

الجمهورية الجزائرية الديمقراطية الشعبية

République Algérienne Démocratique et Populaire

Ministère de l'enseignement supérieur et de la recherche scientifique

Université de 8 mai 1945 – Guelma

Faculté des Mathématiques, d'Informatique et des sciences de la matière

Département d'informatique



16/ 918

Mémoire de Fin d'étude master

Filière : Informatique

Option : Ingénierie des Medias

Thème :

Capture et transformation d'inscriptions en objets virtuels

Encadré par :

Gouasmi Noureddine

Présenté par :

Oumeddour Mouna

Bouchahed Asma

Juin 2016

Remerciement

Au terme de ce travail, nous tenons d'abord à remercier dieu le tout puissant pour nous avoir donné le courage et la volonté d'accomplir ce modeste travail et parce que c'est à lui seul qui retourne le soin de l'achèvement de ce mémoire.

Qu'il nous soit permis ici de remercier très vivement Mr Gouasmi Noureddine de nous avoir encadré et guidé avec autant de sérieux et de gentillesse.

Nous remercions tous nos enseignants du département de l'informatique de l'université 8 mai de Guelma

Enfin, un Remerciement spécial à nos parents, puis que sans eux on n'a jamais pu atteindre nos buts.

Et tous qui nous ont aidé de loin ou de près d'accomplir cette tâche

Je dédie ce modeste travail :

À mes très chers parents, avant tout et partout, que j'adore toujours les avoir à mes côtés.

À la plus merveilleuse mère, que j'adore à en mourir, qui ma mon enfance, qu'Allah te protège et te garde pour nous, comme illumine mon chemin.

Au plus adorable et gentil père « Amar » au monde qui m'a tout donné sans rien recevoir en parallèle, je te remercie infiniment de tes sacrifices et de t'avoir veillé sur moi.

*À mon très cher frère: *Mohamed *inchallah une vie pleine de réussite.*

*À mes très chers sœurs: *Rima * & *Nassima *
longue vie.*

*Et le petit de la famille: *Jyed**

Sans oublier ma grande famille de loin ou de près toute la famille

*À tous mes amies: Lina, Faiza, Meriem, Marwa,
Marwa bos, Asma gh*

À tous les autres amies sections Informatique.

À ma binôme Asma

Mouna

Je dédie ce modeste travail :

Ma grand-mère « Hadda » et mon grand-père « Ahmed » qui m'ont toujours donnés amour, tendresse et affection.

À mon père « Djmal » par sa présence, son sérieux, et ses précieux conseils m'a permis d'arriver là où je suis. Papa qui est toujours là derrière moi, je ne saurais te remercier assez pour tout ce que tu fais pour nous, tu es le meilleur des exemples, Merci.

À ma maman chérie, qui ne cesse de nous pousser d'avantage, qui s'est sacrifiée pour nous et a toujours su nous épauler et nous soutenir dans nos moments difficiles, j'espère être à la hauteur de tes attentes maman

À mon époux « Samir », qui importe tant dans ma vie, qui a et continuera à partager tout ces moment pour la vie Inchallah, tu as cru en moi et tu m'as encouragé à aller de l'avant surtout en cette année, merci.

À mes frères « Islam » « Rida » « Amar » qui apporte le courage et le soutient, merci mes frères.

À ma belle tante et ma la deuxième maman « Nabila » et son mari « Ibrahim » qui m'ont donne le courage dans les moments de faiblesse, merci mes chères.

À ma sœur « Rihana » ma jumele qui apporte tant de joie et d'ambiance dans notre vie, Inchallah tu réussiras dans tes études ainsi que ta vie ma petite ange.

À tous mes amies: Samiha, Soulaf, À tous les autres amies sections Informatique.

À mon binôme Mouna merci pour ta patience avec moi surtout avec mes pleins événement cette années.

Asma

Résumé:

La vision par ordinateur est à l'origine de nombreuses avancées en interaction Homme-machine. Une des possibilités données par l'utilisation de la vision par ordinateur dans l'interaction homme machine est l'interaction à travers l'usage du corps, en particulier les gestes de la main de l'utilisateur. L'interaction gestuelle est déjà utilisée dans plusieurs domaines, tels que les jeux vidéo et les différents dispositifs de communication tactiles.

Nous avons réalisé dans ce projet un système de capture et transformation d'inscriptions en objets virtuels, utilisant des gestes de la main. Ce système permet à un utilisateur d'interagir avec une feuille de papier A4. Il s'agit d'utiliser des gestes de pointage pour sélectionner du texte sur une feuille de papier, filmée par une webcam, et d'exécuter quelques commandes (sélectionner, copier, imprimer) sur le texte sélectionné.

Nous avons testé notre système avec des caméras de résolution différentes, sur des feuilles contenant un texte imprimé à différentes polices et tailles de polices, ainsi que des images, des graphes et des tableaux utilisant plusieurs largeurs de trait. Les tests ont montré que la détection et le suivi du doigt, et donc l'interaction avec la feuille de papier est toujours correcte, par contre, lors de la sélection d'un texte, la lisibilité du texte sélectionné dépend de la résolution de la webcam utilisée.

Mots clés

Vision par ordinateur, Détection de la peau, Interaction gestuelle, Reconnaissance du doigt, Suivi du doigt.

Sommaire

Introduction générale : 7

Chapitre n°1 : vision par ordinateur:

1. Introduction: 9

2. Définition de la vision par ordinateur: 9

3. Système de vision: 10

4. Difficultés liées à la vision par ordinateur: 10

5. Choix des capteurs: 11

6. Système de reconnaissance d'objet: 12

7. Système de reconnaissance de la main: 12

7.1. Détection de la main: 13

7.1.2. Méthode basée sur le filtrage de couleur de la peau: 13

7.1.3. Méthode basée sur l'extraction de caractéristiques géométriques: 17

7.1.4. Approche basée sur le mouvement: 18

7.1.4.1. Différence d'image: 18

7.1.4.2. Calcul du flot optique: 19

7.2. L'étape de reconnaissance de la main: 19

8. La reconnaissance du mouvement: 20

9. Conclusion: 21

Chapitre n°2 : Interaction gestuelle:

1. Introduction: 22

2. Définition de geste: 22

3. Style des gestes: 23

4. Les gestes de la main pour l'interaction Homme-Machine: 24

5. Taxonomie des gestes de Quek: 25

5.1. Les gestes symboliques: 25

5.2. Les gestes actifs:	25
6. Gestes statiques et dynamiques:.....	26
7. Interaction gestuelle:	28
7.1. Avantages des interactions gestuelles:	28
7.2. Dispositifs d'interaction:	28
8. Reconnaissance de gestes humains:.....	31
8.1. Les méthodes de reconnaissance des gestes de la main:	31
9. Conclusion:	35

Chapitre n°3 : Conception:

1. Introduction:.....	36
2. Spécification:	36
3. Conception globale:	36
4. Conception détaillé:	38
4.1. Module de détection:	38
4.1.1. Acquisition:	38
4.1.2. Prétraitement:	38
4.2. Module de reconnaissance:.....	43
4.3. Module d'interaction:	46
5. Conclusion:	47

Chapitre n°4 : Implémentation:

1. Introduction.....	49
2. Environnement de développement matériel et logiciel.....	49
2.1. Environnement de développement matériel	49
2.2. Environnement de développement logiciel	49
3. Description de l'application :.....	51
3.1. Les composants de l'interface :	51

3.2. La description des différent commandes :	53
3.2.1. La commande de la sélection :	53
3.2.2. La commande de la copier :	54
3.2.3. La commande de la imprimer :	55
3.2.4. La commande de La pause :	56
4. Test et Résultat :	57
5. Discussion des résultats :	62
6. Les difficultés et problèmes rencontrés :	62
7. Conclusion	63
Conclusion générale	64
Bibliographie	66

Liste des figures

Figure 1. 1: Modélisation de la vision.....	9
Figure 1. 2: Système à une seule caméra.	12
Figure 1. 3: contour de la main.	13
Figure 1. 4: Sur la gauche, l'image originale sur la droite, l'image binaire résultant après l'application du filtre de couleur de la peau.....	14
Figure 1. 5: Présentation des couleurs dans l'espace RGB.....	15
Figure 1. 6: Présentation des couleurs dans l'espace HSV.....	16
Figure 1. 7: (a) une scène simple avec deux objets, (b) les régions rouges sont la différence entre deux images consécutives.	19
Figure 2. 1: le langage des signes [3]	24
Figure 2. 2: Taxonomie des gestes de Quek.....	25
Figure 2. 3: Exemple de deux gestes symboliques. (a) le geste de continuation, (b) le geste demandant du feu	26
Figure 2. 4: les classifications de harling et edwards des gestes dynamique et statique.....	27
Figure 2. 5: Exemple de gestes statiques et dynamiques, utilisés par QUEK (les pointillés représentent la configuration initiale), extrait de [3].....	27
Figure 2. 6: Le Leaf Menu. a, b, c, d) les différentes étapes de l'interaction novice. e) l'interaction experte [27].	29
Figure 2. 7: Décomposition d'un système interactif	30
Figure 2. 8: les méthodes de reconnaissance de la main	32
Figure 2. 9: reconnaissance de posture de la main	33
Figure 2. 10: a) Segmentation de la couleur de la peau, b) extraction de contour, c) Après l'algorithme de Convex Hull.....	34
Figure 3. 1: Système d'interaction par gestes.	37
Figure 3. 2: Illustration du processus de notre système.	37
Figure 3. 3: Illustration du processus de prétraitement	38
Figure 3. 4: la détection de la peau.....	39
Figure 3. 5: le filtre médian	41
Figure 3. 6: Représentation des régions dans une image binaire	41
Figure 3. 7: Premier balayage	42

