) sur des problèmes de vision. Le partage de poids permet aussi de réduire considérablement le nombre de paramètres libres à apprendre, et ainsi les besoins en mémoire pour le fonctionnement du réseau. La diminution de l'empreinte mémoire permet l'apprentissage de réseaux plus grands donc souvent plus puissants.

Une architecture CNN est formée par un empilement de couches de traitement indépendantes:

1. La couche de convolution (CONV) qui traite les données d'un champ récepteur.
2. La couche de pooling (POOL), qui permet de compresser l'information en réduisant la taille de l'image intermédiaire (souvent par sous-échantillonnage).
3. La couche de correction (ReLU), souvent appelée par abus 'ReLU' en référence à la fonction d'activation (Unité de rectification linéaire).
4. La couche "entièrement connectée" (FC), qui est une couche de type perceptron.
5. La couche de perte (LOSS).

5.3. Apprentissage

Selon [TOU92] L'apprentissage est une phase du développement d'un réseau de neurones durant laquelle le comportement du réseau est modifié jusqu'à l'obtention du comportement désiré, l'apprentissage neuronal fait appel à des exemples de comportement.

Dans le cas des réseaux de neurones artificiels, on ajoute souvent à la description du modèle l'algorithme d'apprentissage. Le modèle sans apprentissage présente en effet peu d'intérêt.

Dans la majorité des algorithmes actuels, les variables modifiées pendant l'apprentissage sont **les poids des connexions**. L'apprentissage est la modification des poids du réseau dans l'optique d'accorder la réponse du réseau aux exemples et à l'expérience. Il est souvent impossible de décider à priori des valeurs des poids des connexions d'un réseau pour une

application donnée. A l'issu de l'apprentissage, les poids sont fixés : c'est alors la phase d'utilisation. Certains modèles de réseaux sont improprement dénommés à apprentissage permanent. Dans ce cas il est vrai que l'apprentissage ne s'arrête jamais, cependant on peut toujours distinguer une phase d'apprentissage (en fait de remise à jour du comportement) et une phase d'utilisation. Cette technique permet de conserver au réseau un comportement adapté malgré les fluctuations dans les données d'entrées.

Au niveau des algorithmes d'apprentissage, deux grandes classes sont été définies selon que l'apprentissage est dit supervisé ou non supervisé.

Dans le cas de l'apprentissage supervisé, les exemples sont des couples (Entrée, Sortie associée) alors que l'on ne dispose que des valeurs (Entrée) pour l'apprentissage non supervisé.

Remarquons cependant que les modèles à apprentissage non supervisé nécessite avant la phase d'utilisation une étape de labélisation effectuée l'opérateur, qui n'est pas autre chose qu'une part de supervision.

5.4. Perceptron multicouche

Le perceptron multicouche (PMC) est la deuxième grande famille de réseaux de neurones. Après qu'on va décrire l'architecture de ces réseaux on va aborder leur principe, et le concept de rétro propagation de l'erreur.

- **Architecture:** Un neurone de perceptron réalise un produit scalaire entre son vecteur d'entrées et un vecteur de paramètres appelé poids, à chaque produit on y ajoute un biais et on utilise une fonction d'activation pour déterminer sa sortie :

$$(1) y = f(x * w + b)$$

Il existe trois types de fonctions d'activation. Mais, seulement deux fonctions d'activation sont souvent utilisées dans la littérature. Ses fonctions sont la tangente hyperbolique $f_1()$ et la Sigmoid standard $f_2()$:

$$(2) f_1 = \tanh(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}}$$

$$(3) f_2 = \frac{\tanh(x)+1}{2}$$

La différence entre ces deux dernières fonctions est le domaine des valeurs de la sortie, qui varient entre $[-1;+1]$ pour la tangente hyperbolique et de $[0;1]$ pour la sigmoïde standard.

La figure 7 montre la courbe de la tangente hyperbolique et le Sigmoid et l'Unités Rectifié Linéaire (ReLU, Rectified Linear Units) :

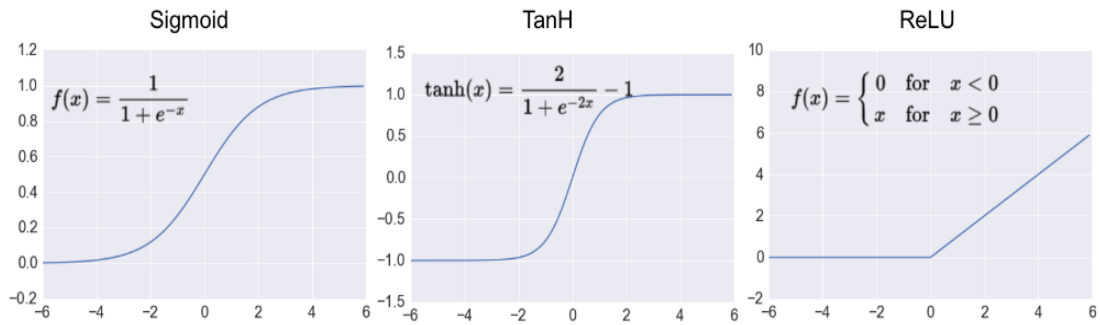


Figure 7 - la courbe de la tangente hyperbolique et le Sigmoid et le ReLU [MOU16]

Le perceptron est organisé en plusieurs couches : la première couche est reliée aux entrées, puis ensuite chaque couche est reliée à la couche précédente. C'est la dernière couche qui produit les sorties du PMC (Perceptron MultiCouche). Les sorties des autres couches ne sont pas visibles à l'extérieur du réseau, et elles sont appelées pour cette raison couches cachées. La figure 8 illustre l'architecture de perceptron multicouche.

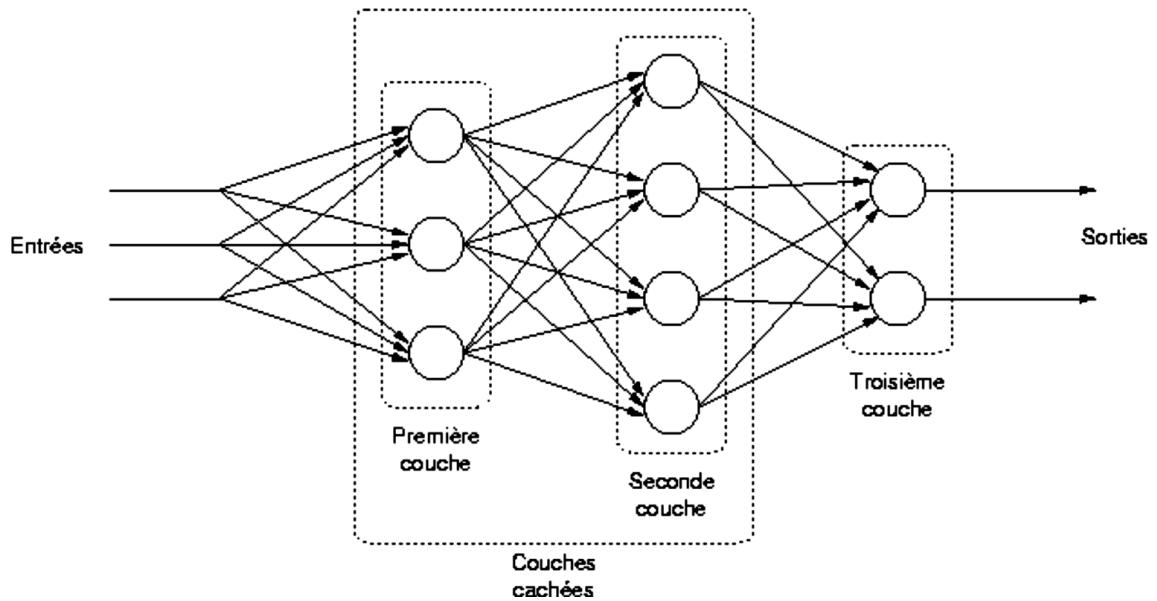


Figure 8 - Perceptron multicouche [TER03]

6. Le Deep Learning

Le Deep Learning ou l'apprentissage profond est une branche de l'intelligence artificielle et du Big Data (grosses données ou mégadonnées), ce domaine connaît un immense développement notamment ces dernières années. Les grandes sociétés comme Google, Facebook, Microsoft et Amazone, travaillent sur ce sujet.

6.1. Concepts et techniques:

Le Deep Learning ou l'apprentissage profond est un ensemble de méthodes d'apprentissage automatique tentant de modéliser, avec un haut niveau d'abstraction, des données grâce à des architectures articulées de différentes transformations non linéaires. Ces techniques ont permis des progrès importants et rapides dans les domaines de l'analyse du signal sonore ou visuel et notamment de la reconnaissance faciale, de la reconnaissance vocale, de la vision par ordinateur ainsi que du traitement automatisé du langage.

Dans les années 2000, ces progrès ont suscité des investissements privés, universitaires et publics importants, notamment de la part du GAFA (Google, Apple, Facebook, Amazon) [WIK18b].

En 1989, le chercheur français Yan LeCun a développé un système de reconnaissance automatique de chèques pour les banques basé sur des algorithmes de Deep Learning. Malgré un franc succès, l'application qui nécessitait trois jours de traitement prenait trop de temps.

C'était tout le problème, même avec la bonne technique, le Deep Learning ne permettait pas d'obtenir des résultats satisfaisants dans un temps raisonnable.

À partir de 2012, l'accès à de quantités de données de plus en plus importantes, les nouveaux algorithmes issus de l'intelligence artificielle, et l'utilisation de GPU (Graphics Processing Units) (cartes graphiques du type Titan X de Nvidia) pour effectuer du calcul très intensif ont complètement changé la donne. Il y a eu vraiment un saut technologique très important qui a transformé les réseaux de neurones en de véritables machines à reconnaître [CLU15].

Les techniques de Deep Learning constituent une classe d'algorithmes d'apprentissage automatique dont les caractéristiques sont les suivantes :

- Elles utilisent différentes couches d'unité de traitement non linéaire pour l'extraction et la transformation des caractéristiques. Chaque couche prend en entrée la sortie de la précédente. Les algorithmes peuvent être supervisés ou non supervisés et leurs applications comprennent la reconnaissance de modèles ou la classification statistique.
- Elles fonctionnent avec un apprentissage à plusieurs niveaux de détails ou de représentations des données. A travers les différentes couches on passe de paramètres de bas niveau à des paramètres de plus haut niveau.
- Ces différents niveaux correspondent à différents niveaux d'abstraction des données.
- Ce nouveau champ d'étude a pour objectif d'avancer davantage vers les capacités d'intelligence artificielle. Ses architectures permettent aujourd'hui de donner du sens à des données sous forme d'image, de son ou de texte.

Par ailleurs la figure 9 illustre un système de Deep Learning pour la reconnaissance du visage. Ce système est basé sur les réseaux de neurones artificiels, qui est constitué d'un ensemble de couches cachées, le mot profond (apprentissage profond) vient de ce nombre des couches cachées, on peut trouver des systèmes qui contiennent un nombre important de couches et aussi de neurones. Ce type d'algorithmes nécessite une capacité de calcul importante, pour cela il est nécessaire d'utiliser les GPU afin d'être capable d'exécuter des opérations complexes en un temps court.

La différence entre un perceptron classique, et un système de Deep Learning est que pour le premier les entrées des réseaux sont les caractéristiques de l'image, mais pour le deuxième ce sont les pixels bruts de l'image d'entrée. En effet dans un système de Deep Learning chaque couche est considérée comme un niveau d'abstraction de l'image.

Sur la figure 9, la couche 1 est capable d'extraire des caractéristiques de niveau d'abstraction inférieure à celui de la couche 2, tandis que la couche 3 a une qualité plus importante. À partir de ces caractéristiques le système est capable de reconnaître les objets avec un taux d'erreur petit.

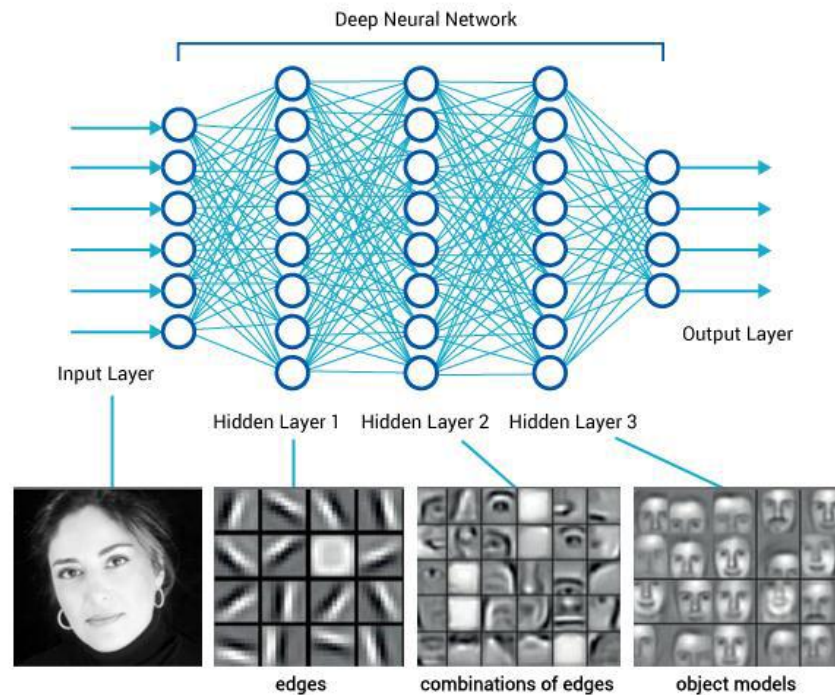


Figure 9 - Exemple d'un système de Deep Learning

[MED17a]

6.2. Les champs d'application

Des millions de personnes profitent déjà des progrès réalisés grâce au Deep Learning. Parmi les champs d'action de cette technologie on trouve : le traitement d'image, de la vidéo, de la voix, et plus spécifiquement de la détection d'objets dans le domaine automobile (piétons, panneaux de signalisation, voitures, bus, marquage au sol...).

De nombreux secteurs d'activité exploitent des programmes de reconnaissance d'image: les sites marchands pour classifier, identifier et trouver des similitudes entre des produits ou des familles de produits , la robotique pour permettre par exemple à un robot-aspirateur de se déplacer en toute autonomie, les réseaux sociaux et autres sites de partage de photos pour la reconnaissance faciale, les moteurs de recherche pour détecter les similitudes entre des images , l'industrie automobile pour le développement des voitures autonomes, la médecine pour la recherche de cellules cancéreuses.

On citant quelques exemples:

a) DeepFace de Facebook

Dévoilé au mois de juin 2014, le système de reconnaissance faciale Deep Facede Facebook se révèle très impressionnant. Le réseau social, qui possède la plus importante galerie photos en ligne du monde (250 milliards de photos !), a développé un réseau de neurones capable de reconnaître le visage d'une personne sur une photo avec un taux de réussite de 97,25%. Pour réussir un tel exploit, Facebook n'a pas fait les choses à moitié et s'est notamment offert les services du chercheur français Yan LeCun pour diriger son laboratoire d'intelligence artificielle basé à Paris. L'équipe de chercheurs a utilisé des techniques d'apprentissage profond et a traité 4 millions de photos représentant 4000 personnes issues de sa base de données [CLU15]. Chaque visage a été modélisé en 3D pour pouvoir être positionné de face puis analysé à l'aide de 120 millions de paramètres variables.

