



République Algérienne Démocratique et
Populaire



Ministère de l'Enseignement supérieur et de
la Recherche Scientifique

Université Echahid Cheikh Larbi Tébessi
Faculté des Sciences Exactes & des Sciences de la Nature et de la Vie
Département de Mathématiques et Informatique
Mémoire de Master

Filière : Informatique

Option : système d'information

**IMPLÉMENTATION D'UN SYSTEME DE
RECONNAISSANCE DE GESTE EN TEMPS REEL**

Présenté par :

BAKHOUCHE Ahmed

Présenté le 06 Juin 2023

Devant le jury :

Dr. BENDIB Issam	MCA	Université de Tébessa Président
Dr. TABET Khaoula	MCB	Université de Tébessa Examineur
Dr. MERZOUG Soltane	MCA	Université de Tébessa Encadreur

Année universitaire : 2022/2023

بِسْمِ اللَّهِ الرَّحْمَنِ الرَّحِيمِ

A mes chers parents,

Que nulle dédicace ne puisse exprimer ce que je leur dois, pour leur bienveillance, leur affection et leur soutien... Trésors de bonté, de générosité et de tendresse, en témoignage de mon profond amour et ma grande reconnaissance.

A mes chers frères,

Je leur dédie ce modeste travail en témoignage de mon grand amour et ma gratitude infinie.

À Mes professeurs

Qui doivent voir dans ce travail la fierté d'un savoir bien acquis.

A tous mes amies

Pour leur aide et leur soutien moral durant l'élaboration du travail de fin d'études.

A tous ma Famille

A tous ceux dont l'oubli du nom n'est guère celui du cœur...

Remerciements

En préambule à ce mémoire on remercie ALLAH qui nous a aidés et nous a donné la patience et le courage durant ces longues années d'études.

On tient d'abord à remercier très chaleureusement

***Mr Dr Merzoug Soltane** qu'il nous ont permis de bénéficier de son encadrement, les conseils qu'on nous a prodigués, la patience, la confiance qu'il nous témoignons ont été déterminant dans la réalisation de notre travail.*

Notre remerciement aussi au les membre de jury pour l'honneur qu'il nous a fait en acceptant d'examiner ce travail.

Un profond respect et un remerciement particulier pour Mr le Chef de département et tous les enseignants de l'Université Echahid Cheikh Larbi Tebessi de Tébessa pour la bonne contribution de ce travail.

Enfin, je remercie tous ceux qui, de près ou de loin, ont contribué à la réalisation de ce travail.

Merci

Résumé

Ce mémoire s'inscrit dans le domaine de la reconnaissance des gestes qui est d'une importance marquée dans la recherche. Cette étude est focalisée précisément sur la reconnaissance des gestes de la main et l'interprétation des signes de ces gestes.

L'approche proposée est composée en deux grandes étapes principales :

La première consiste en l'acquisition de l'image de geste via une web cam en temp réel. Pour assurer la précision dans cette étape on utilise la méthode de détection de zones d'intérêts, qui est la segmentation d'images.

La deuxième consiste en la reconnaissance du geste en utilisant deux algorithmes d'apprentissage automatique pour extraire des caractéristiques et enfin le classificateur pour déterminer le geste.

Mots-clés : reconnaissance des gestes, interaction homme machine, vision par ordinateur, segmentation, classification.

Abstract

This work (memoir) is part of the field of recognition of gesture which is an important area of research. The actual study focused precisely on the recognition of hand gestures and interprets the signs of it.

The proposed approach consists of two major main steps

The first consists in the acquisition of the gesture image via a web cam in real time. To ensure accuracy in this step, we use the method of detecting areas of interest which is image segmentation.

The second is gesture recognition using two machine learning algorithms for feature extraction and finally the classifier to determine the gesture.

Keywords: gesture recognition, human-computer interaction, computer vision, segmentation, classification.

ملخص

هذه الأطروحة هي جزء من مجال التعرف اشارات اليد وهي مجال مهم للبحث ، ركزت دراستنا بدقة على التعرف على إيماءات اليد وتفسر علامات هذه الإيماءة. يتكون نهجنا المقترح من خطوتين رئيسيتين: يتمثل الأول في الحصول على صورة الإيماءة عبر كاميرا الويب في الوقت الفعلي. لضمان الدقة في هذه الخطوة ، نستخدم طريقة الكشف عن مجالات الاهتمام وهي تجزئة الصورة. والثاني هو التعرف على الإيماءات باستخدام خوارزميتين للتعلم الآلي لاستخراج الميزات وأخيرًا المصنف لتحديد الإيماءة.

كلمات مفتاحية: التعرف على الإيماءات ، التفاعل بين الإنسان والحاسوب ، رؤية الكمبيوتر ، التجزئة والتصنيف.

Table des matières

Remerciement	i
Résumé	ii
Tables des matières	v
Liste des figures	viii
Introduction Générale	1
Chapitre 1-L'interaction homme-machine gestuelle	3
Introduction	3
1- Geste comme un moyen d'Interaction Homme - Machine	3
1.1- La communication non verbale ou gestuelle	4
1.1.1- Définition du geste	4
1.1.2- Les trois principale fonctionne de la main	4
1.1.3- Types de geste	4
2- Interfaces gestuelles « Modalité geste »	7
2.1- Interfaces gestuelles Reconnaissance matricielle	7
2.2- Reconnaissance symbolique Reconnaissance symbolique	8
2.3- Reconnaissance mixte Reconnaissance mixte	8
3. Caractériser la gestualité de l'homme (Caractéristique générales)	9
3.1- Interfaces gestuelles Reconnaissance matricielle	9
3.2- Les Interaction entre les deux mains	10
3.3- Différentes classes de signes	10
4. Dispositifs d'interaction et acquisition du geste	12
4.1- Les Difficultés liées à la vision par ordinateur	12

5. Application et Nouvelles possibilités du geste	13
5.1- Réalité virtuelle et La réalité augmentée	13
5.2- Contrôle machine	13
5.3- La reconnaissance de la langue des signes	15
Conclusion	15
Chapitre 2. État de l'art	17
Introduction	18
1. Les techniques d'acquisition d'image pour la reconnaissance des gestes	18
1.1. Caméra	18
1.2. L'utilisation des téléphones intelligents	18
1.3. Tablettes graphiques & écrans tactile	18
1.4. Le gant numérique	18
2. Notion d'Intelligence artificielle	19
2.1. L'apprentissage automatique (Machine Learning)	19
2.2. Apprentissage Profond (Deep Learning)	19
2.3. Quelque type d'algorithmes utilisés en Deep Learning	19
2.3.1. Les réseaux de neurones profonds (DNN)	19
2.3.2. Neurones récurrents (RNN)	19
2.3.3. Les réseaux récurrents à mémoire court et long terme (LSTM)	20
2.3.4. Les réseaux antagonistes génératifs (GAN)	20
2.3.5. Les réseaux de neurones Convolutif (CNN)	21
3. Les techniques IA (segmentation d'image) pour la reconnaissance des gestes	21
3.1. Définition	21
3.2. Différentes approches de segmentation d'images	22
3.2.1. Approche contours	22
3.2.2. Approche régions	23
3.2.3. Approche seuillage	23
4. Reconnaissance du geste état de l'art	24
Conclusion	27
Chapitre3	28
Introduction	29
1. Conception globale	29
1.1. L'acquisition des images (première étape)	30
1.2. La segmentation (deuxième étape)	30
1.2.1. L'algorithme de segmentation	32

1.3. Extraction des caractéristiques (troisième étapes)	32
2. Classification et reconnaissance des gestes	33
3. Les outils de développement utilisés	35
3.1. Python 3.7	35
3.2. Keras 2.1.0	35
3.3. Tensorflow 1.15.0	35
3.4. Numpy 1.21.6	36
3.5. Matplotlib	36
3.6. scikit-image	36
3.7. OpenCV	37
3.8. VGG Image Annotator	37
3.9. Anaconda	38
3.10. PyCharm	38
4. Le matériel	39
5. Importer les bibliothèques	39
Conclusion	43
Conclusion générale & Perspectives	44
Références Bibliographiques	46

Liste des figures

Figure 1.1- (a) Début du geste, (b) image intermédiaire, (c) fin du geste	3
Figure 1.2- Classification les différentes catégories des gestes	4
Figure 1.3- Geste déictique.	5
Figure 1.4- Geste de manipulation	5
Figure 1.5- Geste sémaphorique désignant 'OK'	6
Figure 1.6- Reconnaissance matricielle	7
Figure 1.7- Reconnaissance symbolique	7
Figure 1.8- Reconnaissance mixte.	8
Figure 1.9- Les différents types de dispositifs d'interaction et acquisition du geste	10
Figure 2.1. Les approches de segmentation d'images	17
Figure 3.1. Les approches de segmentation d'images	28
Figure 3.2. Echantillon de donnée	29
Figure 3.3. Figure 3.3: le processus de fonctionnement de segmentation et leur résultat	30
Figure 3.4. le résultat de processus de l'extraction des caractéristiques de geste	31
Figure 3.5. Figure 3.5: l'architecture global de classification et reconnaissance des gestes	32
Figure 3.6. L'architecture détaillé de classification	33
Figure 3.7. Import des bibliographies	37
Figure 3.8. La configuration de modèle	38
Figure 3.9. Accès aux données	38
Figure 3.10. Accès aux web cam de pc pour collecte les images données	39
Figure 3.11. Proseccos de reconnaissance des gestes	39
Figure 3.12. Traitement de résulta de reconnaissance des gestes	40
Figure 3.13. Résulta de reconnaissance des gestes temp de réponse	40
Figure 3.14. Résulta de reconnaissance des gestes	41

Introduction Générale

Lorsqu'un homme (l'être intelligent) est dépourvu d'ouïe ou de parole voire les deux, le geste reste le seul moyen dont on dispose pour établir une communication rituelle ou d'interagir avec l'environnement. De ce fait, la langue des signes est fondée, une langue qui permet de s'exprimer par le biais des gestes de la tête (mimique) et de la main. Les gestes de la main sont un canal de communication naturel et intuitif chez l'homme, pour interagir avec son environnement. Dans ce contexte le sujet du présent mémoire de fin d'études est consacré à développer de méthodes de vision par ordinateur assurant la reconnaissance de gestes de la main.

Le récent avancement marqué aux technologies de l'intelligence artificielle, a permis la mise en œuvre de quelques techniques, tel que la vision par ordinateur. Ce travail consistera à réaliser un système intelligent capable de reconnaître certains gestes de la main comme une première étape, plus précisément l'alphabet de la langue des signes. Les gestes peuvent avoir une signification différente suivant la langue ou la culture : les langues des signes en particulier sont spécifiques à chaque langue.

Ce mémoire sera divisé en trois chapitres :

Dans le premier chapitre, On va présenter en premier lieu la communication gestuelle d'un point de vue général, ensuite on s'intéresse à celle réalisée avec les gestes de la main, enfin on introduira la langue des signes. La vision par ordinateur sera le principal pilier dans la réalisation du travail. Et puis, on va récapituler les différentes techniques développées au cours des dernières années dans le domaine de la reconnaissance des gestes de la main.

Dans le second chapitre, on va établir une étude détaillée comparative des différentes méthodes de détection et reconnaissance de gestes de la main

Dans le troisième chapitre, on va extraire la zone d'intérêt à partir d'images en utilisant la méthode choisie lors du précédent chapitre. Cette dernière contient les données requises par les algorithmes d'apprentissage afin d'effectuer la reconnaissance des gestes de la main.

Chapitre 1

L'interaction homme-machine gestuelle.

Introduction

La communication est l'action de communiquer, d'établir une relation avec autrui, de transmettre quelque chose à quelqu'un. Les besoins des gens nés avec l'incapacité de parler et communiquer par voix verbale ou d'entendre, voire les deux, De ce fait, on se trouve obligés de développer un nouveau moyen de communication, basé seulement sur les gestes, c'est par ce point du départ que la langue des signes a vu le jour. Les gestes sont un des moyens de communication les plus riches que l'être humain possède. Ils permettent d'agir sur le monde physique, et servent aussi à établir une communication cohérente.

Dans ce chapitre on s'intéresser à la communication gestuelle plus précisément à la langue des signes afin de mieux nous familiariser avec le domaine de cette étude.

1- Geste comme un moyen d'Interaction Homme - Machine

1.1- La communication non verbale ou gestuelle :

La communication non verbale ou langage du corps désigne dans une conversation tout échange n'ayant pas recours à la parole.

Elle ne repose pas sur les mots, mais sur les gestes (actions et réactions), les attitudes, les expressions faciales (dont les micro-expressions) ainsi que d'autres signaux, conscients ou inconscients. [1]

1.1.1- Définition du geste :

Un geste est un signe manuel ou corporel qui permet d'illustrer les mots du langage, de les compléter ou de les appuyer. Le geste peut aussi être utilisé pour remplacer la parole, la figure 1 présente un exemple détaillé d'un geste. [2].