Figure 3. 8: Deuxième balayage.....	42
Figure 3. 9: Image avant et après l'étiquetage.....	42
Figure 3. 10: le processus d'étiquetage de l'objet.....	43
Figure 3. 11: les gestes déictiques.....	44
Figure 3. 12: la forme du doigt pour lancer l'action.....	44
Figure 3. 13: le suivi du doigt.....	45
Figure 3. 14: les commandes d'interaction.....	46
Figure 3. 15: Zones de commande.....	47
Figure 3. 16: la sélection d'une phrase.....	47
Figure 4. 1 l'interface principale de l'environnement de Matlab.....	51
Figure 4. 2: l'interface de notre application.....	52
Figure 4. 3: fenêtre principale lors du traitement d'une commande gestuelle.....	53
Figure 4. 4: résultat de la commande de sélection.....	54
Figure 4. 5: résultat de la commande Copier.....	55
Figure 4. 6: résultat de la commande Imprimer.....	56
Figure 4. 7: le résultat de la commande Pause.....	57
Figure 4. 8: résultat de teste sur une image.....	58
Figure 4. 9: le résultat de teste sur un graphe.....	58

Liste des tableaux:

Tableau 4. 1: les caractéristiques de l'environnement matériel	49
Tableau 4. 2: Résultats possibles de la sélection.....	60
Tableau 4. 3: les différents résultats du premier test.....	60
Tableau 4. 4: les caractéristiques de l'environnement de test	61
Tableau 4. 5: les différents résultats du deuxième test.....	61

Introduction générale

La réalité augmentée vise à créer une coopération performante entre l'électronique et les pratiques non informatisées du quotidien. Son objectif est d'estomper la barrière entre les objets informatiques (par exemple, un document présenté à l'écran) et les objets et activités du monde réel (tel le livre ouvert sur la table de travail).

Les technologies de réalité augmentée fonctionnent toutes sur le principe de la vision par ordinateur ou « Computer Vision ». Propulsées par des algorithmes complexes, elles permettent de détecter des éléments distinctifs dans l'univers réel, puis d'y superposer du contenu virtuel.

Dans le domaine de l'Interaction Homme-Machine (IHM), l'utilisation de la vision par ordinateur permet de se libérer de l'interaction traditionnelle (utilisation du clavier et de la souris), en permettant une interaction 'naturelle' à travers l'usage du corps de l'utilisateur comme dispositif d'interaction. La main ou le corps peuvent permettre de communiquer par des gestes avec un ordinateur. Cela est déjà possible avec les interfaces tactiles (stylo, doigts), et peut être étendu à des dispositifs intégrant la vision par ordinateur. C'est le cas dans les jeux vidéo avec la Kinect, qui est basée sur un périphérique d'entrée branché sur la console qui permet d'interagir par commande vocale, reconnaissance de mouvement gestuelle, faciale et d'image. On peut jouer donc sans aucune manette ni périphérique autre que son propre corps.

est-ce que me reconnoît les??

Notre travail a porté sur la réalisation d'une application pour la reconnaissance des gestes de doigt et leur utilisation pour l'interaction avec un ordinateur. Ainsi, l'interaction se fait au doigt, en minimisant le plus possible l'utilisation de périphériques d'entrée, tels que la souris et le clavier.

L'utilisateur pourra sélectionner sur une feuille de papier, filmé avec une webcam, une partie d'un texte, pour pouvoir ensuite appliquer des traitements informatiques sur la sélection. La sélection se fait par l'utilisateur, en pointant du doigt les zones de texte qu'il veut sélectionner. Les traitements (copier, imprimer) sont également commandés par l'utilisateur à travers des gestes dédiés.

Le présent mémoire est composé de quatre chapitres, décrits ci-dessous :

Dans le premier chapitre on a introduit le concept de la vision par ordinateur et sa relation avec les systèmes de reconnaissance qui permettent de reconnaître les caractéristiques d'un objet à partir d'images ou d'un flux vidéo.

Dans le deuxième chapitre, nous avons défini les différentes étapes de la conception d'un système de reconnaissance du geste et aussi les méthodes de reconnaissance des gestes de la main dans le domaine des interactions homme-machine.

Dans le troisième chapitre, nous présentons les étapes franchies pour la conception et la réalisation de notre système.

Dans le quatrième chapitre, on a montré l'implémentation de la conception de notre système, les différents tests réalisés et les résultats fournis par notre système.

Chapitre I

Vision par ordinateur

1. Introduction:

A l'heure actuelle, la recherche sur la reconnaissance d'objet est de plus en plus étendue. La reconnaissance du visage est utilisée, par exemple pour reconnaître des suspects, lors d'actions terroristes ou autres.

Un système de reconnaissance permet de reconnaître les caractéristiques d'un objet à partir d'images ou de flux vidéo issues d'un système de vision par ordinateur.

2. Définition de la vision par ordinateur:

La vision par ordinateur est un processus de traitement de l'information caractérisé par:

- L'entrée est une séquence d'images.
- La sortie est une description de l'entrée en termes d'objets d'intérêt et de relations entre ces objets [1].

L'ordinateur transpose le travail de perception du système de vision humaine en système automatique.

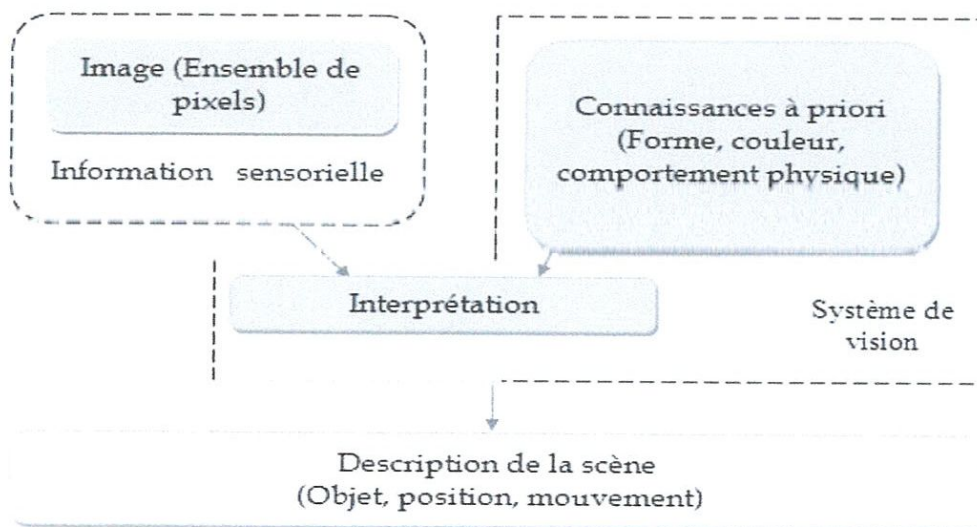


Figure 1. 1: Modélisation de la vision.

3. Système de vision:

Un système de vision comporte en général:

- Une source lumineuse pour éclairer la scène.
- Un capteur.
- Une interface (carte d'acquisition) entre le capteur et l'ordinateur qui convertit l'information analogique en information numérique.
- Un système de traitement (ordinateur et logiciels).

L'utilisation de capteurs de vision est très importante parce qu'ils fournissent une grande richesse d'informations [2].

4. Difficultés liées à la vision par ordinateur:

La conception d'un système de vision par ordinateur nécessite la prise en compte d'un certain nombre de difficultés [3]:

➤ **Éclairage de la scène:** les algorithmes de segmentation et d'extraction de caractéristiques sont influencés par les différences dans l'éclairage de la scène, ce qui provoque des changements de luminosité dans l'image.

➤ **Ombres:** les objets projettent une ombre dans la scène, même si la luminosité est contrôlée. Les ombres sont parfois segmentées avec l'objet, ce qui rend difficile d'utiliser le résultat de la segmentation.

➤ **Objets de la scène:** la scène peut inclure des objets en mouvement. Il faut que l'algorithme de segmentation des objets se réactualise par rapport aux changements du fond de la scène, pour pouvoir la différencier des autres objets.

➤ **Camouflage:** le grand problème lorsque l'objet est proche des objets du fond de la scène, et que celui-ci est difficilement détectable, par exemple, lorsqu'on veut segmenter la main sur un fond contenant un bureau ayant une couleur de teinte chair.

➤ **Occultation:** se fait lorsque l'objet est caché par un autre, par exemple lors de l'utilisation de deux mains, l'une pouvant cacher l'autre. Dans le cas de la main, il y a un problème supplémentaire avec les « auto- occultations », car la main étant un objet complexe et déformable, il arrive souvent que certaines parties, comme les doigts soient cachées par d'autres.

➤ **La disposition des caméras:** la distance entre la caméra et les objets est importante, ainsi que l'angle de vision des caméras qui peut provoquer des distorsions géométriques dans l'image.

5. Choix des capteurs:

Il y a trois systèmes de capture, cités ci-dessous [4]:

➤ **Système stéréographique:** Le système stéréographique fournit des informations très importantes, telle la profondeur pixellisée pour tout point dans le champ de vision des caméras.

Cependant, il faut beaucoup de temps processeur pour calculer les algorithmes stéréo en temps réel, les données 3D ne sont donc pas facilement obtenues ou mises en œuvre.

➤ **Système de vue bidimensionnelle:** Avec ce système, deux ou plusieurs vues 2D du même objet, fournies par des caméras séparées, pourraient être combinées. Bien que chaque vue aurait à souffrir de problèmes similaires, les points de vue combinés de suffisamment de caméras pourraient révéler des données suffisantes pour analyser la scène.