Une fois les données récoltées, le système effectue des comparaisons entre les photos pour trouver des correspondances. Cette fonction a, depuis, été implémentée dans le réseau social partout dans le monde, sauf en Europe où elle a été jugée trop intrusive par les défenseurs du droit à la vie privée [CLU15]. La figure 10 illustre le principe de DeepFace.

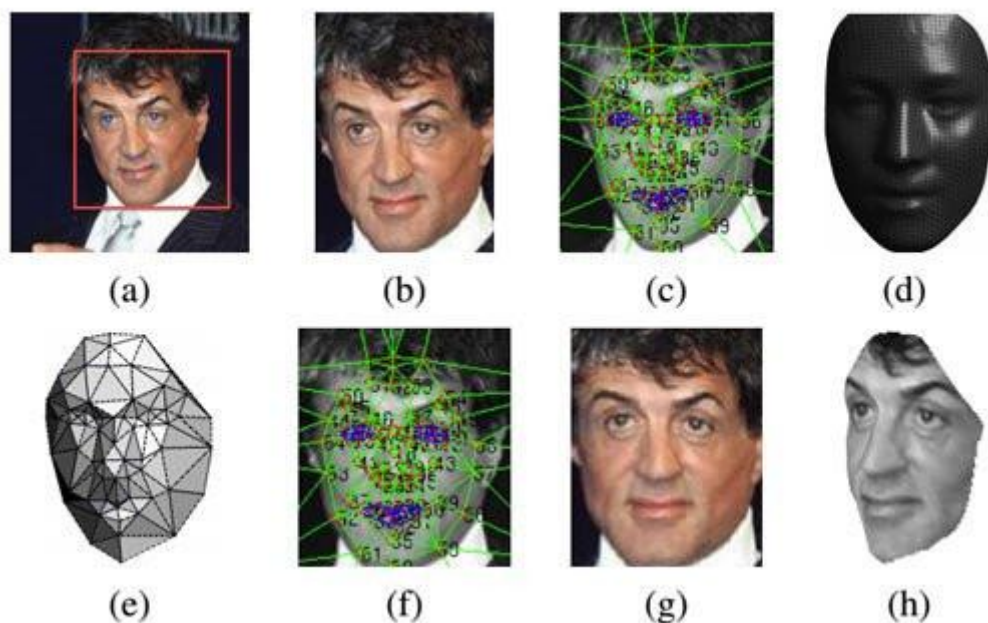


Figure 10 - Principe de DeepFace (de Facebook)
[CLU15]

b) Skype Translator de Microsoft

La firme de Redmond a développé l'une des applications basées sur le Deep Learning et la machine Learning la plus impressionnante. Lancée en version d'essai au mois de décembre 2014, Skype Translator est un système de traduction à la volée de conversations. L'application traduit les appels audio et vidéo quasiment en temps réel dans six langues : anglais, espagnol, français, allemand, italien, et mandarin. Pour les messages écrits, Skype Translator prend en charge 50 langues. Pour l'heure, le système ne fonctionne pas sur les terminaux mobiles, mais uniquement sur PC. Les premiers utilisateurs ne tarissent pas d'éloges à son propos sur les réseaux sociaux [CLU15].

7. Conclusion

Dans ce chapitre, on a montré l'interaction homme machine ensuite on a expliqué la reconnaissance de visage et les grandes difficultés auxquelles elle confronte après on a défini l'émotion et les expressions faciales.

Aussi on a mis l'accent sur les principales notions de réseaux neurones artificiel leurs caractéristiques, finalement on a parlé de Deep Learning ou l'apprentissage profonde, ses procédés, ses techniques, le domaine de son application.

Dans le chapitre suivant on va discuter les travaux connexes et les approches existantes de la détection et reconnaissance de visage, et aussi les projets réaliser des miroirs intelligents.

Chapitre II

Travaux Connexes

Chapitre II : Travaux Connexes

1. Introduction

Afin d'évaluer le miroir intelligent et les techniques utilisées pour sa réalisation il faut qu'on les positionne dans leur axe de recherche et flux des travaux similaires. Dans ce chapitre, on va aborder l'état de l'art des expressions faciales puis on montrera la détection de visage et la classification et aussi les procédés suivis pour la reconnaissance de visage, et les méthodes les plus utilisées pour cela, et aussi les travaux connexes liés au miroir intelligent.

2. Expressions Faciales: Etat de l'Art

Le principe de fonctionnement de base d'un système de reconnaissance faciale (Figure 11) peut être résumé en trois étapes.

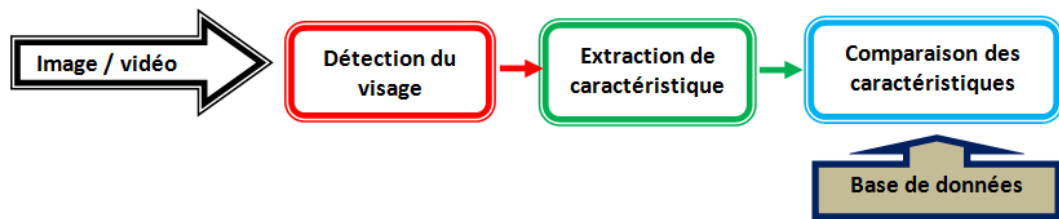


Figure 11 - Le principe de fonctionnement de base d'un système de reconnaissance faciale [OAU12]

Il y a trois approches qui sont proposées dans la littérature pour la reconnaissance automatique d'expression faciale qui est :

L'approche basée modèles, approche basée règles, et approche basée géométrique, chaque approche utilise fréquemment les techniques de réseaux neurones et le flux optique.

2.1. La détection de visage

Il s'agit d'une détection d'un ou plusieurs visages humains par ordinateur dans des images numériques, les procédés de la détection de visage ont été classifiés en 04 catégories par [MHY41].

2.1.1. Approche basée sur les connaissances acquises

D'après [ABA08], cette approche se concentre sur les parties caractéristiques du visage telles que le nez, la bouche et les yeux. Elle est basée sur la définition de règles strictes basées sur la relation entre les caractéristiques du visage. Ces méthodes sont principalement conçues pour la localisation du visage dans une image. Tout d'abord, les caractéristiques du visage sont localisées en utilisant la méthode de projection pour détecter les contours du ou des visages actuelles, la formule dit que Soit $I(x, y)$, l'intensité de la luminance du pixel (x, y) de l'image $m \times n$, les projections horizontale et verticale de cette image sont définies par :

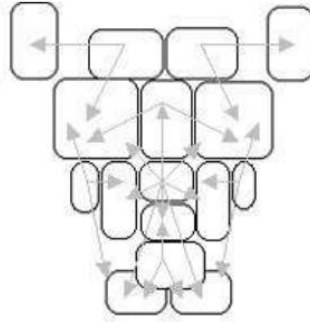
$$HI(x) = \sum_{y=1}^n I(x, y)$$

et

$$VI(x) = \sum_{x=1}^n I(x, y)$$

2.1.2. Approche basée sur "le Template matching " ou le modèle de correspondance

Comme mentionné dans [ABA08], les Templates peuvent être définis soit "manuellement" soit paramétrés à l'aide de fonctions, c'est de calculer le rapport (corrélation) entre l'image candidate et le Template (le modèle), en expliquons plus que Le modèle de correspondance est une technique simple permettant de trouver de petites pièces dans une image qui correspond à une image modèle. Dans cette technique, les images modèles de différents objets sont stockées dans un espace. Lorsqu'une image est mise à l'entrée du système, elle est comparée avec tous les modèles stockés dans l'espace, pour déterminer ou identifier la classe d'appartenance de l'image d'entrée, aussi nous pouvons également le définir comme une technique simple permettant de trouver de petites pièces dans une image qui correspond à une image modèle voir Figure 12.



**Figure 12 - Modèle de visage composé de 16 régions
[ABA08]**

2.1.3. Approche basé sur l'apparence

Selon [ABA08] ces approches utilisent souvent des techniques d'apprentissage automatique de ce fait, les modèles sont appris à partir d'un ensemble d'images représentatives de variabilité de l'aspect du visage, donc ces modèles sont utilisées pour la détection, l'idée principale de ces méthodes est de projeter l'image dans un espaces et à calculer la distance euclidienne entre l'image et sa projection. L'une des approches plus utilisée est l'Eigenface.

2.1.4. Approche basée sur des caractéristiques invariantes

Selon [ABA08] ces approches sont employées généralement pour la localisation des visages, en effet on développe les algorithmes pour trouver les caractéristiques structurales présentes même si la pose, le point de vue, la condition d'éclairage peuvent varier.

2.1.5. Détection des Objets par la méthode de Viola et Jones

C'est une méthode efficace de détection d'objets qui a été proposée par Paul Viola et Michael Jones [PAU01]. C'est une méthode basée sur l'apprentissage de la machine où la fonction cascade est générée à partir d'un grand nombre d'images positives et négatives. La fonction cascade est ensuite utilisée pour détecter des objets pertinents dans d'autres images.

2.1.6. L'algorithme de Viola et Michael Jones (Classificateur Haar-cascade)

Les caractéristiques pseudo-Haar sont décrites pour la première fois dans un article de Paul Viola et Michael Jones paru en 2001(voir figure 13) dans la revue scientifique « International Journal of Computer Vision», dans laquelle ils décrivent une nouvelle méthode de détection du visage. Il est possible de calculer très rapidement les caractéristiques pseudo-Haar à l'aide des images intégrales. Une image intégrale est une table de correspondance 2D, construite

à partir de l'image d'origine, elle est de même taille que l'image d'entrée. Elle contient en chacun de ses points la somme des pixels situés au-dessus et à gauche du pixel courant.

Une fois le calcul de l'image intégrale effectué, la somme des pixels dans n'importe quel rectangle situé dans l'image peut se calculer en seulement 3 opérations et 4 accès à l'image intégrale. Une caractéristique pseudo-Haar à deux rectangles peut alors être déterminée en seulement 6 accès (2 points sont partagés) à l'image, et une caractéristique à 3 rectangles en seulement 8 accès [WIK18c].



Figure 13 - Un exemple des premières caractéristiques pseudo-Haar utilisées par Viola et Jones en 2001 [WIK18c]

Lienhart et Maydt [LIE02] ont introduit en 2002 le concept de caractéristiques pseudo-Haar orientées à 45° afin de mieux prendre en compte dans une image les structures inclinées (par exemple un contour à 45°, voir figure 14). Ils ont montré que ce concept apportait une amélioration de près de 10 % des performances par rapport à l'ensemble de caractéristiques utilisées par Viola et Jones. Ils ont également proposé une extension des images intégrales pour permettre le calcul rapide de ces caractéristiques à 45°.

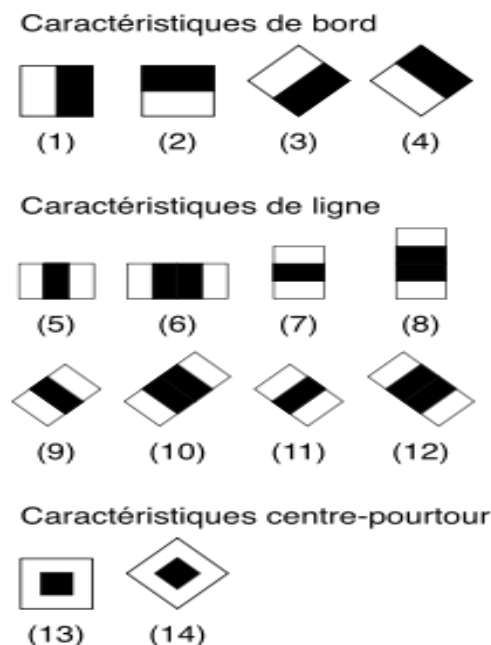
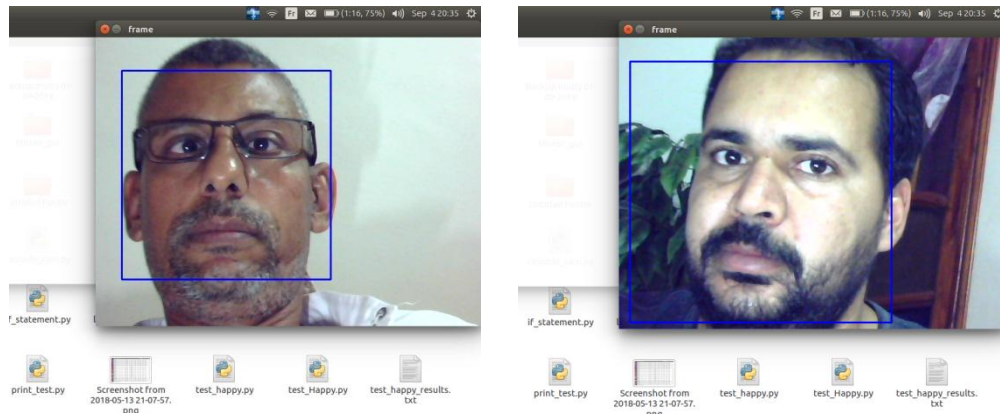


Figure 14 - l'extension du caractéristique pseudo Haar par Lienhart et Maydt [WIK18c]

*Daha*

&

*Medileh***Figure 15 - Le résultat de détection de visage en utilisant le Haar-cascade**

La figure 15 montre le résultat de la détection de visage on utilisant le Haar-cascade

2.2. Les méthodes de reconnaissance faciales

Les méthodes de reconnaissance faciales peuvent être séparées en trois grandes catégories, les méthodes globales, les méthodes locales, et les méthodes hybrides (figure 16). Elle est définie par [WIK15] comme étant "le domaine de la vision par ordinateur consistant à reconnaître automatiquement une personne à partir d'une image de son visage" cela se fait sur une vidéo et sur une photo.

2.2.1. Méthodes globales

Selon [GHA10], les méthodes de cette approche tiennent compte de toute l'information du visage. Cette information est préservée et elle permet au classificateur de découvrir les données pertinentes parmi les données recueillies. Cependant, l'étape de normalisation implique habituellement l'image entière et cela prend généralement beaucoup de temps. Le traitement de tous les Pixels dans l'image est coûteux et un grand espace mémoire est exigé. Ces méthodes sont relativement peu sensibles au mouvement subtil et le résultat peut être facilement affecté par le changement de l'éclairage.

C'est de représenter une image faciale par un seul vecteur de grande dimension, cette représentation est appelée description basée sur l'apparence globale, elle a deux avantages, le premier c'est qu'elle concerne toutes les informations de texture et de forme utiles pour différencier les visages, le deuxième c'est qu'elle tient en compte des aspects d'organisation

structurelle globaux du visage, cependant son inconvénient principale se localise dans la dimension très grande de l'espace de l'image qu'elle a besoin.

Pour se pallier à ce problème on fait appel à la description par visages propres [TUR91] qui est basée sur l'Analyse en Composantes Principales (ACP).

1- Analyse en Composantes Principales (ACP): Son principe se fait à partir de trouver les composantes principales d'un ensemble d'images de visages et cela revient à préciser les vecteurs propres de la matrice de covariance formée par les images en question, l'ACP est une technique ambitieuse car elle permet la reconstruction d'une base de dimension réduite toute fois elle n'est optimisée pour la séparabilité de classe.

Le premier système de reconnaissance de visages qui a permis d'obtenir des résultats significatifs a été réalisé par Turk et Pentland [TUR91] en utilisant la méthode dite des «Eigenfaces ».

Analyse Discriminante Linéaire (ADL): Le but de la majorité des algorithmes basés sur l'ADL [BEL97] et [MAR01] est de trouver les directions de projection les plus discriminantes dans l'espace propre en augmentant le ratio entre les variations inter personnelles et les variations intra personnelles ces dernières peuvent être petites, dans ce cas-là ce ratio n'est pas facile à augmenter car il est déjà grand.