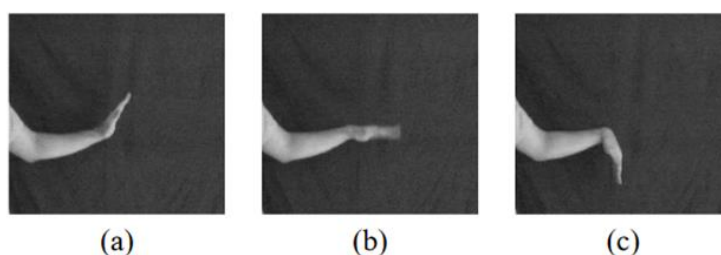


Figure 1.1 : (a) Début du geste, (b) image intermédiaire, (c) fin du geste [2].

1.1.2- Les trois principales fonctions de la main :

Les trois fonctions principales des gestes de la main [3] :

- La fonction argotique : la main joue le rôle d'organe moteur et agit sur le monde physique pour le transformer. Elle applique aux objets des forces pour les déplacer ou les déformer.
- La fonction épistémique : la main joue le rôle d'organe de perception. Le sens du toucher (sens tactilo-proprio-kinesthésique) donne des informations sur la forme, l'orientation, la distance, la grandeur, le poids, la température, les mouvements des objets, etc.
- La fonction sémiotique : la main joue le rôle d'organe d'expression pour l'émission d'informations visuelles. Cela comprend la langue des signes, le geste Co-verbal, qui accompagne la parole, ou les gestes permettant une communication basique lorsqu'on ne peut pas utiliser la parole, comme dans un environnement bruyant ou en plongée sous-marine.

1.1.3- Types de geste :

Il existe plusieurs classifications de gestes. la figure 2 présente les différentes catégories

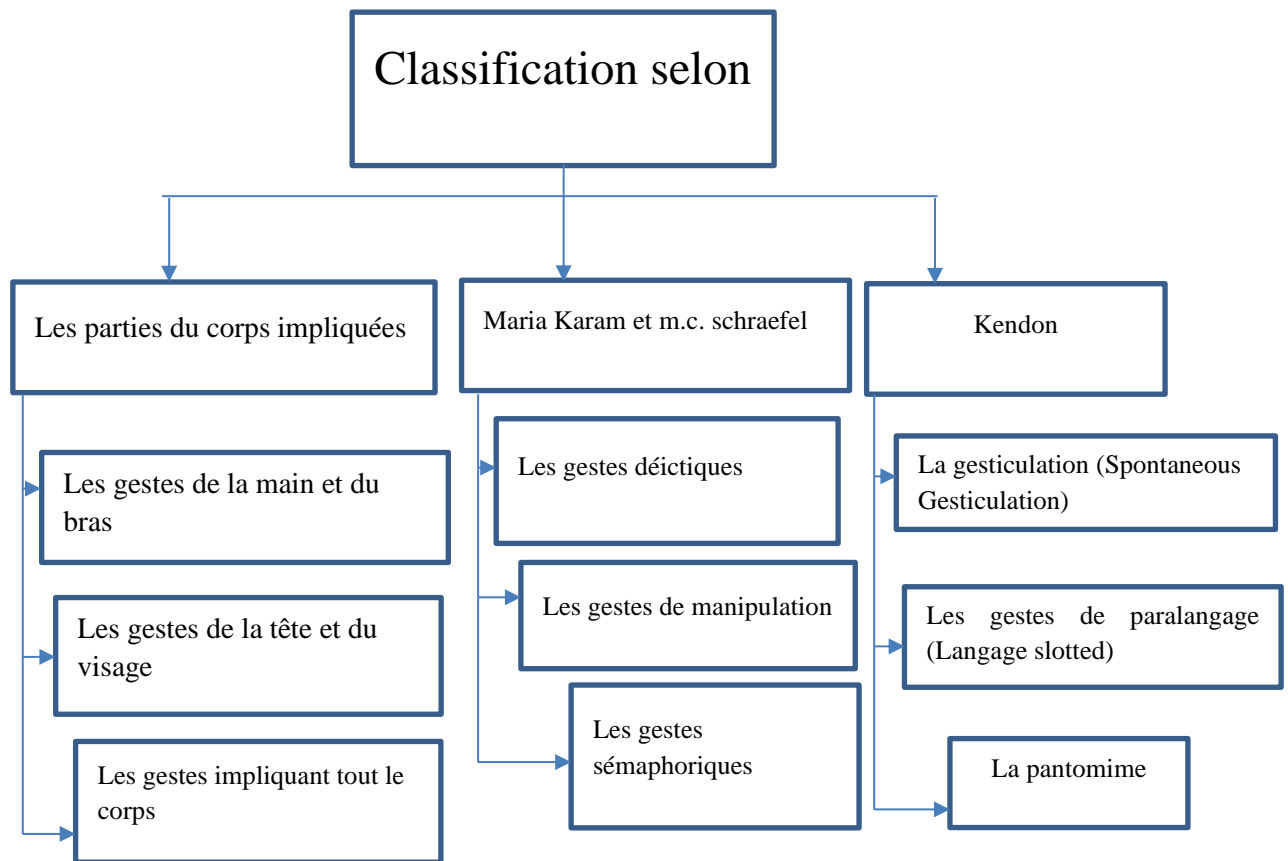


Figure 1.2 : Classification les différentes catégories des gestes.

1.1.3.1- Classification selon les parties du corps impliquées : [4]

On distingue généralement trois types de gestes :

- **Les gestes de la main et du bras** : Ils forment la principale catégorie de gestes interactifs. La main permet de réaliser des gestes précis et complexes. Les recherches autour de ces gestes concernent principalement la reconnaissance de positions de la main, l'interprétation du langage des signes et le développement d'interface homme-machine permettant la manipulation et l'interaction avec des données ou des éléments d'un environnement virtuel.
- **Les gestes de la tête et du visage** : Peu de gestes de la tête ont une signification spécifique ; l'orientation de la tête est quant à elle très utile pour la détection du champ de vision. Les recherches dans ce domaine s'intéressent à la reconnaissance faciale comme moyen d'authentification biométrique, comme soutien à d'autres systèmes de reconnaissance tels

que la reconnaissance de la parole ; l'analyse des gestes faciaux est également utile pour la réalisation d'avatars virtuels réalistes.

- **Les gestes impliquant tout le corps** : Les recherches dans ce domaine s'intéressent à tout le corps en interaction avec son environnement (analyse des gestes d'un danseur afin de générer de la musique idoine ; analyse des gestes d'un athlète pour améliorer ses performances).

1.1.3.2- Classification selon Maria Karam et m.c. schraefel : [5]

Karam et schraefel, présentent trois styles de gestes, principalement de la main et du bras, synthétisant les types d'interaction décrits à travers la littérature scientifique :

- **Les gestes déictiques** : Sont des gestes de pointage permettant d'identifier un objet ou son emplacement. Ces gestes sont typiquement utilisés dans des environnements virtuels. Ces gestes peuvent être considérés comme implicites dans d'autres formes de gestes (par exemple lorsque l'on pointe des doigts un objet à manipuler).



Figure 1.3 : Geste déictique. [5]

- **Les gestes de manipulation** : Dont le but est de contrôler une entité en appliquant une relation étroite entre le mouvement du geste et l'entité qui est manipulée.



Figure 1.4 : Geste de manipulation. [5]

- **Les gestes sémaphoriques** : Ils font partie de tout système gestuel basé sur un catalogue conventionnel de gestes statiques ou dynamiques (par exemple : le geste (statique) pour le signe « ok » le signe (dynamique) de la main pour dire « au revoir »).



Figure 1.5 : Geste sémaphorique désignant 'OK'. [5]

1.1.3.3- Classification selon Kendon :

- **La gesticulation (Spontaneous Gesticulation)** : Il faut entendre par cela l'ensemble des gestes co-verbaux (accompagnant la parole).
- **Les gestes de paralangage (Langage slotted)** : Résulte de la traduction de l'anglais *langage-like gestures*. En réalité, ils ne constituent nullement un langage car, à la différence des langues des signes, ils apparaissent plutôt comme un simple système de codage au vocabulaire très restreint. Ce sont par exemple les gestes utilisés pour l'arbitrage de certains sports.

- **La pantomime** : vise à créer une image gestuelle d'un objet, d'un événement, d'un état d'esprit. Elle est fortement iconique.
- **Les emblèmes** : sont des gestes complétant les mots ou utilisés à leur place. Ils sont utilisés au sein d'une société, dans laquelle ils ont une signification fixe et communément acceptée. Le froncement des sourcils est ainsi largement répandu pour exprimer le doute, le désaccord ou une difficulté de compréhension. En revanche, la représentation des chiffres avec les doigts de la main diffère selon les cultures.
- **Les langues des signes** : (Signes) sont les seules à pouvoir être considérées comme de véritables langues. Elles possèdent certaine règle de formation et d'ordonnancement, et une combinatoire permettant de générer de nouveaux items lexicaux selon les besoins.

2- Interfaces gestuelles « Modalité geste »

2.1- Interfaces gestuelles Reconnaissance matricielle :

Reconnaissance basée sur l'analyse d'une matrice Facilité : quelques points significatifs suffisent Restrictions : sens précis (vertical), taille prédéfini (taille de la case) Reconnaissance dans un univers fini (monde limité) : possibilité de calculer « la distance » entre gestes pour une bonne qualité de reconnaissance. [6]

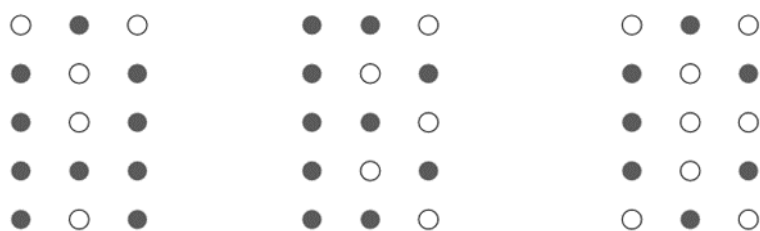


Figure 1.6- Reconnaissance matricielle. [6]

2.2- Reconnaissance symbolique Reconnaissance symbolique :

Reconnaissance basée sur des segments Facilité : identification de constituants simples (segments, arcs de cercle,) significatifs Problème : découpage en constituants simples souvent difficile (besoin de lissage) Généralité : orientation quelconque, taille quelconque.

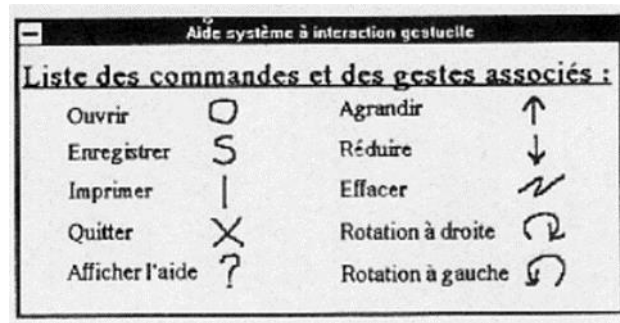


Figure 1.7- Reconnaissance symbolique. [6]

2.3- Reconnaissance mixte Reconnaissance mixte :

Dans une matrice, mais avec le point de départ, point d'arrivée, forme (verticale, horizontale, complète) Reconnaissance optimiste (globalement satisfaisant : départ – arrivée & forme) Reconnaissance pessimiste (robuste).

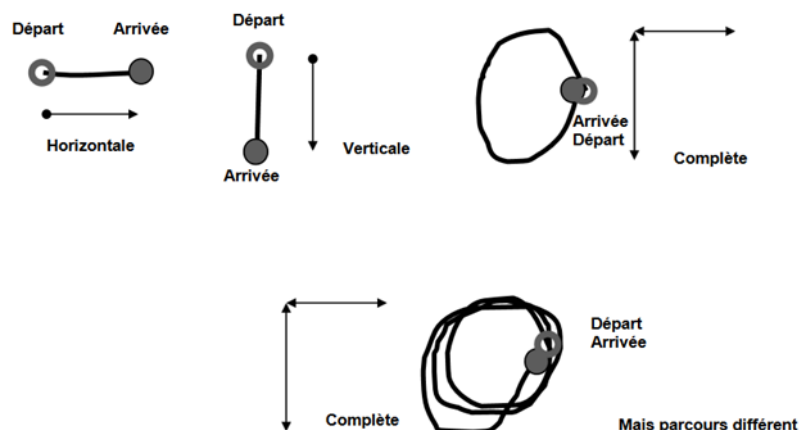


Figure 1.8- Reconnaissance mixte. [6]

3- Caractériser la gestualité de l'homme (Caractéristiques générales)

Cette section présente les différentes caractérisations générales du geste. Ensuite, on va présenter les systèmes de représentation et de transcription du geste.

La Langue des Signes est le moyen d'expression utilisé par les communautés de sourds ou malentendants pour communiquer entre eux. La Langue des Signes est une véritable langue à part entière (avec un lexique, une syntaxe...) et constitue de ce fait la forme la plus évoluée en communication gestuelle. L'expression de phrases en LS, Langue des signes, ne se réduit pas aux gestes produits par les deux mains, c'est le corps tout entier qui peut être mis à contribution pour exprimer une phrase. On peut distinguer trois principales parties qui interviennent : les mains, la tête (mimique et regard) et le buste. Dans cette étude on se restreint uniquement aux mains, c'est pourquoi les autres parties du corps ne seront pas abordées par la suite. Chaque phrase est constituée d'une suite de gestes des mains que l'on appelle signes et qui sont agencés suivant une syntaxe régie par une logique spatiale et temporelle. [7].

