



REPUBLIQUE ALGERIENNE DEMOCRATIQUE ET POPULAIRE



MINISTERE DE L'ENSEIGNEMENT SUPERIEUR ET DE LA RECHERCHE SCIENTIFIQUE

UNIVERSITE DE 20 AOUT 1955 SKIKDA

FACULTE DES SCIENCES

DEPARTEMENT INFORMATIQUE

Mémoire

En vue de l'obtention du diplôme de

Master en informatique

Spécialité : Réseaux et systèmes distribués

Etude de méthode de reconnaissance de geste

Réalisé par : Khettaf Yasmine

Encadré par : Mr. Bouremel Abdelhakim

Année Universitaire 2021- 2022

Remerciements :

Je remercie Dieu le tout puissant.

En premier lieu, je tiens à remercier et exprimer ma profonde gratitude au chef de département de l'informatique ainsi l'ensemble des enseignants de mon cursus universitaire pour leurs encouragements, et leurs précieux conseils.

Je tiens à remercier très sincèrement, Mr. Bouremel de m'avoir donné cette chance aujourd'hui ainsi, tous les membres du jury d'avoir bien voulu accepter de faire partie de la commission d'examineur de mon travail.

Enfin, je souhaite particulièrement remercier mes parents Khettaf Malik, Saad rouana Mounira et ma sœur Hadjer qui ont toujours été la pour moi. Leur soutien inconditionnel et leurs encouragements ont été d'une grande aide.

Résumé

Dans de ce mémoire, nous nous sommes perchés sur le domaine de reconnaissance de geste et plus précisément de la reconnaissance de la langue des signes dans le but de d'intégrer les communautés des sourds et muets à la communication homme machine.

L'objectif de ce travail est d'étudier les différentes étapes et méthodes qui permettant d'obtenir un système de reconnaissance de geste. L'approche ici se déroule en deux étapes consécutives essentielles dans :

La première consiste à étudier le domaine de l'interaction homme machine qui nous permet d'obtenir un système de communication gestuelle et les différents gestes intégrer.

La deuxième consiste en l'acquisition du geste via une caméra d'un ordinateur par la méthode de segmentation du geste afin d'obtenir un système de reconnaissance du geste ensuite l'utilisation de l'apprentissage automatique dans ce système de reconnaissance.

Abstract

In this dissertation, we perched on the field of gesture recognition and more precisely sign language recognition with the aim of integrating the communities of the deaf and dumb into human-machine communication.

The objective of this work is to study the different steps and methods that make it possible to obtain a gesture recognition system. The approach here takes place in two essential consecutive steps in:

The first is to study the field of human-computer interaction that allows us to obtain a gestural communication system and the different gestures integrate.

The second consists of the acquisition of the gesture via a camera of a computer by the method of segmentation of the gesture in order to obtain a system of recognition of the gesture follows the use of machine learning in this recognition system.

Sommaire

Introduction générale

I. Chapitre1 : Une Interaction Homme-Machine gestuelle ?	
A. Introduction	2
B. L'interaction homme-machine :	2
C. Vers une interaction Homme-Machine gestuelle :	3
D. Les dispositifs d'interaction :	3
1. Dispositif d'entrée :	3
2. L'écran tactile :	4
3. Gant électronique :	5
4. Les caméras vidéo :	6
E. Applications et nouvelles possibilités d'interaction :	6
1. La Réalité Virtuelle :	6
2. Réalité augmentée :	7
3. Réalité mixte :	8
4. Le contrôle des robots :	8
5. Reconnaissance des nombres :	9
6. Reconnaissance de la langue de signe (LS) :	9
F. Applications grand public :	10
G. Les gestes :	11
1. Définition du geste :	11
2. Comment reconnaître un geste :	11
3. Catégorie de geste :	12
4. Classification des gestes de la main :	13
a) Selon Maria Karam et m.c schraefel :	13
(1) Les gestes déictiques :	13
(2) Les gestes sémaphoriques :	13

Sommaire

(3) Les gestes de manipulation :.....	14
b) Selon Quek:	14
c) Selon Kendon :.....	14
5. Fonctionnalité du geste :.....	15
H. Conclusion :.....	16
II. Chapitre2 : Système de reconnaissance de geste et la vision par ordinateur	
A. Introduction	17
B. La vision par ordinateur	17
C. Les images numériques et leurs traitements :.....	17
1. Définition d'images :	17
2. Types d'images :	18
3. Codage des couleurs d'une image numérique :.....	18
4. Caractéristiques d'une image numérique :.....	19
D. Etat de l'art de la reconnaissance des gestes de la main	19
1. Acquisition du geste :	20
a) La convolution d'image :	22
b) Filtrage d'une image :.....	22
(1) Le filtre Gaussien :.....	22
2. Traitement de données :.....	23
a) La segmentation :	23
(1) La segmentation par la détection de la couleur de peau :	23
(2) Seuillage (thresholding) :	25
(3) Le contour :	26
(4) Le codage :	28
b) Extraction des caractéristiques :.....	28
c) La reconnaissance :.....	28

Sommaire

(1) La modélisation :	29
(2) L'apprentissage :	29
E. Conclusion	30
III. Chapitre 3 : L'implémentation et Résultat	32
A. Introduction	32
B. La reconnaissance de la langue des signes :	32
1. Les caractéristiques :	32
2. Les paramètres :	33
C. L'intelligence artificielle est la reconnaissance gestuelle :	33
1. L'apprentissage automatique :	34
2. Types d'apprentissage :	34
a) Apprentissage symbolique :	34
b) Apprentissage numérique :	35
c) L'apprentissage selon l'intervention :	35
(1) Apprentissage supervisé :	35
(2) Apprentissage non-supervisé :	36
D. Environnement de développement :	36
E. Reconnaissance des langues des signes :	37
F. Conclusion	39
Conclusion Général :	40
IV. Webographie :	

Liste des figures

Liste des figures

Figure 1 : Dispositif d'entrée	4
Figure 2 : L'écran tactile	5
Figure 3 : Les gants numériques.....	5
Figure 4 : La réalité virtuelle.....	7
Figure 5 : la réalité augmentée	8
Figure 6 : Contrôle robot.....	9
Figure 7 : Reconnaissance de nombre.....	9
Figure 8 : Les gestes de la main.....	11
Figure 9 : Geste statique et dynamique	14
Figure 10 : codage d'une image numérique.....	19
Figure 11 : La reconnaissance de geste de la main	20
Figure 12 : L'exemple de kincet.....	21
Figure 13 : Traitement de donnée.....	23
Figure 14 : Le seuillage avec la méthode d'Otsu	26
Figure 15 : Segmentation de la main.....	28
Figure 16 : La langue des signes	32
Figure 17 : Reconnaissance du mot "Okey"	37
Figure 18 : La reconnaissance de la phrase "its good"	38
Figure 19 : Reconnaissance de la phrase "see you later"	38
Figure 20 : Reconnaissance de la phrase "call me"	39



Introduction Générale

Introduction générale

Depuis quelques années, l'intelligence artificielle (IA) connaît un regain d'intérêt sans précédent grâce à d'importantes avancées technologiques. IA fait référence à des systèmes ou des machines qui imitent l'intelligence humaine afin d'effectuer des tâches qui peuvent être améliorées en fonction des informations collectées grâce à l'itération, de plus elle est notamment liée à des domaines de l'apprentissage machine, qui étendent les capacités des ordinateurs et accroissent leurs performances. De nombreuses revues sont basées sur l'IA tel que (la reconnaissance de geste et d'image, traitement du langage, compréhension de la parole, robotique, etc.). Dans notre projet de fin d'étude nous allons étudier la reconnaissance de gestes de la main.

Les gestes de la main est défini comme un moyen de communication avec l'environnement dans l'intention d'interagir avec des objets (prendre, déplacer). Le gestuelle désigne l'ensemble des éléments d'informations transmis lors d'une communication face à face, c'est une manière naturelle et intuitif chez l'homme pour renforcer la parole en véhiculent d'une manière congruente le message. La recherche scientifique à permet d'imiter l'homme en réutilisant le mécanisme de geste et de reconnaissance du geste, afin de pouvoir réaliser l'interaction avec la machine ou entre les machines. Cependant, l'utilisation de la technologie progresse et à ainsi cédé la place à la mise en œuvre des nouveaux systèmes dédiés à la société sourde et auditive. Par conséquent, cette communauté à enfin trouvé sa place dans le domaine d'interaction et cela grâce aux différents systèmes de reconnaissance. Ce dernier a permis de représenter la langue des signes en prenant considération qu'il est un langage à part entière c'est-à-dire contient un alphabet, un lexique et une syntaxe. Depuis, l'homme cherche à interpréter les images et les vidéos afin que l'ordinateur ait la capacité de se voir (la vision par ordinateur).

En effet, Ce projet vise essentiellement à étudier la mise en œuvre des systèmes qui détectent les gestes de la main mais plus précisément la langue des signes, prédéfinies en temps réel grâce aux gestes produits par la main. Par conséquent, ces systèmes à pour objectif de donner la capacité aux utilisateurs de construire leurs propres gestes personnels afin que les problèmes rencontrés par ces personnes incapables de s'exprimer puissent être résolus avec une assistance technologique la chose qui aidera à franchir la barrière d'expression.



Chapitre I
Une Interaction
Homme-Machine gestuelle

Chapitre1 : Une Interaction Homme-Machine gestuelle

A. Introduction

Dans ce chapitre nous allons présenter d'une manière globale l'étude de la démarche stratégique de ce projet visant à connaître le lien entre le domaine d'interaction homme-machine et la communication gestuelle.

«Un système informatique peut-il utiliser les gestes comme périphérique d'interaction ?»

«Quelle sont les différents dispositifs classiques qui permet d'interagir avec un ordinateur ?»

«Qu'est-ce qu'un geste ? » «Comment reconnaître un geste ? ».

Ce sont les questions auxquelles nous allons tenter de répondre. En premier temps, nous allons d'abord présenter le domaine de l'interaction homme-machine gestuelle d'une façon général et les différents dispositifs qui permettent cette interaction. Ensuite, nous définissons le geste et comparons les différentes fonctionnalités associées avec ce canal de communication

B. L'interaction homme-machine :

L'interface entre l'homme et la machine prend une place grandissante dans la production connectée. Pour une collaboration étroite et sûre, et grâce à un langage commun la communication ce fait d'une manière intuitive et sans ambiguïté. Les humains interagissent avec les ordinateurs qui les entourent de plusieurs manières, l'interaction nécessite des interfaces graphiques s'inspirent principalement sur le paradigme WIMP, la croissance dans le domaine de l'interaction homme-machine a conduit à une augmentation de la qualité de l'interaction et a donné lieu de nombreux nouveaux domaines de recherche au-delà.