Pour remédier ce problème, on emploi d'abord l'ACP puis l'ADL, cette méthode se nomme Fisher faces, l'inconvénient majeur des méthodes globales est son utilisation des photos 2D d'apparence faciale uniquement, la solution de ce problème est de recourir aux représentations faciales locales.

Belhumeur et Col. ont proposé la méthode des "Fisherfaces" pour résoudre le problème de la robustesse face aux variations de pose, d'illumination et d'expressions mettant en difficultés la PCA ,l'inconvénient de ces méthodes, utilisent uniquement les photos 2D [BEL97].

2.2.2. Méthodes locales

Selon [GHA10] Les méthodes dans cette approche suivent la position de quelques composants faciaux comme les yeux et la bouche, et supposent que le mouvement relatif de ces composants est lié à l'intensité de l'expression. En utilisant des méthodes de cette approche le nombre des entrées est considérablement réduit ce qui provoque une réduction du temps de

calcul et minimisation de la complexité. Dans ce cas l'efficacité du traqueur devient plus importante.

Elles sont classées en deux catégories, celles basées sur les point d'intérêts et celles basées sur l'apparence du visage.

1- Méthodes locales basées sur les points d'intérêts: on extrait des caractéristiques localisées sur les points d'intérêts à partir la détection de ces derniers, d'autre manière on dit comme suite :

- Détection des points d'intérêts.
- Extraction des caractéristiques localisées sur ces points d'intérêts.

Les méthodes ont deux inconvénients :

- les caractéristiques géométriques sont difficiles à extraire dans le cas ou des occultations de visages sont présentes.
- les caractéristiques géométriques ne sont pas suffisantes pour représenter réellement un visage.

2- Méthodes locales basées sur l'apparence du visage: par ces méthodes on divise le visage en petites régions sur lesquelles les caractéristiques locales sont extraites directement.

- Divisions du visage en petites régions.
- Extraction des caractéristiques locales.

Les méthodes locales sont les plus difficiles à mettre en place, mais ils sont les plus robustes. Car, ils appliquent les transformations autour des points caractéristiques des images.

2.2.3. Méthodes hybrides

Elles utilisent à la fois des caractéristiques globales et les caractéristiques locales, cette particularité peut donner des informations complémentaires utiles à la tâche de classification

Notons aussi que d'un certain point de vue, les méthodes locales peuvent être considérées comme des méthodes hybrides car l'information globale est généralement prise en compte.

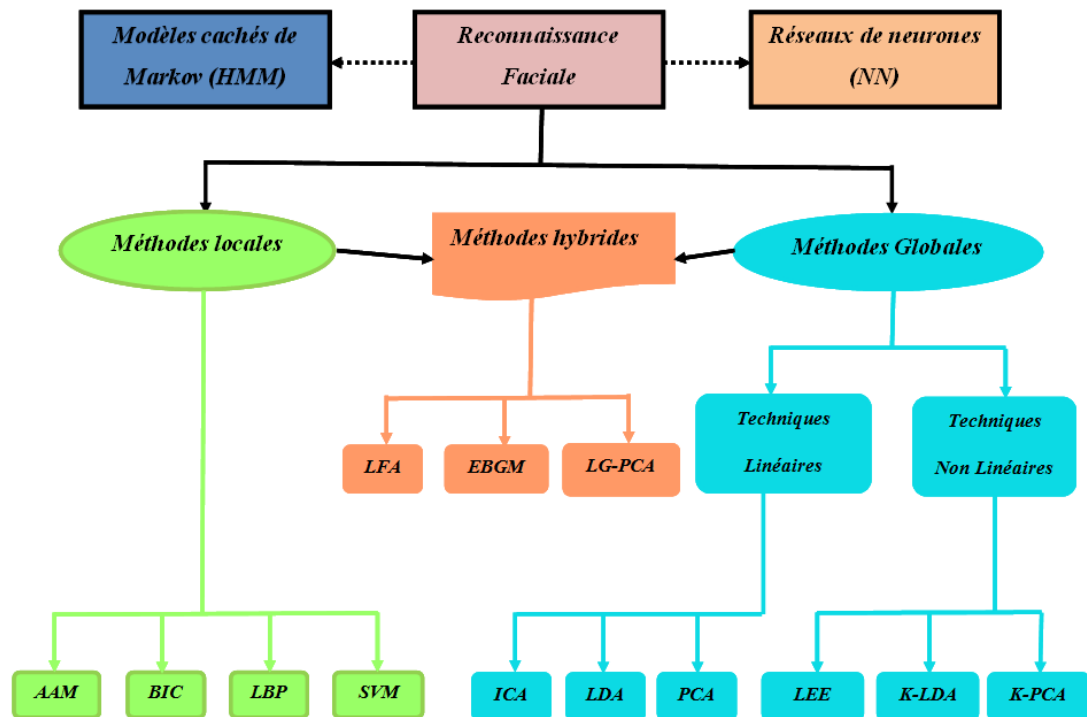


Figure 16 - Classification des algorithmes principaux utilisés en reconnaissance faciale.
[OUA12]

2.3. Le développement émotionnel

2.3.1. Méthode subjective

Elle est plus utilisée, en effet, cette méthode considère la face comme étant un stimulus engendrant un jugement chez les observateurs sur l'état émotionnel, lorsqu'on recourt aux jugements, l'observateur doit répondre librement ou bien répondre en choisissant une catégorie d'émotion parmi une liste proposée [LAP07].

2.3.2. Méthode objective :

Elle consiste à déterminer la quantité et la qualité des paramètres faciaux. Il y'a beaucoup de chercheurs qui ont utilisé cette méthode dont Darwin, Ekman et Friesen, ces derniers ont proposé une liste d'expressions faciales qui sont décrites à partir d'un system de codification de changement faciaux (Facs) ce système contient 44 composantes faciales de base nommées unités d'actions , chaque unité se rapporte à un changement de l'expression de visage sous l'effet d'un muscle ou d'un groupe musculaire au niveau de la face.

2.3.3. Neuraux imagerie fonctionnelle

Cette méthode vise à comparer les changements transitoires d'états émotionnels aux variations d'activité du système nerveux, de ce fait on recourt à des paradigmes « d'activation émotionnelle » ainsi on distingue l'induction externe et interne d'état émotionnel (voir figure 17).

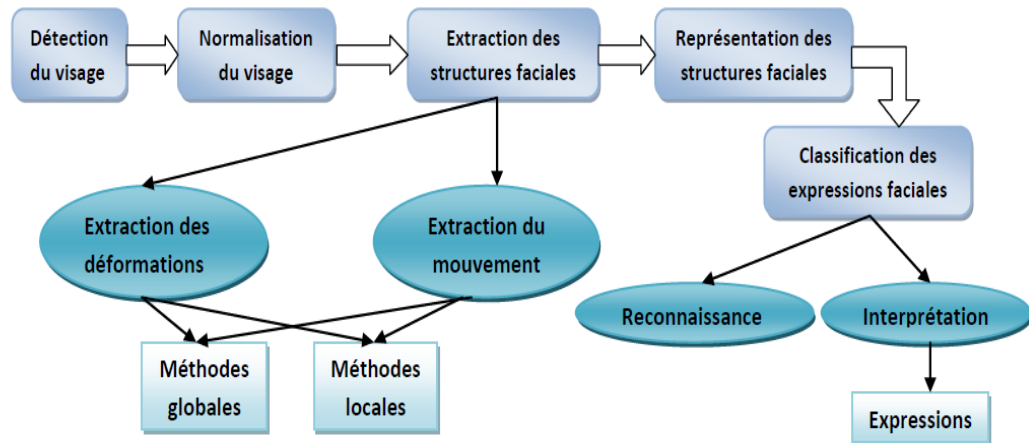


Figure 17 - Méthodes d'analyse automatique d'un système de reconnaissance d'expressions faciales [OUA12]

3. Smart Mirrors: Etat de l'Art

Le miroir intelligent ou *smart mirror*, il s'agit d'un miroir connecté capable d'interpréter les émotions puis il réagit suivant les traits de visage, dans cette section on va présenter quelques travaux liés.

3.1. Panasonic Smart Mirrors

En septembre 2014, lors des salons industriels IFA (International Funkausstellung Berlin), Panasonic a présenté pour la première fois un miroir intelligent prototypique, qui applique le maquillage numérique pour que les utilisateurs puissent découvrir de nouveaux looks.

En utilisant l'aide des capteurs et du logiciel de reconnaissance faciale, le miroir crée un hologramme élargi du visage de l'utilisateur, puis diagnostique les zones faibles (rides, sourires, imperfections), puis, lorsque c'est fait, il fait des recommandations personnalisées basées sur l'état de la peau [ENG14].

Après deux ans, en janvier 2016, au CES 2016 (Consumer Electronics Show) au Las Vegas Convention Center à Las Vegas, Nevada, aux États-Unis, Panasonic a représenté la deuxième version de ses miroirs intelligents. Le miroir a été équipé d'une caméra et de capteurs qui aident à l'analyse de la peau comme le premier modèle et a ajouté un nouveau maquillage différent. Sachant que les deux miroirs sont équipés d'écrans tactiles.

En plus des fonctions précédentes, les miroirs étaient assez intelligents pour voir à travers les épaisses couches de maquillage et si la peau avait besoin d'un peu plus d'hydratation.

D'un autre côté, le miroir cosmétique permet aux utilisateurs d'appliquer toutes sortes de maquillage, des ombres à paupières aux teintes de lèvres peachy. Ils peuvent même affiner le maquillage en apportant une palette particulière dans Photoshop et l'améliorer, en sélectionnant juste un maquillage pour essayer, et optez pour des looks personnalisés [ENG16].

3.2. The Wize Mirror

En novembre 2015, Sara Colantonio et onze chercheurs ont publié un article de recherche intitulé «Un miroir intelligent pour promouvoir un style de vie sain» dans le journal Elsevier.

L'idée derrière le Wize Mirror est que le visage est le canal de communication par excellence entre les humains: c'est un miroir du statut, des émotions et de l'humeur. En tant que tel, il constitue la base des séméiotiques médicales, qui révèlent l'état de santé d'un individu grâce à une combinaison de signes physiques du visage (par exemple, la couleur de la peau, la graisse sous-cutanée) et des expressions faciales. Datant d'aussi loin qu'Aristote, les séméiotiques médicales sont encore utilisées aujourd'hui par les médecins. Le Wize Mirror transforme les séméiotiques médicales dans le domaine numérique: il traduit le code sémiotique du visage en descripteurs de calcul liés au risque cardio-métabolique et des mesures extraites automatiquement des vidéos, des images et des scans en 3D du visage des personnes devant le miroir [COL15].

3.3. Google Android Smart mirror

Le 19 janvier 2016, un développeur de logiciels chez Google nommé Max Braun a publié un article sur Medium.com, intitulé: "Mon miroir de salle de bain est plus intelligent que le vôtre".

Braun a créé un miroir qui fonctionne sur Google Android et indique les nouvelles, l'heure et les informations météorologiques. Il a été inspiré par Arnold Schwarzenegger Film "Le sixième jour - 2000", Braun a essayé de réaliser l'idée futuriste à ce moment-là, et il continue d'expérimenter avec des mises à jour de trafic, des rappels quotidiens et des recherches vocales [QUA16].

3.4. Microsoft IoT Magic Mirror

En mai 2016, lors de l'Innovation Festival Unbound (conférence sur la technologie numérique) à Singapour, Microsoft a dévoilé son Magic Mirror pour la première fois.

Ce Magic Mirror utilise un Raspberry Pi qui fonctionne sur la technologie Windows 10 IoTCore et Windows Hello pour reconnaître et saluer les utilisateurs, et lire les émotions en utilisant une caméra intégrée avec un système de reconnaissance faciale, qui peut reconnaître huit émotions différentes, telles que la colère et le bonheur. En outre, le miroir affiche la météo, l'heure, les infos de Facebook et Twitter, et d'autres informations pertinentes.

Ce qui rend ce miroir intelligent, c'est sa capacité à reconnaître les utilisateurs et lire leurs émotions et de se comporter en fonction de l'identité de l'utilisateur [DAI16].

3.5. Perseus, le premier miroir intelligent au monde

Perseus est le premier miroir commercial le plus intelligent au monde. C'est un miroir interactif à commande vocale qui est aussi un écran numérique pour aider les gens à rester connectés, de garder le contrôle et d'être ridiculement beau chaque fois qu'ils sortent. En moyenne, les gens passent chaque jour plus de 30 minutes devant le miroir, Nous voulons donc que ces 30 minutes soient aussi utiles, efficaces et amusantes que possible [BLO17].

4. Conclusion

Dans ce chapitre, on a abordé les expressions faciales et on a montré la détection de visage et la classification de détection de visage, également-on vu les procédés suivis pour la reconnaissance de visage et les méthodes les plus employées, finalement on a parlé de l'efficacité de système de reconnaissance de visage tout en citant les travaux connexe dans le domaine.

Dans le chapitre suivant nous allons donner une vue sur la conception et la méthodologie de travail.

Chapitre III

Conception

1. Introduction

Dans ce chapitre, on va expliquer la méthodologie de travail suivie et les étapes nécessaires pour atteindre les résultats souhaités, on va citer les algorithmes de Deep Learning et leurs utilités dans la réalisation d'un miroir intelligent, on va évoquer les moyens utilisés pour la conception et la classification procédée à l'apprentissage automatique et l'apprentissage profonde et les étapes suivis pour comparer entre les deux, et la manière de faire un classificateur des expressions faciales en utilisant un réseau de neurones convolutif avec leNet, aussi le processus de l'apprentissage et finalement, on va présentera l'architecture de miroir intelligent.

2. Méthodologie du travail

Les algorithmes de Deep Learning nécessitent des performances importantes. Malgré L'évolution des ordinateurs, ces derniers restent incapables d'exécuter les algorithmes du Deep Learning en un temps très court. L'utilisation d'un GPU est la seule solution pour améliorer les processus d'exécution du Deep Learning.

Le Raspberry Pi est un micro-ordinateur de petite taille donc on peut facilement l'intégrer dans des grands systèmes (par exemple les robots). Malheureusement les micro-ordinateurs disponibles comme Raspberry Pi ou Banana n'ont pas les performances nécessaires, donc ne peuvent pas être utilisés pour exécuter les algorithmes de Deep Learning. Pour remédier à ce problème nous avons proposé une méthode que nous décrivons dans ce qui suit.

Notre travail est divisé en deux processus, le premier est de créer un modèle CNN qualifié avec Keras et Tensorflow en utilisant un PC, et le deuxième est d'utiliser ce modèle dans le Raspberry Pi, qu'il va utiliser une webcam pour capturer les images, un modèle Haarcascade sera utilisé pour détecter le visage de l'utilisateur, puis le recadrer, le redimensionner en 48x48 pixels et le convertir en une image en niveaux de gris à utiliser avec le modèle entraîné.

3. Classification en utilisant l'apprentissage automatique traditionnel par rapport à l'apprentissage en profondeur

La classification à l'aide d'un algorithme d'apprentissage automatique comporte deux phases (figure 18).

Phase d'apprentissage: Dans cette phase, on forme un algorithme d'apprentissage automatique en utilisant un ensemble de données comprenant les images et leurs étiquettes correspondantes.

Phase de prédiction: Dans cette phase, nous utilisons le modèle formé pour prédire les étiquettes d'images invisibles.