3.1- Interfaces gestuelles Reconnaissance matricielle :

Chaque geste d'une main peut être décomposé en quatre paramètres qui sont indépendants et peuvent être aussi bien dynamiques qu'invariants durant l'émission du signe [8].

- **La configuration** : forme de la main (exemples de noms de configurations : les noms de l'alphabet dactylogique, bec de canard, index).
- **L'orientation** : l'orientation de la main est la position de la paume de la main par rapport au signeur. Il y a 6 directions de la paume de la main et une direction pour les deux paumes. Ces directions peuvent varier durant l'exécution du signe. La paume peut être orientée vers : Le haut, Le bas, La droite, La gauche Vers soi Devant soi.
- **Le mouvement** : Il peut y avoir mouvement sur les doigts, la main, le bras, le corps et la tête en jouant sur l'ampleur, la vitesse ou la répétition.

- **L'emplacement** : L'emplacement de la main représente l'endroit où l'action est effectuée ou un endroit générique. Elle se trouve en générale dans l'espace, devant le signeur, dans la zone d'à côté de la tête du signeur et sur le corps du signeur. Chacun de ces paramètres est porteur d'information et contribue au sens d'un signe.

3.2- Les Interaction entre les deux mains :

Un signe peut aussi bien faire intervenir une main que deux mains et les deux mains ont différentes façons d'interagir. Lorsque les deux mains sont impliquées dans un signe, deux cas se présentent. Dans le premier cas on voit apparaître un rôle pour chaque main. Une main est dite dominante et a pour rôle de décrire "l'action", tandis que l'autre main qui est appelée main dominée sert de référence à cette action. En général la main dominante se déplace au cours du geste tandis que la main dominée reste statique. Dans le deuxième cas, les deux mains sont complètement synchronisées : leurs paramètres sont identiques ou symétriques [9].

3.3- Différentes classes de signes :

Dans les langues des signes, on peut distinguer différentes classes de signes, chacune correspondant à un usage particulier.

L'ensemble des signes dits :

- **Standards** : correspond à des mots (nom, verbe, adjectif...) ayant un sens bien établi. On peut leur associer une correspondance relativement précise dans les langues orales.
- **Directe** : ce sont les signes qui ont été le plus étudiés en reconnaissance de geste.

Il existe deux autres classes de signes qui eux ne permettent aucune correspondance directe avec des mots du langage oral. Il s'agit des spécificateurs de forme et de taille, et des classificateurs.

Les spécificateurs permettent de décrire un objet, un animal, une scène. C'est à l'aide de la forme des mains, de leur orientation et de leur mouvement que le

signeur décrit une forme et les dimensions d'un objet. Les classificateurs sont similaires aux spécificateurs dans le sens où ils représentent également un objet (ou personne, animal...) et donc la forme de la main est en rapport avec celle de l'objet ou de sa fonction. Mais ils ont un tout autre rôle car ils servent en quelque sorte de pronom. Lorsqu'un objet a été cité dans une phrase à l'aide d'un signe du vocabulaire standard (ou de spécificateurs), un classificateur peut ensuite être utilisé pour représenter cet objet dans le reste de la phrase. On peut avec ce classificateur préciser la position de l'objet ou décrire une trajectoire qu'il a emprunté.

4- Dispositifs d'interaction et acquisition du geste

L'interaction par le geste est possible grâce aux systèmes de capture de mouvement du corps ou des gants de données. La modélisation des différentes parties du corps peut être très précise, en temps-réel.



Figure 1.9- les différents types de dispositifs d'interaction et acquisition du geste. [10]

L'acquisition du geste par vision artificielle a été le sujet de plusieurs travaux de recherches, ces derniers peuvent être classés en deux principales catégories, d'une part l'analyse des aspects d'une image et d'autre part la modélisation 3D de leur contenu.

Les méthodes basées sur la 3D tirent avantage de la connaissance préalable de la forme d'un modèle et de la possibilité d'appliquer des transformations géométriques afin de le déformer, il est alors nécessaire d'utiliser la stéréovision 1 ou plus de deux caméras pour acquérir l'information ce qui constitue un

procédé. Les méthodes basées sur la 2D quant à elles reconnaissent un nombre limité de gestes après un procédé d'apprentissage. [10].

4.1- Les Difficultés liées à la vision par ordinateur :

Contrairement aux gants numériques, qui permettent d'obtenir directement des informations sur la position et la configuration de la main, la conception d'un système de vision par ordinateur nécessite la prise en compte d'un certain nombre de difficultés [11] :

- **Eclairage de la scène** : les variations de l'éclairage de la scène ont un impact important sur les algorithmes de segmentation et d'extraction des caractéristiques, en provoquant des variations de luminosité dans l'image.
- **Ombres** : la main projette une ombre dans la scène, même si la luminosité est contrôlée. Les ombres sont parfois segmentées avec la main, suivant la méthode utilisée, ce qui rend le résultat de segmentation difficilement exploitable.
- **Objets de la scène** : la scène peut comporter des objets mobiles tels que le clavier, la souris ou des documents. L'algorithme de segmentation de la main doit pouvoir s'adapter à des changements du fond de la scène et distinguer la main des autres objets.
- **Camouflage** : lorsque les caractéristiques de l'objet sont proches de celles du fond, celui-ci est difficilement détectable. Dans le cas de la main, ce problème se produit par exemple lorsque le bureau a une couleur de teinte chair.
- **Occultations** : les occultations se produisent lorsqu'un objet est masqué par un autre, ce qui peut arriver lorsqu'on utilise les deux mains. Dans le cas de la main, il existe un problème supplémentaire avec les « Auto occultations » : la main étant un objet complexe et déformable, il arrive fréquemment que certaines parties, telles que les doigts, soient cachées par d'autres.

- **Disposition des caméras** : la distance entre la caméra et les objets joue un rôle important, de même que l'angle de vue des caméras qui peut causer des déformations géométriques dans l'image.

5- Application et Nouvelles possibilités du geste

Cette section présente différentes applications, et notamment les surfaces interactives, ainsi que quelques applications récentes destinées au grand public.

5.1- Réalité virtuelle et La réalité augmentée :

La réalité virtuelle consiste à plonger l'utilisateur dans un environnement d'images de synthèse. Cette immersion dans un environnement virtuel peut se faire grâce à un casque, ou dans une pièce dédiée munie de plusieurs écrans ou d'un vidéoprojecteur [12] [13]. L'utilisateur est donc en immersion complète dans un environnement dans lequel il peut interagir, notamment par le biais de gestes. La réalité augmentée mélange le monde physique et des informations virtuelles, en superposant ces informations dans le champ de vision de l'utilisateur [11].

Ce type de système est fondé sur une analyse par l'ordinateur du monde environnant l'utilisateur, au moyen d'un système de vision, de type caméra vidéo par exemple, ou de capteurs spécifiques. Des informations virtuelles sont alors projetées sur des éléments réels, par exemple par l'intermédiaire d'un vidéo projecteur. Une des applications principales de la reconnaissance de gestes en réalité augmentée est de rendre une surface interactive : une table, un tableau ou un bureau, sur laquelle des images sont projetées. L'utilisateur peut alors interagir avec des objets réels ou virtuels.

5.2- Contrôle machine :

L'interaction gestuelle joue un rôle important pour le contrôle des machines, ce qui s'explique par son efficacité pour la commande. Plusieurs travaux ont été développés dans ce contexte, parmi eux un travail sur l'exploitation du geste de la main pour remplacer la souris afin de transformer un écran classique en un écran tactile et de permettre à l'utilisateur de se servir de son doigt comme dispositif de pointage. Dans le but d'établir une communication

naturelle entre l'homme et un robot personnel, fut développée une interaction basée sur le geste de la main (geste de pointage), et la parole pour informer un robot mobile, qui est doté de capteur sonore et d'un système de stéréovision, sur la localisation d'un objet dans son environnement [11]

5.3- La reconnaissance de la langue des signes :

La reconnaissance de la langue des signes est une application naturelle de la reconnaissance de gestes. En effet, un système de traduction automatique des signes est particulièrement intéressant pour les sourds et muets. Braffort [13...5] présente une étude détaillée de la (LSF) Langue des Signes Française.

Conclusion

Afin de pouvoir traiter convenablement le thème abordé par ce projet de fin d'étude, on a parcouru les différentes notions liées à la vision par ordinateur, d'où on a cité l'utilisation d'un nouveau moyen de communication lors des interactions Homme-Machine, qui sont reparties sous le terme « le geste ». La main est l'outil servant couramment à exprimer le geste, qui représente une communication naturelle et l'une des méthodes les plus utilisées uniquement chez les personnes ayant des problèmes d'élocution ou éventuellement troubles d'audition.

L'utilisation du geste ne s'appartient seulement qu'aux gens atteints de troubles physiologiques, cependant tant d'activités humaines font appel à déployer un certain raisonnement basé sur le geste, comme par exemple aux activités menées sous-marin, où la communication par voix ou ne peut pas eu lieu.

On a conclu par mentionner les problèmes rédhibitoires imposées lors du développement du système de reconnaissance de geste réalisé. Cela implique le bon choix de configuration du système, ainsi la bonne maîtrise de différentes méthodes disponible pour accomplir.

Dans le chapitre qui suit, on va décrire en détail les méthodes qui forment l'approche de ce travail.

Chapitre 2

État de l'art

Introduction

La plupart des approches de détection et reconnaissance de formes reposent sur des techniques d'apprentissage automatique ou apprentissage profonds. Dans cette partie, on essaiera de donner une vue générale sur la notion d'IA dans le domaine de reconnaissance des gestes de la main en citant les outils et les algorithmes mis en œuvre pour l'IA. Premièrement on va aborder les techniques de l'apprentissage, bien précisément l'apprentissage automatique.

Ensuite, citer ses différents types utilisés lors de la conception d'un système doté d'une intelligence artificielle et parler plus sur les méthodes de la détection de zone d'intérêt, parmi lesquelles on trouve la segmentation d'image qui s'avère être très efficace pour le domaine de reconnaissance des gestes. Mais avant de commencer ce chapitre plus en détail, on va brièvement signifier que des images peuvent être capturées, en fait, il existe plusieurs techniques pour l'acquisition des images.

1- Les techniques d'acquisition d'image pour la reconnaissance des gestes

La vision par ordinateur est une branche de l'intelligence artificielle dont le principal but est de permettre une machine d'analyser, traiter et comprendre une ou plusieurs images prises par un système d'acquisition (dans notre système nous allons opter pour une Webcam). La plupart des systèmes de reconnaissance du geste utilisent cette technique, car son aspect intuitif et simple reste un moyen plus approprié pour l'interaction Homme-Machine. La reconnaissance de gestes à l'aide de la webcam est une option attrayante pour remplacer l'interaction homme-machine à l'aide d'une souris. De nos jours, trois méthodes sont utilisées principalement pour effectuer la reconnaissance gestuelle.

1.1- Caméra :

Tirer a fourni une partie de la puissance de la vision artificielle pour l'agriculture repose sur la sélection de la bonne caméra. Tout d'abord, il est important d'identifier le problème à résoudre car cela a déterminé la région de longueur d'onde, la résolution requise, la vitesse et l'interface de la caméra [14]. Les conditions environnementales pour l'imagerie seront également importantes. L'imagerie dans la plage visible avec des conditions d'éclairage variables nécessite une caméra avec une plage dynamique élevée pour capturer des images sans sursaturation ni perte de détails dans le bruit de lecture [14].

1.2- L'utilisation des téléphones intelligents :

Le produit technologique le plus vendu aujourd'hui est sans conteste le téléphone intelligent ou les ordiphones. Ces appareils embarquent des caméras avec de plus en plus de capteurs puissants. Il n'est pas donc inutile de vouloir exploiter ces caractéristiques technologiques. Pour certaines tâches de capture, les téléphones portables sont l'alternatif le plus efficace qui peut effectivement remplacer d'autres appareils. Les images et les commandes utilisent leurs autres avantages, tels que le GPS intégré, leur Connectique (GSM, WIFI, Bluetooth...etc.) [15].

1.3- Tablettes graphiques & écrans tactile :

Ces techniques sont destinées à la reconnaissance de l'écriture manuscrite, le geste de dessins et les gestes de pointage. Ces techniques ont pour objectif de remplacer les traditionnels menus. Le crayon et la tablette graphiques sont d'usage courant en Conception Assistée par Ordinateur (CAO) [15].

1.4- Le gant numérique :

Les gants spéciaux principalement utilisé dans l'industrie du cinéma. Il fonctionne bien mais est coûteux [16].