➤ **Système de caméra unique:** Ce système donne beaucoup moins d'informations sur les objets. Certaines informations (comme le doigt sur un fond de la peau) seraient très difficiles à distinguer car il ne peut pas récupérer l'information de profondeur. Par contre, l'information "silhouette" pourrait être extraite avec précision. Les données silhouette seraient relativement sans bruit et nécessitent moins de temps processeur à calculer.

Le système de caméra unique est moins bruyant, et moins coûteux en temps de calcul. Il est possible, par exemple, de détecter un grand sous-ensemble de gestes en utilisant des informations de silhouette. L'inconvénient de ce système est qu'il présente plus d'ambiguïté qu'avec l'un des autres systèmes.

La figure 2 présente un exemple de système utilisant une seule caméra, montée au-dessous d'une surface de bureau de couleur neutre (dispositif présenté dans [4]).

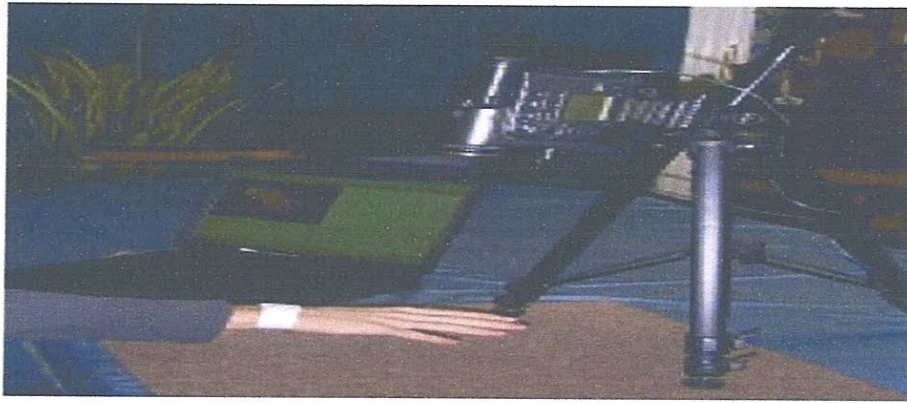


Figure 1. 2: Système à une seule caméra.

6. Système de reconnaissance d'objet:

La reconnaissance automatique d'objets est classée dans les problèmes les plus difficiles en vision par ordinateur. Elle est une étape importante pour la mise en œuvre de plusieurs applications actuelles qui nécessitent une interprétation de haut niveau d'images [5]

Ce système exploite des traitements pour l'extraction des caractéristiques sur chaque image, Ces caractéristiques peuvent être la couleur, la taille, la forme, la texture, etc., puis leur exploitation pour reconnaître un objet [6].

La reconnaissance d'objet peut être la reconnaissance d'un visage, d'un corps humain ou la reconnaissance de la main, etc.

Les étapes d'un système de reconnaissance des objets sont les suivantes: la détection de l'objet, le suivi de l'objet et, enfin, la reconnaissance.

Nous allons nous intéresser par la suite à la reconnaissance d'un objet spécifique: la main.

7. Système de reconnaissance de la main:

Pour la reconnaissance de la main, les étapes d'un système de reconnaissance sont comme suit:

7.1. Détection de la main:

La détection correspond au repérage de la main sur une image par segmentation de la zone de l'image où elle se trouve. Pour cela, il existe de nombreuses méthodes basées sur la couleur de la peau, la forme, le mouvement, des modèles anatomiques, etc. [7].

7.1.1. Méthode basée sur le contour:

Elle consiste à détecter et à extraire la région de la main de l'arrière-plan et la segmentation de l'image de la main. Les caractéristiques du modèle de la main, extraites de la région de la main segmenté, sont représentées par son contour [8]. La figure 3 présente un exemple du contour d'une main.



Figure 1. 3: contour de la main.

7.1.2. Méthode basée sur le filtrage de couleur de la peau:

La détection de couleur de peau humaine est largement utilisée et s'est avérée être une caractéristique efficace dans plusieurs applications, parmi lesquelles, la détection du visage et le suivi des mains. Bien que la couleur de la peau puisse largement varier, les récentes recherches montrent que la différence principale est plutôt dans l'intensité que dans la chrominance [9].

Plusieurs espaces de couleur sont utilisés pour étiqueter les pixels comme pixels de couleur de peau: RGB, RGB normalisé, HSV, YCrCb, YIQ,... De plus, des méthodes sont proposées pour construire un modèle de couleur de peau: méthode utilisant la tonalité de pixel, méthode basée sur l'histogramme [10], méthode utilisant une fonction gaussienne [11].

La figure 4 illustre le résultat d'une détection de la peau de la main.

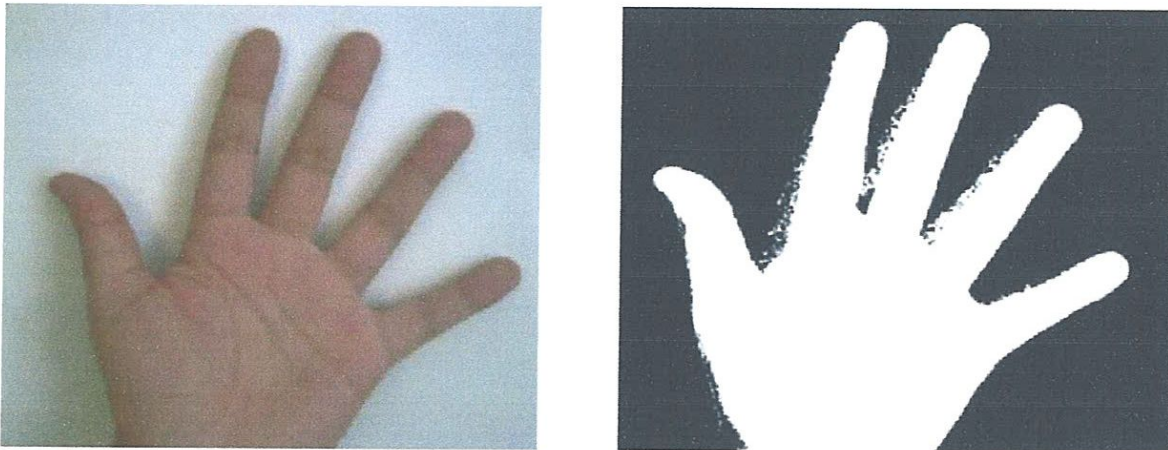


Figure 1. 4: Sur la gauche, l'image originale sur la droite, l'image binaire résultant après l'application du filtre de couleur de la peau.

La couleur de peau est représentée dans plusieurs espaces chromatiques:

1. Espace de couleur RVB:

Le codage RGB (en français RVB), mis au point en 1931 par la Commission Internationale de l'Eclairage (CIE) consiste à représenter l'espace des couleurs à partir de trois rayonnements monochromatiques de couleurs [12]:

- Rouge (de longueur d'onde égale à 700,0 nm).
- Vert (de longueur d'onde égale à 546,1 nm).
- Bleu (de longueur d'onde égale à 435,8 nm).

Etant donné que le codage RGB repose sur trois composantes proposant la même gamme de valeur, on le représente généralement par un cube dont chaque axe correspond à une couleur primaire (Figure 5).

Les paramètres ci-dessous sont utilisés pour détecter la couleur de la peau humaine [13]:

Cas éclairage fort:

$$(R > 95) \text{ et } (G > 40) \text{ et } (B > 20) \text{ et } ((\max(R, G, B)) - \min([R, G, B])) > 15 \text{ et } (\text{ABS}(R - G) > 15) \text{ et } (R > G) \text{ et } (R > B)$$

Cas éclairage faible:

$$(R > b) \text{ et } (G > B) \text{ ou } ((R > 220) (G > 210) \text{ et } (B > 170) \text{ et } (ABS(R-G) \leq 15))$$

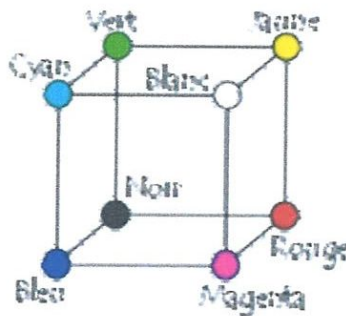


Figure 1. 5: Présentation des couleurs dans l'espace RGB.

2. Espace de couleur HSV: (Hue, Saturation, Value).

C'est l'espace le plus proche de la perception physiologique de la couleur par l'œil humain. Il consiste à décomposer la couleur selon des critères physiologiques:

- **la teinte**, correspondant à la perception de la couleur,
- **la saturation**, décrivant la pureté de la couleur, c'est-à-dire son caractère vif ou terne,
- **la luminance**, indiquant la quantité de lumière de la couleur, c'est-à-dire son aspect clair ou sombre [13].

Dans l'espace HSV, l'information d'intensité est représentée par le canal V, pour cette raison, ce canal doit être négligé dans le processus de détection de la peau, on ne considère que les canaux H et S qui représentent l'information chromatique.

Dans l'espace HSV, on a: $H \in [0 - 360^\circ]$ et $S \in [0 - 1]$ et $V \in [0 - 1]$ (Figure 6).