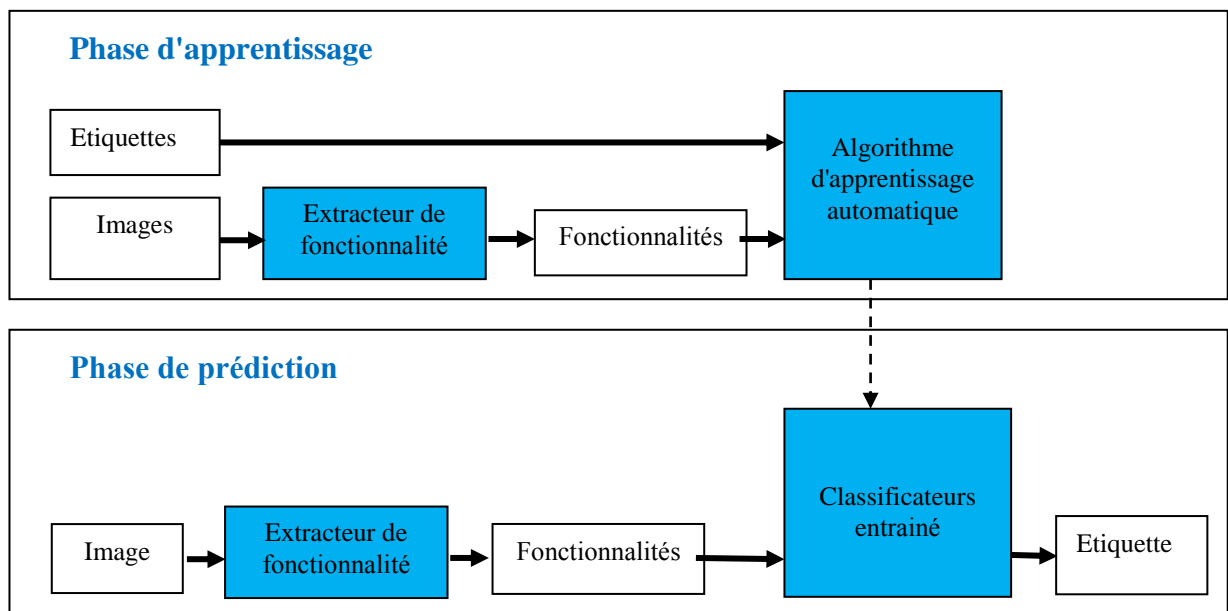


Figure 18 – les phases de Machine Learning
[MOU16]

La phase d'apprentissage d'un problème de classification d'image comporte deux étapes principales:

Extraction de caractéristiques: Dans cette phase, on utilise la connaissance du domaine pour extraire de nouvelles fonctionnalités qui seront utilisées par l'algorithme d'apprentissage automatique. HoG (Histograms of Oriented Gradients) et SIFT (Scale Invariant Feature Transform) sont des exemples de fonctionnalités utilisées dans la classification d'images.

Modèle d'apprentissage: Dans cette phase, on utilise un ensemble de données propre composé des caractéristiques des images et les étiquettes correspondantes pour former le modèle d'apprentissage de machine.

Dans la phase de prédiction, on applique le même processus d'extraction de caractéristiques aux nouvelles images et on transmet les caractéristiques à l'algorithme d'apprentissage automatique pour prédire l'étiquette.

La différence principale entre l'apprentissage automatique traditionnel et les algorithmes d'apprentissage profond (figure 19), réside dans l'ingénierie des fonctions. Dans les algorithmes traditionnels d'apprentissage automatique, nous devons façonner les caractéristiques à la main. En revanche, dans les algorithmes d'apprentissage en profondeur, l'ingénierie des caractéristiques est effectuée automatiquement par l'algorithme. L'ingénierie des fonctionnalités est difficile, prend beaucoup de temps et nécessite une expertise de domaine.

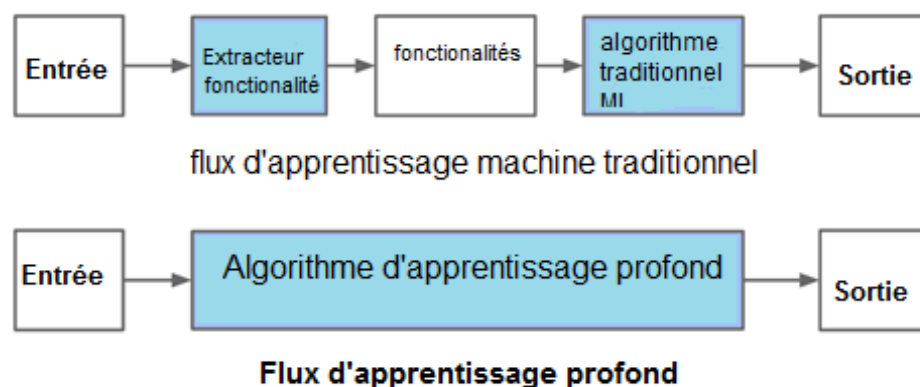


Figure 19 – La différence principale entre l'apprentissage automatique traditionnel et les algorithmes d'apprentissage profond
[MOU16]

La promesse de l'apprentissage en profondeur est plus précise des algorithmes d'apprentissage de la machine par rapport à l'apprentissage traditionnel de la machine avec moins ou pas d'ingénierie de caractéristiques.

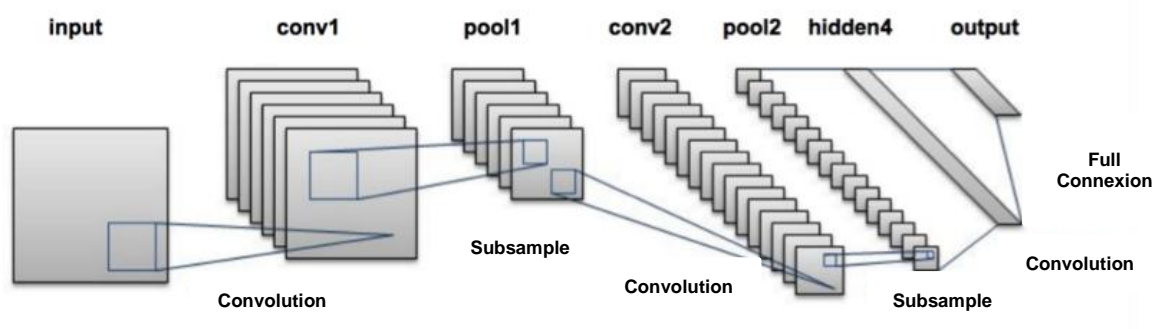
4. Construire un classificateur des expressions faciales en utilisant un réseau de neurones convolutif

On va implémenter un classificateur des expressions faciales en utilisant un réseau de neurones convolutionnels.

En raison de faibles performances du Raspberry Pi par rapport à un ordinateur, nous avons décidé d'utiliser une architecture CNN simple "LeNet".

4.1. LeNet:

Le LeNet, est un réseau convolutif à 7 niveaux pionnier de LeCun et Al en 1998 (figure 20), qui classifie les chiffres, il a été appliqué par plusieurs banques pour reconnaître des numéros manuscrits sur des chèques numérisés en 32x32 pixels en niveau de de gris. La possibilité de traiter des images à plus haute résolution nécessite des couches plus volumineuses et plus convolutives. Cette technique est donc limitée par la disponibilité des ressources informatiques.



**Figure 20 – Architecture de LeNet
[MED17b]**

L'architecture de LeNet est simple et petite (en termes d'empreinte de mémoire), et elle fonctionne sur le processeur (CPU).

5. Les phases de l'apprentissage et de la prédiction

5.1. La phase de l'apprentissage

La phase de l'apprentissage commencera par la préparation des données à utiliser (les images de l'apprentissage), les images doivent être classifiées par class, et on va appliquer

l'algorithme de Deep learning LeNet qui va donner comme sortie un classificateur entraîné et prêt à être utilisé par la phase de prédiction. (voir figure 21)

5.2. La phase de la prédiction

Cette phase va commencer par l'acquisition d'image avec une webcam, ensuite on va utiliser le modèle de Haar-Cascade pour détecter le visage de l'utilisateur et le recadrer, puis le convertir en une image en niveaux de gris et le redimensionner en 48x48 pixels pour l'utiliser avec le classificateur qui va faire la prédiction. (voir figure 21)

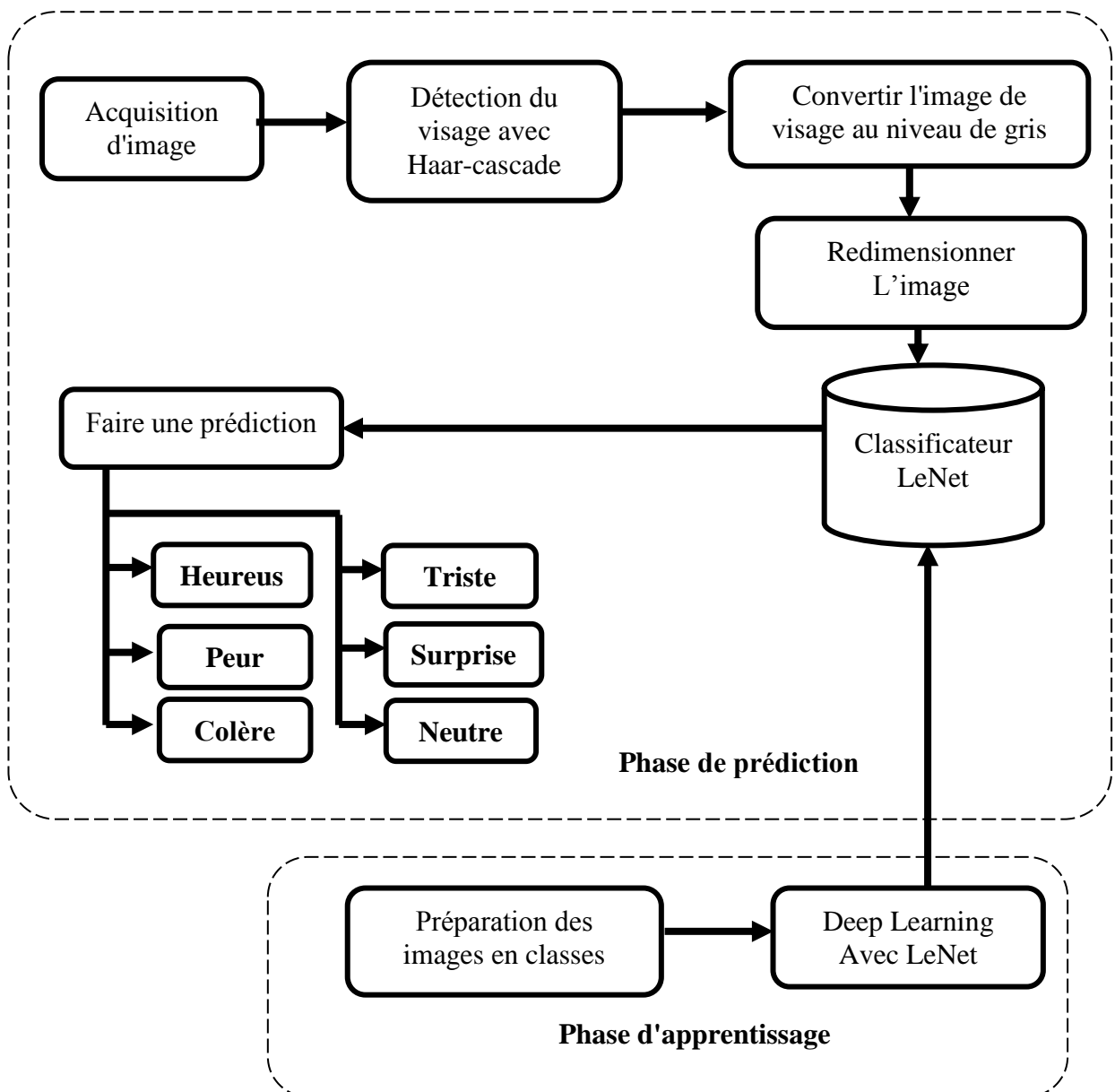


Figure 21 – Les phases de l'apprentissage et de la prédiction

6. Architecture de miroir intelligent

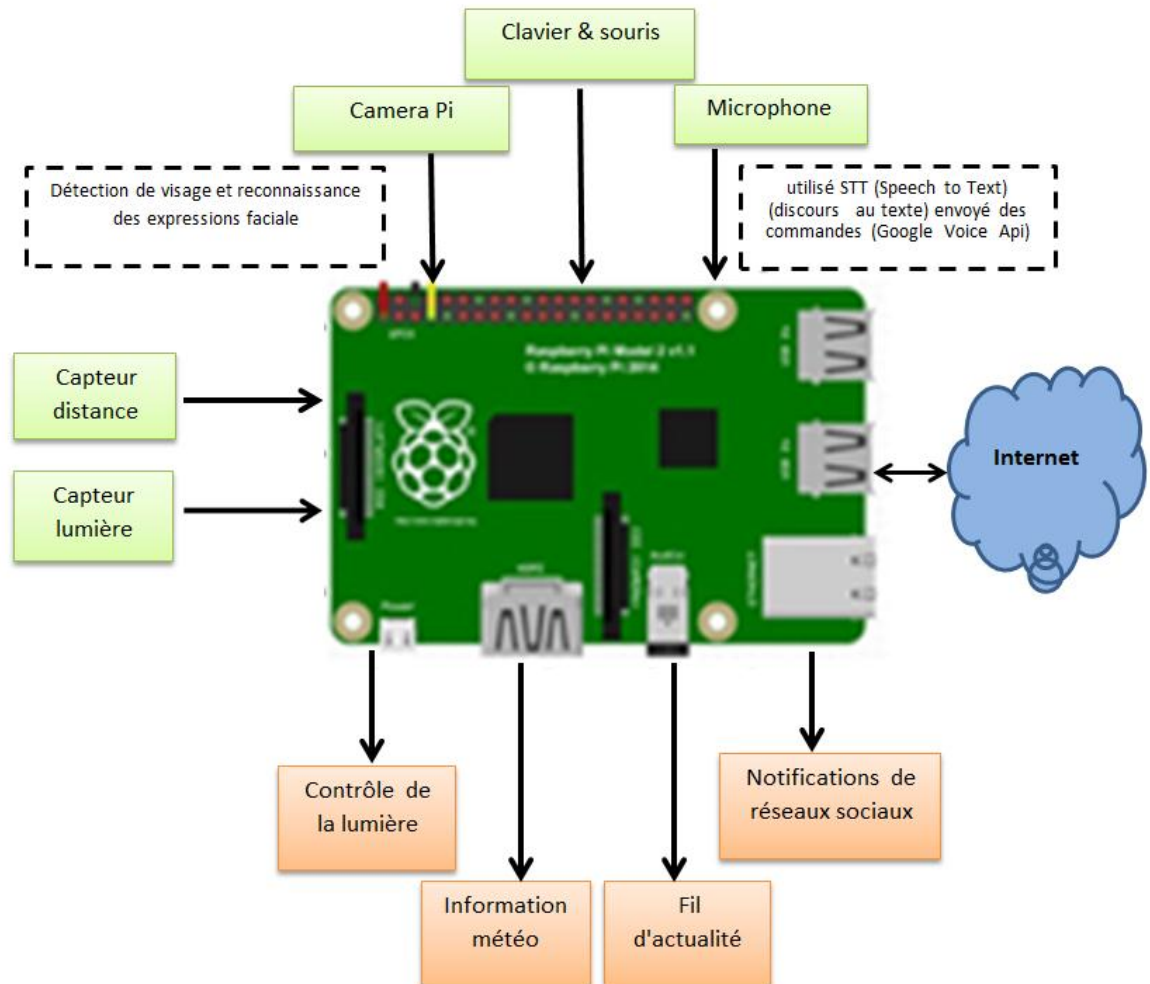


Figure 22 – Conception du miroir

Ce miroir va permet de fournir des informations pratiques utiles telles que: la météo, des actualités, l'heure, la date actuelle ... etc.