Ces recherches à pour but de trouver les moyens les plus efficaces, les plus accessibles et les plus intuitifs afin qu'un utilisateur puisse compléter une tâche le plus rapidement et le plus précisément possible. En imaginant de nouveaux moyens de communication avec un système informatique ou l'utilisateur n'est plus réduit au couple clavier/souris. Par conséquent, de la traditionnelle interface utilisateur graphique 'graphical user interface' (GUI) utilisé dans des applications de bureau, les navigateurs internet, les ordinateurs de poche et les kiosques informatiques, à l'interface perceptuelle 'perceptual user interface' (PUI) qui inclut l'utilisation des gestes de la main.

Chapitre1 : Une Interaction Homme-Machine gestuelle

C. Vers une interaction Homme-Machine gestuelle :

Depuis quelques années, les recherches effectuées sur le domaine de l'interaction homme machine visent à améliorer les performances des utilisateurs d'un système informatique en s'intéressant à son utilisation et non à lui-même.

Intuitive, identifiable, ergonomique, simple et efficace telles sont les exigences imposées aux concepteurs d'Interfaces Homme-Machine. L'ambition d'interactions avec des machines de manière intuitive et naturelle est plus grande que jamais à tel point que, la recherche en reconnaissance de gestes s'est attelée peu à peu au développement de systèmes capables de reconnaître les gestes de l'homme et de les interpréter. Les gestes apparaissent comme un moyen spontanés pour qu'une personne puisse communiquer avec son environnement, ils sont faciles à utiliser, rapides et correspondent à une réalité humaine. Il semble donc admissible de faire évoluer les systèmes informatiques pour prendre en compte ce moyen de communication plus naturels que l'utilisation des périphériques classiques (clavier et souris).

D. Les dispositifs d'interaction :

La plupart des systèmes d'exploitation sont basés sur le paradigme WIMP qui signifie (Window, Icon, Menu, Pointing device) d'autres extensions sont parfois utilisées telles que la substitution de « souris » pour le menu, ou « menu déroulant » et « pointage » pour le pointeur. Il existe aussi des interfaces haptiques, résultent de la stimulation de la peau résultant des mouvements actifs d'exploration de la main entrant en contact avec des objets en permettant un retour d'information à l'utilisateur, avec un retour de toucher ou un retour d'effort.

Les différents dispositifs qui créent un point de jonction et d'échange entre l'utilisateur et la machine sont : Les dispositifs d'entrée (clavier, souris et les caméras vidéo), les dispositifs d'entrée visuelle 2D (écran tactile), les dispositifs d'entrée visuelle 3D (gants de données).

1. Dispositif d'entrée :

Le périphérique d'entrée est un appareil physique avec laquelle l'utilisateur a le contact le plus direct. Le couple clavier /souris sont des dispositifs d'entrée standard génériques qui permettent de communiquer avec une grande variété d'applications, à taper sur un clavier et à cliquer sur des boutons, ouvrir des menus et même déplacer des icônes avec une souris. Il

Chapitre1 : Une Interaction Homme-Machine gestuelle

existe aussi la boule de commande, ou « trackball ». Le pavé tactile ou « touchpad » est un dispositif de pointage spécifique aux ordinateurs portables, permettant de remplacer la souris. Il s'agit d'une surface sensible de faible dimension, utilisant la capacité électrique.

Les joysticks et manettes sont deux périphériques d'entrée utilisés dans des jeux vidéo, les manettes permettent de contrôler le mouvement et les vues grâce à des curseurs et boutons, alors que le joystick permet d'agir en le bougeant la manche dans une direction.



Figure 1: Dispositif d'entrée

2. L'écran tactile :

Un écran tactile (*touch screen*, en anglais) est un dispositif informatique qui gère à la fois l'affichage et le pointage en remplaçant la souris par le doigt ou un stylet. Il permet ainsi de sélectionner certaines de ses zones par contact. Ces écrans se trouvent en générale sur les bornes de ventes de tickets de métro, les GPS ou bien-sûr les smartphones.

Son fonctionnement repose sur le fait que le corps humain peut emmagasiner une charge électrostatique qui est capable de modifier le champ électrique appliqué sur l'écran tactile. L'écran voit alors sa capacité électrique (quantité de charges) modifié localement (attirés par le doigt). Il existe différentes technologies pour les écrans tactiles : capacitive, résistive, infrarouge, à ondes de surface... De manière générale, les écrans tactiles souffrent de différents inconvénients : ils sont onéreux, limités en surface, sujet à l'usure et sensibles aux rayures (suivant la technologie employée).



Figure 2 : L'écran tactile

3. Gant électronique :

Un gant électronique (aussi appelé gant numérique ou gant de réalité virtuelle, ou encore, plus rarement, gant de données) est un périphérique intrusif le plus caractéristique du domaine du geste. Le gant comportant des capteurs, qui permet à un utilisateur de saisir presque naturellement un objet virtuel et de le manipuler, en numérisant en temps réel les mouvements de la main. Il est utilisé pour l'interface homme-machine dans la réalité virtuelle. Certains gants électroniques sont dotés d'un dispositif à retour tactile qui permet aux utilisateurs de ressentir ce qu'ils touchent.

Les gants de données sont utilisés de longue date pour la reconnaissance de la langue des signes car ils fournissent les positions précises et fiables des articulations de la main, mais l'inconvénient c'est qu'ils sont extrêmement couteux.



Figure 3 : Les gants numériques

Chapitre1 : Une Interaction Homme-Machine gestuelle

4. Les caméras vidéo :

Les caméras vidéo captent les mouvements de la main sans que l'utilisateur ne soit contraint de porter un équipement particulier, ou d'utiliser un périphérique dédié. Toutefois, pour certains systèmes, des marqueurs ou un gant coloré sont utilisés pour faciliter la détection des différentes parties de la main. La difficulté de cette approche est de mettre au point des traitements robustes pour interpréter le flux vidéo et extraire l'information utile de la grande quantité d'information disponible. Par ailleurs, une caméra ne fournit qu'une information 2d. Pour obtenir des informations en 3d, il faut utiliser deux ou plusieurs caméras, ou une modélisation 3d.

Par conséquent, les occultations sont un problème important, inhérent à la projection de l'espace 3d dans une image. Avec l'augmentation de la puissance des ordinateurs de bureau et l'apparition de caméras bon marché, il est désormais possible de développer des systèmes de reconnaissance de gestes fonctionnant en temps réel. Les caractéristiques telles que le taux de rafraîchissement ou la résolution varient d'une caméra à l'autre. Des valeurs élevées de ces caractéristiques sont avantageuses pour avoir une mise à jour fréquente des images et un niveau de détail important. Toutefois, un compromis est nécessaire car des valeurs trop importantes augmentent la complexité et le temps de traitement des données. Par ailleurs, les caractéristiques du capteur (CCD ou CMOS) et de l'optique ont une influence primordiale sur la qualité de l'image obtenue et la sensibilité à l'éclairage.

E. Applications et nouvelles possibilités d'interaction :

Cette section décrit diverses applications, en particulier les surfaces Interactives gestuelles et certaines applications récentes destinées au public.

1. La Réalité Virtuelle :

La réalité virtuelle est une simulation artificielle générée par ordinateur. Cela peut être une simulation d'un environnement, d'une activité, d'une situation. La réalité virtuelle se réfère à la création d'un monde virtuel dans lequel l'utilisateur peut interagir avec son environnement, cela peut être réalisé dans une pièce dédiée munie de plusieurs écrans ou d'un vidéoprojecteur. Le but est d'immerger l'utilisateur au sein d'un monde virtuel de façon qu'il puisse interagir avec des gestes et à ce que ses sens ne puissent plus réellement faire la différence entre la réalité et le virtuel.

Chapitre1 : Une Interaction Homme-Machine gestuelle

Le casque de réalité virtuelle tel que l'Oculus Rift ou autre sont indispensables, et pour que l'expérience soit optimale un casque audio serait également nécessaire. Il existe deux utilisations principales de la réalité virtuelle :

- Pour créer et améliorer l'immersion dans un jeu vidéo, pour le divertissement et tout simplement pour jouer (jeux vidéo, film 3D...).
- Pour améliorer les conditions d'entraînement pour des situations réelles. C'est en effet une façon de s'entraîner sans aucun risque et cela est très utile par exemple pour les pilotes de chasse ou les pilotes d'avion en général.



Figure 4 : La réalité virtuelle

2. Réalité augmentée :

La réalité augmentée va quant à elle permettre à un utilisateur d'enrichir et de compléter en quelque sorte ce qu'il voit, avec des éléments virtuels afin de se projeter et d'imaginer plus facilement une situation. Elle est utilisable sur téléphone portable ou tablette par exemple pour intégrer des meubles ou objets décoration à son intérieur en réalité augmentée, mais aussi directement sur des lunettes permettant une utilisation plus professionnelle de la réalité augmentée.

Le but de cette technologie est de rajouter une part de virtuel à notre réalité existante. Ainsi, la réalité augmentée est également utilisée dans l'application de la reconnaissance de gestes afin 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.

Chapitre1 : Une Interaction Homme-Machine gestuelle

- La combinaison de la réalité augmentée et de la réalité virtuelle a donné naissance à une nouvelle technologie hybride : la réalité mixte.



Figure 5 : la réalité augmentée

3. Réalité mixte :

Le développement Du nouveau casque Hololens par la collaboration Microsoft et la NASA à donner lieu à la naissance de cette réalité mixte. Cette réalité est définie par la capacité du casque Hololens à prendre en compte l'espace réel, et ce, grâce aux capteurs intégrés au casque, la position de l'utilisateur est également calculée en temps réel. De plus, l'ensemble permet d'interagir physiquement avec les éléments en 3D grâce à une série de gestes à effectuer, il s'agit d'une version améliorée de la réalité augmentée. Plus simplement, une couche de virtuelle se plaque sur le réel.

4. Le contrôle des robots :

Le contrôle des robots à l'aide de gestes est considéré comme l'une des applications les plus populaires. Plus intéressant à cet égard. Ils ont inventé un système utilisant les signes de numérotation, le système compte les cinq doigts de la main pour contrôler le robot. Cette Ordonnez au robot d'effectuer une tâche spécifique, où chaque signe a un Signification spéciale, indiquant des fonctions spéciales, telles que le nombre "un" signifie "avant", "cinq" signifie "arrêt", etc.