La segmentation de la couleur de peau selon l'espace HSV nécessite la conversion de l'image originale de l'espace RGB vers l'espace HSV dans lequel la couleur de peau se représente dans les intervalles suivants:

$$0 < H < 50 \text{ et } 0.23 < S < 0.68$$

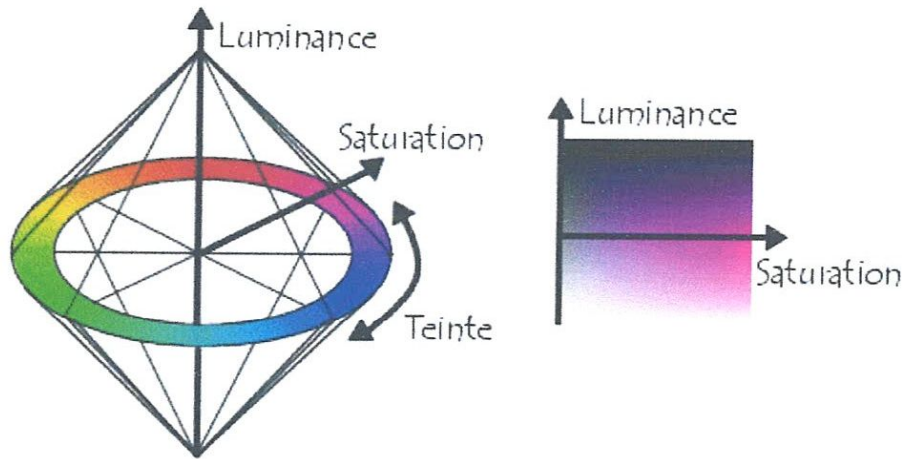


Figure 1. 6: Présentation des couleurs dans l'espace HSV.

3. Espace de couleur YCbCr:

La luminance (appelée aussi Luma) est séparée de la couleur, présentée par la valeur Y et calculée par une somme pondérée des composantes de R, G et B [12]:

$$Y = 0,299 * R + 0,587 * G + 0,114 * B$$

Les deux autres composantes de cet espace représentent l'information chromatique et elles sont calculées à partir de *Luma*:

$$Cr = R - Y$$

$$Cb = B - Y$$

Comme pour l'espace HSV, seul les deux canaux Cb et Cr sont retenus à l'application du seuillage.

Deux seuillages représentent la couleur de la peau [13]:

Seuillage YCbCr(1):

$$((Cb \geq 85) \text{ et } (Cb \leq 135)) \text{ Et } ((Cr \geq 135) \text{ et } (Cr \leq 180))$$

Seuillage YCbCr(2):

$$((Cb \geq 77) \text{ et } (Cb \leq 127)) \text{ Et } ((Cr \geq 133) \text{ et } (Cr \leq 173))$$

7.1.3. Méthode basée sur l'extraction de caractéristiques géométriques:

La description des objets à l'aide de la forme nécessite l'extraction de quantités sans dimension indépendantes de sa taille qui sont représentatifs de la forme de l'objet, également appelées: les acteurs de la forme. Ceux-ci sont généralement calculées en utilisant des mesures de l'objet segmenté telles que le diamètre, la région, etc., et de représenter leur ressemblance avec des objets idéaux, par exemple ellipse, cercle, etc. Souvent, les facteurs de forme sont normalisés de sorte que les quantités de similarité varie de 0 à 1, lorsque 1 correspond à la similitude maximale. Ils sont applicables à toutes formes géométriques [14].

Dans la littérature de nombreux algorithmes utilisent des facteurs de forme pour extraire des caractéristiques de forme de la main. On va présenter ici certains facteurs de forme couramment utilisés pour la représentation de la forme de la main:

- **Centre coordonnées (x, y):** Cette position permet de déterminer différents aspects spatiaux de l'objet. Par exemple, le centre de gravité de la région, la position du centre de la paume, le centre de la zone de délimitation contenant la région, etc. Cela dépend directement de l'algorithme utilisé pour la détection et le suivi des objets et le modèle choisis pour la détection et le suivi.
- **Zone A et périmètre P:** la zone quantifie la mesure d'une surface 2D. Dans l'idéal la surface d'objet doit avoir une forme prédéfinie. En traitement d'image, la zone d'une région irrégulière est calculée en utilisant les pixels appartenant à la région d'une image binaire.

Ceci est défini comme suit:

$$a = \sum_{\{x,y\} \in I} p_{x,y}$$

Où:

- $\{x, y\}$ correspond à la position de chaque pixel dans l'image binaire I et $p_{x,y}$ La valeur du pixel 0 ou 1.

- le périmètre correspond à la quantité de longueur de chemin entourant une zone. C'est le nombre de pixels constituant le contour de l'objet.

7.1.4. Approche basée sur le mouvement:

L'utilisation de l'information de mouvement peut être un moyen simple pour mettre en œuvre une technique rapide de détection de peau dans une vidéo. Généralement, cette approche est utilisée conjointement avec l'information de couleur de peau, car elle suppose que la scène est statique et la peau (visage, main) est en mouvement [3]. On trouve dans cette approche plusieurs techniques:

7.1.4.1. Différence d'image:

Ces approches basiques considèrent la caméra comme fixe durant l'acquisition de la séquence d'images.

- **Soustraction de fond:** Cette méthode est très populaire et elle est utilisée par plusieurs applications. Elle détecte la région de mouvement en soustrayant pixel par pixel l'image courante à l'image de fond.

Soit I_t l'image à l'instant t . B_t est l'image de fond à l'instant t . L'objet en mouvement se compose des pixels qui satisfont l'équation suivante [15]:

$$\max (|I_t(x, y) * c - B_t(x, y) * c|, c = (R, G, B)) \geq S(x, y)$$

Où $S(x, y)$ est le seuil de la position (x, y) à l'instant t . Cette méthode nous permet d'obtenir la forme complète de l'objet et de compter exactement les objets du mouvement parce que l'image de fond n'est pas l'objet mouvant. Cependant, dans un environnement réel, le changement de l'espace a lieu souvent, ainsi que le changement de la lumière: le matin, le soir, il y a du soleil, il pleut, etc.

- **Différence entre deux images successives:** Dans cette méthode, les pixels dont l'intensité résultante est proche de zéro sont assimilés comme étant les pixels du fond.

Le choix du seuil de décision (proche de zéro) n'est pas évident et doit tenir compte du bruit et des changements de luminosité. De plus, ce seuil n'a aucune raison d'être global sur toute l'image. En effet, des objets et des mouvements différents entraînent des variations d'intensité différentes. De nombreuses méthodes, utilisant deux ou trois images successives, permettent de décider si un pixel a bougé [16]. La figure 7 (tirée de [15]) présente un exemple d'utilisation de cette méthode.



Figure 1. 7: (a) une scène simple avec deux objets, (b) les régions rouges sont la différence entre deux images consécutives.

7.1.4.2. Calcul du flot optique:

Un flot optique peut provenir du mouvement relatif des objets à un spectateur pour qu'il puisse donner des informations importantes sur l'arrangement spatial des objets vus et le taux de variation de cet arrangement [17].

Le flot optique est un champ de vecteurs à deux dimensions représentant la projection sur le plan image du mouvement réel observé (tridimensionnel). Cette méthode est très utilisée mais elle reste une opération très coûteuse en temps de calcul. Le temps-réel peut néanmoins être atteint en sous-échantillonnant les trames et en choisissant un algorithme rapide [18].

L'un des algorithmes les plus rapides et les plus populaires est celui de Lucas et Kanade (1981) [19]. Il s'agit de calculer au temps t , le déplacement $d = (d_x, d_y)^t$ du point $x = (x, y)^t$. L'hypothèse sur laquelle se base la méthode est celle de la conservation de l'intensité lumineuse.

7.2. L'étape de reconnaissance de la main:

Elle permet de regrouper les données spatio-temporelles extraites dans les étapes de détection, suivi et d'assigner aux groupes résultant des étiquettes associées à des catégories particulières de gestes [2].

?? mouvement par caméra

8. La reconnaissance du mouvement:

La reconnaissance du mouvement est un sujet informatique qui a comme but d'interpréter les gestes de l'humain via des algorithmes mathématiques. Les gestes peuvent provenir de n'importe quel mouvement du corps ou d'une pose, mais les centres d'intérêt dans ce domaine sont le visage, et plus précisément la reconnaissance des émotions qui s'expriment à travers le visage, et les gestes faits avec la main.

La reconnaissance des gestes peut être considérée comme un moyen pour les ordinateurs de commencer à comprendre le langage du corps humain, créant ainsi une liaison plus forte entre les machines et les humains, autre que celle offerte par les interfaces primitives en mode texte ou même des interfaces graphiques, qui limitent encore la communication au clavier et à la souris.

Les mouvements des doigts de la main sont régis par des contraintes biomécaniques (Ces contraintes sont statiques ou dynamiques) qui rendent certaines postures irréalisables [20, 21].

En ce qui concerne la programmation de la reconnaissance du mouvement, celle-là peut être accomplie avec des techniques de vision par ordinateur et de traitement d'images.

Baudel et Beaudouin-Lafon [22] présentent un certain nombre d'avantages offerts par l'utilisation des gestes pour l'interaction:

- Une interaction naturelle: les gestes sont une forme naturelle de l'interaction, et sont faciles à utiliser ;
- Simple et puissant: un geste simple peut être utilisé pour spécifier à la fois une commande ainsi que ses paramètres ;
- Une interaction directe: la main utilisée en tant que périphérique d'entrée élimine le besoin d'objets intermédiaires entre la personne et la machine.

L'interaction gestuelle sera détaillée dans le chapitre suivant.

9. Conclusion:

Dans ce chapitre, nous avons présenté les systèmes de vision par ordinateur. Ces systèmes utilisent des capteurs qui donnent des images, à partir desquelles sont détectées des objets qui sont extraits à partir d'un certain nombre de caractéristiques utilisées pour reconnaître ces objets.