Au début, pour connaître la forme finale du miroir intelligent, il faut d'abord visualiser le schéma prototype et mettre les touches et les perceptions dans le protocole expérimental du miroir utilisé et connecté à la machine (figure 22).

Toute en mettant d'abord le formulaire et le mode suivant:

En termes d'entrées (les inputs)

Au début, il est nécessaire de connaître le nombre d'entrées nécessaires pour cette conception associée à la machine Raspberry Pi sachant que:

- **La WebCam:** est déjà connecté à la machine Raspberry Pi au-dessus de miroir car est un élément essentiel pour la détection des visages et de connaissance des expressions faciales pour donner un résultat fiable et convainquant.
- **Le capteur de distance:** est nécessaire car la distance pour capturer l'image du visage est importante pour que la caméra assure la bonne capture de l'image de visage d'une façon correct, et aussi de ne pas activer la camera si il n'y a pas de présence de l'utilisateur, le capteur va aider d'économiser l'énergie par l'activation de l'écran et la camera quand il y a un utilisateur devant le miroir.
- **Détecteur de lumière:** détermine la qualité et l'intensité de la lumière de l'environnement où se trouve le miroir.
- **Clavier et souris:** l'utilisateur va les utiliser pour configurer le système du miroir.
- **Le microphone:** l'utilisateur va l'utiliser pour envoyer des commandes vocales au miroir.

En termes des sorties (les outputs)

- **Contrôle de la lumière:** selon la valeur donné du détecteur de lumière si elle est appropriée ou non de prendre une photo du visage, le contrôle de la lumière va allumer ou éteindre les lampes associées au miroir.
- **Informations sur les conditions météorologiques:** ces informations concernent l'endroit où se trouve le miroir ainsi que les informations sur la date du jour et l'heure.
- **Les notifications sur les réseaux sociaux :** servent à afficher les messages et les chats. Toutes ces différentes fonctionnalités de la recherche miroir sont connectées à Internet via Wifi pour faciliter et activer l'utilisation du miroir intelligent pour l'utilisateur.

Le diagramme ci-dessous (figure 23) montre comment le miroir fait la perception et l'interaction dans son environnement à l'aide des capteurs et des actuateurs.

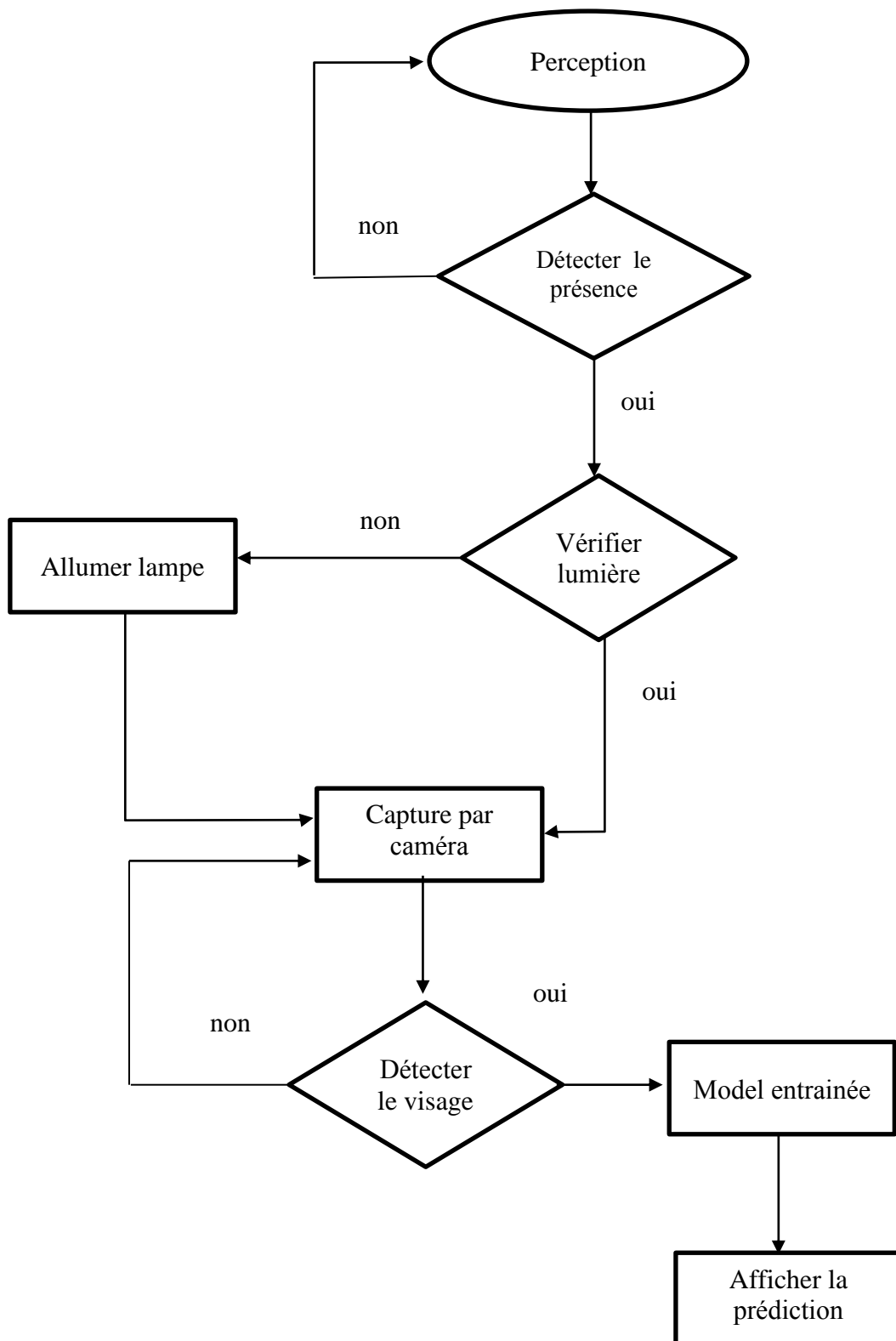


Figure 23 - Graphe de flux

7. Conclusion

Dans ce chapitre, on a parlé de la méthodologie de travail, les algorithmes de Deep Learning et leurs utilisations au développement de miroir intelligent et leurs précisions et exactitudes, également, on a présenté la classification suivie à l'apprentissage automatique et l'apprentissage profond et les étapes suivies et comment produire un modèle entraîné d'expression faciale basée sur LeNet, et aussi l'architecture du miroir intelligent.

Dans le chapitre suivant, on va aborder l'implémentation et discuter les résultats obtenus et d'autres choses plus sur le miroir et sa fonctionnalité.

Chapitre IV

Implémentation et résultats

1. Introduction

Dans les chapitres précédents on a travaillé sur le côté théorique du projet, en revanche dans ce dernier chapitre on travaille sur le côté hardware. En effet, on va créer un miroir intelligent qui peut lire l'humeur de la personne qui le regarde embarquée dans Raspberry Pi.

Un micro-ordinateur comme le Raspberry Pi n'a pas les performances suffisantes pour exécuter des algorithmes de Deep Learning, notre idée est d'utiliser un ordinateur portable, qui a les performances nécessaires d'exécuter les algorithmes de Deep Learning pour servir de microordinateur, également on va parler du hardware et expliquer l'environnement de développement puis on va présenter l'environnement hardware et par la suite on explique l'environnement software et les programmes utilisés, leurs fonctions et caractéristiques et conclure par les résultats obtenus.

2. Environnement de développement

Dans cette partie, on va utiliser les ressources software et hardware suivantes :

2.1. Environnement Hardware

2.1.1. Configuration de la machine

Sachant que le Raspberry Pi a une faible performance par rapport à un pc, nous avons décidé d'utiliser un ordinateur pour effectuer l'apprentissage.

Pour former des réseaux de neurones convolutifs, nous avons besoin d'une machine avec une unité de traitement graphique puissante. Ces unités de traitement graphique sont excellentes pour exécuter les algorithmes de Deep Learning à cause de leur architecture de traitement parallèle. En fait, de nos jours, il existe de nombreux GPU conçus spécifiquement pour l'apprentissage en profondeur. Ils sont utilisés en dehors du domaine des jeux sur ordinateurs.

Notre machine est un lap top Dell Latitude E6410 qui ne possède pas d'une carte graphique.

Les Specs de la machine sont:

- **CPU:** Intel® Core™ i5 CPU M 560 @ 2.67GHZ
- **Ram:** 4 Go
- **OS:** Ubuntu 16.04 LTS 64 bits

Pour que la machine être prête pour l'apprentissage, il faut installer les outils suivants sur l'ordinateur et le Raspberry Pi :

- Dépendances du système Ubuntu
- Python
- Compiler et Install OpenCV
- Tensorflow
- Keras

2.1.2. Le Raspberry Pi 3

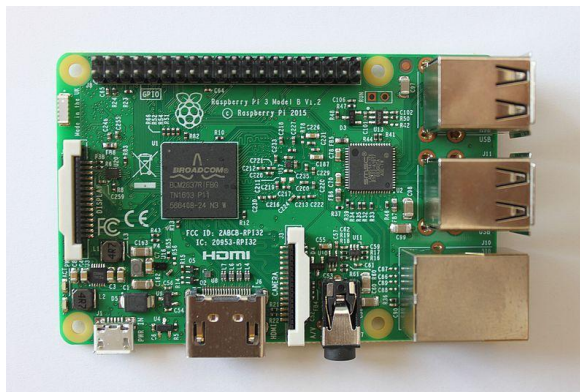


Figure 24 - Raspberry Pi 3
[WIK18d]

Le Raspberry Pi (figure 24) est un micro-ordinateur mono carte à processeur ARM (Advanced RISC Machine) conçu par le créateur de jeux vidéo David Braben, dans le cadre de sa fondation Raspberry Pi2.

Cet ordinateur, qui a la taille d'une carte de crédit, est destiné à encourager l'apprentissage de la programmation informatique, il permet l'exécution de plusieurs variantes du système d'exploitation libre GNU/Linux et des logiciels compatibles. Il est fourni nu (carte mère seule, sans boîtier, alimentation, clavier, souris ni écran) dans l'objectif de diminuer les coûts et de permettre l'utilisation de matériel de récupération [WIK18d].

Dans notre cas nous allons utiliser le Raspberry Pi 3 qui a la configuration suivante:

- Système d'exploitation Raspbian Stretch
- Quad Core Broadcom BCM2836 CPU avec 1GB de Ram
- 40 Pins GPIO externe (General-Purpose Input/Output)
- Micro SD externe comme une mémoire
- 4 ports USB, et Full Size HDMI, 4 pole Stéréo output et un Composite vidéo Port
- Micro USB pour l'alimentation.
- Interface de réseaux sans fil WiFi et Bluetooth.

2.2. Les capteurs et les modules

2.2.1. Module de détection aux ultrasons

Le capteur HC-SR04 (figure 25) utilise les ultrasons pour déterminer la distance d'un objet. Il offre une excellente plage de détection sans contact, avec des mesures de haute précision et stables. Son fonctionnement n'est pas influencé par la lumière du soleil ou des matériaux sombres, bien que des matériaux comme les vêtements puissent être difficiles à détecter [BAC15].

On a utilisé ce module come capteur de présence qui détecte la présence de l'utilisateur.



Figure 25 - Module de détection aux ultrasons (HC-SR04)

2.2.1. Photorésistance

Une photorésistance (figure 26) est un composant électronique dont la résistivité fluctue (fréquemment une diminution) selon l'augmentation de lumière qui l'atteint. On peut aussi le nommer résistance photo-dépendante (Light-Dependent Resistor (LDR)) ou photoconducteur [COM18].



Figure 26 - photorésistance

2.2.2. Convertisseur analogique-numérique (MCP3008)

Le MCP3008 (figure 27) est un convertisseur analogique / numérique 10 bits à 8 canaux. Il est facile à connecter et ne nécessite aucun composant supplémentaire. Il utilise le protocole de bus SPI (Serial Peripheral Interface) qui est pris en charge par les entrées GPIO du Raspberry Pi [RAS13].

Pour quoi le MCP3008?

Contrairement au microcontrôleur Arduino, le Raspberry Pi ne peut pas lire les entrées de capteurs analogiques, c'est la raison pour laquelle nous avons besoin du MCP3008 pour convertir ces valeurs entrées et les envoyer au Raspberry Pi.

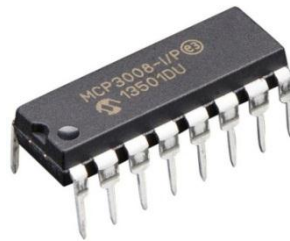


Figure 27 – Convertisseur analogique-numérique (MCP3008)

2.3. Le câblage

La figure 28 montre comment faire le câblage entre les capteurs et le Raspberry Pi.

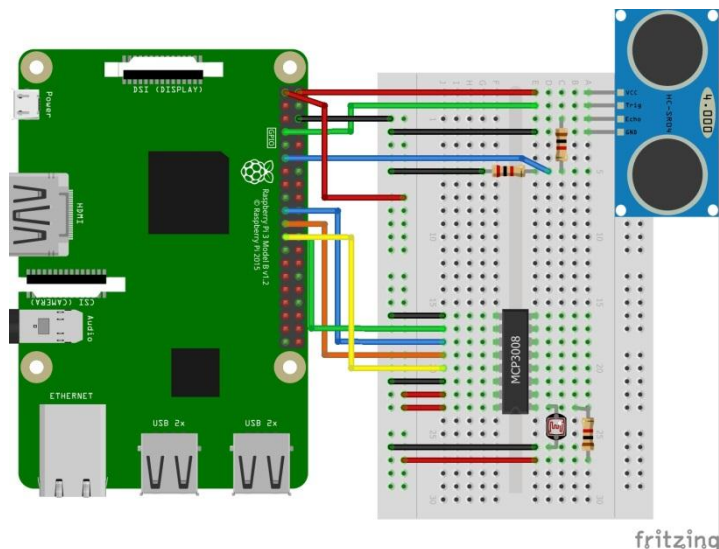


Figure 28 – le câblage entre les composants et le Raspberry Pi

2.4. Environnement software

2.4.1. Open CV

Open CV pour Open Computer Vision est une bibliothèque graphique développée sous licence BSD. BSD pour "Berkeley Software Distribution License" est une licence libre utilisée pour la distribution de logiciels. Elle permet de réutiliser tout ou une partie du logiciel sans restriction, qu'il soit intégré dans un logiciel libre ou propriétaire. Open CV est initialement développée par Intel et est spécialisée dans le traitement d'images en temps réel. Elle est développée en C, C++, Java, Androïde, Python et implémentable sur toutes ces plateformes.

2.4.1.1. Les fonctionnalités

Open CV propose diverses fonctionnalités dans les domaines des calculs matriciels, du traitement d'images, traitement vidéo, des algorithmes d'apprentissage.

2.4.1.2. Fonctionnalités obtenus dans les calculs matriciels

A l'aide de cette API, il est possible d'effectuer des opérations de :

- Calcul du déterminant d'une matrice.
- Calcul de la transposée d'une matrice.
- Calcul des valeurs propres.