2- Notion d'Intelligence artificielle

L'intelligence artificielle (IA) renvoie à "une application capable de traiter des tâches qui sont, pour l'heure, réalisées de manière plus satisfaisante par des êtres humains dans la mesure où elles impliquent des processus mentaux de haut niveau comme l'apprentissage perceptuel, l'organisation de la mémoire et l'esprit critique" [17].

Les scientifiques et afin de créer une intelligence artificielle, utilisent un ensemble de techniques différentes regroupées sous une même discipline qu'on appelle l'apprentissage machine (ou machine Learning, en anglais). Donc dire intelligence artificielle revient automatiquement à dire "Machine Learning", ou l'apprentissage automatique [17].

2.1- L'apprentissage automatique (Machine Learning) :

L'apprentissage automatique est considéré comme un sous-domaine de l'intelligence artificielle. Son objectif majeur est de comprendre la structure des données et de les faire intégrer dans des modèles simples que tout le monde peut comprendre et utiliser [18].

Bien que l'apprentissage automatique soit un domaine de l'informatique, il diffère de ces approches traditionnelles. [18]

De nos jours, l'apprentissage automatique est omniprésent dans toutes nos interactions avec les banques, nos achats en ligne et notre utilisation des médias sociaux. [18] Il existe plusieurs méthodes d'apprentissage automatique, mais les plus largement utilisées sont l'apprentissage supervisé où les algorithmes sont basés sur des données d'entrées et de sorties bien spécifiées par l'homme, et l'apprentissage non supervisé qui ne fournit pas à l'algorithme des données étiquetées afin de lui permettre de définir une structure et de découvrir une logique dans les données entrées. Auxquelles on ajoute l'apprentissage par renforcement.

2.2- Apprentissage Profond (Deep Learning) :

L'apprentissage profond (dl), est l'une des principales technologies de Machine Learning et d'intelligence artificielle. Découvrez en quoi consistent cette technologie, son fonctionnement, et ses différents secteurs d'application [18]. Le Deep Learning ou apprentissage profond est un type d'intelligence artificielle, dérivé du *machine Learning* qui a été développé dans le but de créer des algorithmes capables d'apprendre et de s'améliorer de manière autonome, contrairement à la programmation où la machine se contente d'exécuter à la lettre des règles prédéterminées [18]. Profonde de la formation n'est pas seulement une question de technologie, mais aussi d'une vision globale de l'homme. L'un des premiers

objectifs d'un apprentissage en profondeur est la lecture des capacités de l'homme et même la création d'une version améliorée [18].

2.3- Quelque type d'algorithmes utilisés en Deep Learning :

Les architectures les plus utilisées en réseaux de neurones profonds sont : [19]

2.3.1- Les réseaux de neurones profonds (DNN) :

Les réseaux de neurones profonds (Deep Neural Networks : DNN) sont des réseaux à flux direct dans lesquels les données circulent de la couche d'entrée à la couche de sortie sans revenir en arrière et les liens entre les couches sont à sens unique et ils ne touchent jamais un nœud de nouveau [19].

2.3.2- Neurones récurrents (RNN) :

Un réseau neuronal récurrents (Récurrent Neural Network : RNN) est une classe de réseaux neuronaux artificiels dans lesquels les connexions entre les nœuds forment un graphique vectoriel le long d'une séquence qui leur permet de montrer un comportement dynamique temporel d'une séquence temporelle. Les RNN peuvent utiliser leur état interne (mémoire) pour gérer la séquence d'entrée. Cela le rend applicable à des tâches telles que la reconnaissance de l'écriture manuscrite ou la reconnaissance vocale. [19]

2.3.3- Les réseaux récurrents à mémoire court et long terme (LSTM) :

Le réseau récurrent à mémoire court et long terme (Long short-term memory : LSTM) sont un type particulier de RNN, capable d'apprendre les dépendances à long terme qui permettent à RNN de se souvenir de choses qui se sont produites dans le passé et de trouver des modèles à travers le temps pour donner un sens à ses prochaines suppositions. [27] Les LSTM ont battu des records pour l'amélioration de la traduction automatique, de la modélisation linguistique et du traitement linguistique multilingue. [19]

2.3.4- Les réseaux antagonistes génératifs (GAN) :

C'est le profond algorithme d'apprentissage derrière le générative de l'IA. Entre autres choses, Gan peut créer de nouvelles données de cas qui sont similaires aux données qu'ils ont été formés. Les réseaux de génératives antagonistes se composent d'un discriminateur de générateur. Ils sont utilisés à des fins diverses, telles que la création ou l'amélioration des images, la génération de texte, etc. [19]

2.3.5- Les réseaux de neurones Convolutif (CNN) :

Les réseaux de neurones Convolutif (CNN ou ConvNets) sont une classe de réseaux de neurones profonds couramment appliqués à l'analyse d'images visuelles. Leurs autres applications incluent la compréhension vidéo, la reconnaissance vocale et la compréhension du traitement du langage naturel. [19]

3- Les techniques IA (segmentation d'image) pour la reconnaissance des gestes

La segmentation est une étape primordiale, Il existe de nombreuses méthodes de segmentation pour la reconnaissance des gestes de la main, plus ou moins performantes suivant les suppositions faites sur la scène, et suivant l'environnement (intérieur, extérieur, condition d'illumination...) [20]. De plus, nous souhaitons non seulement détecter la main, mais aussi la segmenter afin d'obtenir son contour, avec la meilleure précision possible, pour en extraire des informations sur sa forme et sa position. Par ailleurs, il faut prendre en compte la contrainte temps réel, car la segmentation de la main n'est que la première étape d'un système de reconnaissance de gestes, qui en compte plusieurs autres. Il faut donc trouver un compromis entre les performances de la segmentation et la rapidité d'exécution

3.1- Définition :

La segmentation d'images consiste à regrouper les pixels des images qui partagent une même propriété pour former des régions connexes. Il existe plusieurs définitions de la segmentation parmi lesquelles nous trouvons celle de Zucker [20], qui l'a défini comme le partitionnement de l'ensemble des pixels d'une image I en n sous ensembles appelés régions R_i , $i=1..n$: $I = R_1, R_2, R_3, \dots, R_n$ Tel que :

Ce qui se traduit par l'algorithme :

$$\bigcup_{i=1}^n R_i = I \quad \text{Avec } R_i \cap R_j = \emptyset \quad i \neq j$$

$$\text{ET : } \begin{cases} P(R_i) = \text{vrai} & \forall i = 1..n \\ P(R_i \cap R_j) = \text{faux} & R_i \text{ Adjacente à } R_j \end{cases}$$

Avec :

- L'union de toutes les régions R_i , avec $i=1..n$, forme l'image I .
- L'intersection de deux régions R_i et R_j est un ensemble vide pour $i \neq j$.

- Soit P un prédicat d'homogénéité, soient R_i et R_j deux régions adjacentes de l'image. Si la région R_i satisfait le prédicat P , l'intersection de R_i et R_j ne le satisfera pas.

3.2- Différentes approches de segmentation d'images :

La segmentation d'image est un domaine vaste où l'on trouve de très nombreuses approches. Toutes ces approches visent à extraire des indices visuels et cela de différentes façons, le figure de schéma suivent en présente les plus connues :

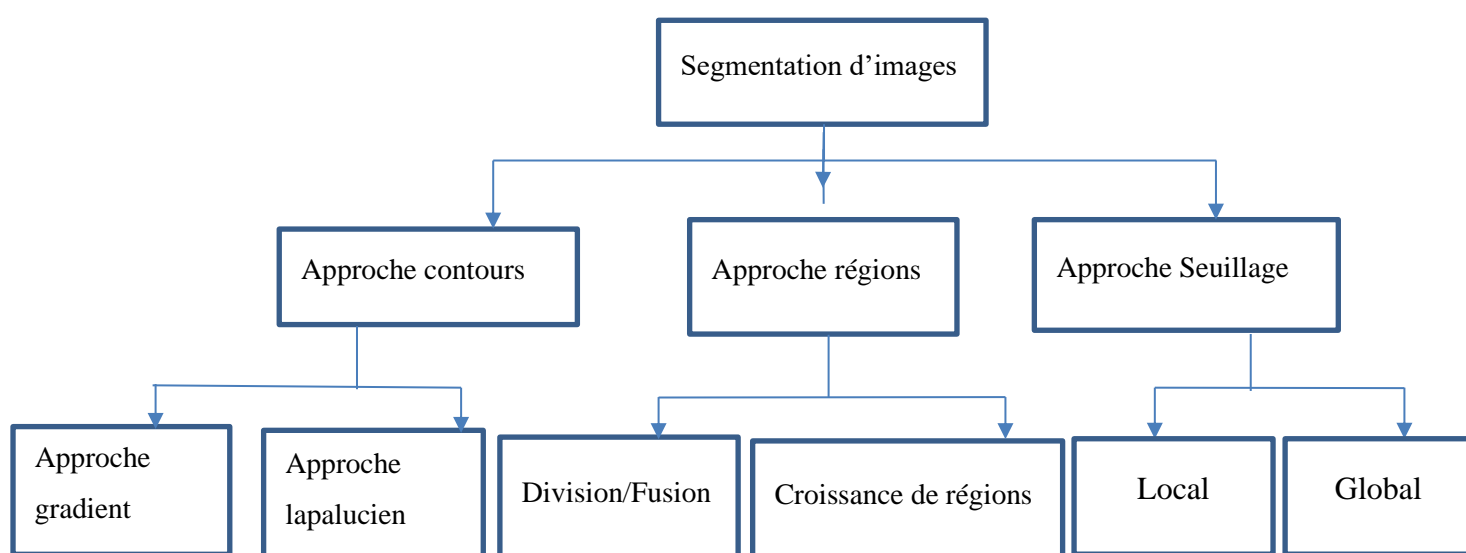


Figure2.1- les approches de segmentation d'images.

3.2.1- Approche contours :

Les contours sont les lieux de variation significative de l'information niveau de gris.

La notion de contours étant relié à celle de la variation, il est évident qu'une telle variation nous amène vers une évaluation de la variation en chaque point. Ce qui se traduit par le calcul de deux grandeurs qui sont le gradient 1 et lapalucien (le deuxième dérivé du gradient). Les contours sont alors repérés là où le gradient présente un maximum local ou lors des passages à zéro de lapalucien. [21]

3.2.2- Approche régions :

Les méthodes appartenant à cette famille manipulent directement des régions. Soit-elle partie d'une première partition de l'image, qui est ensuite modifiée en divisant ou regroupant des régions, et on parle alors de méthodes de division/fusion (ou split and merge en anglais) ; soit elles partent de quelques régions, qui sont amenées à croître par incorporation de pixels jusqu'à ce que toute l'image soit couverte, et on parle alors de méthodes par croissance de régions. [21]

3.2.3- Approche seuillage :

Le seuillage est une technique de segmentation très populaire à cause de sa facilité de mise en œuvre et sa rapidité. Elle permet d'extraire les objets du fond de l'image. Dans le cas le plus classique, les pixels de l'image sont classés en deux classes par l'intermédiaire d'un niveau de gris S appelé seuil. La première classe regroupe les pixels du fond et la deuxième classe les pixels de l'objet. Cette opération s'appelle binarisation.

Pour ce seuillage, la classification d'un pixel dépend non seulement de son niveau de gris mais aussi de ses informations locales c'est-à-dire des niveaux de gris de ses voisins. On définit alors pour chaque pixel un seuil selon sa position. Le premier à proposer une technique donnant de bons résultats fut Bernsen.[21]

Le principe du seuillage global est d'utiliser une valeur seuil à partir de laquelle on peut choisir à quelle classe le pixel appartient. Les méthodes utilisant ce type de seuillage reposent sur l'exploitation de l'histogramme de toute l'image qui caractérise la distribution des niveaux de gris. En général une méthode de seuillage consiste à déterminer la valeur optimale du seuil S en se basant sur certain critère.

4- Reconnaissance du geste état de l'art

Les études réalisées dans le domaine de la reconnaissance automatique des gestes sont relativement récentes, les premiers étant liés à l'apparition des gants numériques en 1987 avec le Data Glove de VPL par la suite, nombre d'entre elles se sont d'abord attachées à la reconnaissance de gestes dans le cadre d'interaction homme-machine.

D'autres ont porté sur la classification par MMC, Modèle de Markov Caché, de mouvement de tennis en s'appuyant sur des techniques de quantification vectorielle dans des séquences d'images binaires. [22]

Dans le domaine plus spécifique de la reconnaissance du langage des sourds, Braffort a présenté un système pour la langue des signes française basé sur l'acquisition de paramètres à partir de gant instrumentés, les mots y sont analysés et caractérisés selon plusieurs critères, le système de reconnaissance étant basé sur des MMC, Modèle de Markov Caché, divisée en deux modules, l'un d'une part pour classifier les signes conventionnels et le second d'autre part pour les signes variables et non conventionnels, les taux de reconnaissance obtenus sont respectivement de 96% et de 92% parmi un dictionnaire de 44 phrases et un vocabulaire de 7 signes. [22]

Starner et Pentland, ont développé un système basé sur la vidéo pour la reconnaissance des phrases de la langue des signes américaine, à partir d'un vocabulaire de 40 signes, les gestes sont modélisés par un MMC, Modèle de Markov Caché, à 4 états, une seule caméra est utilisée pour l'enregistrement des images, le taux de reconnaissance est compris entre 75% et 99%. [22]

Vogler et Metaxás ont décrit un système basé sur les MMC et destiné à la reconnaissance de l'ASL, *Américain Signe Langage*, d'un vocabulaire de 53 signes, 3 caméras sont utilisées pour l'acquisition de paramètres 3D des mouvements des bras et des mains des signeurs, 97 phrases de test ont été utilisées, les résultats obtenus varient entre 92.1% et 95.8%.