Pour un objet spécifique, la main, cette reconnaissance permet le développement de système d'interaction par geste, développées dans le chapitre suivant.

Chapitre II

Interaction gestuelle

1. Introduction:

L'objectif de recherche dans le domaine des interactions homme-machine est de développer des modèles, des concepts, des outils et des méthodes pour réaliser des systèmes qui répondent aux besoins et aux aptitudes des utilisateurs. Une des aptitudes des utilisateurs est l'emploi des gestes.

La reconnaissance de gestes humains est l'un des nombreux défis posés dans le domaine de la vision par ordinateur et dans le domaine des interactions humains-machines. Cette problématique qui a gagné en attention au cours des dernières décennies, faisant suite aux nouveaux développements technologiques et nouveaux résultats scientifiques obtenus dans les communautés du traitement du signal et de l'apprentissage automatique, entre autres.

Par exemple, la Kinect de Microsoft permet à l'utilisateur de contrôler un jeu sans utiliser de manette mais uniquement grâce aux mouvements du corps ou n'importe quelle geste avec la main, rendant l'interaction avec le jeu plus directe et plus naturelle qu'avec une manette.

2. Définition de geste:

Un geste, d'après l'Oxford English Dictionary, est le mouvement du corps ou de n'importe quelle partie du corps (par exemple la main) pour exprimer une pensée ou un sentiment. C'est-à-dire qu'il n'y a geste que lorsqu'un mouvement du corps est effectué avec une intention bien précise.

Le geste peut à la fois être considéré d'un point de vu physiologique ou du point de vue de l'interaction [23].

Du point de vu physiologique, un geste est le résultat de contractions ou détractions musculaires visibles, réflexes ou volontaires.

Du point de vue de l'interaction, un geste est une forme de communication non verbale.

Pour De Marconnay [1] « le geste englobe tous les mouvements des mains permettant de communiquer des informations significatives et pertinentes ».

3. Style des gestes:

Dans [24], Karam et al. Présentent cinq styles de gestes, principalement de la main et du bras, résumant les types d'interaction décrits par la littérature scientifique:

1. les gestes déictiques: sont des mouvements de pointage qui sont en général exécutés avec l'index tendu et les autres doigts pliés, mais parfois aussi avec d'autres parties du corps (tête, nez, menton...) ou par l'intermédiaire d'artefacts (règle, stylo...). Ils désignent un objet qui est simultanément référencé dans le discours [24].

2. Les gestes de manipulation: le but de ces gestes est de contrôler une entité, impliquant une relation étroite entre le mouvement du geste et l'entité qui est manipulée [23].

3. Les gestes sémaphoriques: ils sont basés sur un catalogue conventionnel de gestes statiques ou dynamiques [23].

Ex: - le geste statique pour signifier « ok ».

- le signe dynamique de la main pour dire « au revoir ».

4. Les gesticulations: ce sont les gestes les plus naturels, ils sont également nommés: les gestes co-verbal. L'interprétation de ce type de gestes est du domaine de la recherche en reconnaissance de geste le plus ambitieux car contrairement aux gestes sémaphoriques, leur signification ne peut pas être dérivée directement à partir d'un dictionnaire de gestes et doivent être mis en relation avec d'autres modalités telles que la parole [23].

5. le langage des signes: désignent les langues gestuelles. Les gestes qui sont utilisés dans ce langage sont souvent considérés indépendamment des autres types de gestes car ils sont basés sur des principes de linguistique et qu'ils permettent une combinaison de gestes et signes pour former des structures grammaticales utiles à la conversation. La figure suivante montre un exemple de ce type de langage (figure 2.1) [3].

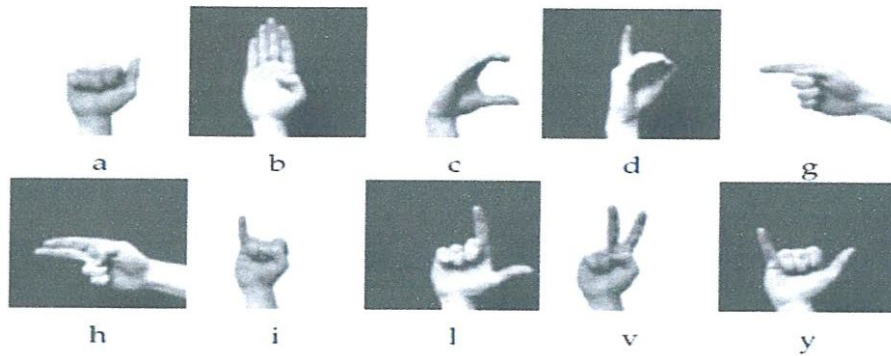


Figure 2. 1: le langage des signes [3]

4. Les gestes de la main pour l'interaction Homme-Machine:

Dans le domaine des Interfaces Homme-Machine (IHM), la main peut servir à pointer (pour remplacer la souris), à manipuler des objets (pour la réalité augmentée ou virtuelle), ou à communiquer par gestes avec un ordinateur. Par rapport à la richesse de l'information véhiculée par les gestes de la main, les possibilités de communication avec les ordinateurs sont aujourd'hui réduites avec la souris et le clavier.

Le choix des gestes de la main à utiliser pour l'interaction Homme-Machine est important car les gestes doivent pouvoir être réalisés intuitivement, ou avec une période d'apprentissage limitée.

Cadoz définit trois fonctions principales de la main [3]:

- **La fonction ergotique:** la main joue le rôle d'organe moteur, elle intervient sur le monde physique pour le transformer. Par l'intermédiaire de l'ensemble de la structure osseuse et des muscles, elle applique à un objet des forces, pour le déplacer ou le déformer.

- **La fonction épistémique:** dans ce cas, la main joue le rôle d'organe de perception. Par l'intermédiaire du sens tactilo-proprio-kinesthésique, des informations sur la température, l'état de surface, le poids, la forme, ou les mouvements des objets, peuvent être obtenues.

- **La fonction sémiotique:** la main joue alors le rôle d'organe d'émission d'informations à destination de l'environnement. Elle s'adresse à la perception visuelle d'un ou de plusieurs interlocuteurs. 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 bruyé ou en plongée sous-marine.

La fonction sémiotique du geste est celle qui est la plus riche et la plus complexe. Elle peut être décomposée en plusieurs catégories, et différentes classifications ont été proposées.

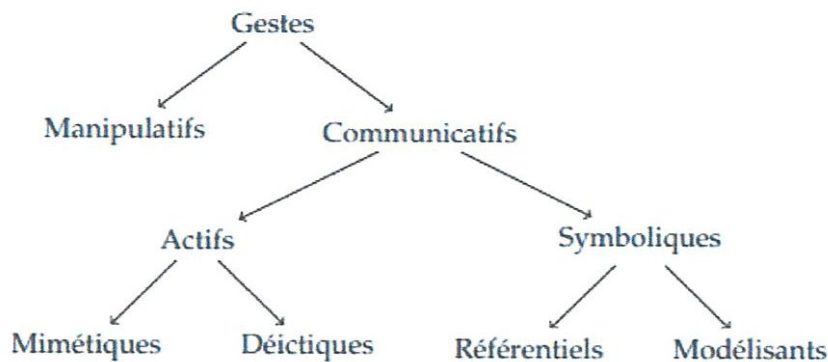


Figure 2. 2: Taxonomie des gestes de Quek

5. Taxonomie des gestes de Quek:

La taxonomie de Quek [24] est une classification bien adaptée au domaine de l'interaction homme-machine qui décompose les gestes en gestes manipulatifs, correspondant aux fonctions ergotique et épistémique, et en gestes communicatifs, correspondant à la fonction sémiotique. Les gestes communicatifs sont décomposés en gestes actifs et en gestes symboliques (figure 2.2).

5.1. Les gestes symboliques: demandent un apprentissage pour comprendre leur signification. Il s'agit par exemple des gestes des langues des signes.

Ils sont décomposés en deux sous-ensembles de gestes:

- Geste référentiels, faisant directement référence à un objet ou à un concept (par exemple, le frottement du pouce et de l'index pour évoquer l'argent),
- Gestes modélisant, qui modélisent un état ou une opinion et s'emploient souvent en addition à d'autres moyens de communication (exemple pour donner une idée de la taille d'un objet). Ces gestes peuvent avoir un sens différent selon la culture.

5.2. Les gestes actifs: sont directement liés à leur interprétation et sont utilisés en complément de la parole. Ils sont décomposés en deux gestes:

- Gestes mimétiques, consistant à mimer une action.

- Gestes déictiques, ou gestes de pointage, qui sont très utilisés pour l'interaction homme-machine.

La figure suivante (figure 2.3), extraite de [1], montre deux types de geste symboliques:

(a) le geste de continuation: mouvements circulaires répétés de la main pointant sur le côté. Ce geste a un sens différent suivant le contexte pouvant indiquer « continue de me donner des exemples », « continue de faire défiler la page », « continue d'avancer », « fais tourner »,

(b) le geste demandant du feu: mouvement de pliage/dépliage du pouce imitant l'utilisation d'un briquet.



Figure 2. 3: Exemple de deux gestes symboliques. (a) le geste de continuation, (b) le geste demandant du feu

6. Gestes statiques et dynamiques:

On différencie également les gestes dynamiques des gestes statiques (figure 2.5).

Un geste statique, également appelé posture, concerne la configuration du corps ou d'une partie du corps à un moment fixe dans le temps alors que le geste dynamique désigne une succession continue de postures.