2.4.1.3. Fonctionnalités obtenus pour le traitement d'image et vidéo

Parlant du traitement d'image et vidéo, nous pouvons réaliser des opérations de :

- Lecture, écriture et affichage d'une image.
- Lecture, écriture et affichage d'une vidéo (depuis un fichier ou une caméra).
- Calcul de l'histogramme des niveaux de gris ou d'histogrammes couleurs.
- Détection de visages par la méthode de Viola et Jones.
- Détection de points d'intérêts (yeux par exemples) [OPE18].

2.4.2. TensorFlow

TensorFlow est un Framework de programmation pour le calcul numérique qui a été rendu Open Source par Google en Novembre 2015. Depuis son release, Tensor Flow n’a cessé de gagner en popularité, pour devenir très rapidement l’un des Framework les plus utilisés pour le Deep Learning et donc les réseaux de neurones. Son nom est notamment inspiré du fait que les opérations courantes sur des réseaux de neurones sont principalement faites via des tables de données multidimensionnelles, appelées Tenseurs (Tensor). Un Tensor à deux dimensions est l’équivalent d’une matrice. Aujourd’hui, les principaux produits de Google sont basés sur TensorFlow: Gmail, Google Photos, Reconnaissance de voix [TEN18].

2.4.3. Keras

Keras est une API de réseaux de neurones de haut niveau, écrite en Python et capable de fonctionner sur TensorFlow ou Theano. Il a été développé en mettant l’accent sur l’expérimentation rapide. Être capable d’aller de l’idée à un résultat avec le moins de délai possible est la clé pour faire de bonnes recherches. Il a été développé dans le cadre de l’effort de recherche du projet ONEIROS (Open-ended Neuro-Electronic Intelligent Robot Operating System), et son principal auteur et mainteneur est François Chollet, un ingénieur Google.

En 2017, l’équipe TensorFlow de Google a décidé de soutenir Keras dans la bibliothèque principale de TensorFlow. Chollet a expliqué que Keras a été conçue comme une interface plutôt que comme un cadre d’apprentissage end to end. Il présente un ensemble d’abstractions de niveau supérieur et plus intuitif qui facilitent la configuration des réseaux neuronaux indépendamment de la bibliothèque informatique de *Backend*. Microsoft travaille également à ajouter un *Backend CNTK* à Keras aussi [KER18].

2.4.4. Python

Python est un langage de programmation de haut niveau interprété (il n’y a pas d’étape de compilation) et orienté objet avec une sémantique dynamique. Il est très sollicité par une large communauté de développeurs et de programmeurs. Python est un langage simple, facile à apprendre et permet une bonne réduction du cout de la maintenance des codes. Les bibliothèques (packages) python encouragent la modularité et la réutilisabilité des codes. Python et ses bibliothèques sont disponibles (en source ou en binaires) sans charges pour la majorité des plateformes et peuvent être redistribués gratuitement [PYT18].

3. L'apprentissage de classificateur d'images

3.1. Préparation des données pour l'apprentissage (l'entraînement)

L'ensemble de données que nous avons utilisé pour l'entraînement du modèle provient d'un défi de Reconnaissance d'Expression Faciale *Kaggle.com* dans l'année 2013 (FER2013-*"Facial Expression Recognition"*). Il comprend un total de 35887 images en niveaux de gris, pré-recadrées de 48x48 pixels de visages chacun étiquetés avec l'une des 7 classes d'émotion suivantes:

- 0: 4593 images → **colère**
- 1: 547 images → **dégoût**
- 2: 5121 images → **peur**
- 3: 8989 images → **heureuse**
- 4: 6077 images → **triste**
- 5: 4002 images → **surprise**
- 6: 6198 images → **neutre**

Sachant que nous n'avons pas d'ordinateur puissant pour former notre réseau convolutif, nous avons décidé de poursuivre l'apprentissage avec six expressions: colère, peur, heureuse, triste, surprise et neutre, avec toutes les images sauf le dégoût, par ce qu'il y a un petit nombre des images comparent avec les autres expressions.

L'ensemble de données contient les images sous forme de chaînes de valeurs de niveaux de gris, nous les avons séparées par expression, et nous avons utilisé un programme Python pour extraire ces images et les sauvegarder au format ".jpg" pour la phase d'entraînement. Les images sont enregistrées dans six dossiers différents nommés *Angry*, *Fear*, *Happy*, *Neutral*, *Sad*, *Surprise*, (voir les figures 29, 30, 31).

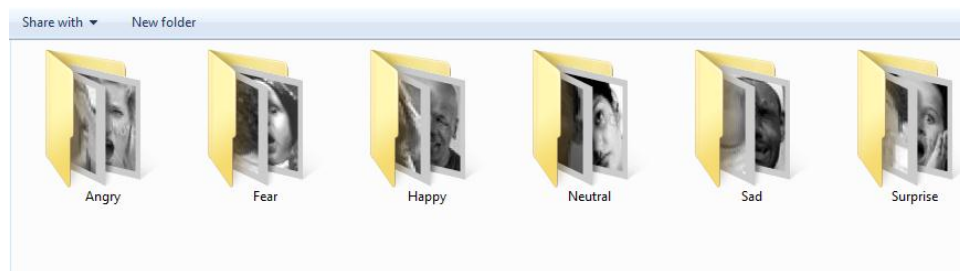


Figure 29 – Les dossiers des images

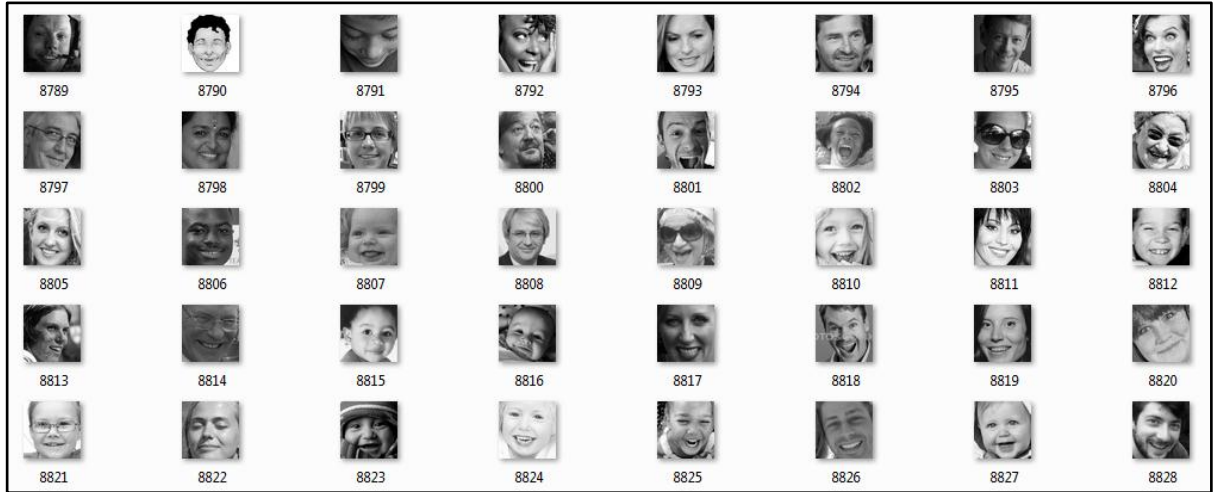


Figure 30 – un ensemble des images représentent l'expression heureuse (Happy)

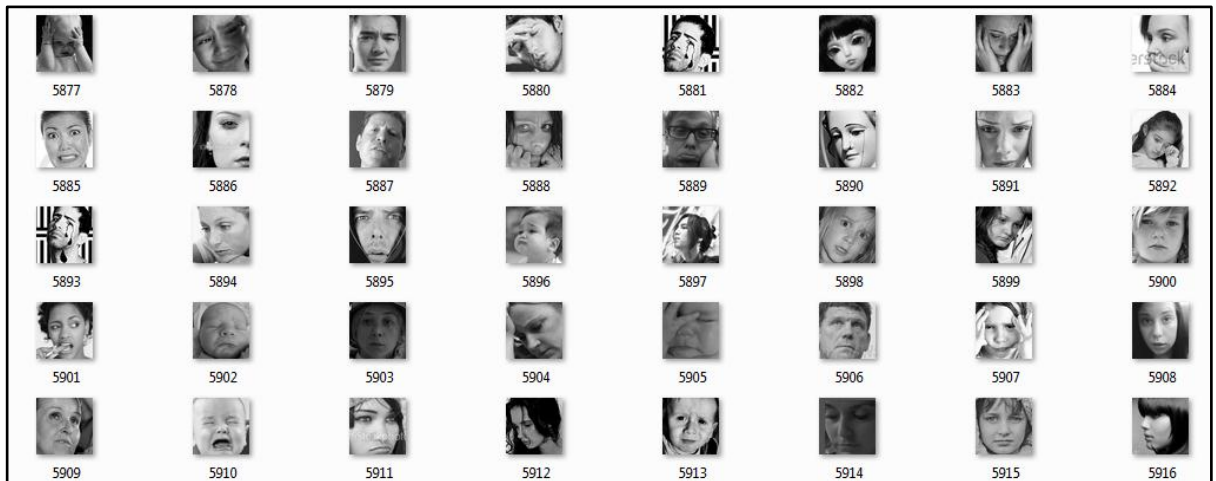


Figure 31 – un ensemble des images représentent l'expression triste (Sad)

3.2. L'apprentissage

Après la préparation des données, on a procédé à l'apprentissage en utilisant l'architecture CNN simple de LeNet, et on a utilisé une classe avec le nom *lenet.py* qui implémente l'architecture LeNet du CNN.

3.2.1. L'apprentissage de classificateur d'images de réseau neuronal convolutionnel avec Keras et Tensorflow

Nous commençons à former le classificateur d'images en utilisant l'apprentissage profond, avec Keras et Tensorflow à l'aide du Python en suivant les étapes suivantes:

1. Chargez l'ensemble de données d'image à partir du disque

On commence par initialiser les listes de données et d'étiquettes. Ces listes seront chargées les images à partir du disque avec leurs étiquettes de classe correspond. La classification est basée sur les noms de dossier, les images du dossier Happy seront classées comme Happy, et la même chose pour les autres.

Cette méthode de classification nous permet d'organiser efficacement l'ensemble de données et d'analyser les étiquettes de classes facilement.

2. Instancier le réseau de neurones convolutif

Nous créons un modèle Séquentiel en transmettant une liste d'instances de couche au constructeur. Nous effectuons ensuite un partage d'apprentissage et tests sur les données en utilisant 75% des images pour l'apprentissage et 25% pour les tests.

Maintenant, nous commençons le processus de l'apprentissage en appelant la classe *lenet.py* en suivant l'algorithme suivant:

Algorithme de l'apprentissage:

Début

Chargement de la liste des dossiers des images;

Pour chaque Dossier *faire*

Pour chaque image dans le dossier *faire*

 Lire(image) ;

 Image.étiquette \leftarrow nom_de_dossier;

 Appliquer le LeNet;

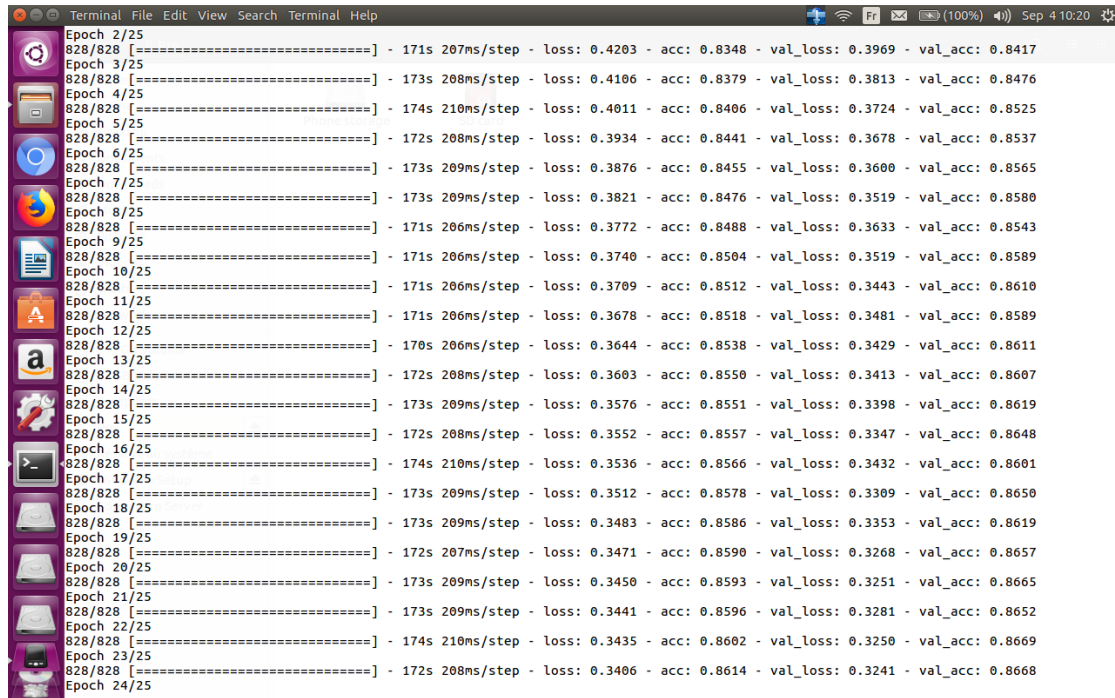
 Créer le Model de classificateur;

Fin pour

Fin

3.3. Résultat de l'apprentissage

Sachant que nous n'avons pas utilisé une unité de traitement graphique, le temps de l'apprentissage a été si longtemps et a pris plus de 71 minutes pour terminer le processus de l'apprentissage (figure 32).



Epoch	Time	Step	loss	acc	val_loss	val_acc
Epoch 2/25	828/828	=====	0.4203	0.8348	0.3969	0.8417
Epoch 3/25	828/828	=====	0.4106	0.8379	0.3813	0.8476
Epoch 4/25	828/828	=====	0.4011	0.8406	0.3724	0.8525
Epoch 5/25	828/828	=====	0.3934	0.8441	0.3678	0.8537
Epoch 6/25	828/828	=====	0.3876	0.8455	0.3600	0.8565
Epoch 7/25	828/828	=====	0.3821	0.8476	0.3519	0.8580
Epoch 8/25	828/828	=====	0.3772	0.8488	0.3633	0.8543
Epoch 9/25	828/828	=====	0.3740	0.8504	0.3519	0.8589
Epoch 10/25	828/828	=====	0.3709	0.8512	0.3443	0.8610
Epoch 11/25	828/828	=====	0.3678	0.8518	0.3481	0.8589
Epoch 12/25	828/828	=====	0.3644	0.8538	0.3429	0.8611
Epoch 13/25	828/828	=====	0.3603	0.8550	0.3413	0.8607
Epoch 14/25	828/828	=====	0.3576	0.8551	0.3398	0.8619
Epoch 15/25	828/828	=====	0.3552	0.8557	0.3347	0.8648
Epoch 16/25	828/828	=====	0.3536	0.8566	0.3432	0.8601
Epoch 17/25	828/828	=====	0.3512	0.8578	0.3309	0.8650
Epoch 18/25	828/828	=====	0.3483	0.8586	0.3353	0.8619
Epoch 19/25	828/828	=====	0.3471	0.8590	0.3268	0.8657
Epoch 20/25	828/828	=====	0.3450	0.8593	0.3251	0.8665
Epoch 21/25	828/828	=====	0.3441	0.8596	0.3281	0.8652
Epoch 22/25	828/828	=====	0.3435	0.8602	0.3250	0.8669
Epoch 23/25	828/828	=====	0.3406	0.8614	0.3241	0.8668
Epoch 24/25	828/828	=====				

Figure 32 – Le processus de l'apprentissage

On peut voir que le réseau s'est entraîné pendant 25 époques et nous avons atteint une grande précision (86,69%) et une faible perte qui suit la perte d'entraînement, comme le montre le tableau ci-après.