Figure 6 : Contrôle robot

5. Reconnaissance des nombres :

Une autre application récente de geste de la main est la reconnaissance des chiffres. Ils ont proposé un système automatique qui pourrait isoler et de reconnaître la signification du geste de mouvement de la main des chiffres arabes de 0 à 9 en temps réel. Le système est composé de deux réseaux de neurones, un premier pour la prédiction de présence et de localisation des nombres, puis un second qui prédit quel est le nombre détecté (de 0 à 9) dans chacune de ces positions.

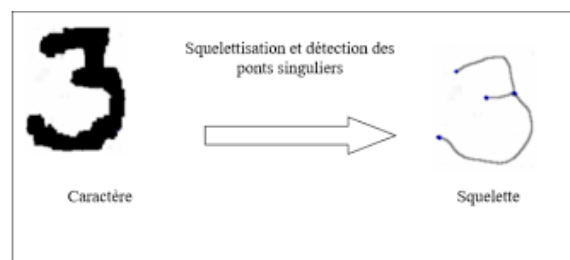


Figure 7 : Reconnaissance de nombre

6. Reconnaissance de la langue de signe (LS) :

La reconnaissance de la langue de signe est un système de traduction automatique des signes. En effet, La langue de signe est un moyen d'expression utilisé par des personnes qui souffrent de trouble de la parole notamment les sourdes et les malentendantes de façon

Chapitre1 : Une Interaction Homme-Machine gestuelle

qu'ils puissent réagir et communiquer entre eux. Elle possède sa propre syntaxe, sa propre grammaire et constitue de ce fait la forme la plus évoluée en communication gestuelle.

F. Applications grand public :

Ces dernières années, plusieurs applications grand public ont émergé. La possibilité d'interaction a permis de capter les mouvements de l'utilisateur afin de les placer au centre d'un jeu vidéo, ou à permettre la manipulation de données numériques avec une interface interactive, Cela peut être fait par plusieurs personnes en même temps.

Par exemple, **Le système HandVu de Kolsch**, Est un système de pointage par reconnaissance de mouvement plus précisément une interface gestuelle basée sur la vision. Ce système fonctionne avec un casque intégrant une caméra et un dispositif de visualisation ainsi qu'un microphone. Il démontre la faisabilité d'une interface gestuelle utilisant la vision.

Le système EyeToy, Développé par Sony en 2003 utilisé dans les jeux vidéo comme la console Playstation dans certain jeux comme d'EyeToy, il utilise une petite caméra USB pour capturer le mouvement du corps humain et modéliser la tête de l'utilisateur en 3D.

Elle peut aussi être utilisée comme webcam sur un ordinateur grâce à sa connectique USB et les pilotes appropriés.

Le système Surface de Microsoft, Microsoft Surface comprend une série de PC à écran tactile et des tableaux blancs interactifs, conçus spécifiquement pour des méthodes de travail évolutives. Il utilise des caméras pour détecter des gestes de la main, ou des objets. Les résultats sont affichés en surface par rétroprojection. Ainsi, les utilisateurs peuvent interagir avec leurs données numériques avec leurs mains. Le système peut être utilisé simultanément par plusieurs personnes qui peuvent se regrouper autour de la surface interactive. Des objets physiques peuvent également être placés sur des surfaces, par exemple en utilisant des appareils numériques pour transmettre des données.

G. Les gestes :

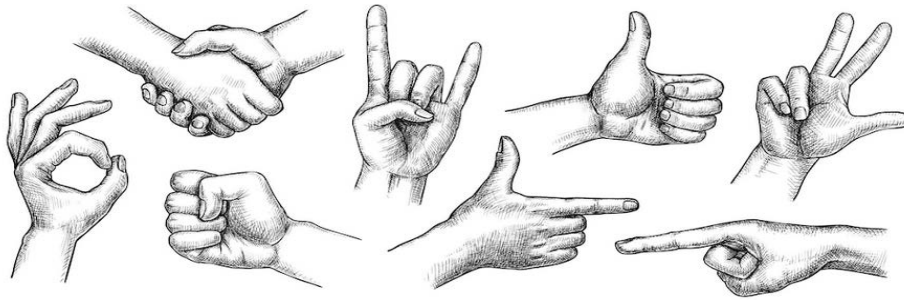


Figure 8 : Les gestes de la main

1. Définition du geste :

Dans notre vie de tous les jours, les gestes font partie intégrante de notre communication qu'ils soient intentionnels ou non. On peut définir un geste comme un mouvement du corps, un signe manuel ou corporel afin qu'ils permettant de communiquer des informations significatives et pertinentes, d'une autre façon le geste permet d'illustrer les mots du langage, de les compléter ou les appuyer. La gestuelle est un langage de communication non-verbal et inconscient que l'on peut apprendre à décrypter et qui peut être même combiné avec une communication verbale ou la remplacer. Elle exprime ce que l'on ressent, dévoile qui on est et trahit ce que l'on cache. Parmi les cinq modes de communication (l'ouïe, la vue, la parole, le toucher et le geste) le geste semble être l'un des plus riches des canaux de communication.

2. Comment reconnaître un geste :

Il existe différents techniques d'analyses qui nous permettent de reconnaître un geste, nous déterminent trois techniques principales dans Les techniques visuelles (la vision par ordinateur), Les techniques basées sur des gants et enfin Les techniques utilisant des gestes de dessin.

- Les techniques utilisant des gestes de dessin : se base généralement sur Les écrans graphiques qui sont également des périphériques d'interaction en permettant la reconnaissance de gestes de dessins, l'utilisation d'une tablette graphique permet de mesurer la position du stylo et retourne ses différentes coordonnées à l'ordinateur. Les gestes effectués avec ces périphériques sont des marqueurs qui définissent des

Chapitre1 : Une Interaction Homme-Machine gestuelle

commandes Leur but est de remplacer les menus traditionnels dans différentes applications Surtout dans les systèmes avec un espace limité, tels que exemple d'assistant personnel.

- Les techniques basées sur des gants : pour présenter une application utilisant un gant numérique, introduisons brièvement Comment ça marche Des fibres optiques placées sur chaque doigt sont utilisées pour mesurer L'angle auquel les deux doigts sont pliés L'intensité du signal lumineux envoyé à la fibre Et sa force au bout des doigts permet de déterminer ces angles. De plus, des capteurs de position et d'orientation des mains sont ajoutés. Il se compose de deux modules. Premièrement, l'émetteur est positionné dans une position fixe et émet trois champs électromagnétiques directionnels. Deuxièmement, le récepteur est fixé au gant et la réception de ces champs magnétiques génère trois courants induits dont les valeurs sont proportionnelles à La position et l'orientation du récepteur par rapport à l'émetteur.

Dans notre étude, nous allons s'intéressé particulièrement aux techniques visuelles à l'aide d'une webcam.

3. Catégorie de geste :

Il existe trois (3) types des gestes selon les parties du corps impliquées nous citons :

- **Les gestes impliquant tout le corps** : les recherches dans ce domaine s'intéressent à tout le corps en interaction avec son environnement. (l'exemple d'analyse des gestes d'un athlète pour améliorer ces performances).
- **Les gestes de la tête et du visage** : l'expression faciale comprennent tous les mouvements lisibles sur le visage d'une personne est souvent considéré comme un miroir des émotions décrypter par les grimaces et les mimiques du visage.
En Revenge, les peu gestes de la tête ont signification spécifique. (L'exemple de l'orientation de tête qui est considéré comme un moyen utile pour la détection du champ de vision).

Chapitre1 : Une Interaction Homme-Machine gestuelle

- **les gestes de la main et du bras** : ils forment la principale catégorie de gestes interactifs. Les mains sont utilisées pour effectuer des gestes précis et complexes. Lorsqu'on se place dans l'informatique, la gestuelle a été introduite comme moyen d'interaction en complément des interfaces traditionnelles.

En 1990 Kurtenbach a défini les gestes comme des mouvements corporels contenant des informations [Kurtenbach, 1990]. En effet, pour Kurtenbach, dire au revoir dans un mouvement latéral avec la main est un geste, mais appuyer sur une touche d'un clavier n'est pas un geste, car dans la phase finale de l'action, le mouvement du doigt n'est ni observé ni important pour le système : « Ce qui compte, c'est quelle touche est enfoncée » [Kurtenbach, 1990]. En réalité virtuelle, les gestes permettent en effet de transmettre des informations à l'environnement virtuel.

4. Classification des gestes de la main :

Il existe plusieurs classifications des gestes de la main parmi lesquels nous citons :

a) Selon Maria Karam et m.c schraefel :

(1) Les gestes déictiques :

Sert à identifier des objets ou leur places typiquement utilisés dans les environnements virtuels, ces gestes peuvent être considéré comme implicite dans d'autres formes de gestes. Dans l'immense majorité des cas, ils sont exécutés grâce à l'index tendu et parfois un faible mouvement. (Par exemple, lorsque on pointe des doigts un objet à manipuler ou lorsque on indique la direction).

(2) Les gestes sémaphoriques :

Ils font partie de tout système gestuel basé sur des catalogues conventionnel de gestes statiques ou dynamiques traditionnels (par exemple : geste (statique) pour le signe « ok » geste (dynamique) pour dire « au revoir »).

Chapitre1 : Une Interaction Homme-Machine gestuelle

(3) Les gestes de manipulation :

Son but est d'appliquer La relation étroite entre le mouvement du geste et l'entité manipulée.

b) Selon Quek :

Le geste est un moyen d'expression ou l'action de parler est souvent accompagné par des mouvements des mains, Quek décompose ces gestes en deux (2) classes : gestes de manipulation qui correspondant aux fonctions ergotique et épistémique et en gestes de communication qui correspondant à la fonction sémiotique. Les gestes communicatifs sont décomposés en gestes actifs et en gestes symboliques.

1. Les gestes actifs : sont directement liés à leur interprétation et en complément de la parole. Ils sont décomposés en gestes mimiques, pour imitez une action et en gestes déictiques (geste de pointage) qui sont largement utilisés dans l'interaction homme-machine puisque le doigt Représente un dispositif de pointage naturel.

2. Les gestes communicatifs : ces gestes ne peuvent pas être compris directement car ils peuvent avoir un sens différent selon la culture donc, il faut être initié afin de comprendre leur signification (Par exemple, la langue des signes). Ils sont décomposés en geste référentiel faisant directement référence à un objet ou à un concept et en geste modélisant qui modélise un état ou une opinion (Par exemple, pour donner une idée de la taille d'un objet).