La majorité des travaux sur la reconnaissance du geste utilise les méthodes de Markov caché, basé sur l'apprentissage de *Baum-Welche* et *l'algorithme de Viterbe* pour approximer la vraisemblance d'une observation étant donné le modèle. [22]

Ji et al. [23] ont proposé une méthode basée sur des réseaux de neurones convolutifs 3D (Convolution Neural Networks ou CNN 3D) en étendant les CNN 2D, utilisés pour les images, aux vidéos en prenant en compte la dimension temporelle. La méthode a été utilisée pour reconnaître des actions filmées par des caméras de vidéos surveillances.

En 2014, Karpathy et al. [24] ont également utilisé des réseaux de neurones convolutifs pour reconnaître des actions dans des vidéos. Ils testent différentes méthodes de fusion de données pour prendre en compte plusieurs trames pour entraîner leurs CNN. Les meilleurs résultats sont obtenus avec une "*slow fusion*", c'est à dire les premières couches de convolutions prennent en compte des informations contenues dans 4 trames consécutives, puis ces couches sont ensuite fusionnées entre elles de manière à ce que la couche résultante contienne des informations provenant de 10 images successives

Wang et al.[25] par exemple. Simonyan et al. [26] fusionnent deux CNN : un CNN spatial entraîné sur des images issues de vidéos et un CNN temporel entraîné sur des flux optiques issus de ces mêmes vidéos. Cela leur permet d'avoir des informations sur le mouvement ainsi que sur l'apparence. La fusion de données des sorties des deux CNN se fait soit par SVM, soit par moyennage. Ils obtiennent de meilleurs résultats que Karpathy et al. [24] et des résultats similaires à ceux obtenus par Wang et al.[25]. Les données provenant des caméras de profondeur ont dans un premier temps été utilisées pour reconnaître des objets dans des images.

Socher et al. [27] ont notamment utilisé les données provenant du capteur RGB et du capteur de profondeur de la *Kinect* pour reconnaître des objets en utilisant un système composé de CNN et de RNN.

Les auteurs Du et al. [28] ont utilisé des RNN hiérarchiques pour reconnaître des actions dans des vidéos. Pour cela ils utilisent les squelettes des personnes filmées, divisent les parties du corps en 5 groupes qui alimentent chacun un RNN et qui sont finalement fusionnés pour prendre une décision finale. Les auteurs utilisent notamment une base de données créée à partir de squelettes calculés avec des données de profondeur provenant de la *Kinect*.

Kang et al. [29] ont utilisé des CNN pour reconnaître les lettres de l'alphabet du langage des signes en utilisant des données de profondeur. Néanmoins, dans ce cas, il s'agit plus de la reconnaissance de postures que de gestes dynamiques.

Molchanov et al. [30] ont utilisé des CNNs et des données provenant du capteur RGB et *depth* d'une *Kinect* pour reconnaître des gestes afin de limiter l'utilisation de bouton dans l'habitacle d'une voiture. Néanmoins, la combinaison des données provenant des caméras de profondeur pour faire de la reconnaissance de gestes avec des méthode de Deep Learning a été encore un domaine peu étudié mais qui semble être prometteur.

Gillian et al. [31] proposent une méthode, appelée ND-DTW, pour calculer le chemin optimal entre deux signaux à N-dimensions. Une première utilisation pour la reconnaissance des

gestes a été faite par Gavrilu et Davis [32]. Après avoir déterminé la posture 3D d'une personne portant des marqueurs, ils utilisent les angles 3D des articulations comme caractéristiques pour reconnaître des gestes grâce aux DTW.

Liu et al. [33] utilisent les DTW pour reconnaître des actions en utilisant des données issues d'accéléromètres. Pour permettre au système de s'adapter à l'utilisateur, ils permettent de modifier le geste "modèle" par une mise à jour de celui-ci lorsqu'il est mal reconnu.

Sempena et al. [34] reconnaissent des gestes comme "taper des mains", "boxer" et "courir" en utilisant des DTW. Chaque geste modèle est décrit par une suite de quaternions représentant les angles de chacune des 15 articulations issues du squelette calculé grâce au SDK de la *Kinect*. Reyes et al [35] utilisent également des données issues de la *Kinect* pour reconnaître des gestes avec des DTW. Ils utilisent directement les positions 3D des articulations. Pour chaque DTW, les positions des articulations sont pondérées en fonction de leur impact dans le geste effectué. Cette méthode permet d'augmenter jusqu'à 13% le taux de reconnaissance par rapport à un DTW classique, c'est à dire sans pondération.

Calinon et Billard [36] ont proposé d'utiliser la reconnaissance de gestes, avec des HMMs, pour permettre aux robots d'apprendre des gestes en observant des exemples. Ils proposent une méthodologie pour reproduire des gestes appris. Les auteurs utilisent des capteurs inertiels fixé à l'utilisateur/le démonstrateur pour capturer le geste, la reconnaissance se fait avec l'algorithme Forward-Backward. L'algorithme de Viterbi est ensuite utilisé pour générer à nouveau des gestes.

Conclusion

Au cours de ce chapitre, on a examiné la notion d'IA qui caractérise la reconnaissance automatique des gestes, tout en évoquant les détails de l'apprentissage profond (Deep Learning), et puis encore passant à la notion des réseaux de neurones artificiels et ses différents types et leurs principes. Ensuite, on a donné quelques détails concernant les différents concepts de la reconnaissance des gestes tel que la segmentation et leur approche. Ces techniques deviennent essentielles car au cours de dernières années il y avait eu un regain énorme de l'utilisation des techniques de segmentation classification implémentées par plusieurs techniques de L'IA, et en plus beaucoup de recherches ont été faites dans le domaine de la reconnaissance automatique des gestes et la vision artificielle.

La reconnaissance automatique ne présente entièrement pas un processus parfait et intégrale, qui seulement à partir le modèle proposé est capable de traiter et reconnaître un spectre si variant de gestes complexe. Ce domaine reste toujours le sujet d'y mener d'améliorations requises pour atteindre des résultats plus précises et convenables.

Il faudra donc introduire une nouvelle approche qui sert à explorer de nouvelles solutions au problème consistant à détecter les gestes complexe.

Le chapitre suivant sera une présentation détaillée de l'architecture globale du système.

Chapitre 3

Contribution et Expérimentation.

Introduction

Le chapitre précédent est une introduction aux différentes techniques utilisées dans le traitement d'images et plus particulièrement la segmentation d'images et la reconnaissance des gestes dans le domaine vision par ordinateur.

Les dernières années ont connu un regain énorme de l'utilisation des techniques de IA tel que la segmentation et la reconnaissance implémentées par l'approche de Deep Learning comme les réseaux de neurones et beaucoup d'autres recherches sont faites dans le domaine de la vision artificielle.

De plus, le domaine vision par ordinateur comme ce cas d'étude de la reconnaissance des gestes de la main en temps réel n'échappe pas à la règle. Cela étant dit, il est toujours intéressant d'explorer de nouvelles solutions afin de les mener à ce problème consistant à détecter les gestes d'une façon rapide et précise.

Le présent chapitre sera dédié à la présentation de l'architecture du modèle de reconnaissance des gestes, ainsi que l'illustration de différents outils qui sont à l'aide d'exécuter les expérimentations prises en œuvre lors de la réalisation de ce projet de recherche.

1-Conception globale

Le but souhaité de ce projet consiste en une implémentation d'un système de reconnaissance des gestes de la main, et cela est afin d'améliorer l'interaction homme-machine ou entre hommes en facilitant particulièrement la communication entre un handicapé et un homme normal.

Ce dernier chapitre illustre le choix qu'on a proposé, et qui accumule les techniques et les méthodes utilisées pour la réalisation de ce système. Ci-après est une représentation générale du système de reconnaissance des gestes de la main qu'on a réussi à construire.

Dans cette section, on va mettre en forme généralement le système réalisé. On va traiter les données "images" qui seront divisées en deux parties 80% données d'entraînement et 20% données de validation. Ensuite, un modèle d'apprentissages sera formulé, qui prend les entrées de données traitées à l'étape passée et s'entraîne sur une partie des données (données d'entraînement) la figure suivante présente une vue globale du système réalisé.

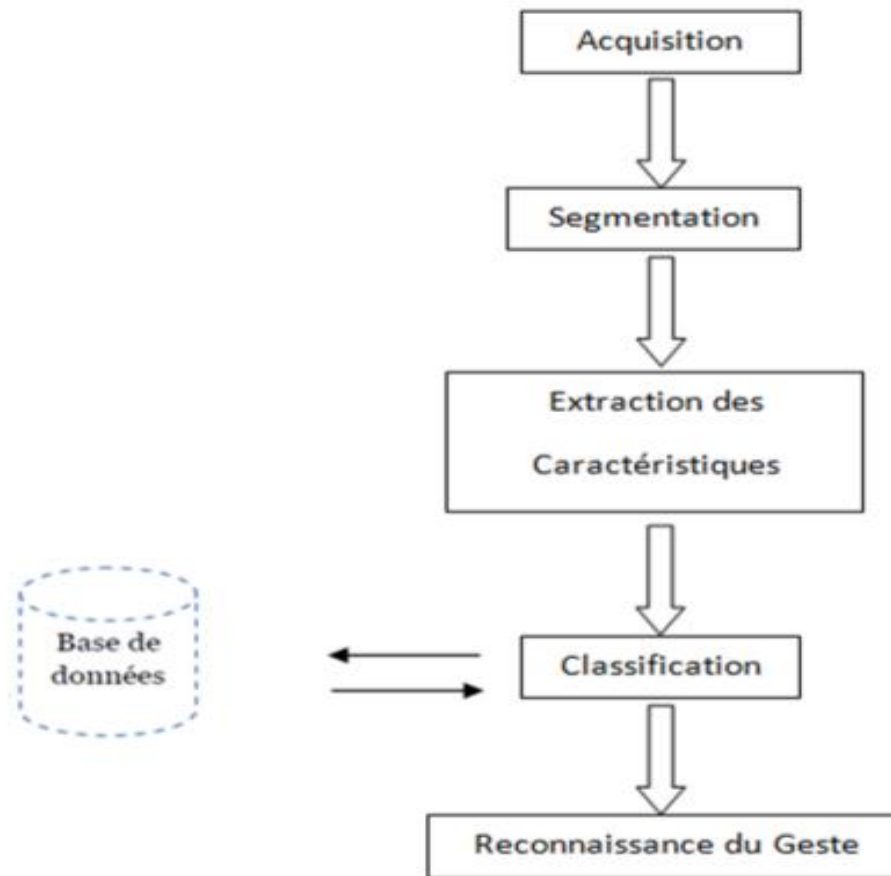


Figure 3.1: schéma générale de notre système

1.1-L'acquisition des images (première étape)

Au temps contemporain les ordiphones, ou communément connus sous le nom **les smartphones** sont les produits technologiques les plus répandus aux marchés. Ces appareils sont dotés des capteurs gagnant de plus en plus de puissance et précision, ce qu'il se rend essentiels pour certains d'en prendre profit.

La base de données du travail réalisé est formulée à l'aide d'un smartphone, d'où on a pris des images en gestes issues de l'anglais, avec haute résolution afin de rendre facile la détection de tel gestes.

Cependant, le fonctionnement de l'application de reconnaissance des gestes de la main dont on a construit, repose sur des images en haute qualité prises en temps réel à l'aide d'un Webcam. Le processus de l'identification commence par la prise d'une image par la webcam, ensuite on la traite pour la détection d'un geste de la main, tout en analysant ses caractéristiques, pour arriver finalement à la reconnaître.

On est arrivé finalement à identifier sept types de gestes convenablement y compris : *Stop, help me, I am lost.*

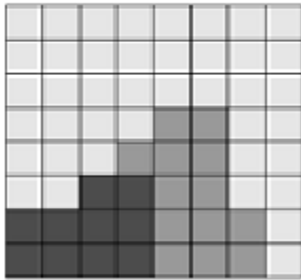


Figure 3.2.: Echantillon de donnée

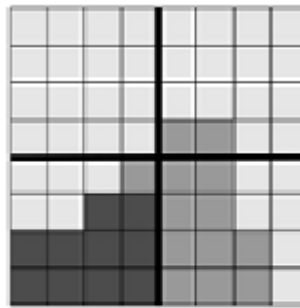
1.2-La segmentation (deuxième étape)

On a introduit la méthode de segmentation de division/fusion (ou split and merge en anglais) appartenant à cette famille manipulant directement des régions. Soit-elle partie d'une première partition de l'image, qui est ensuite modifiée en divisant ou regroupant des régions, ; soit elles partent de quelques régions, qui sont amenées à croître par incorporation de pixels jusqu'à ce que toute l'image soit couverte, et on parle alors de méthodes par croissance de régions. La figure suivante montre la fonctionnent de cette méthode.