Epoch	perte d'apprentissage	précision d'apprentissage	valeur perdu	valeur de précision
1	0.4349	0.8331	0.4127	0.8346
2	0.4203	0.8348	0.3969	0.8417
3	0.4106	0.8379	0.3813	0.8476
4	0.4011	0.8406	0.3724	0.8525
5	0.3934	0.8441	0.3678	0.8537
6	0.3876	0.8455	0.3600	0.8565
7	0.3821	0.8476	0.3519	0.8580
8	0.3772	0.8488	0.3633	0.8543
9	0.3740	0.8504	0.3519	0.8589
10	0.3709	0.8512	0.3443	0.8610
11	0.3678	0.8518	0.3481	0.8589
12	0.3644	0.8538	0.3429	0.8611
13	0.3603	0.8550	0.3413	0.8607
14	0.3576	0.8551	0.3398	0.8619
15	0.3552	0.8557	0.3347	0.8648
16	0.3536	0.8566	0.3432	0.8601
17	0.3512	0.8578	0.3309	0.8650
18	0.3483	0.8586	0.3353	0.8619
19	0.3471	0.8590	0.3268	0.8657
20	0.3450	0.8593	0.3251	0.8665
21	0.3441	0.8596	0.3281	0.8652
22	0.3435	0.8602	0.3250	0.8669
23	0.3406	0.8614	0.3241	0.8668
24	0.3386	0.8622	0.3311	0.8644
25	0.3387	0.8622	0.3270	0.8663

Tableau 1 – Résultats de l'apprentissage

Après l'analyse des résultats obtenus, On constate les remarques suivantes :

D'après les graphes ci-dessous (figure 33) La précision de l'apprentissage et de test augmente avec le nombre d'époque, ceci reflète qu'à chaque époque le modèle apprend plus d'informations. Si la précision est diminuée alors on aura besoin de plus d'information pour faire apprendre notre modèle et par conséquent on doit augmenter le nombre d'époque et vice versa.

De même, l'erreur d'apprentissage et de la validation diminue avec le nombre d'époque.

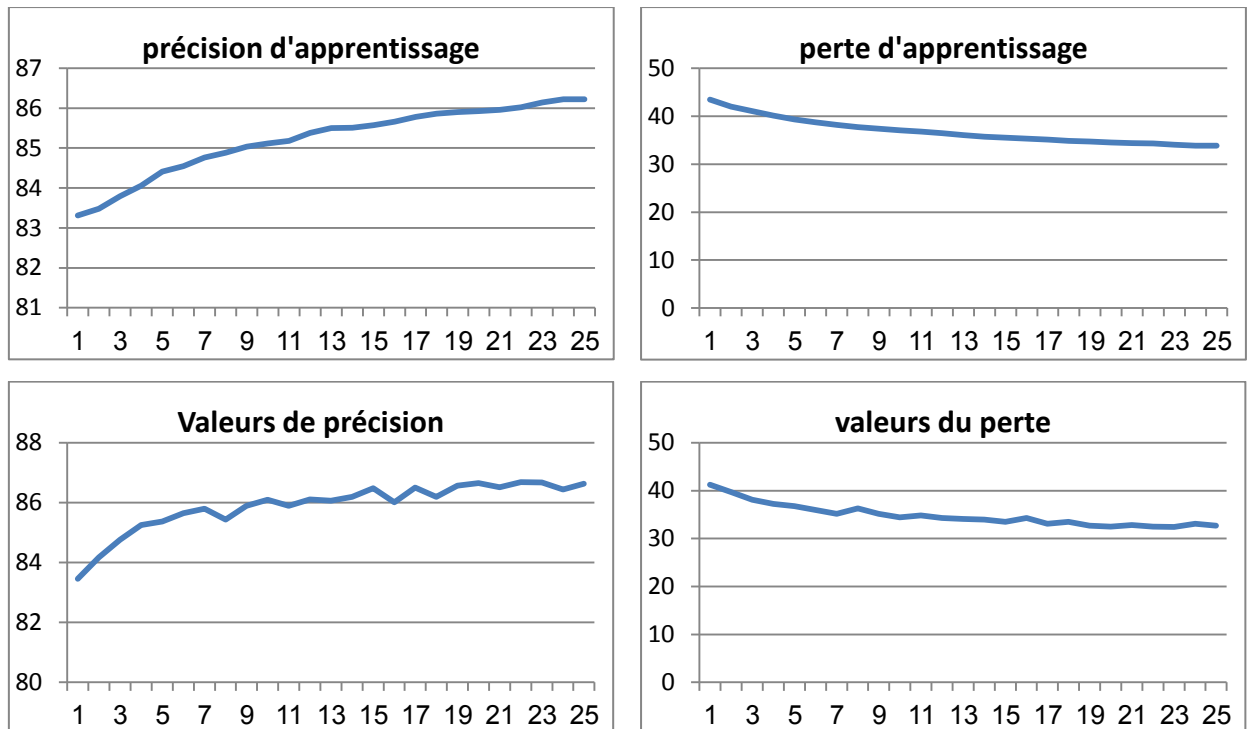


Figure 33 – Les graphes du résultats

4. Évaluation de notre classificateur d'images

L'étape suivante consiste à évaluer notre classificateur, s'il fonctionne bien ou non, on va utiliser une webcam pour capturer nos visages en émettons les expressions faciale qu'on a utilisé pendant la phase d'apprentissage et on va analyser les résultats obtenus après les tests.

Nous avons utilisé un programme Python pour tester notre classificateur.

L'algorithme du test:

Début

```

Acquisition de l'image;
Détecter le visage;
Convertir l'image en niveau de gris;
Redimensionné l'image;
classificateur ← charger_classificateur ;
valeur (colère, peur, heureux, triste, surprise, neutre) ← classificateur.prediction(image) ;
Prédiction ← Max(valeur);





```

Fin

4.1. Les tests

4.1.1. Test en direct avec la webcam

Le tableau N° 2 ci-dessous, montre quelques tests et leurs résultats en utilisant notre classificateur d'expressions faciales :

N°	Tests	Résultats	Observation
1.	Neutre	 Neutre 74% de précision	Positif
2.	Heureux	 Heureux 72% de précision	Positif
3.	Heureux	 Heureux 32% de précision	Positif
4.	Heureux	 Peur 26% de précision	Négatif






La suite de tableau N° 2			
N°	Tests	Résultats	Observation
5.	En colère	 <p>Heureux 64.32% de précision</p>	Négatif
6.	En colère	 <p>En colère 41.83% de précision</p>	Positif
7.	Peur	 <p>Peur 42.04% de précision</p>	Positif
8.	Triste	 <p>Peur 42 % de précision</p>	Négatif
9.	Surpris	 <p>Surpris 60 % de précision</p>	Positif

Tableau 2 – Tests et résultats avec webcam

Examinons de plus près les prévisions relatives aux émotions individuelles, le tableau numéro 02 ci-dessus montre les résultats obtenus du classificateur, sur l'ensemble de test en direct sur la webcam, le tableau donne le nombre de prédictions d'émotion et quelques indications sur le modèle de classification des expressions faciales.

Le modèle fonctionne très bien sur la classification des émotions dont les 09 essais qu'on a fait 06 positives et 03 négatifs, ce qui se traduit par des scores de précision relativement élevés pour l'heureuse et la neutre et la peur et la colère puisque l'utilisateur ne peut pas rester devant la webcam une durée très longue, ça donne une modification et une variation de changement des muscles de visage suivi de changement de pose, et changement de l'éclairage de la pièce, alors on a décidé de faire une autre série de tests à partir des images enregistrées sur le disque.

4.1.2. Test à partir des images enregistrées sur disque

Pour avoir une bonne vue sur les résultats obtenus à partir des tests, on a décidé d'utiliser des images enregistrées sur le disque (Figure 34) et obtenu aussi toutes les valeurs de chaque expression pour chaque image testée, pour nous aider à analyser et comparer les résultats obtenus.

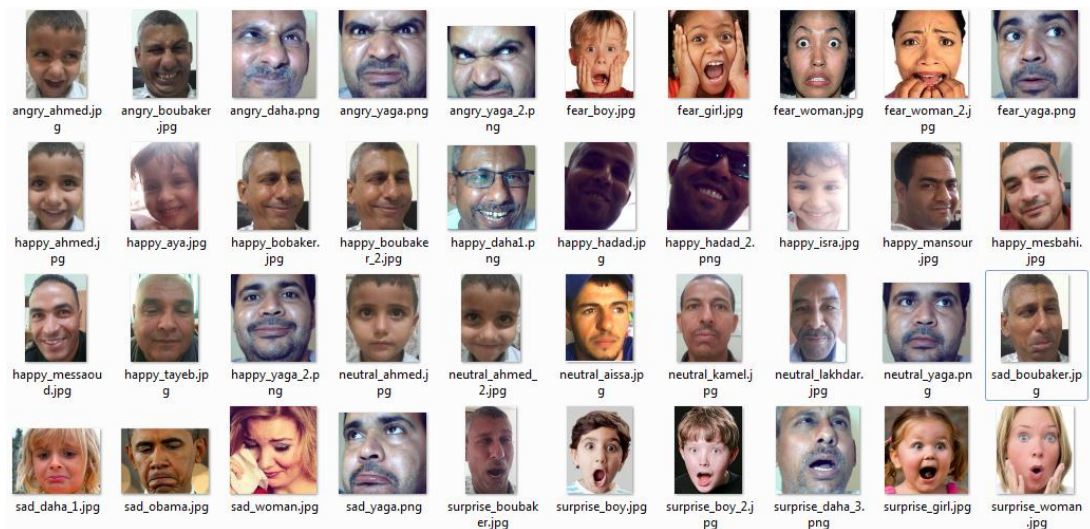


Figure 34 – l'ensemble des images enregistrées sur le disque

On a préparé un ensemble des images des visages de nous-même et aussi nos amis et enfants, et quelques images des visages de l'internet.

Le tableau suivant, montre les résultats obtenus des tests.

N°	test	Colère	Heureux	Neutre	Triste	Surprise	Peur	Résultats	Précision	obs
1.	Colère	3.85	71.76	10.95	6.98	1.56	4.89	Heureux	71.76%	négatif
2.	Colère	18.25	58.36	14.55	4.16	1.07	3.61	Heureux	58.36%	négatif
3.	Colère	23.63	12.11	14.87	16.8	10.72	21.87	Colère	23.63%	positif
4.	Colère	29.19	33.98	20.53	5.73	2.17	8.39	Heureux	33.98%	négatif
5.	Colère	48.31	3.58	9.75	13.02	3.57	21.77	Colère	48.31%	positif
6.	Peur	45.34	7.35	10.16	8.05	3.37	25.74	Colère	45.34%	négatif
7.	Peur	14.15	27.55	0.8	1.29	22.97	33.24	Peur	33.24%	positif
8.	Peur	0.95	12	1.54	0.85	53.81	30.86	Surprise	53.81%	négatif
9.	Peur	0.08	0.02	0.15	0.04	2.34	97.37	Peur	97.37%	positif
10.	Peur	7.92	3.96	6.81	6.1	18.32	56.9	Peur	56.90%	positif
11.	Heureux	0.06	98.22	1.01	0.13	0.4	0.18	Heureux	98.22%	positif
12.	Heureux	4.2	69.46	13.57	8.54	0.32	3.91	Heureux	69.46%	positif
13.	Heureux	11	30.24	38.14	9.87	1.28	9.47	Neutre	38.14%	négatif
14.	Heureux	10.54	16.62	49.36	15.31	0.89	7.29	Neutre	49.36%	négatif
15.	Heureux	4.48	78.63	5.31	1.48	1.98	8.12	Heureux	78.63%	positif
16.	Neutre	1.73	4.16	67.49	9.97	5.05	11.59	Neutre	67.49%	positif
17.	Neutre	2.11	76.44	9.78	7.3	0.56	3.81	Heureux	76.44%	négatif
18.	Neutre	8.32	6.64	42.39	15.26	14.09	13.29	Neutre	42.39%	positif
19.	Neutre	7.91	1.79	80.22	4.67	0.47	4.94	Neutre	80.22%	positif
20.	Neutre	16.61	6.23	22.6	6.63	14.45	33.48	Peur	33.48%	négatif
21.	Triste	19.33	13.95	2.17	3.23	33.67	27.65	Surprise	33.67%	négatif
22.	Triste	28.16	5.74	42.57	13.1	0.35	10.08	Neutre	42.57%	négatif
23.	Triste	25.85	1.57	18.97	35.5	0.79	17.33	Triste	35.50%	positif
24.	Triste	21.65	4	20.38	33.19	0.46	20.32	Triste	33.19%	positif
25.	Triste	12.05	8.32	30.75	35.66	0.96	12.27	Triste	35.66%	positif
26.	Surpris	13.89	8.43	4.7	7.45	30.01	35.53	Peur	35.53%	négatif
27.	Surpris	28.86	4.17	14	12.5	10.33	30.13	Peur	30.13%	négatif
28.	Surpris	12.11	43.33	2.55	6.44	12.21	23.36	Heureux	43.33%	négatif
29.	Surpris	19.08	17.3	0.16	3.02	30	30.44	Peur	30.44%	négatif
30.	Surpris	0.31	0.11	0.05	0.06	85.01	14.46	Surpris	85.01%	positif

Tableau 3 – Résultats des tests sur images enregistré sur disque

Examinons de plus près les prévisions relatives aux émotions individuelles, le tableau 3 ci-dessus montre les résultats obtenus après le test sur le classificateur. On a utilisé cinq images pour chaque expression pour le test.

Le tableau donne les valeurs des prédictions d'émotion.

On peut voir que notre classificateur fonctionne très bien sur la classification les expressions: heureux, peur et neutre, avec des bonnes précisions 98.22%, 97.37% et 80.22%.

Mais par contre, le classificateur semble faible pour les expressions: colère et triste, avec des faibles précisions 48.31% et 35.66%.

Mais pour l'expression de surpris, la plupart des prédictions ont été mal prédit comme peur, à cause de la ressemblance de ces expressions faciales.

5. Le fonctionnement du miroir

Le miroir intelligent est fabriqué avec un cadre en bois (figure 35) qui tient un verre à sens unique qui reflète les objets environnants comme un miroir et peut afficher le contenu de l'écran de l'ordinateur en dessous.

Il y a aussi la webcam, le détecteur de lumière et le capteur de présence.



Figure 35 – une vue sur le miroir

Le miroir affichera:

- l'heure et la date actuelles
- les informations de la météo
- fil d'actualités

Lorsque l'utilisateur est devant le miroir, le miroir va détecter son présence à l'aide du capteur de présence et quitte le mode économiseur d'écran, après il va vérifier la luminosité de la pièce, si la lumière est appropriée ou non pour prendre une photo du visage le contrôle de la lumière va allumer ou éteindre les lampes associées au miroir selon l'état de lumière, la caméra va démarrer la capture de visage de l'utilisateur et vérifier son expression faciale, et selon son expression, il imprimera une citation correspondante.