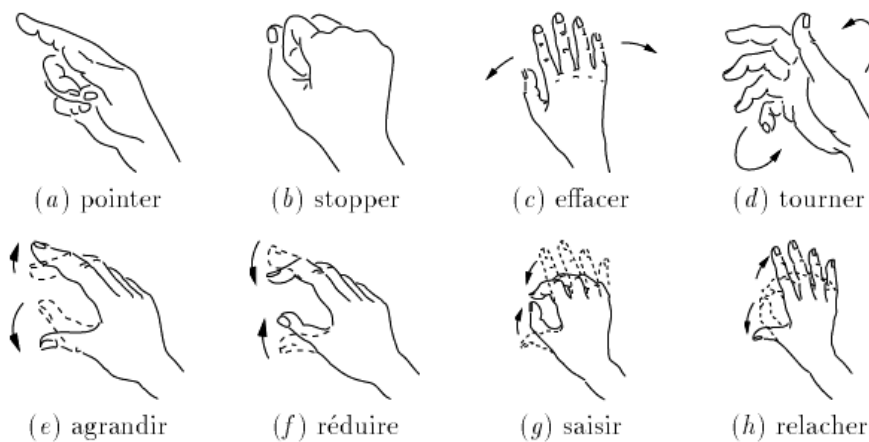


Figure 9 : Geste statique et dynamique

Chapitre1 : Une Interaction Homme-Machine gestuelle

c) Selon Kendon :

Le geste et le langage ne sont pas équivalents mais complémentaires, les gestes qui accompagnent les paroles ou qui s'y substituent (la langue des signes) pour nous inciter à distinguer Les différents degrés de régularité impliqués dans la communication gestuelle, Kendon propose le continuum suivant :

1. La gesticulation : il s'agit de mouvements spontanés de la main ou du bras qui accompagne toujours la parole.
2. Les gestes de paralangage : Ils ne constituent en aucun cas un langage, car à la différence des langues des signes ils apparaissent plutôt comme un simple système de codage au vocabulaire très restreint.
3. La pantomime : il s'agit d'un geste ou d'une séquence de gestes qui ont une fonction narrative, vise à créer une image gestuelle d'un objet, d'un événement ou d'un état d'esprit.
4. Les emblèmes : ce sont des gestes qui présentent un fort degré de conventionalité car ils remplacent souvent la parole et correspondent à des codes culturels (Par exemple, le geste salut et la représentation des chiffres avec les doigts de la main).
5. Les langues des signes : sont les seules à pouvoir être considérées comme de véritables langues. ce type de geste assure toutes les fonctions remplies par la parole et dépend d'un code gestuel très précis.

5. Fonctionnalité du geste :

A l'égard de la nature expressive du canal gestuel, on distingue trois fonctions du geste qui sont présentés comme suit :

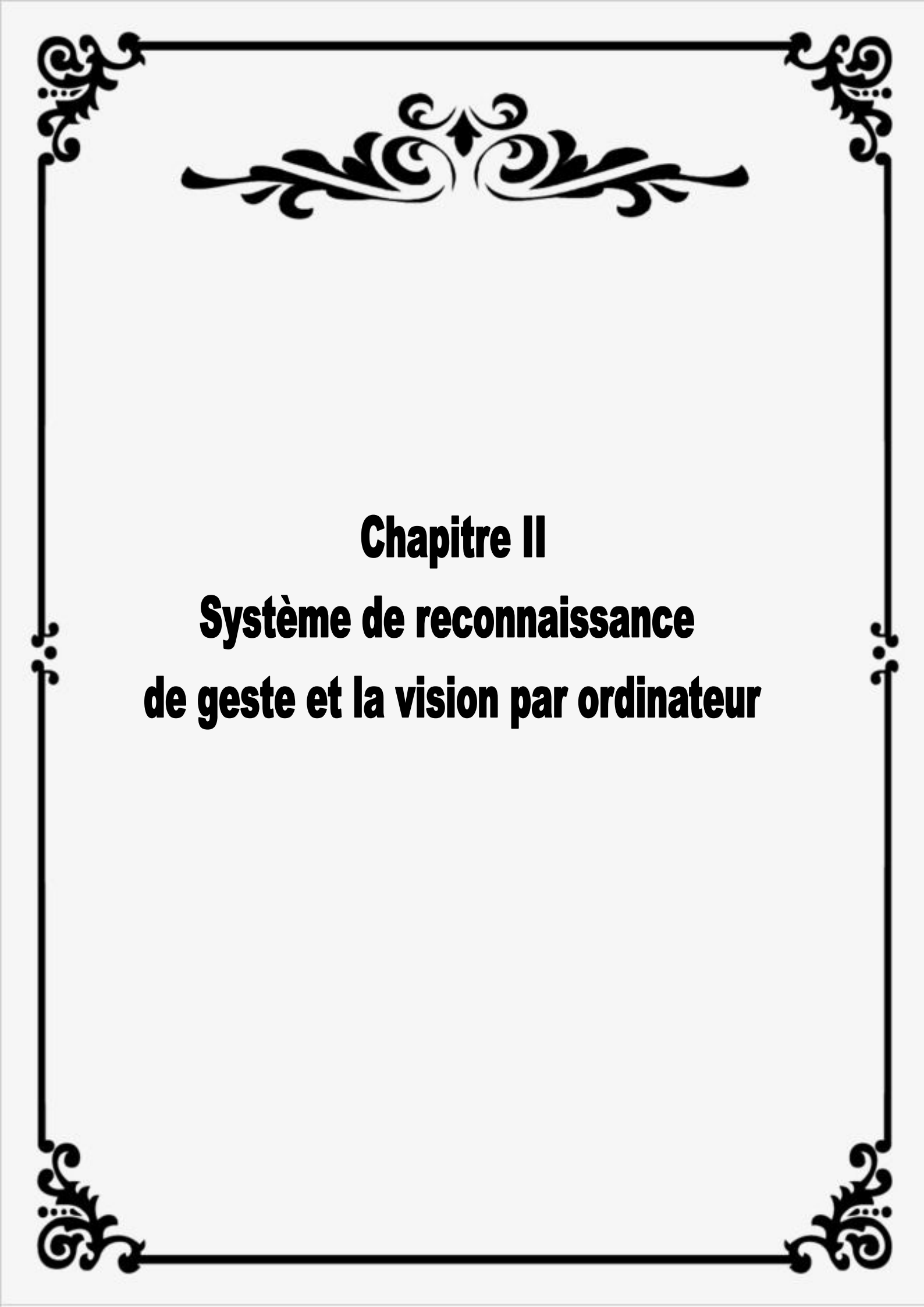
- **Fonction ergotique** : la main agit sur le monde physique c'est-à-dire elle est en prise direct avec la matière qui peut être transformé, brisé ou déformé.
- **Fonction épistémique** : la main est un organe de perception, elle donne des informations relatives à la température, la pression, la distance, la grandeur, le poids donc il s'agit du sens de toucher.
- **Fonction sémiotique** : la main est un organe d'expression autant dire qu'elle joue le rôle de producteur de l'information. Elle s'adresse à la perception visuelle d'un ou de plusieurs interlocuteurs.

Chapitre1 : Une Interaction Homme-Machine gestuelle

Dans le cadre de la communication homme-machine la fonction sémiotique est la plus complexe et la plus riche étant donné qu'elle regroupe les gestes qui accompagnent la langue des signes en forment une véritable langue dotée d'une syntaxe et permettant même la création dynamique de signes non standard en fonction des besoins.

H. Conclusion :

Dans ce chapitre nous nous sommes intéressés au domaine de l'interaction homme-machine gestuelle. Nous avons conclu que l'objectif de la recherche dans le domaine des interactions homme-machine est le développement de modèles de concepts d'outils et de méthodes pour la réalisation de systèmes répondant aux besoins et aux aptitudes des utilisateurs. Ainsi, nous avons défini le geste comme une forme très expressive de communication non verbale ce qui permet l'interaction le plus naturellement en complètent le canal oral, et ces différentes utilisations.



Chapitre II
Systeme de reconnaissance
de geste et la vision par ordinateur

I. Chapitre2 : Système de reconnaissance de geste et la vision par ordinateur

A. Introduction

L'humain est une source d'inspiration pour l'intelligence artificielle en imitant nos sens et notre vision afin qu'il puisse voir et identifier les choses. Ce qui a permis au processus de reconnaissance de geste d'être aussi proche du raisonnement de l'être humain. Cette reconnaissance de geste est effectuée par plusieurs méthodes, dans la vision par ordinateur qui consiste à apprendre aux ordinateurs à «voir» des images numériques telles que des photos et des vidéos, est une base de plusieurs travaux de reconnaissance comme la reconnaissance de la langue des signes.

B. La vision par ordinateur

La vision par ordinateur (aussi appelée vision artificielle ou vision numérique) est une branche de l'intelligence artificielle qui traite de la façon dont les ordinateurs peuvent acquérir une compréhension de haut niveau à partir d'images ou de vidéo numérique. Cette vision cherche à exécuter et à automatiser les tâches qui répliquent les capacités humaines. Dans ce cas, la vision par ordinateur cherche à répliquer la manière dont les êtres humains voient et la façon dont les êtres humains donnent un sens à ce qu'ils voient.

La reconnaissance d'objets est considérée comme un sous-domaine de la vision par ordinateur. Est une méthode de détection de la présence d'instances ou de classes dans des images numériques. Les deux phases qui permettent d'obtenir des meilleurs résultats sont : la phase d'acquisition qui assure les transformations d'images par l'interprétation des images traitées et la phase de détection d'images qui assure le traitement d'images.

C. Les images numériques et leurs traitements :

1. Définition d'images :

Une image est la projection et la représentation d'un ensemble de rayons lumineux provenant du monde

extérieur, elle est constituée d'un ensemble de points appelés pixels. Le pixel représente le

Chapitre 2 : Système de reconnaissance de geste et la vision par ordinateur

plus petit élément constitutif de cette image, précisément chaque point d'une image possède deux coordonnées (X, Y) de tel sorte que chaque couple de coordonnées (X, Y) représente l'emplacement d'un point dans cette image sur un plan et dont l'image de ce couple correspond à la valeur du rayon lumineux capté. Une image est avant tout un signal 2D (x, y), souvent, cette image représente une réalité 3D (x, y, z).

D'un point de vue mathématique : une image est une matrice de nombres, représentant un signal.

D'un point de vue humain : une image contient plusieurs informations sémantiques. Il faut interpréter le contenu au-delà de la valeur des nombres.

2. Types d'images :

Il existe deux (2) types d'image numérique :

Image matricielle (ou image bitmap) : s'agit d'images pixellisées composée d'une matrice (tableau) de points à plusieurs dimensions, chaque dimension représentant une dimension spatiale (hauteur, largeur, profondeur) et chacun de ces points possédant une ou plusieurs valeurs décrivant sa couleur.

Image vectorielle : permet de représenter les données de l'image par des formules géométriques telles qu'un cercle, un rectangle ou un segment.