Il faut donc distinguer la position et la configuration de la main. En combinant ces deux aspects, on obtient la classification proposée par Harling et Edwards [25] (figure 2.4):

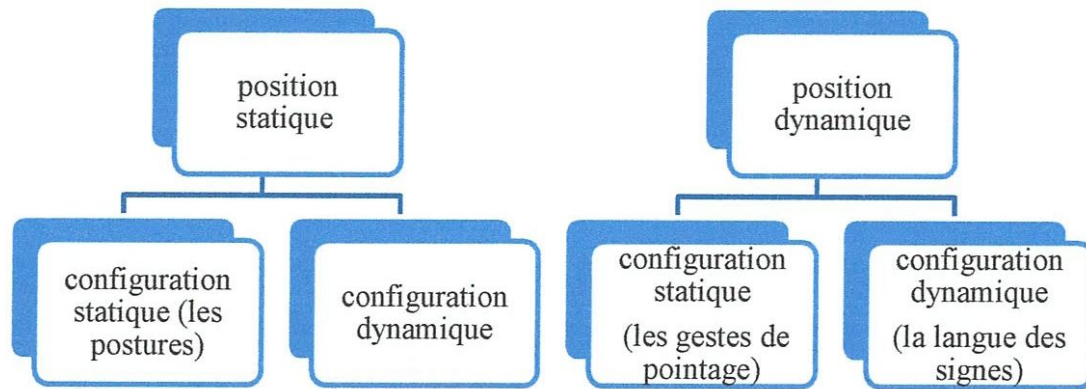


Figure 2. 4: les classifications de harling et edwards des gestes dynamique et statique

Mais il est aussi possible de regrouper la position et la configuration de la main dans un vecteur de mesures à n dimensions. Ainsi, un geste est représenté par une trajectoire dans l'espace des mesures.

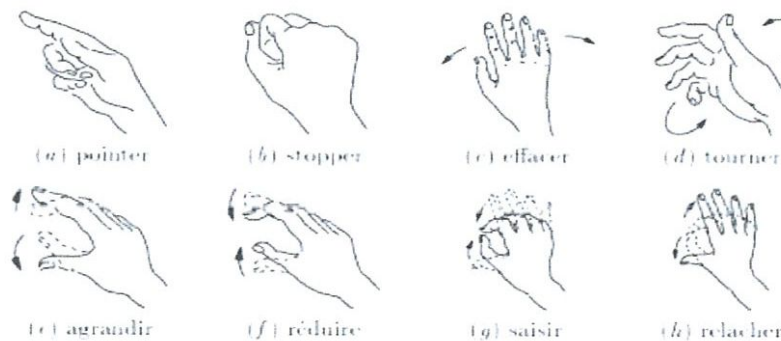


Figure 2. 5: Exemple de gestes statiques et dynamiques, utilisés par QUEK (les pointillés représentent la configuration initiale), extrait de [3].

Deux réalisations d'un même geste ne donnent pas exactement le même vecteur de paramètres, mais les valeurs sont suffisamment proches pour que le geste soit identifiable.

7. Interaction gestuelle:

La richesse des interactions gestuelles permet aux concepteurs d'interfaces de proposer de nouvelles modalités pour interagir avec le système, au-delà des interactions plus répandues comme pointer, déplacer, tourner et redimensionner des objets graphiques [26].

7.1. Avantages des interactions gestuelles:

Les interactions gestuelles tendent à enrichir la communication Homme-Machine par les améliorations suivantes [2]:

- Interaction plus naturelle: la communication gestuelle fait appel aux moyens de communication humains naturels.
- Interaction concise et efficace: Les utilisateurs peuvent exprimer une commande avec une seule posture.
- Interaction directe: la main devient elle-même le périphérique d'interaction, dans les interfaces gestuelles.
- Interaction standard: une interaction gestuelle est manipulable par toutes les catégories d'utilisateurs. Elle ne nécessite aucun entraînement.

7.2. Dispositifs d'interaction:

La majorité des systèmes d'exploitation repose sur le paradigme WIMP, avec un dispositif de pointage, généralement la souris, qui permet d'interagir avec des éléments graphiques tels que des fenêtres, des icônes et des menus. Il existe aussi des interfaces haptiques, qui permettent un retour d'information à l'utilisateur, avec un retour de toucher ou un retour d'effort. La perception tactilo-kinesthésique, ou haptique, résulte de la stimulation de la peau par le contact avec des objets [3].

Nous présentons deux dispositifs d'interaction par geste: les écrans tactiles et les systèmes de vision par ordinateur.

a. Ecran tactile:

Les écrans tactiles équipent les dispositifs commerciaux depuis l'arrivée des PDAs (personnal digital assistant) en 1993 mais se sont surtout répandus de manière exponentielle depuis l'arrivée de l'iPhone en 2007. D'abord résistifs et nécessitant l'utilisation d'une pression suffisante sur l'écran, ils ont ensuite été produits avec différents moyens technologiques par exemple:

Capacitif: quand le doigt, conducteur d'électricité touche l'écran, des charges électrique lui sont transférées.

Caméra infrarouge: telle que sur la tablette multitouche Microsoft Surface, la majorité des tablettes interactives et certains tableaux interactifs.

Acoustique: les tableaux interactifs ultra-soniques.

Pendant, les technologies utilisées sont dépendantes de la taille du dispositif. Ainsi, comme il est encore aujourd'hui difficile de réaliser des écrans capacitifs de grande taille, les technologies infrarouges sont préférées. Par contre, la plupart des écrans destinés aux smartphones sont aujourd'hui des écrans tactiles capacitifs [27]. La figure suivante (figure 2.6) présente un exemple d'utilisation d'un écran tactile.

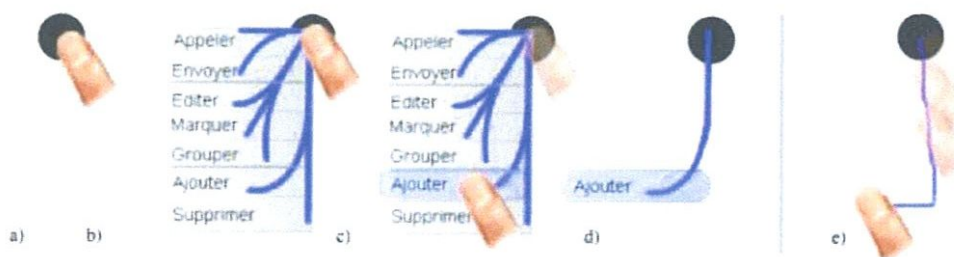


Figure 2. 6: Le Leaf Menu. a, b, c, d) les différentes étapes de l'interaction novice. e) l'interaction experte [27].

b. Les systèmes de vision par ordinateur:

Un système interactif est généralement composé de deux parties: l'interface utilisateur et le noyau fonctionnel (figure 2.7) [28].

L'interface utilisateur est constituée des éléments logiciels et matériels qui sont mis en œuvre lors de la capture des entrées de l'utilisateur et lors de la restitution des sorties du système.

Le noyau fonctionnel représente le système de traitement et de stockage de l'information.

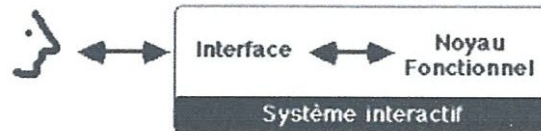


Figure 2. 7: Décomposition d'un système interactif

Dans l'interaction homme-machine, la vision par ordinateur permet [29]:

- d'étendre les capacités visuelles par délégation à la machine de tâches d'observation hors champ de vision ou difficiles pour le système visuel et cognitif: l'analyse de scène et la télésurveillance sont des exemples où la vision par ordinateur intervient comme prothèse visuelle.

- d'interagir avec le système de manière non intrusive, c'est-à-dire sans "fil à la patte", ouvrant ainsi la voie à la disparition de la station de travail telle que nous la connaissons aujourd'hui ;

- de servir au plus près la notion d'engagement dans l'action par élimination des dispositifs intermédiaires lorsqu'ils s'avèrent inutiles. Par exemple, le doigt, dont la trajectoire est suivie par un système de vision, peut agir en direct sur la représentation d'un concept. Dans les interfaces à manipulation directe usuelles, la souris est un intermédiaire imposé. Avec un système de suivi par vision artificielle, le doigt est le dispositif d'action. L'intermédiaire physique, surplus encombrant, est éliminé.

8. Reconnaissance de gestes humains:

La reconnaissance de gestes humains regroupe l'ensemble des techniques visant à capturer des informations caractérisant les mouvements d'un corps humain dans l'espace et à détecter / reconnaître des gestes significatif d'une action, d'une intention ou caractérisant une expression spécifique.

Les applications sont nombreuses [30]:

- Reconnaissance automatique des langues signées: reconnaissance des phrases, stockage sous une forme symbolique compacte en vue, par exemple, de les transmettre sur un réseau ou de les synthétiser via un avatar virtuel.

- Interactions humain-machine: conception de nouvelles interfaces par association geste / commande, réactions pertinentes de la machine vis-à-vis des actions effectuées par un humain, etc.

- Apprentissage d'activité assistée par ordinateur, connaissant une séquence de gestes efficace pour effectuer une certaine activité (musique, sport, ...), la machine peut être à même de détecter des écarts par rapport à un geste expert ou souhaité.

- Vidéo-surveillance: détection automatique de certaines classes de gestes (menaces, attaques, ...).