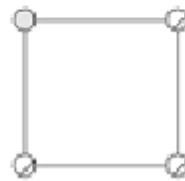
Image d'origine
division



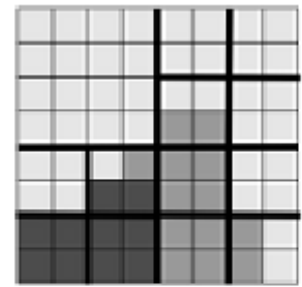
Première division (Split)



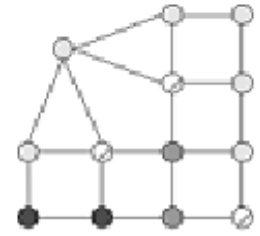
B



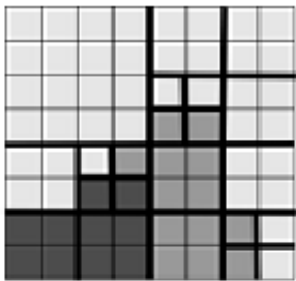
Deuxième



C



Troisième division



D

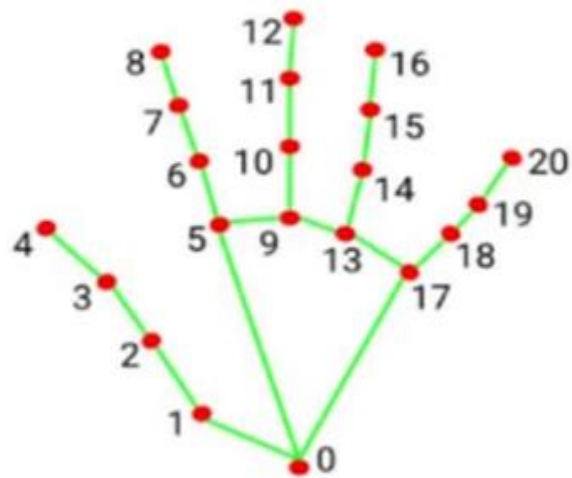
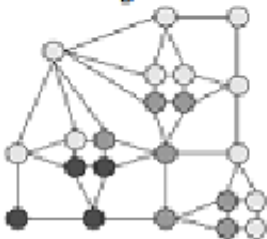


Figure 3.3: le processus de fonctionnement de segmentation et leur résultat

1.2.1-L'algorithme de segmentation

Créer une liste S de points de départ.

Pour chaque pixel P dans la liste S

 Si le pixel P est déjà associé à une région, alors prendre le pixel P suivant

 Créer une nouvelle région [R]

 Calculer la valeur moyenne de [R]

 Créer la liste [N] des pixels voisins du pixel P

 Pour chaque pixel 'Pn' dans la liste [N]

 Si ('Pn' n'est pas associé à une région et R+Pn est homogène) alors

 Ajouter le pixel Pn dans la région R

 Ajouter les pixels voisins de 'Pn' dans [N]

 Recalculer la valeur moyenne de [R]

 Fsi

 Fin pour

 Fsi

Fin pour.

1.3-Extraction des caractéristiques (troisième étapes)

Le processus de l'extraction des caractéristiques renvoie un total de 21 points clés pour chaque main détectée la figure suivante résumé les 21 point

- | | |
|-----------------------|-----------------------|
| 0. WRIST | 11. MIDDLE_FINGER_DIP |
| 1. THUMB_CMC | 12. MIDDLE_FINGER_TIP |
| 2. THUMB_MCP | 13. RING_FINGER_MCP |
| 3. THUMB_IP | 14. RING_FINGER_PIP |
| 4. THUMB_TIP | 15. RING_FINGER_DIP |
| 5. INDEX_FINGER_MCP | 16. RING_FINGER_TIP |
| 6. INDEX_FINGER_PIP | 17. PINKY_MCP |
| 7. INDEX_FINGER_DIP | 18. PINKY_PIP |
| 8. INDEX_FINGER_TIP | 19. PINKY_DIP |
| 9. MIDDLE_FINGER_MCP | 20. PINKY_TIP |
| 10. MIDDLE_FINGER_PIP | |

Figure 3.4 : le résultat de processus de l'extraction des caractéristiques de geste

2-Classification et reconnaissance des gestes

Le modèle de classification et reconnaissance des gestes basés sur CNN sont principalement applicables aux systèmes de recommandation. Les modèles sont utilisés pour la détection d'objets haute performance et la reconnaissance des objets à partir des images. Cette modelé divise l'image en un système de grille, et chaque grille détecte les objets en elle-même. Ils peuvent être utilisés pour découvrir des objets en temps réel sur la base de flux de données. Il nécessite très peu de ressources de calcul.

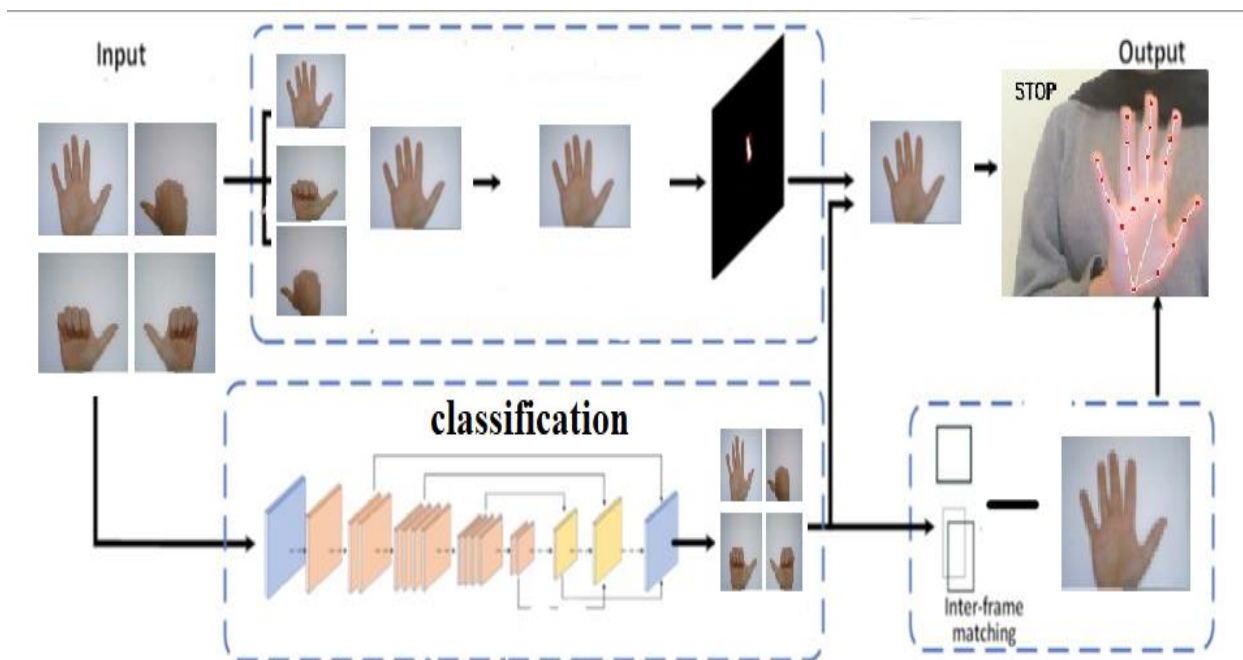


Figure 3.5: l'architecture global de classification et reconnaissance des gestes

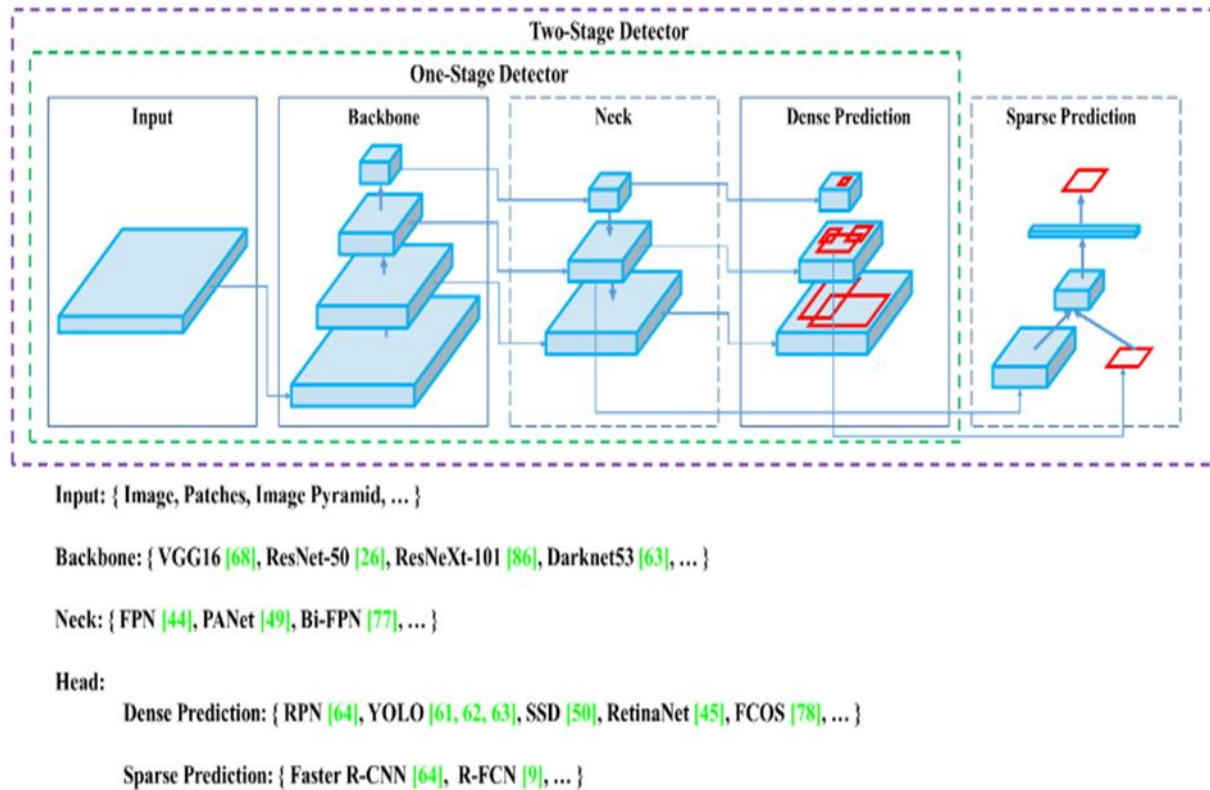


Figure 3.6 : l'architecture détaillée de classification

L'architecture détaillée de classification. Il se compose de trois parties : (1) Backbone : CSP Darknet, (2) cou (Neck) : PANet et (3) Tête (Head) . Les données sont d'abord entrées dans CSP Darknet pour l'extraction de caractéristiques, puis transmises à PANet pour les caractéristiques la fusion.

➤ Backbone

Le modèle de backbone est principalement utilisé pour extraire les caractéristiques clés de l'image d'entrée. CSP (Cross-Phase Partial Networks) est utilisé comme épine dorsale dans notre modèle pour extraire de riches propriétés utiles de l'image d'entrée.

➤ Cou (neck)

Le modèle de cou est principalement utilisé pour créer des pyramides distinctives. Des modèles pyramidaux distinctifs aident à généraliser avec succès lorsqu'il s'agit de mesurer des objets. Il aide à reconnaître le même objet de différentes tailles et échelles. Les pyramides distinguées sont très utiles pour aider les modèles à fonctionner efficacement sur des données inédites. D'autres modèles, tels que FPN, BiFPN et PANet, utilisent différents types d'approches de hiérarchie des fonctionnalités.

➤ Tête (Head):

La tête du modèle est principalement responsable de l'étape de détection finale. Utilise des boîtes liées pour créer des vecteurs de sortie finaux avec des probabilités de classe, des scores d'objet et des boîtes englobantes.

3-Les outils de développement utilisés

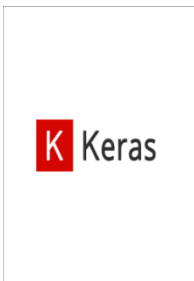
Dans cette partie, on présente en première partie le langage de programmation python utilisé dans l'implémentation de model, ainsi que les outils logiciel tel que les bibliographies importées : Keras, Tensorflow, Matplotlib, Numpy, Sklearn, tensorboard, OpenCV,... Et le Matériel utilisé pour implémenter notre modèle.