3. Codage des couleurs d'une image numérique :

Une image numérique utilise plus ou moins de mémoire selon le codage des informations de couleurs qu'elle possède, l'image peut être alors en :

Noir et blanc (Images binaires) : le pixel peut prendre uniquement la valeur noir ou blanc.

Gris : le pixel est noir, blanc, ou a un niveau de gris entre les deux. Il est codé sur 1 octet (8 bit), ce qui lui donne 256 valeurs possibles, dont le blanc(255) et le noir(0).

Couleur : la couleur d'un pixel est obtenue en mélangeant les trois couleurs fondamentales à savoir le Rouge, Vert et Bleu. Chaque pixel est codé sur 3 octets ce qui lui donne $(2^3)^3$ valeurs possibles.



Figure 1 : codage d'une image numérique

4. Caractéristiques d'une image numérique :

1. Définition (Taille) :

Le nombre de pixels constituant l'image, c'est-à-dire sa «dimension informatique» (le nombre de colonnes de l'image que multiplie son nombre de lignes). Une image possédant 640 pixels en largeur et 480 en hauteur aura une définition de 640 pixels par 480, notée 640x480.

2. Résolution :

Terme souvent confondu avec la "définition", détermine par contre le nombre de points ou pixels par unité de surface, exprimé en *points par pouce* (**PPP**, en anglais **DPI** pour *Dots Per Inch*) un pouce représentant 2.54 cm. Elle permet ainsi d'établir le rapport entre le nombre de pixels d'une image et la taille réelle de sa représentation sur un support physique (impression d'une photo).

3. Profondeur :

La profondeur est le nombre de bits utilisés pour représenter la couleur d'un pixel dans une image.

D. Etat de l'art de la reconnaissance des gestes de la main

Les progrès scientifiques et technologiques dans la recherche en vision par ordinateur se sont orientés vers la compréhension des scènes comportant tout type d'objets, mais en particulier vers l'établissement de la langue des signes dans le domaine de la reconnaissance des formes à l'aide de la webcam afin de remplacer la souris dans l'interaction homme-machine.

Chapitre 2 : Système de reconnaissance de geste et la vision par ordinateur

La reconnaissance des formes est l'étape qui permet à l'information de passer De notre univers (vue, ouïe, sens) à l'univers des machines. Ce processus n'inclut pas seulement la transformation d'une source d'information analogique en binaire, mais aussi l'interprétation de cette source comme symboles. Cela se fait en différentes étapes acquisition, traitement des données, reconnaissance. Nous allons décrire ces différentes étapes Dans le cadre de la reconnaissance gestuelle.

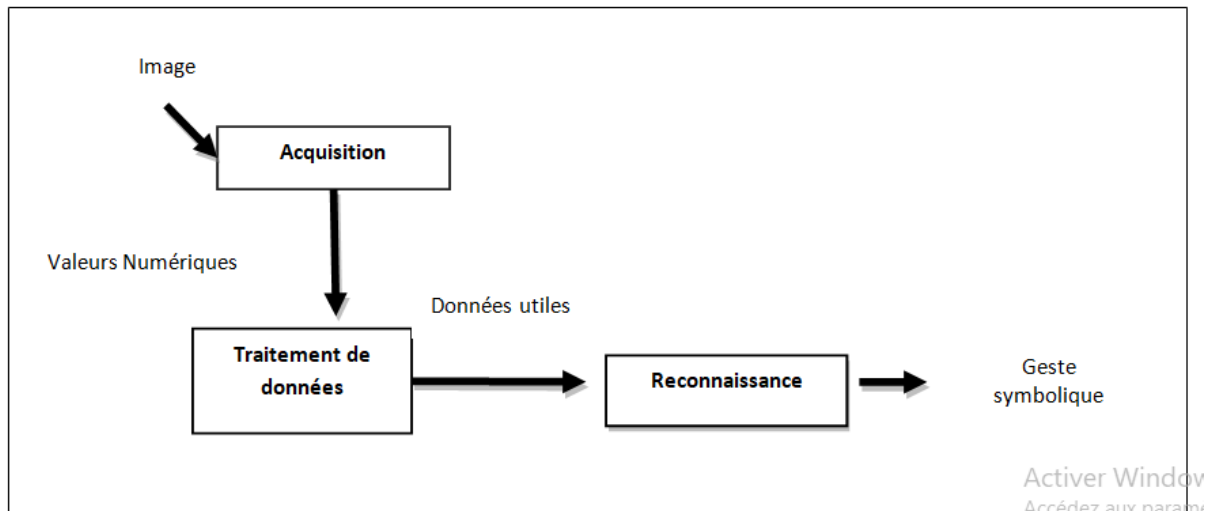


Figure 2: La reconnaissance de geste de la main

1. Acquisition du geste :

Cette première étape consiste à capter les données physiques qui sont sous forme signaux lumineux provenant de la scène à analyser et les encoder sous forme numérique. Cette étape peut se produire en deux (2) manières soit par :

L'acquisition intrusive : fonctionne via une modalité mécanique d'où en suit le déplacement des muscles et/ou des os de la personne. Le dispositif de gant numérique nous permet le suivi des mouvements de la main et cela grâce à des marqueurs ou capteurs qui seront fixés sur la main de la personne.

L'acquisition non-intrusive : l'utilisateur dans ce cas sera libre de tout système, cette acquisition fonctionne via des systèmes à base d'une ou plusieurs caméras. Ces dernières années le Kinect est le capteur le plus utilisé dans l'acquisition non intrusive.

Chapitre 2 : Système de reconnaissance de geste et la vision par ordinateur

➤ Kinect :

Kinect c'est la dernière technologie de la reconnaissance de geste son lancement a eu lieu le 10 novembre 2010 par Microsoft, connu sous le nom Project Natal, est un périphérique destiné à la console de jeux vidéo Xbox 360 permettant de contrôler des jeux vidéo sans utiliser de manette. C'est un petit appareil constitué de différents composants : d'un émetteur et un capteur infrarouge permettant de suivre les mouvements d'un ou plusieurs joueurs en temps réel, d'un microphone destiné à la commande vocale et d'une caméra vidéo. Un logiciel s'occupe ensuite d'interpréter ces données pour les transposer à l'écran, afin de reproduire le squelette d'un joueur via une vingtaine de points. Avec kinect y'aurai plus besoin d'utilisé une manette ou une souris pour donner une commande, il suffit de tendre notre main devant nous et celle-ci apparaît immédiatement à l'écran sous forme de curseur.

En effet, dans notre cas d'étude la webcam est la meilleure solution pour la détection des gestes de la main, et cela grâce à sa mobilité, sa facilité d'utilisation et son prix raisonnable.



Figure 3 : L'exemple de kincet

L'étape de l'acquisition du geste nécessite un traitement afin d'étudier les données obtenues comme étant des images numériques dans le but d'améliorer leur qualité (Convolution et Filtrage) ou d'en extraire de l'information (Segmentation).

a) La convolution d'image :

La convolution d'image est le traitement d'une matrice sur une autre, appelée matrice de convolution (ou noyau). La plupart des filtres de traitement d'image utilisent des matrices de convolution. Un filtre ou matrice de convolution est utilisé comme première matrice de l'image, une collection de pixels en coordonnées cartésiennes 2D, et un noyau variable selon l'effet recherché. Les tailles de matrice de convolution les plus courantes sont 3x3 ou 5x5. Le filtre étudie tour à tour chaque pixel de l'image, et pour chaque pixel que nous appelons le "pixel initial", il multiplie la valeur de ce pixel et la valeur de chaque pixel qui l'entoure par la valeur correspondante au noyau. Il additionne tous les résultats puis le pixel initial prend la valeur du résultat enfin.

b) Filtrage d'une image :

Le filtrage est utilisé dans le but de détecter les contours d'une image mais principalement permet d'éliminer les bruits (parasites) sur l'image (lissage). Nous distinguons deux types de filtrage :

Filtrage global : chaque pixel de la nouvelle image est calculé en prenant en compte la totalité des pixels de l'image de départ.

Filtrage Local : chaque pixel de la nouvelle image est calculé en prenant en compte seulement un voisinage du pixel correspondant dans l'image d'origine. Il est d'usage de choisir un voisinage carré et symétrique autour du pixel considéré. Ces voisinages sont donc assimilables à des tableaux à deux dimensions (matrices) de taille impaire.

Le filtrage le plus utilisé est Le filtre Gaussien (Gaussian Blur).

(1) Le filtre Gaussien :

Il appartient au deuxième type de filtrage, c'est-à-dire au filtrage local. Le principe est le suivant : Calcule une moyenne pondérée des valeurs de voisinage. Pixels de voisinage Pixels proches Poids centraux (= plus d'influence) que ceux plus éloignés. Il utilise une matrice de convolution (également appelée noyau) pour effectuer le filtrage.

2. Traitement de données :

Une fois les données sont collectées par le système d'acquisition, deux traitements vont devoir être appliqués : la segmentation et l'extraction des caractéristiques afin de collecter des données représentatives.

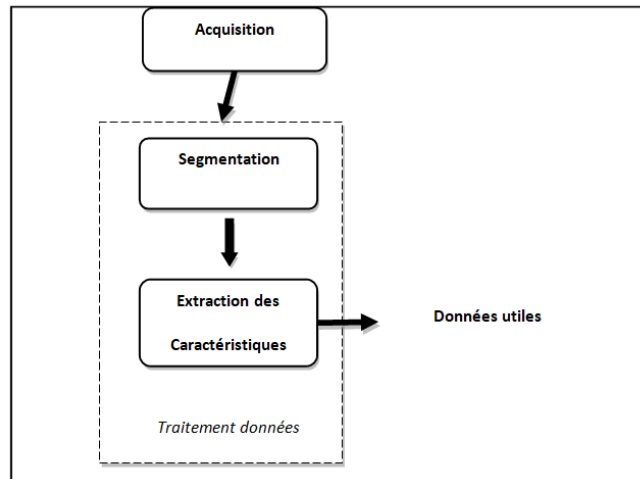


Figure 4: Traitement de donnée

a) La segmentation :

Cette étape est primordiale, elle représente une méthode de détection de zone d'intérêt d'une image numérique, c'est-à-dire mettre en évidence des zones de cette image jugée pour l'analyse. En vision par ordinateur, la segmentation d'image est le processus de partitionner une image numérique en plusieurs segments ou régions homogènes (ensembles de pixels, également appelés super-pixels). Le but de la segmentation est de simplifier et / ou changer la représentation d'une image en quelque chose de plus significatif et plus facile à analyser. Dans le cadre de la reconnaissance du geste la segmentation est une approche qui permet de différencier les pixels de peau des autres et cela par la localisation des doigts (lignes, courbes, etc.), obtenir les contours afin d'extraire des informations sur sa forme et sa position. En effet, le résultat obtenu est utilisé dans la phase de reconnaissance.