8.1. Les méthodes de reconnaissance des gestes de la main:

La plupart des systèmes de reconnaissance des gestes utilisent 3 méthodes:

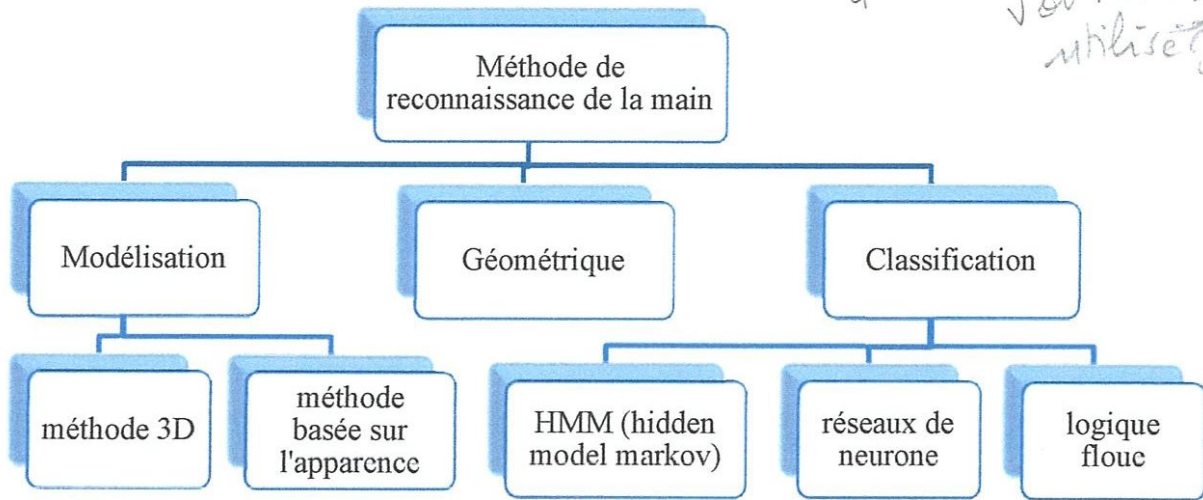


Figure 2. 8: les méthodes de reconnaissance de la main

a. Méthodes par modélisation:

Ces méthodes peuvent être divisées en deux catégories: approches 3D fondées sur un modèle 3D de la main [31] et des approches basées sur l'apparence de la main [32].

➤ Approches 3D:

Cette approche est définie par la construction d'un modèle 3D de la main. Une ou plusieurs caméras fournissent à la station de travail une séquence d'images de la main de l'utilisateur. Pour en déduire la position relative de cette main par rapport aux caméras, on la compare avec le modèle 3D, et on va essayer de calquer ce modèle sur l'image de la main de l'utilisateur dans la séquence d'images.

Par exemple, Digiteyes [31] est un système de suivi de la main à base de modèle, qui peut récupérer l'état d'un modèle de la main avec 27 degrés de liberté, à partir d'images d'échelle de gris ordinaires à des vitesses allant jusqu'à 10 Hz.

Cette approche présente plusieurs inconvénients qui ont empêché son utilisation réelle:

- Le processus d'ajustement est sensible au bruit (par exemple des aberrations de l'objectif, le bruit du capteur) dans le processus d'imagerie.
- L'approche ne peut pas gérer l'auto-occlusion inévitable de la main.

➤ **Approches par apparence:**

Dans cette approche, l'aspect visuel de la main est modélisé et comparé avec les caractéristiques d'image extraites de l'entrée vidéo. D'une manière générale, les méthodes par apparence présentent l'avantage de performance en temps réel élevées, en raison des caractéristiques simples des images 2D qui sont utilisés.

Dans [32], une méthode pratique pour déterminer la position de la main et reconnaître ensuite un certain nombre de poses dans des séquences d'images.

L'approche consiste à identifier les emplacements de main probables dans l'image en fonction des caractéristiques discriminantes de la couleur et du mouvement. Un ensemble de modèles est stocké en mémoire et un classificateur du plus proche voisin est ensuite utilisé pour estimer la pose.

La figure suivante (figure 2.10) montre dix postures de la main, enregistrées sur un fond neutre, avec des résultats de reconnaissance superposés.

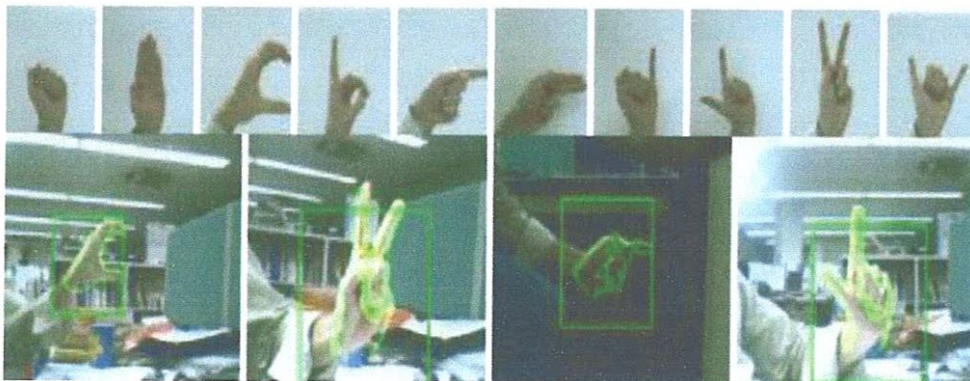


Figure 2. 9: reconnaissance de posture de la main

b. Méthode géométrique:

Les méthodes géométriques sont toutes les approches qui utilisent des paramètres extraits puis traité pour décrire la main, comme les contours, bouts des doigts, le centre de la main et plusieurs autres paramètres.

Par conséquent, les méthodes géométriques ne sont pas assez solides pour la mise en œuvre en temps réel. Pour surmonter ce problème, ils sont développé ces méthode pour la reconnaissance des gestes simples de la main qui dépendent purement sur la simple segmentation et techniques.

Dans [33], est présentée une méthode qui consiste à extraire le plus grand contour de l'image, on suppose que c'est celui de la main, ensuite, le centre de coordonnées de la main serait calculé. La taille de la main serait alors déterminée en traçant un cercle centré sur le centre de coordonnées. Ensuite, l'algorithme de Convex Hull est utilisé pour résoudre le problème de trouver le plus grand polygone dont les sommets sont supposés être les bouts de doigts, et reconnaître si un doigt est plié ou non.

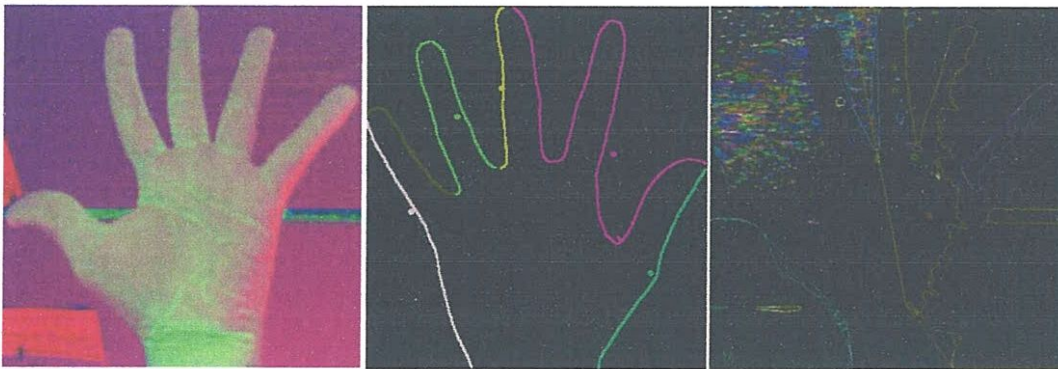


Figure 2. 10: a) Segmentation de la couleur de la peau, b) extraction de contour, c) Après l'algorithme de Convex Hull

c. Méthode de classification:

Il y a plusieurs des méthodes pour la classification:

- **Réseaux de neurone (RDN):** Le principe consiste à apprendre à classifier correctement des données à partir d'un jeu d'exemples déjà classifiés, c'est-à-dire

l'apprentissage par l'expérience. Un réseau de neurones est constitué d'un graphe pondéré orienté dont les nœuds symbolisent les neurones.

• **Logique floue (fuzzy logic):** est la représentation des connaissances imprécises, des raisonnements approchés et elle cherche à modéliser les notions vagues du langage naturel pour pallier l'inadéquation (surmonter le décalage) de la théorie des ensembles classiques dans ce domaine.

• **HMM:** et ont été largement utilisés avec succès dans la reconnaissance vocale et de reconnaissance d'écriture manuscrite. Par conséquent, ils semblent être efficace pour la reconnaissance visuelle des gestes de la main, structurés et complexes comme la reconnaissance du langage des signes.

Un HMM peut être employé pour représenter le comportement statistique d'une séquence de symboles observables en termes d'un réseau d'États. Pour chaque symbole observable, elle peut être modélisée comme l'un des États de la HMM, et ensuite le HMM soit reste dans le même état ou se déplace dans un autre état, fondé sur un ensemble de probabilités de transition associée à l'état.

Le HMM utilise seulement trois paramètres [34]:

- a. le vecteur de probabilité initiale d'État,
- b. la matrice de probabilité de transition de l'état,
- c. la matrice de probabilité de symbole observable.

L'analyse des images dynamiques donnera une reconnaissance plus précise que celle d'une seule image statique.

9. Conclusion:

Dans ce chapitre, nous avons défini les gestes humains et les styles de gestes, ensuite nous avons présenté une classification des gestes de la main (taxonomie de Quek). Les gestes sont utilisés dans l'interaction Homme-Machine, d'où la présentation de dispositifs d'interaction utilisant les gestes (les écrans tactiles et la vision par ordinateur).