3.1-Python 3.7



Python est un langage de programmation interprété, multiparadigme et multiplateformes. Il favorise la programmation impérative structurée, fonctionnelle et orientée objet. Il est conçu pour optimiser la productivité des programmeurs en offrant des outils de haut niveau et une syntaxe simple à utiliser

3.2-Keras 2.1.0



Keras est une bibliothèque open source écrite en python, La bibliothèque Keras permet d'interagir avec les algorithmes de réseaux de neurones profonds et d'apprentissage automatique, notamment Tensorflow3, Theano, Microsoft Cognitive Toolkit4 ou Plaid ML Conçue pour permettre une expérimentation rapide avec les réseaux de neurones profonds, elle se concentre sur s ergonomie, sa modularité et ses capacités d'extension

3.3-Tensorflow 1.15.0



Tensorflow est un outil open source d'apprentissage automatique développé par Google. Le code source a été ouvert le 9 novembre 2015 par Google et publié sous licence Apache. Il est fondé sur l'infrastructure DistBelief, initiée par Google en 2011, et est doté d'une interface pour Python, Julia et R2.

3.4-Numpy 1.21.6



Numpy est une bibliothèque pour langage de programmation Python, destinée à manipuler des matrices ou tableaux multidimensionnels ainsi que des fonctions mathématiques opérant sur ces tableaux. Plus précisément, cette bibliothèque logicielle libre et open source fournit de multiples fonctions.

3.5-Matplotlib



Matplotlib est une bibliothèque du langage de programmation Python destinée à tracer et visualiser des données sous formes de graphiques. Elle peut être combinée avec les bibliothèques python de calcul scientifique Numpy et SciPy.

3.6-scikit-image



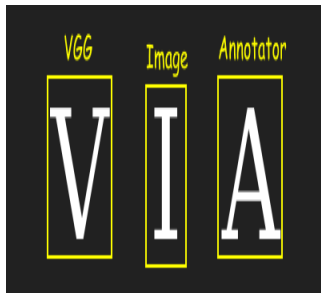
scikit-image est une collection d'algorithmes pour le traitement d'images. Il est disponible gratuitement et sans restriction

3.7-OpenCV



Open CV est une bibliothèque graphique libre, initialement développée par Intel, spécialisée dans le traitement d'images en temps réel. La société de robotique Willow Garage et la société ItSeez se sont succédé au support de cette bibliothèque.

3.8-VGG Image Annotator



VGG Image Annotator est un logiciel d'annotation manuelle simple et autonome pour l'image, l'audio et la vidéo. VIA s'exécute dans un navigateur Web et ne nécessite aucune installation ou configuration. Le logiciel VIA complet tient dans une seule page HTML autonome d'une taille inférieure à 400 kilo-octets qui s'exécute comme une application hors ligne dans la plupart des navigateurs Web modernes.

3.9-Anaconda



Anaconda est une distribution scientifique de Python : c'est-à-dire qu'en installant Anaconda, vous installerez Python, Jupyter Notebook (que nous présenterons plus en détail au prochain chapitre) et des dizaines de packages scientifiques, dont certains indispensables à l'analyse de données !

3.10-PyCharm



PyCharm est un environnement de développement intégré utilisé pour programmer en Python. Il permet l'analyse de code et contient un débogueur graphique. Il permet également la gestion des tests unitaires, l'intégration de logiciel de gestion de versions, et supporte le développement web avec Django.

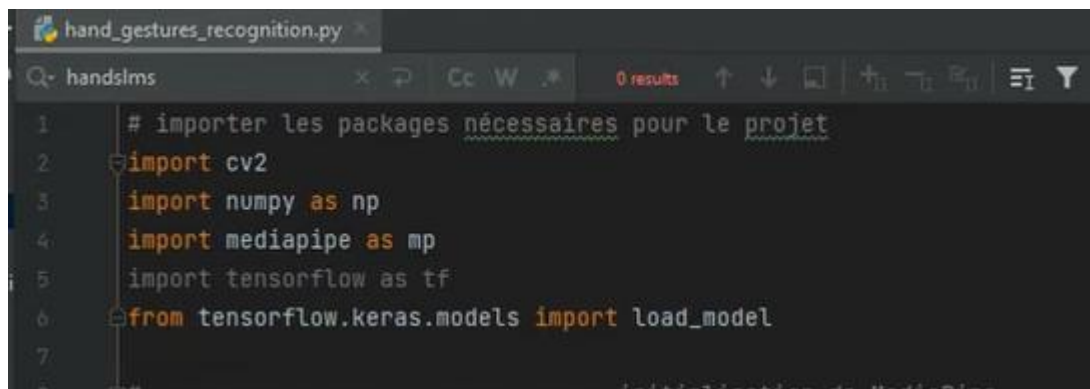
4-Le matériel

Le *Deep Learning* est un domaine qui nécessite énormément des calculs intenses, et la disponibilité de ressources dédiées, notamment de GPU, a un impact fondamental sur l'expérience de l'utilisateur. En l'absence de ces ressources, il faudra beaucoup de temps pour apprendre de ses erreurs, ce qui peut être décourageant. C'est pourquoi on a choisi la plateforme *Anaconda* avec l'*IDE PyCharm* pour mener les expérimentations. La configuration matérielle utilisée comprend un CPU i7, 12 Go de RAM et un GPU de 6 Go.

Après avoir défini la conception du système visé, ainsi que les architectures de modèles et les technologies utilisées dans cette recherche, on est passés à la phase d'implémentation. Dans cette phase, on va détailler la mise en pratique de ces technologies, et exploiter leurs configurations pour effectuer la détection et la reconnaissance des gestes.

5-Importer les bibliothèques

On a utilisé le modèle prédéfini *mediapipe* en y appliquant des petites modifications au sein de ses propres paramètres, et cela pour répondre aux besoins souhaités.

A screenshot of a code editor window titled 'hand_gestures_recognition.py'. The editor shows a search bar with 'handslms' and '0 results'. The code content is as follows:

```
1 # importer les packages nécessaires pour le projet
2 import cv2
3 import numpy as np
4 import mediapipe as mp
5 import tensorflow as tf
6 from tensorflow.keras.models import load_model
7
```

Figure 3.7: Import des bibliographies

La configuration appliquée sur notre modèle de reconnaissance des gestes de la main.

```
#-----initialisation de MediaPipe
#Exécuter l'algorithme de reconnaissance de la main
mpHands = mp.solutions.hands
#Configurer le modèle (détecter une seule main)
hands = mpHands.Hands(max_num_hands=1, min_detection_confidence=0.7)
#dessinez les points clés détectés pour la main
mpDraw = mp.solutions.drawing_utils

#----- initialisation de TensorFlow
#Charger le modèle de reconnaissance de geste pré-entraîné par TensorFlow
model = load_model('mp_hand_gesture')
```

Figure3.8: La configuration de modèle

```
#----- Charger les noms de classe
# Le fichier gesture.names contient le nom des classes de mouvements.
f = open('gesture.names', 'r')
classNames = f.read().split('\n') # lire le fichier en utilisant la fonction read().
f.close()
print(classNames)
```

Figure 3.9 : Accès aux données

```

# #----- Initialisation du caméra the webcam
cap = cv2.VideoCapture(0)

while True:
    # Lire chaque image de la webcam
    _, frame = cap.read()
    x, y, c = frame.shape

    # Retournez le cadre verticalement
    frame = cv2.flip(frame, 1)

    # Convertir le format des images RGB en format RVB
    framergb = cv2.cvtColor(frame, cv2.COLOR_BGR2RGB)

    # Obtenir la prédiction du point de repère de la main
    result = hands.process(framergb)

    className = ''

```

Figure 3.10: Accès aux web cam de pc pour collecte les images données

```

# -----reconaitre geste de main

# Predict gesture
prediction = model.predict([landmarks])
# print(prediction)
classID = np.argmax(prediction)
className = classNames[classID]

# montrer la prédiction sur le cadre
cv2.putText(frame, className, (50, 50), cv2.FONT_HERSHEY_DUPLEX,
            1, (0, 0, 0), 2, cv2.LINE_AA)

# Afficher la sortie finale
cv2.imshow("Hand gesture", frame)

if cv2.waitKey(1) == ord('q'):
    break

# relâchez la webcam et détruisez toutes les fenêtres actives
cap.release()

cv2.destroyAllWindows()

```

Figure 3.11: Proceccos de reconnaissance des gestes

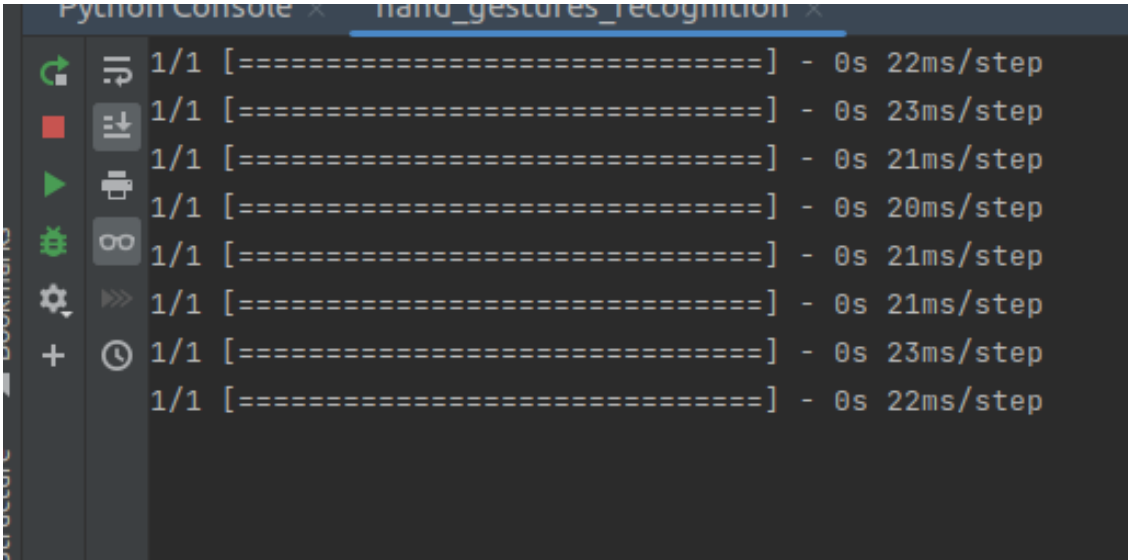
```
#-----traiter le resultat
#Vérifier si une main est détectée
if result.multi_hand_landmarks:
    landmarks = []
    #Parcourir chaque détection
    for handlist in result.multi_hand_landmarks:
        for lm in handlist.landmark:
            # print(id, lm)

            lmy = int(lm.y * y)
            lmx = int(lm.x * x)

            #Stocker les coordonnées sur une liste
            landmarks.append([lmx, lmy])

#Dessiner des repères sur des cadres
mpDraw.draw_landmarks(frame, handlist, mpHands.HAND_CONNECTIONS)
```

Figure 3.12.: Traitement de résultat de reconnaissance des gestes



```
Python Console x hand_gestures_recognition x
1/1 [=====] - 0s 22ms/step
1/1 [=====] - 0s 23ms/step
1/1 [=====] - 0s 21ms/step
1/1 [=====] - 0s 20ms/step
1/1 [=====] - 0s 21ms/step
1/1 [=====] - 0s 21ms/step
1/1 [=====] - 0s 23ms/step
1/1 [=====] - 0s 22ms/step
```