(1) La segmentation par la détection de la couleur de peau :

La segmentation de la couleur de peau (couleur chair) est une étape préliminaire dans plusieurs applications tel que la détection de mouvement humain, l'interaction homme

Chapitre 2 : Système de reconnaissance de geste et la vision par ordinateur

machine et la vidéosurveillance. Un point commun entre ces trois axes est la détection des visages et/ou les mains où la détection de la couleur de peau va permettre une réduction considérable de l'espace de recherche qui va se limiter uniquement aux sections de couleur chair. Les espaces colorimétriques utilisés sont le RGB, HSV et YCbCr.

1. RGB :

RVB est un pixel d'espace colorimétrique 3D où chaque pixel a une combinaison de trois couleurs Rouge, vert et bleu à un endroit précis. Cette technique largement utilisée en traitement d'image pour la distinction entre les pixels qui correspondent à la couleur de la peau et de celle appartenant au fond de la scène.

Le cas d'éclairage est faible :

((R>B) et (G>B)) ou ((R>220) (G>210) et (B>170) et (ABS (R-G) <=15))

Le cas d'éclairage est fort :

La caméra est sensible aux conditions d'éclairage et la présence d'éléments ayant des couleurs proches de la couleur de la peau donc :

Si (R>95) et (G>40) et (B>20) et ((max[R,G,B])– (min [R,G,B]) >15)

Et (ABS (R-G) >15) et (R>G) et (R>B)

Alors (Image(R, G, B))=0 ; Sinon niveau de gris (Image(R, G, B))=255

2. YCbCr chrominance de luminance :

Cet espace colorimétrique est utilisé dans la vidéo numérique. Les informations de couleur représentent deux couleurs Cb et Cr. Cb est la différence entre le bleu et Cr est la différence entre les références des composants rouges. Il s'agit essentiellement d'une transformation RVB en YCbCr pour la séparation de la luminance et chrominance pour la modélisation des couleurs.

3. HSV (Teinte, Saturation et Valeur) :

En HSV, La teinte détecte la couleur dominante et la saturation définit la couleur tandis que la valeur mesure l'intensité ou la luminosité. C'est assez bien pour choisir une seule couleur

Chapitre 2 : Système de reconnaissance de geste et la vision par ordinateur

mais cela ignore la complexité de l'apparence des couleurs. Il compromet la vitesse de calcul par calcul pertinence coûteuse et perceptuelle.

(2) Seuillage (thresholding) :

Le seuillage est une technique de segmentation elle permet d'extraire la main de l'arrière-plan. 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 la main cette opération s'appelle binarisation. Plus précisément, le seuillage produira une image binaire avec tous les pixels de 1 représentant la main. Cela se produit lorsque le pixel dont l'intensité est inférieure à la valeur seuil est défini sur 0, tandis que tout pixel dont l'intensité est supérieure à la valeur seuil qui est défini sur 1.

Le seuillage d'une image gris I consiste à associer la valeur 0 (noir) à tous les pixels dont la valeur est inférieure au seuil α et la valeur 1 (blanc) à tous les autres. On obtient donc une image seuillée S :

$$\forall(x, y), S(x, y) = \begin{cases} 1, & \text{si } I(x, y) \geq \alpha \\ 0, & \text{sinon} \end{cases}$$

L'avantage du seuillage est dans sa facilité de mise en œuvre et sa rapidité. Par conséquent, le seuil est déterminé manuellement ce qui rend la tâche compliquée de plus il est très difficile de trouver un seuil qui donne de bons résultats dans toutes les situations.

Otsu a proposé une méthode plus robuste dans seuil est calculé automatiquement en fonction de l'image, elle consiste à faire l'hypothèse que l'image contient deux classes, décrites chacune par une partie de l'histogramme de l'image $h(i)$. La qualité du seuillage est quantifiée en mesurant la variance des niveaux de gris I de chaque classe. On cherche ensuite la valeur de seuil S qui minimise les variances des deux classes en fonction du seuil T , et on utilise cette valeur pour binariser l'image. Cette méthode est adaptée à des images dans lesquelles les deux classes sont bien définies.

$$\mu_1 = \frac{1}{T} \sum_{i=0}^{T-1} h(i) \quad \sigma^2 = \frac{1}{T} \sum_{i=0}^{T-1} (h(i) - \mu_1)^2$$

$$\mu_2 = \frac{1}{256-T} \sum_{i=T}^{255} h(i) \quad \sigma^2 = \frac{1}{256-T} \sum_{i=T}^{255} h(i) = 1 - p_1$$

Si l'image comporte N pixels, on a les probabilités des deux classes :

$$p_1 = \frac{1}{N} \sum_{i=0}^{T-1} h(i) \quad p_2 = \frac{1}{N} \sum_{i=T}^{255} h(i) = 1 - p_1$$

La figure suivante représente le résultat du seuillage avec la méthode d'Otsu et avec un seuil fixe :

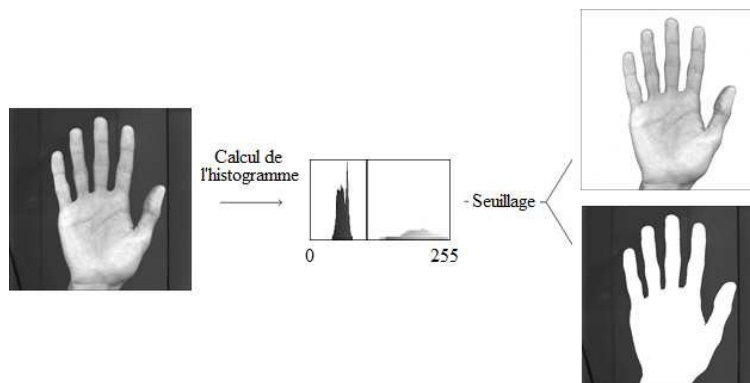


Figure 5 : Le seuillage avec la méthode d'Otsu

(3) Le contour :

Une fois obtenue la masse uniforme correspondant à l'extraction de la main (segmentation), nous étudierons principalement son contour afin de les détecter et éliminer les bruits (parasites) sur l'image (lissage). La résolution du problème est l'utilisation des approches basées filtrage. Le principe de ces filtres de détection de contour, repose sur l'étude des dérivés de la fonction d'intensité de l'image telle que : le gradient de la fonction d'intensité et le passage à zéro de laplacien.

(a) Approche gradient :

La première étape est de représenter les variations de valeur dans l'image. Si l'image est une fonction, et qu'elle est considérée différentiable, il est possible d'utiliser le gradient pour déterminer les variations de valeur en chaque point.

Chapitre 2 : Système de reconnaissance de geste et la vision par ordinateur

Le gradient d'une fonction est un vecteur qui représente les variations de la fonction par rapport à ses différents paramètres. La direction et le sens du gradient indiquent le sens dans lequel la fonction varie le plus et sa norme indique l'intensité de la variation. Afin d'approximer les dérivées par différence le filtre gradient consiste à :

Avec $n=1$.

$$\nabla_u I(u, v) = I(u, v) - I(u - n, v),$$

Ou

$$\nabla_u I(u, v) = I(u + n, v) - I(u - n, v)$$

Ces dérivées sont calculées par une convolution de l'image en utilisant un masque de :

L'opérateur de Sobel :

Est le masque le plus couramment utilisé

$$h1 = 1/4 \begin{bmatrix} -1 & 0 & 1 \\ -2 & 0 & 2 \\ -1 & 0 & 1 \end{bmatrix} \quad h2 = 1/4 \begin{bmatrix} -1 & -2 & -1 \\ 0 & 0 & 0 \\ 1 & 2 & 1 \end{bmatrix}$$

(b) Approche Laplacien :

Comme son nom l'indique la recherche des passages par zéro du laplacien, c'est une approche pour la détection de contour. La convolution avec le masque suivant est la méthode la plus simple :

$$m1 = \begin{bmatrix} 1 & 1 & 1 \\ 1 & -8 & 1 \\ 1 & 1 & 1 \end{bmatrix} \quad \text{Ou} \quad m2 = \begin{bmatrix} 0 & 1 & 0 \\ 1 & -4 & 1 \\ 0 & 1 & 0 \end{bmatrix}$$

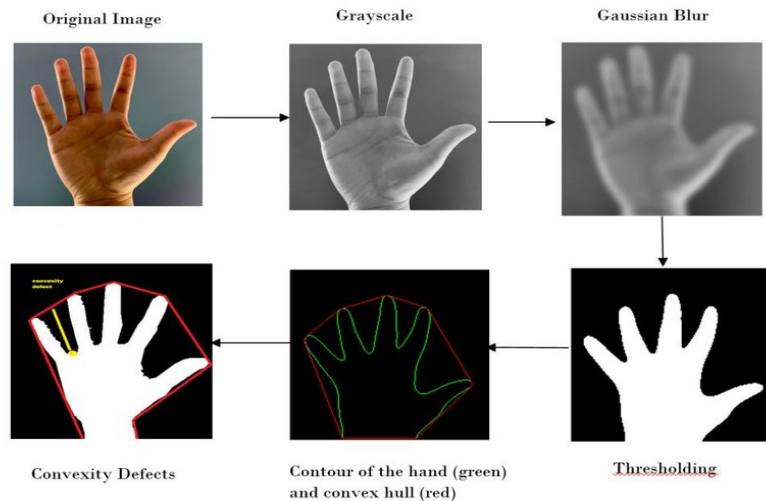


Figure 6 : Segmentation de la main

(4) Le codage :

Le codage consiste à représenter les informations graphiques de la main segmentée de manière à détecter les différentes caractéristiques.

Le Codage de Freeman (codage d'orientation) :

Le principe dans ce codage est de pouvoir décrire le contour par un vecteur qui permet la représentation de l'orientation, donc chaque pixel est codé par une valeur comprise entre 0 et 7.

b) Extraction des caractéristiques :

L'extraction de caractéristiques est une composante du traitement de l'image qui permet d'extraire des données utiles. Par exemple, dans le cas de capture par caméra, on va devoir dégager les informations pertinentes de l'image. Il faut donc arriver à extraire le geste de la main du fond de l'image.

c) La reconnaissance :

Cette étape consiste à identifier ou classer les gestes en se basant sur certaines de leurs caractéristiques. La modélisation des données sous forme de classes est effectuée par plusieurs méthodes basées sur les probabilités ou la statistique mises en œuvre. Pour chaque élément du vocabulaire que nous voulons identifier, nous aurons une classe qui le

Chapitre 2 : Système de reconnaissance de geste et la vision par ordinateur

représente sous une forme utilisable par une technique de reconnaissance. On distingue principalement deux approches de la reconnaissance des formes :

- ♣ Les méthodes statistiques : l'extraction des caractéristiques produit des valeurs numériques qui sont confrontées aux modèles statistiques caractérisant chaque classe
- ♣ Les méthodes structurelles (ou syntaxiques) : l'extraction des primitives produit des valeurs symboliques et des relations qui font l'objet d'une analyse structurelle ou syntaxique.