6. Comment se fait l'interaction du miroir?

Lorsque la Compagnie Apple a voulu rendre l'iPhone plus intelligent que les autres téléphones, elle a créé Siri, un assistant intelligent pour l'iPhone. Cet assistant répondra aux questions des utilisateurs et faire des appels téléphoniques, régler l'alarme et envoyer des courriels, prendre des notes, et faire des recherches sur Internet C'était une grande évolution pour la compagnie Apple à cette époque.

Notre but de cette étude était de créer un miroir intelligent qui interagit avec l'utilisateur selon son expression faciale. Jusqu'à présent, nous avons pu créer un classificateur capable de reconnaître l'expression faciale de l'utilisateur.

Mais pour que le miroir interagisse intelligemment avec l'utilisateur, il doit pouvoir communiquer avec lui, en utilisant des commandes vocales ou un écran tactile, et aussi il devrait être capable de répondre aux questions des utilisateurs avec une voix. Ainsi, l'utilisation de fonctionnalités de synthèse vocale (Text to speech) et de reconnaissance automatique de la parole (Speech to text) va le rendre plus intelligent par rapport à un miroir normal.

Donc, pour que nous puissions créer un assistant capable de parler à l'utilisateur, nous devons créer un Chatbot pour cela. Et pour y arriver, nous avons besoin d'un grand ensemble de dialogues pour pouvoir former l'assistant miroir et faire parler le miroir à l'utilisateur en se basant sur son expression faciale.

Malheureusement, nous ne pouvions pas atteindre cet objectif, car nous n'avions ni le temps ni l'expérience pour cela, nous avons passé beaucoup de temps à essayer de comprendre les nouveaux concepts pour créer le classificateur des expressions faciales. Mais nous avons donné au miroir une capacité minimale d'interaction avec l'utilisateur.

7. Conclusion

Dans ce dernier chapitre on a projeté la lumière sur l'environnement de travail, et les composants électroniques qu'on va utiliser avec le Raspberry Pi pour réaliser le miroir intelligent. On a créé un classificateur d'expressions faciales à l'aide de Deep learning avec une architecture CNN simple "LeNet", qu'il va être utilisé par le miroir. On a analysé les résultats des tests sur ce classificateur qu'on a fait en direct de la webcam et sur des images enregistrés sur le disque, et on a parlé du fonctionnement de notre miroir intelligent.

Conclusion générale

Conclusion générale

L'objectif principal de notre travail est de réaliser un miroir intelligent interactif basé sur le Raspberry Pi, qui peut détecter l'humeur de la personne qui le regarde et réagit en fonction du résultat obtenu. Les autres objectifs sont de faire le miroir interagit selon l'environnement où se trouve, de minimiser la consommation d'énergie par le contrôle de la lumière.

Le miroir utilise un classificateur d'expressions facial créé à l'aide de Deep learning avec une architecture CNN simple "LeNet", on a choisi cette architecture parce qu'elle est plus simple et n'a pas besoin du matériel très puissant pour se servir, sachant que le Raspberry Pi n'est pas un micro-ordinateur puissant. Le miroir se compose de quelques composants électroniques (capteurs et actionneurs) qui les aides pour avoir un bon fonctionnement.

Ce miroir est capable de :

- Afficher l'horloge et la date actuel.
- Afficher les informations de la météo.
- Interagit selon l'humeur de l'utilisateur (à partir de son expression facial).
- Afficher des infos et des actualités.
- Contrôler la lumière.
- Détecter la présence de l'utilisateur.

En parcourant les différents chapitres, nous avons décrit et clarifié l'architecture du système, les limitations rencontrées, et les objectifs atteints.

Ce miroir conviendra parfaitement aux propriétaires des salons de coiffure et de maquillage pour leur permettre de voir si leurs clients sont satisfaits de leur service ou non, sans demander, juste à partir de leur expression faciale, ou bien pour l'utilisation personnelle.

Perspectives

Perspectives

Dans le futur, nous allons essayer de rendre le miroir plus intelligent en ajoutant un assistant intelligent basé sur le Chatbot qui peut faire des dialogues avec l'utilisateur, et ce sera une bonne idée de créer un service de Cloud qui pourra apprendre et évoluer à partir de l'expérience de l'interaction entre le miroir et les utilisateurs, et aussi pour l'assistant de miroir de pouvoir prédire la raison pour laquelle l'utilisateur est heureux ou en colère ou triste par ses interactions et ses activités dans les réseaux sociaux.

Bibliographie

Bibliographie

- [ABA08] SOUHILA GUERFI ABABSA. Authentification d'individus par reconnaissance de caractéristiques biométriques liées aux visages 2D/3D. PhD thesis, Université D'EVRY VAL D'ESSONNE, 03 Octobre 2008.
- [ADI97] Y. Adini, Y. Moses, S. Ullman, Face recognition: The problem of compensating for changes in illumination direction. *IEEE Trans. Patt. Anal. Mach. Intell.* 19, 721–732, 1997.
- [BAC15] Lucien Bachelard - *HC-SR04 - Module de détection aux ultrasons - Utilisation avec Picaxe*. **[En ligne] disponible sur:** <https://www.gotronic.fr/pj2-hc-sr04-utilisation-avec-picaxe-1343.pdf>
Publier le: 28 novembre 2015.
- [BEL97] P. N. Belhumeur, J. P. Hespanha, and D. J. Kriegman. Eigenfaces vs. Fisherfaces : Recognition using class specific linear projection. *IEEE Trans. PAMI*, 1997.
- [BLO17] The Blog – Meet Perseus: The World's First Smart Mirror
[En ligne] disponible sur: https://www.huffingtonpost.com/harold-stark/meet-perseus-the-worlds-f_b_11546776.html/ **éditer par:** Harold Stark, le 06/12/2017
Accès: Février 2018
- [CLU15] Clubic – Deeplearning : quand l'intelligence artificielle tente d'imiter le cerveau humain, **[En ligne] disponible sur:**
<https://www.clubic.com/technologies-d-avenir/intelligence-artificielle/deep-learning/article-783038-2-promesses-deep-learning.html>
publier par: Jérôme Cartegini, **le:** 24/11/2015
Accès: Mars 2018
- [COL15] Sara Colantonio, Giuseppe Coppini, Danila Germanese, Daniela Giorgi, Massimo Magrini, Paolo Marraccini, Massimo Martinelli, Maria Aurora Morales, Maria Antonietta Pascali, Giovanni Raccichini, Marco Righi, Ovidio Salvetti, A smart mirror to promote a healthy lifestyle National Research Council of Italy, Pisa, Italy
Published by Elsevier Ltd. 2015
- [COM18] Composant électronique – Photorésistance
[En ligne] disponible sur: <http://www.composelec.com/photoresistance.php>
Accès: Mars 2018
- [DAI16] Dailymail – Microsoft reveals magic mirror that knows how you're feeling – and can show you everything from your emails to the weather in the morning
[En ligne] disponible sur: <http://www.dailymail.co.uk/sciencetech/article-3620487/Microsoft-reveals-magic-mirror-knows-feeling-emails-weather-morning.html>
publier par: Stacy Liberatore, **le:** 02/06/2016
Accès: Février 2018
- [EKM76] Ekman P, Friesen WV, Press CP: Pictures of Facial Affect. Consulting psychologists press; 1976.

- [ENG14] Engadget – Panasonic's smart mirror turned me into a 1970s space princess
[En ligne] disponible sur: <https://www.engadget.com/2014/09/06/panasonic-smart-mirror-hands-on/> **publier par:** Dana Wollman, **le:** 06/09/2014
Accès: Mars 2018
- [ENG16] Engadget – Panasonic's smart mirrors gave me a makeover [En ligne] disponible sur: <https://www.engadget.com/2016/01/10/panasonics-smart-mirrors-gave-me-a-makeover/> **publier par:** Nicole Lee, **le:** 10/01/2016
Accès: Mars 2018
- [GHA10] KHADOUDJA GHANEM, Reconnaissance des Expressions Faciales à Base d'Informations Vidéo ; Estimation de l'Intensité des Expressions Faciales, **Soutenu le :** 11/10/2010, Université Mentouri de Constantine.
- [GRE01] H. Greenspan, J. Goldberger, I. Eshet. Mixture model for face color modeling and segmentation, *Pattern Recognition Lett.* 22 (14) 1525–1536, 2001.
- [KER18] Keras documentation – [En ligne] **disponible sur:** <https://keras.io/#keras-the-python-deep-learning-library>
Accès: Mars 2018
- [KRI07] David Kriesel, A Brief Introduction to Neural Networks : 2007
- [LAP07] LAPIERRE Lauréline – Construction d'une échelle d'évaluation des capacités de reconnaissance des expressions faciales émotionnelles, Université Paul Sabatier Juin 2007.
- [LIE02] Rainer Lienhart and Jochen Maydt. An Extended Set of Haar-like Features for Rapid Object Detection. *IEEE ICIP* 2002
- [MAR01] A. M. Martinez and A. C. Kak. Pca versus lda. *IEEE Trans. PAMI*, 23(2) :228-233, 2001
- [MAR98] A.M Martinez, R. Banavente. The AR face database. Tech. Report 24 CVC Barcelone, Espagne, June 1998.
- [MAT88] Matsumoto D, Ekman P: Japanese and Caucasian facial expressions of emotion and neutral faces (JACFEE and JACNeuF). *Hum Interact Lab UnivCalif San Franc* 1988, 401.
- [MCC43] W.S. McCulloch and W. Pitts. A logical calculus of the ideas immanent in nervous activity. *Bulletin of Mathematical Biology*, 5(4):115–133, 1943
- [MED17a] Medium – Deep Learning for Computer Vision for the average person,
[En ligne] disponible sur: <https://medium.com/diaryofawannapreneur/deep-learning-for-computer-vision-for-the-average-person-861661d8aa61>
publier par: Aashay Sachdeva, **le:** 06/03/2017
Accès: Mars 2018
- [MED17b] Medium – Comment les Réseaux de neurones à convolution fonctionnent
Traduction de l'article original de Brandon Rohrer (Senior Data Scientist à Facebook), "How Convolutional Neural Networks Work"
[En ligne] disponible sur: https://medium.com/@siddharthdas_32104/cnns-architectures-lenet-alexnet-vgg-googlenet-resnet-and-more-666091488df5
Publier par: Siddharth Das, **le:** 16/11/2017
Accès: Mars 2018
- [MEH96] Mehrabian, A. Pleasure-arousal-dominance: A general framework for describing and measuring individual differences in temperament. *Current Psychology*, 14 (1996),261-292.

- [MHY41] Ming-Hsuan, Yang, Kriegman David, J, and Ahuja Narendra. Detecting faces in images : A survey. Dans IEEE Transactions on Pattern Analysis and Machine Intelligence, pages 34–58, Volume 24(1).
- [MIN69] M. Minsky and S. Papert. Perceptrons. MIT Press, Cambridge, Mass, 1969
- [MOK17] Mokri Mohammed Zakaria, Classification des images avec les réseaux de neurones Convolutionnels,
(Soutenu le : 03/07/2017)
- [MOR09] Nicolas MORIZET. "Reconnaissance Biométrique par Fusion Multimodale du Visage et de l'Iris". 2009
- [MOU16] Adil Moudjahid – A Practical Introduction to Deep Learning with Caffe and Python **[En ligne] disponible sur:**
<http://adilmoudjahid.com/posts/2016/06/introduction-deep-learning-python-caffe/>
Publier le: 26/06/2016
Accès: Mars 2018
- [OMK15] Deep Face Recognition Omkar M. Parkhi, Andrea Vedaldi, Andrew Zisserman BMVC 2015
- [OPE18] OPencv – **[En ligne] disponible sur:** <https://opencv.org/>
Accès: Mars 2018
- [OUA12] OUAMANE Hanane: Identification de reconnaissance faciale avec des expressions (Soutenu le : 07 Juin 2012)
- [PAU01] Paul Viola et Michael Jones, « Rapid Object Detection using a Boosted Cascade of Simple Features », IEEE CVPR, 2001
- [PHI03] P. J. Phillips, P. J. Grother, R. J. Micheals, D. M. Blackbrun, E. Tabassi, and J. M. Bone. Face recognition vendor test 2002: Evaluation report. NISTIR 6965. Available En ligne at <http://www.frvt.org>, 2003.
- [PYT18] Python documentation – **[En ligne] disponible sur:** <https://www.python.org/doc/>
Accès: Mars 2018
- [QUA16] Quartz – A Google engineer invented this smart mirror that tells you the news while you admire your looks **[En ligne] disponible sur:** <https://qz.com/609042/a-google-engineer-invented-this-smart-mirror-that-tells-you-the-news-while-you-admire-your-looks/> **publier par:** Adam Epstein, le 03/02/2016
Accès: Février 2018
- [RAS13] Raspberry Pi Spy - Analogue Sensors On The Raspberry Pi Using An MCP3008 **[En ligne] disponible sur:** <https://www.raspberrypi-spy.co.uk/2013/10/analogue-sensors-on-the-raspberry-pi-using-an-mcp3008/>
Publier le: 20/10/2013 , **Accès:** Mars 2018
- [RUM86] D. Rumelhart, G. Hinton, and R. Williams. Learning representations by back propagating errors. Nature, 323:533–536, October 1986.
- [SOU08] Souhila Guerfi Ababsa –Authentification d'individus par reconnaissance de caractéristiques biométriques liées aux visages 2D/3D – Docteur de l'Université Evry Val d'Essonne Spécialité-03 octobre 2008.
- [TEN18] Tensorflow – **[En ligne] disponible sur:** <https:// tensorflow.org>
Accès: Mars 2018

- [TER03] Sylvain Tertois. Réduction des effets des non-linéarités dans une modulation multiporteuse à l'aire de réseaux de neurones. Domain_other. Université Rennes 1, 2003. Français. <tel-00004015>
- [TOU92] Claude Touzet. Les reseaux de neurones artificiels, introduction au connexionnisme: cours, exercices et travaux pratiques. Ec2, 1992, collection de l'EERIE, N. Giambiasi. <hal-01338010>
- [TUR91] M. A. Turk and A. P. Pentland, Face Recognition using Eigenfaces, Proc. IEEE, 1991, 586-591.
- [WIK15b] Wikipedia – Reconnaissance de visage [En ligne] disponible sur: https://fr.wikipedia.org/wiki/Reconnaissance_de_visage.
Accès: Mars 2018
- [WIK18a] Wikipedia – Réseau de neurones artificiels [En ligne] disponible sur: https://fr.wikipedia.org/wiki/Réseau_de_neurones_artificiels
Accès: Mars 2018
- [WIK18b] Wikipedia – Apprentissage profond [En ligne] disponible sur: https://fr.wikipedia.org/wiki/Deep_learning
Accès: Mars 2018
- [WIK18c] Wikipedia – Caractéristiques pseudo-Haar [En ligne] disponible sur: https://fr.wikipedia.org/wiki/Caractéristiques_pseudo-Haar
Accès: Mars 2018
- [WIK18d] Wikipedia – Raspberry Pi [En ligne] disponible sur: https://fr.wikipedia.org/wiki/Raspberry_Pi
Accès: Mars 2018
- [WIL06] Wilson, J. M., Fernandez, M., & Hadaway, N. (2006). Mathematical problem solving. URL. <http://jwilson.coe.uga.edu/emt725/PSsyn/PSsyn.html>
- [ZHO14] B. Zhou, A. Lapedriza, J. Xiao, A. Torralba, and A. Oliva :Learning Deep Features for Scene Recognition using Places Database. , Advances in Neural Information Processing Systems 27 (NIPS) spotlight, 2014.