Figure 3.13: Résultats de reconnaissance des gestes temps de réponse

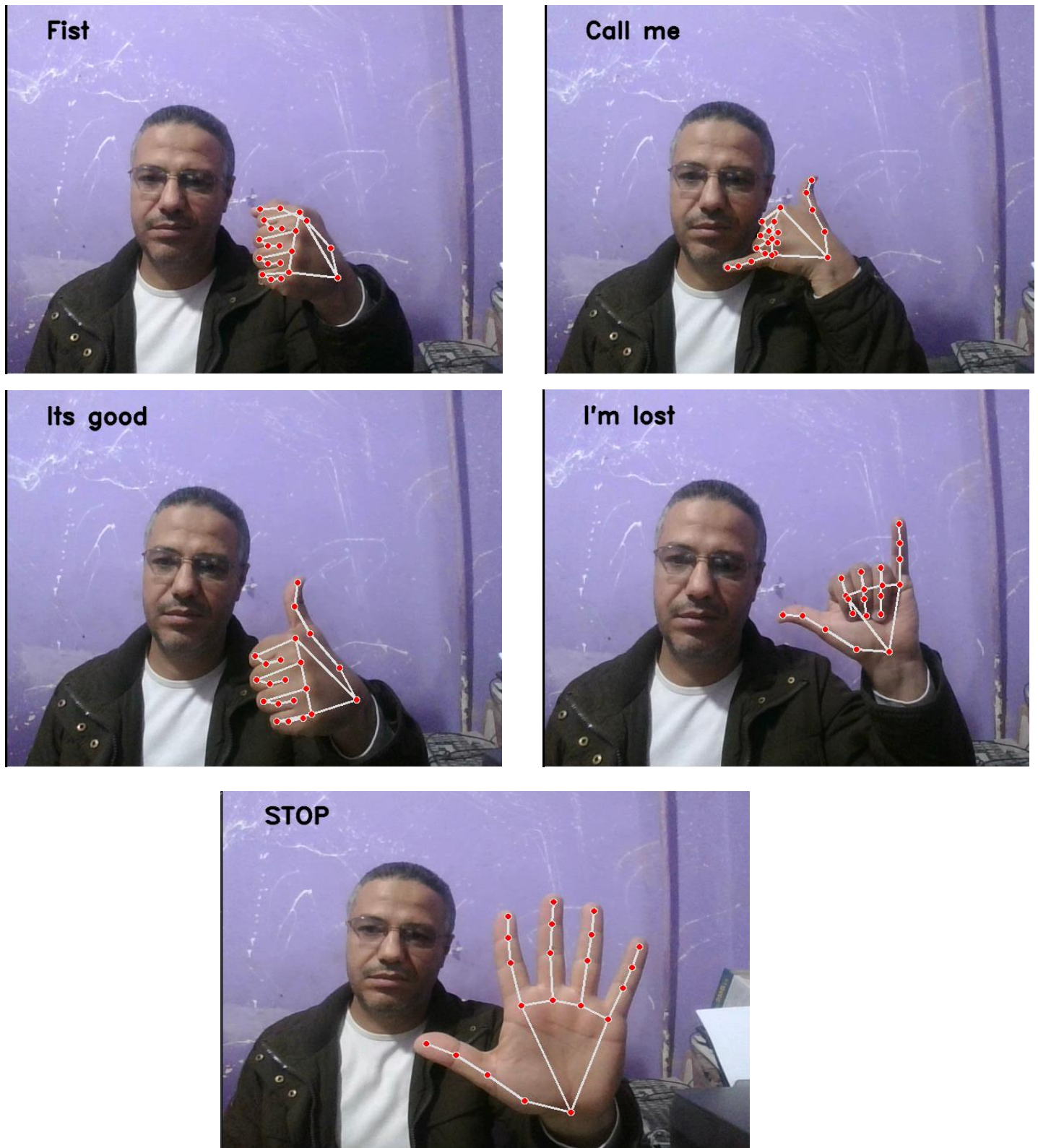


Figure 3.14: Résultats de reconnaissance des gestes

Conclusion

Au cours de ce chapitre on a entamé la mise en œuvre de la conception des systèmes proposés, ainsi la formulation de l'architecture de modèles utilisés lors de l'établissement de cette contribution à la recherche.

La méthodologie suivie lors de la réalisation du présent travail s'appuie en premier lieu sur une vue d'ensemble profonde de l'architecture proposé. Ensuite, on passe en revue les différentes parties composant l'architecture proposée. On conclut par la présentation de différents outils ayant avoir exploitées durant la réalisation du travail.

La partie d'expérimentation comprit une phase d'implantation, d'où on présente en détail la mise en pratique de ces technologies, tout en exploitant les données obtenues, pour qu'on fasse reconnaître les gestes de la main en temp réel.

Conclusion Générale

Le traitement des images (machine Learning ou Deep Learning) a connu récemment un grand essor. En effet il s'intervient extensivement dans plusieurs domaines et spécialités, quelles se soient liés à la recherche ou dans monde professionnel. Le domaine médical bénéficie vivement de cette progression technologique, d'où on distingue des applications fondées sur l'intelligente précision, qui est dédiées aux gens dépourvus de quelques caractéristiques physiologique courantes, tel que les sourds mués.

Cette recherche s'inscrit dans le cadre d'une contribution basée sur des algorithmes d'intelligence artificielle construites dans le but de reconnaître des gestes de main, qui permettent éventuellement la reconnaissance des objets et la traitement des images.

L'actuelle recherche se focalisera sur un des problèmes majeurs de traitement d'images. Le système qu'on va construire permettra la reconnaissance de la langue des signes par exemple le signes stop, call me ...etc.

On va commencer par fournir un aperçu complet sur langue des signes et de ses types exister, sa prévalence. Ensuite, on va effectuer une étude détaillée sur les dernières méthodologies et applications actuelles des chercheurs dans l'IA tel que la segmentation, la classification, DL.

On va détailler le modèle proposé, qui est fondé à partir deux algorithmes qui sont les plus répandues : la segmentation, la classification. Tout cela est consacré à construire un système automatique plus précis. Malgré la petite taille de données utilisé, les résultats obtenus sont très encourageants.

Ce travail établit un point de départ d'un projet d'intérêt général. Outre, on peut fournir les idées suivantes, qui peuvent s'intervenir à la réalisation des futurs œuvres, qui sont :

- L'amélioration de la reconnaissance des gestes de la main, tout en les transférant sous forme de Song ;
- La reconnaissance de mots et de phrases de la langue de signes et leur retranscription en langage écrit ;
- Le processus inverse qui consiste à traduire des mots et des phrases en langue des signes ;
- Le dépoilement de l'application sur mobiles et tablettes et sa mise en disponibilité sur le marché algérien.

Références Bibliographiques

- [1] : Decrouez, M., Dupont, R., Gaspard, F., Devernay, F., & Crowley, J. L. (2011, June). Modélisation explicite des objets et de l'environnement en combinant les approches topologique et métrique pour la localisation. In ORASIS 2011-13e Congrès des jeunes chercheurs en vision par ordinateur (p. 8p).
- [2] : BERRACHED, C. (2014). SYSTEME DE RECONNAISSANCE DE Gestes de la Main (Doctoral dissertation).
- [3]: Cadoz, C. (1994). Le geste canal de communication homme/machine: la communication" instrumentale". Revue des Sciences et Technologies de l'Information-Série TSI: Technique et Science Informatiques, 13(1), 31-61.
- [4]: Thomet, J. (2009). Une vue d'ensemble de la reconnaissance de gestes. Département d'informatique-Université de Fribourg, Séminaire «Gesture recognition», Semestre d'automne.
- [5]: Karam, M. (2005). A taxonomy of gestures in human computer interactions.
- [6]: Granier, O., Bertrand, R., & Ochs, M. (2021). La persuasion dans les interactions humain-humain et humain-machine: une revue de la littérature. WACAI 2021.
- [7]: T. Starner, J. Weaver et A. Pentland : Real-time american sign language recognition using desk and wearable computer based video. IEEE trans. on Pattern Analysis and Machine Intelligence, 20(12):1371–1375, 1998.
- [8]: J. Segen et S. Kumar : Fast and accurate 3D gesture recognition interface. In Proc. of the IEEE Int. Conf. on Pattern Recognition, 1998
- [9]: N. Otsu : A threshold selection method from gray level histograms. IEEE trans. on Systems, Man, and Cybernetics, 9:62–66, mars 1979.
- [10] : Jégou, J. F. (2013). *Interaction basée sur des gestes définis par l'utilisateur: Application à la réalité virtuelle* (Doctoral dissertation, Ecole Nationale Supérieure des Mines de Paris).
- [11] : Simon Conseil, thèse de doctorat, Université paul cézanne faculté des sciences et techniques, université paul cézanne aix-marseille iii, Année 2008
- [12] : Billinghamurst, Marc, & Bill Buxton. 2011. "Gesture Based Interaction." In Haptic Input, Cambridge University Press, 35
- [13] : Bouisset, Simon, Eric Berton, Guillaume Rao, & David Amarantini. 2006. "Étude et Modelisation Du Mouvement Humain." In Le traité de la Réalité Virtuelle Volume 1, Paris, France: Les presses des Mines, 101–16
- [14] : Blandine Bril and Valentine Roux. Le geste technique. Réflexions méthodologiques et anthropologiques. 2002.

- [15] : Mertz, C., Vinot, J. L., & Etienne, D. (2002, October). Entre manipulation directe et reconnaissance de l'écriture: les gestes écologiques. In *Ergo-IHM 2002, Ergonomie et informatique Avancée, Interaction Homme-Machine* (pp. pp-145).
- [16] : Ong, S. C., & Ranganath, S. (2005). Automatic sign language analysis: A survey and the future beyond lexical meaning. *IEEE Transactions on Pattern Analysis & Machine Intelligence*, 27(06), 873-891.
- [17] : R\U00E9Daction, L. (s. d.). Intelligence artificielle : définition, conseils, comparatifs, témoignages. . . intelligence artificiel. [https://www.journaldunet.com/solutions/dsi/1493139-intelligence-artificielle/#:%7E:text=L%27intelligence%20artificielle%20\(IA\),de%20la%20m%C3%A9moire%20et%20l%27](https://www.journaldunet.com/solutions/dsi/1493139-intelligence-artificielle/#:%7E:text=L%27intelligence%20artificielle%20(IA),de%20la%20m%C3%A9moire%20et%20l%27)
- [18] : La Classification d'images d'insectes ravageurs en utilisant le Deep Learning. (2019–2020).<https://www.ummo.dz/dspace/bitstream/handle/ummo/13208/Lounis%20K.%3B%20Moussi%20D..pdf?sequence=1&isAllowed=y>
- [19] : SPRH LABS. (2021, 9 décembre). Understanding Deep Learning : DNN, RNN, LSTM, CNN and R-CNN. Medium. <https://medium.com/@sprhlab/understanding-deep-learning-dnn-rnn-lstm-cnn-and-r-cnn-6602ed94dbff>
- [20] : Zucker, S. W. (1976). Region growing: Childhood and adolescence. *Computer graphics and image processing*, 5(3), 382-399.
- [21] : Bernsen, J. (1986). Dynamic thresholding of grey-level images fcV. In *Proceeding of the 8 International Conference O11 Pattern Rec-Gn Ition* (pp. 1251-1255).
- [22] : Deslandes, A., Dorizzi, B., & Horain, P. (2002). Reconnaissance du Geste Humain par Vision Artificielle: Application à la Langue des Signes. *rapport de stage dans le cadre du DEA IARFA: intelligence artificielle, reconnaissance de formes et application de l'université Paris6, Pierre et Marie Curie*.
- [23] : huiwang Ji, Wei Xu, Ming Yang, and Kai Yu. 3D Convolutional Neural Networks for Human Action Recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 35(1) :221–231, jan 2013
- [24] : Andrej Karpathy, George Toderici, Sachin Shetty, Tommy Leung, Rahul Sukthankar, and Li Fei-Fei. Large-scale video classification with convolutional neural networks. In *Computer Vision and Pattern Recognition (CVPR), 2014 IEEE Conference on*, pages 1725–1732, 2014
- [25] : Heng Wang, Alexander Kläser, Cordelia Schmid, and Cheng-Lin Liu. Dense Trajectories and Motion Boundary Descriptors for Action Recognition. *International Journal of Computer Vision*, 103(1) :60–79, mar 2013
- [26] : Karen Simonyan and Andrew Zisserman. Two-Stream Convolutional Networks for Action Recognition in Videos. 2014
- [27] : Richard Socher, Brody Huval, Bharath Bhat, Christopher D Manning, and Andrew Y Ng. Convolutional-Recursive Deep Learning for 3D Object Classification. In *Advances in Neural Information Processing Systems*, pages 665–673, 2012
- [28] : Yong Du, Wei Wang, and Liang Wang. Hierarchical Recurrent Neural Network for Skeleton Based Action Recognition. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1110–1118, 2015.
- [29] : Byeongkeun Kang, Subarna Tripathi, and Truong Q. Nguyen. Real-time sign language fingerspelling recognition using convolutional neural networks from depth

- map. In 2015 3rd IAPR Asian Conference on Pattern Recognition (ACPR), pages 136–140. IEEE, nov 2015
- [30]** : Pavlo Molchanov, Shalini Gupta, Kihwan Kim, and Jan Kautz. Hand Gesture Recognition With 3D Convolutional Neural Networks. In The IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops, pages 1–7, 2015
- [31]** : Juan Antonio Corrales Ramón, Gabriel Jesús García Gómez, Fernando Torres Medina, and Véronique Perdereau. Cooperative tasks between humans and robots in industrial environments. InTech, 2012.
- [32]** : D. M. Gavrila and L. S. Davis. Towards 3-D model-based tracking and recognition of human movement : a multi-view approach. In International Workshop on Automatic Face- and Gesture-Recognition. IEEE Computer Society, pages 272–277, 1995
- [33]** : Jiayang Liu, Lin Zhong, Jehan Wickramasuriya, and Venu Vasudevan. uWave : Accelerometer-based personalized gesture recognition and its applications. *Pervasive and Mobile Computing*, 5(6) :657–675, dec 2009
- [34]** : Samsu Sempena, Nur Ulfa Nur Ulfa Maulidevi, and Peb Ruswono Peb Ruswono Aryan. Human action recognition using Dynamic Time Warping. In Proceedings of the 2011 International Conference on Electrical Engineering and Informatics, pages 1–5. IEEE, jul 2011
- [35]** : Miguel Reyes, Gabriel Dominguez, and Sergio Escalera. Featureweighting in dynamic timewarping for gesture recognition in depth data. In 2011 IEEE International Conference on Computer Vision Workshops (ICCV Workshops), pages 1182–1188. IEEE, nov 2011
- [36]** : Sylvain Calinon and Aude Billard. Stochastic Gesture Production and Recognition Model for a Humanoid Robot. In Intelligent Robots and Systems, 2004.(IROS 2004). Proceedings. 2004 IEEE/RSJ International Conference on, pages 2769–2774, 2004