(1) La modélisation :

La modélisation des différentes classes se fait généralement à travers divers Structures telles que matrices, vecteurs moyens, graphiques, etc. La modélisation permet fournir au système de reconnaissance des ensembles d'exemples des gestes qui vont lui permettre de créer ces structures afin de composer des classes. Ainsi, les classes vont devenir plus claires et plus précise par l'ajout de nouveaux exemples. On voit donc que la qualité de l'exemple est Modélisez de manière concluante la classe correctement.

Dans un premier temps, l'utilisateur du système doit lui même fournir des exemples et décide les différentes classes des gestes que l'on doit recevoir dans la sortie. Ensuite, le système lui-même doit dégager les différentes classes et effectuer l'apprentissage de ces classes et cela à partir de la masse de données reçues en entrée. Une fois l'apprentissage effectué, nous serons en mesure d'identifier les données fournies au système. L'étape de décision est chargée de choisir la classe correspondant aux données.

(2) L'apprentissage :

La phase d'apprentissage consiste à effectuer une comparaison des données reçues en entrée avec toutes les classes stockées dans le système de reconnaissance de geste. La ressemblance est évaluée pour chaque comparaison entre les données et la classe choisie, la classe à sélectionner comme Les données représentatives sont celles pour lesquelles la meilleure évaluation est obtenue.

Chapitre 2 : Système de reconnaissance de geste et la vision par ordinateur

Dans le système de reconnaissance de geste plusieurs techniques ont été testées. Parmi toutes les techniques existantes, trois d'entre-elles sont les plus utilisées : le modèle statistique, le réseau de neurone et le modèle de Markov cachés.

(a) Le modèle statistique :

Cette approche consiste à modéliser les caractéristiques extraites dans la phase d'apprentissage sous forme vecteurs afin d'estimer la différence entre les données en entrée et les différents vecteurs appris. Les vecteurs qui ressemblent le plus aux données en entrée sont choisis comme étant gestes reconnu. Ce modèle se caractérise par sa facilité de mise en œuvre et le temps de réponse qui est très rapide.

(b) Le réseau de neurone :

Ce modèle s'inspire du fonctionnement du réseau de neurones du cerveau. La possibilité d'apprendre sur la base constitue l'une des nombreuses fonctionnalités des réseaux de neurones ce qui permet à l'utilisateur de modéliser ses données et établir des règles précises. Les algorithmes d'apprentissage vont apprendre automatiquement la structure des données grâce à la collecte des données représentatives. Ce modèle nécessite une modélisation avant d'effectuer une reconnaissance des gestes fournis en entrée du réseau de neurones.

(c) Le modèle de Markov cachés :

Ce modèle consiste à la modélisation d'un processus à l'aide d'un graph d'états. Le passage d'un état à un autre se fait en fonction des probabilités basées sur le déroulement des états précédents. En conséquent, la succession d'état qui représente chaque partie du geste permet d'obtenir une modélisation d'un geste. En utilisant un modèle discret Starner utilise des modèles de Markov pour reconnaître 40 mots de la langue des signes américaines. Dans ce système les mains sont suivies en temps réel par une camera couleur. Le processus de suivi extrait la position des mains leur forme et leur trajectoire.

E. Conclusion

Dans ce chapitre nous avons étudié les différentes étapes dans le système de reconnaissance de geste. Tout d'abord, nous avons pu concevoir que l'utilisation de la détection visuelle en appliquant la notion de vision par ordinateur, ce qui permet l'extraction de différentes caractéristiques. Ensuite, nous avons pu conclure que l'interprétation de ces caractéristiques passe par plusieurs phases dans le traitement d'image dans le choix de méthode dépend de la pertinence des informations qu'elle révèle. Dans l'étape suivante nous allons étudier l'implication de la langue des signes dans le système de reconnaissance de geste.



Chapitre III

L'implémentation et Résultat

I. Chapitre 3 : L'implémentation et Résultat

A. Introduction

Dans nos jours, la langue des signes est le principal moyen de communication pour la communauté des sourds. Pour des millions des personnes, l'interaction avec les malentendants est assez difficile, c'est pourquoi l'objectif principal de développement de système qui permet d'introduire la langue des signes dans la reconnaissance des gestes basé sur la vision par ordinateur est la meilleure solution d'interaction.

B. La reconnaissance de la langue des signes :

On conclut que la langue des signes dans l'interaction homme machine a une très grande importance par rapport aux autres types et cela par sa richesse et sa précision de son sens. Nous allons donc présenter les différentes caractéristiques et paramètres de cette langue.

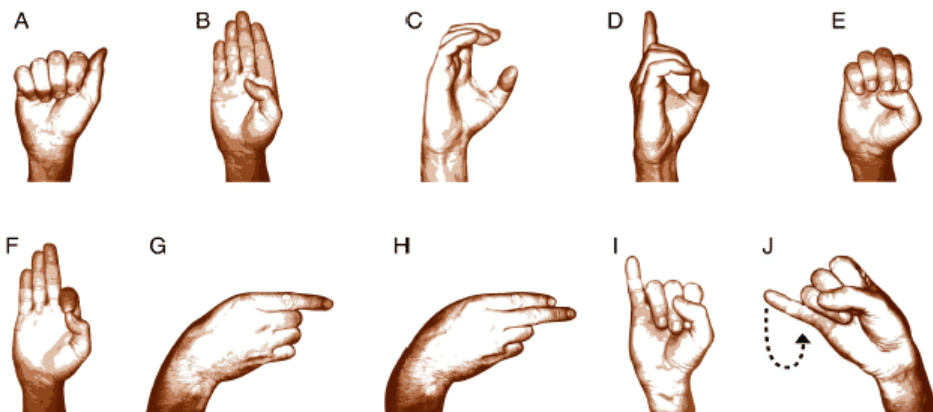


Figure 1 : La langue des signes

1. Les caractéristiques :

La langue des signes est considérée comme une véritable langue, l'expression de phrase ne se réduit pas juste aux gestes produits par la main mais aussi le corps entier qui peut être un moyen d'expression. Dans notre étude on se restreint uniquement aux gestes appelés aussi signes.

2. Les paramètres :

Chaque geste de la main se décompose des paramètres, le signe gestuelle de la langue des signes se compose d'une similitude simultanée de cinq paramètres dans :

La configuration :

Correspond à la forme de la main qui compose le signe, la configuration est noté selon (l'aide d'index, le bec de canard et les lettres évoqué par la forme de la main.

Le mouvement :

C'est le paramètre le plus complexe, il permet la représentation des actions de la main on peut y trouver une forte signature phonémique. Chaque changement de mouvement effectuer modifier le sens d'un mot.

L'emplacement :

Correspond à l'indication d'endroit ou les actions qui sont effectuées.

L'orientation :

Se situe par rapport à la paume de la main permet de caractériser le mouvement, préciser l'orientation des objets. C'est un paramètre qui reste flou et discutable dans la mesure où il est difficile de décrire de manière précise l'orientation de la paume de la main. En effet, cette orientation permet la conjugaison des verbes dans la langue des signes tels que les verbes directionnels (se conjuguent dans l'espace) et les verbes non- directionnels.

L'expression faciale :

Pour exprimer le mode du discours, les mouvements de sourcils, les yeux ou encore de la bouche vont être importants pour exprimer des notions : interrogation, sentiments, intensité... Une expression du visage permettra même, dans certains cas, de faire la différence entre deux mots signés de la même façon.

C. L'intelligence artificielle est la reconnaissance gestuelle :

L'intelligence artificielle vise à programmer des machines capable d'effectuer plusieurs tâches telles que l'analyse automatique des photos satellite, la reconnaître de monde vocale, faciale et gestuelle.

Chapitre 3 : L'implémentation et Résultat

Dans notre projet les programmes d'IA sont aujourd'hui capables de reconnaître les gestes de la main en s'adaptant à toutes les situations auquel il peut être exposé. D'ailleurs, IA possède un module d'apprentissage qui va permettre d'apprendre nos gestes à une machine en utilisant des différents microcapteurs afin de transmettre des données, Un sous ensemble du machine learning (algorithme d'IA) va indiquer donc à la machine comment acquérir les images comme données et savoir à quel moment les utiliser, ils seront donc reconnu par le système, mémoriser et interpréter.

Les approches d'intelligence artificielle permet de tirer parti du potentiel d'interaction offert par la reconnaissance de geste 2D et même 3D, en effectuant des gestes plusieurs commandes gestuelle offrent à l'utilisateur la possibilité d'exécuter de nombreuses actions.

En effet, IA possède un module d'apprentissage automatique / artificielle dans le programme de reconnaissance gestuelle est fondé sur ces algorithmes. Donc nous allons avoir un aperçu de cet aspect.

1. L'apprentissage automatique :

L'apprentissage automatique est un champ d'étude d'IA, se fonde sur des approches mathématiques et statistiques. Elle englobe toute méthode qui permet la construction d'un modèle de la réalité à partir des données prédéfini soit pour l'amélioration d'un modèle déjà existant ou soit par la création complète.

L'apprentissage donc donne la capacité à l'ordinateur d'apprendre et cela dans le but d'améliorer ces performances, la prise de décision afin de résoudre des tâches. Plus précisément, cet apprentissage concerne la conception, l'analyse, l'optimisation, le développement et l'implémentation.

2. Types d'apprentissage :

Il existe trois (3) types d'apprentissage : selon la nature de donnée traitée (symbolique/numérique) et l'apprentissage selon intervention.

a) Apprentissage symbolique :

Repose sur des règles définies explicitement par le programmeur afin de guider la machine dans sa prévision. Cet apprentissage est dans le but de reproduire le raisonnement de

Chapitre 3 : L'implémentation et Résultat

l'homme en le modélisant par un ensemble de symboles, la machine learning exploite les algorithmes sans utiliser des instructions explicites.

En effet, l'idée principale est donc l'élaboration des méthodes qui permettant l'extraction des connaissances structurelle.

b) Apprentissage numérique :

Comme son nom l'indique l'apprentissage numérique désigne la présence du numérique dans le contenu de formation, on parle donc du traitement des valeurs quantitatives appelé aussi le digital learning ce qui fait référence à un domaine plus vaste que le symbolique grâce à s'adaptative.

Elle utilise tous les outils numériques et la production des procédures de classification en boîte noire, ce qui permet de favoriser l'apprentissage et la montée de compétence.

c) L'apprentissage selon l'intervention :

(1) Apprentissage supervisé :

Est une méthode de machine learning qui s'appuie sur des données ou exemples labellisés (étiquetés ou annotés) pour entraîner des modèles intelligence artificielle prédictifs. Le système apprend à classer selon un modèle de classification ou de classement grâce à des classes prédéterminées (associée à chaque objet) ou des exemples connus.

Elle désigne la situation d'apprentissage automatique où les données ne sont pas étiquetées. Il s'agit donc de découvrir les structures sous-jacentes à ces données non étiquetées. Puisque les données ne sont pas étiquetées, il est impossible à l'algorithme de calculer de façon certaine un score de réussite.

L'absence d'étiquetage ou d'annotation caractérise les tâches d'apprentissage non supervisé et les distingue donc des tâches d'apprentissage supervisé.

L'introduction dans un système d'une approche d'apprentissage non supervisé est un moyen d'expérimenter l'intelligence artificielle. En général, des systèmes d'apprentissage non supervisé permettent d'exécuter des tâches plus complexes que les systèmes d'apprentissage supervisé, mais ils peuvent aussi être plus imprévisibles.

(2) Apprentissage non-supervisé :

Désigne la situation d'apprentissage automatique où les données ne sont pas étiquetées. Il s'agit donc de découvrir les structures sous-jacentes à ces données non étiquetées. Puisque les données ne sont pas étiquetées, il est impossible à l'algorithme de calculer de façon certaine un score de réussite.

L'absence d'étiquetage ou d'annotation caractérise les tâches d'apprentissage non supervisé et les distingue donc des tâches d'apprentissage supervisé.

L'introduction dans un système d'une approche d'apprentissage non supervisé est un moyen d'expérimenter l'intelligence artificielle. En général, des systèmes d'apprentissage non supervisé permettent d'exécuter des tâches plus complexes que les systèmes d'apprentissage supervisé, mais ils peuvent aussi être plus imprévisibles.

D. Environnement de développement :

Le Language de programmation utilisé dans l'étude des différents systèmes de reconnaissance de geste est python avec un environnement de développement pycharm d'où plusieurs bibliothèques sont intégrer telles que : Média pipe, TensorFlow, keras, OpenCV, Keras, etc. Ce système sera réalisé sur un environnement matériel PC à l'aide d'une webcam.

Python :

Est un langage de programmation interprété, multi paradigme et multiplateformes. Il favorise la programmation impérative structurée, fonctionnelle et orientée objet. Il est doté d'un typage dynamique fort, d'une gestion automatique de la mémoire par ramasse-miettes et d'un système de gestion d'exceptions.

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.

OpenCV :

Est un framework de vision par ordinateur et de traitement d'images en temps réel construit sur C/C++, mais il peut être utilisée sur python via un package OpenCv-python.

MediaPipe :

Est un framework de solutions d'apprentissage automatique livré avec des solutions pré-entraînées comme la reconnaissance faciale, l'estimation de pose, la reconnaissance de la main, la détection d'objet (voiture, avion,...), etc.

Tensorflow :

Est une bibliothèque open source pour la machine learning et le deep learning développée par l'équipe Google utilisé particulièrement sur les réseaux de neurones.

Keras :

Est une bibliothèque open source écrite en python, elle permet d'interagir avec les algorithmes des réseaux neurones et l'apprentissage automatique.

E. Reconnaissance des langues des signes :

Voici quelques modèles d'un système qui permet l'intégration de quelques mots et phrases de la langue des signes dans le système de reconnaissance de geste :



Figure 17 : Reconnaissance du mot "Okey"

Chapitre 3 : L'implémentation et Résultat

Ce mouvement de geste représente le mot okay qui est utilisé dans la langue des signes et même dans notre Language actuel.

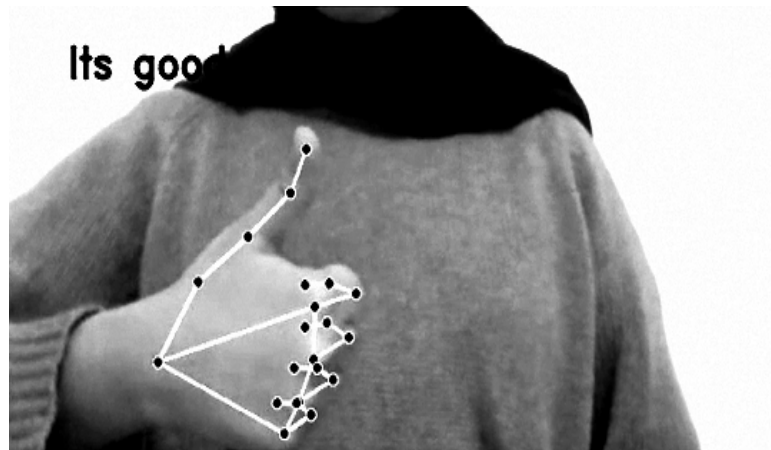


Figure 18 : La reconnaissance de la phrase "its good"

Ce mouvement de geste représente la phrase its good qui est utilisé dans Le langue des signes et même dans notre Language actuel.

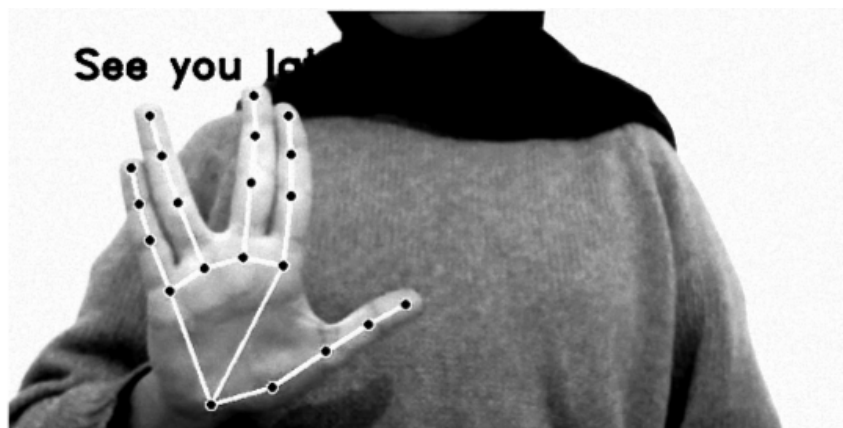


Figure 19 : Reconnaissance de la phrase "see you later"

Ce mouvement de geste représente la phrase au revoir qui est utilisé uniquement dans la langue des signes.

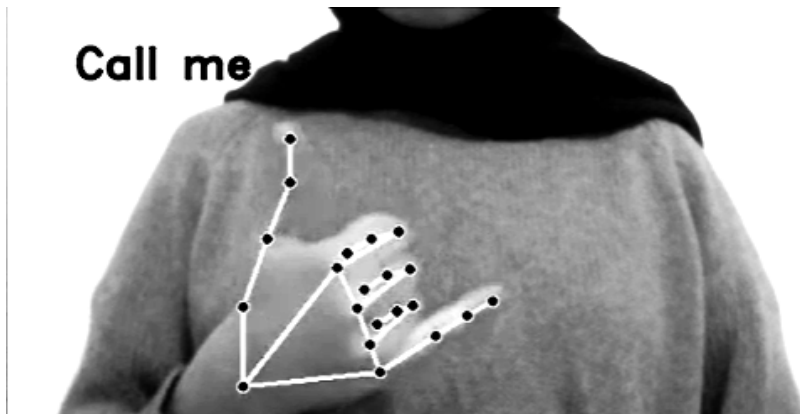


Figure 20 : Reconnaissance de la phrase "call me"

Ce mouvement de geste représente la phrase appel moi qui est utilisé dans la langue des signes et même dans notre Language actuel.

F. Conclusion

Dans ce chapitre, nous avons défini la relation entre la reconnaissance de geste et l'intelligence artificielle et en passe aux différents types d'apprentissage automatique qui permet la création des systèmes de reconnaissance. Ensuite, nous avons d'abord présenté la langue des signes avec ces différentes paramètres et caractéristiques, et enfin nous avons données quelques modèles de l'intégration de langue dans le système de reconnaissance de geste basé sur la vision par ordinateur et réalisé dans un environnement de développement.

Conclusion Général

Conclusion Général :

Dans nos jours, la portée par les progrès de l'intelligence artificielle et du traitement du langage naturel dans la notion de vision par ordinateur ne cesse de progresser dans plusieurs domaines scientifiques tels que la reconnaissance par des gestes de la main.

La reconnaissance de geste est un domaine très vaste et riche de méthodes qui permettant de reconnaître les différents mouvements de la main, l'intégration de plusieurs langues comme la langue des signes, qui est défini comme une langue entières dans ces systèmes de reconnaissance est la meilleur solution pour l'intégration dans le monde numérique.

La réalisation de système de reconnaissance se base principalement sur plusieurs méthodes dans la méthode acquisition de la main à l'aide d'un web Cam, ensuite la méthode de segmentation qui permet la localisation de la main par la détection de la couleur de peau et le seuillage ainsi l'extraction des caractéristique et en enfin le traitement de ces données.

Finalement, après avoir arrivé à étudier ces différentes étapes de la reconnaissance de geste, nous prévoyons après avoir terminé les recherches nécessaires d'améliorer cette thèse par la réalisation d'une application qui va nous permettre de communiquer par nos gestes avec le monde des sourds-muets.

Webographie

I. Webographie :

<http://dspace.univ-tlemcen.dz/bitstream/112/6044/1/SYSTEME-DE-RECONNAISSANCE-DE-Gestes-de-la-Main.pdf>. [En ligne]

<https://en.wikipedia.org/wiki/PyCharm>. [En ligne]

[https://en.wikipedia.org/wiki/Python_\(programming_language\)](https://en.wikipedia.org/wiki/Python_(programming_language)). [En ligne]

https://fr.wikipedia.org/wiki/Traitement_d%27images. [En ligne]

<https://tel.archives-ouvertes.fr/tel-00006749/file/tel-00006749.pdf>. [En ligne]

<https://tel.archives-ouvertes.fr/tel-00307197/document>. [En ligne]

https://www.irit.fr/publis/TCl/Dalle/these_gianni.pdf. [En ligne]

https://www.irit.fr/publis/TCl/Dalle/these_gianni.pdf. [En ligne]

univ-tiaret.dz/images/1GI/2020/S2/TAI/TAI_COURS_GI_2019_2020.pdf. [En ligne]

<www-sop.inria.fr/members/Pierre.Kornprobst/journal/2006-aubert-vuibert-encyclopedia.pdf>. [En ligne]