Dans le chapitre suivant, nous présentons la conception détaillée de notre système d'interaction par geste en temps réel.

Handwritten notes:
C'est un système
de gestion de
la production
dans
le
secteur
industriel

Chapitre III

Conception du système

1. Introduction:

Ce chapitre concerne la présentation de la conception d'un système pour l'interaction à travers des gestes de la main. Notre système permet de numériser les inscriptions contenues sur un support papier pour les manipuler (sélectionner, imprimer, copier). Toutes ses actions sont commandées par l'utilisateur à travers des positions de doigt, filmés par une webcam.

2. Spécification:

Le système développé doit pouvoir capturer les inscriptions « montrées » par l'utilisateur sur une feuille de papier, puis les transformer en objets virtuels. Ensuite, ces objets virtuels seront manipulés par l'utilisateur (imprimer ou copier sur disque dur), à travers des gestes dédiés.

Notre système a pour ces objectifs:

1. Réalisation d'un prototype prenant en charge la capture d'inscriptions sur une feuille de papier au format A4 avec une webcam.
2. Reconnaître et suivre le déplacement d'un doigt sur la feuille de papier contenant du texte.
3. Exécuter des actions liées à la position du doigt sur la feuille de papier.

3. Conception globale:

La conception globale de notre système consiste sur trois modules: module de détection, module de reconnaissance et module d'interaction (voir figure 3.1).

*plus votre système et
qui n'existant pas avant.*

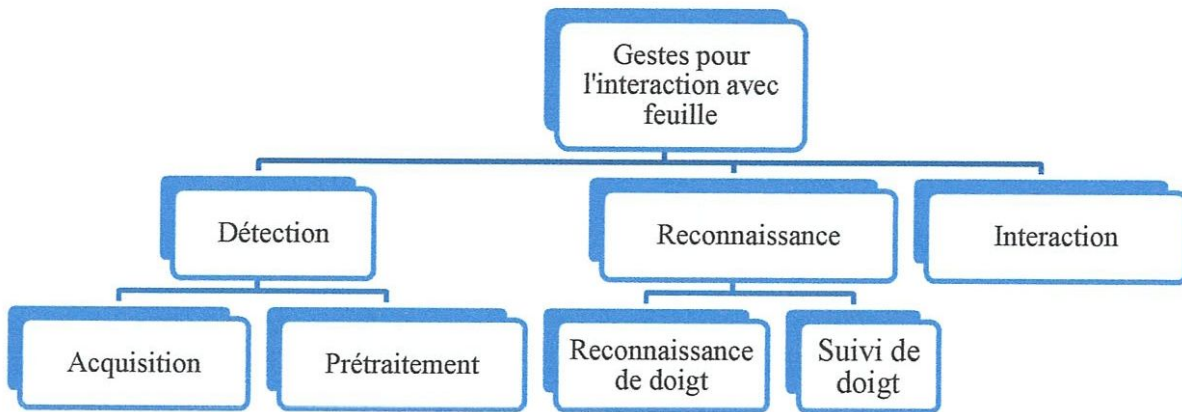


Figure 3. 1: Système d'interaction par gestes. *[cref]*

Module de détection: ce module est utilisé pour la détection de la main.

Module de reconnaissance et suivi du doigt: pour reconnaître le bout du doigt. Il contient 2 sous-modules.

Module d'interaction: il contient les différentes opérations d'interaction du système avec l'utilisateur.

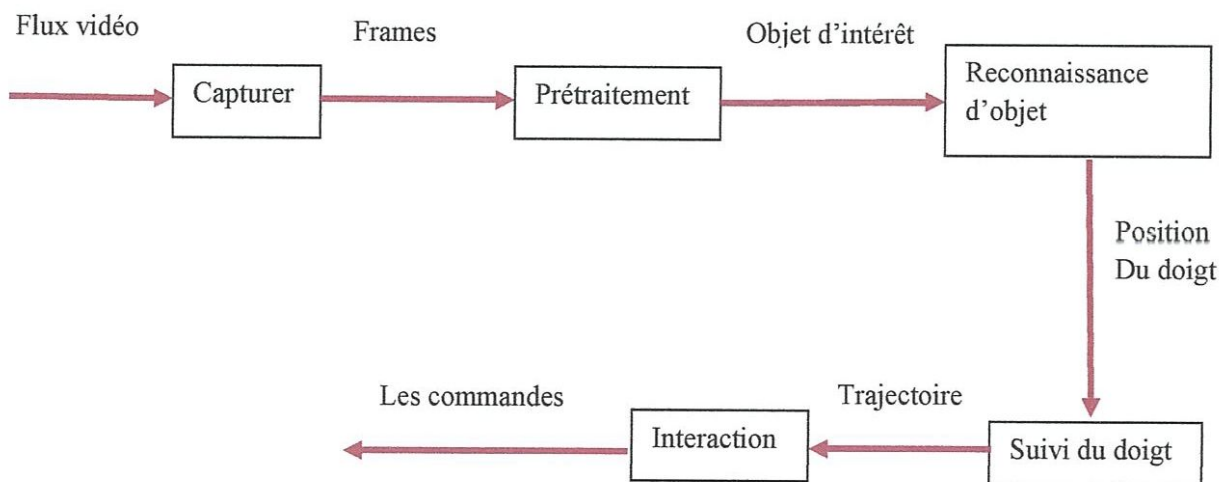


Figure 3. 2: Illustration du processus de notre système. *[cref]*

c-par votre proposition

4. Conception détaillé:

4.1. Module de détection:

Il est composé de deux sous-modules:

4.1.1. Acquisition:

On a utilisé dans cette partie un flux vidéo acquis par une caméra.

L'acquisition se fait en temps réel, chaque frame capturé sera prétraité puis analysé pour détecter le doigt de l'utilisateur, et déterminer sa position, dans les modules suivants.

4.1.2. Prétraitement:

Ce module a pour objectif d'analyser le frame acquis pour extraire l'objet d'intérêt. Le processus suivi par ce sous-module est le suivant (figure3. 3):

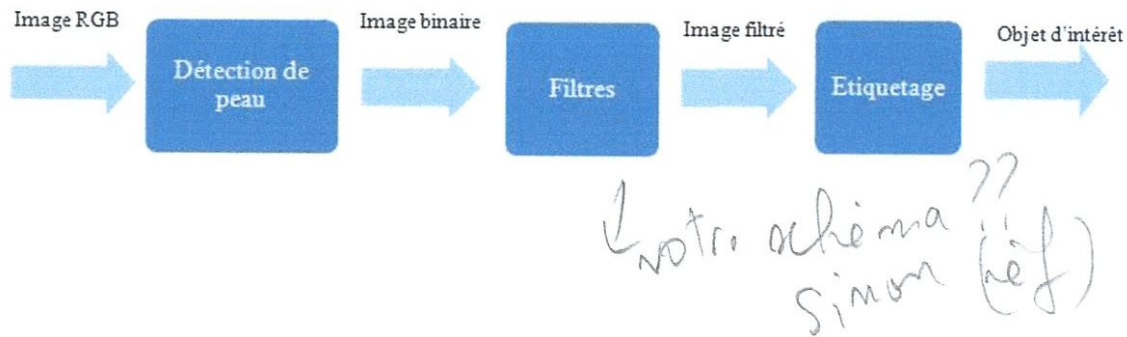


Figure 3. 3: Illustration du processus de prétraitement

✓ Détection de peau:

Dans le cadre de notre projet, on a travaillé par la méthode de détection de la couleur de peau. Cette méthode est caractérisée par un temps de calcul faible, mais son inconvénient est sa sensibilité à l'éclairage.

- **La détection de la couleur de peau:**

Nous avons travaillé dans l'espace RGB (cf. 7-1-2) qui ne nécessite aucun modèle de peau et aucune transformation des couleurs. Il suffit de vérifier une série de contraintes pour décider si un triplé de couleur (R, G, B) représente une couleur de peau ou non.

Les contraintes sont divisées en deux, selon l'état de l'éclairage, fort (**jour**) ou faible (**nuite**).

Eclairage fort:

$$(R > 45) \text{ ET } (G > 30) \text{ ET } (B > 10) \text{ ET } (MAX(R, G, B) - MIN(R, G, B) > 15) \text{ ET}$$

$$(ABS(R-G) > 15) \text{ ET } (R > G) \text{ ET } (R > B)$$

Eclairage faible:

$$((R > B) \text{ ET } (G > B)) \text{ OU } ((R > 220) \text{ ET } (G > 210) \text{ ET } (B > 170) \text{ ET } (ABS(R-G) \leq 15))$$

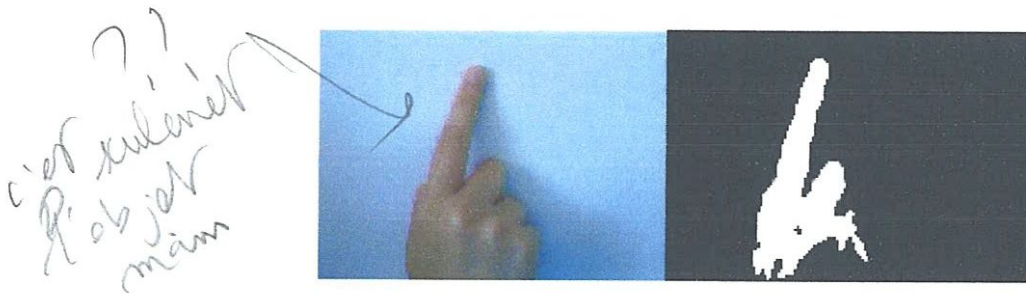


Figure 3. 4: la détection de la peau

- **Filtrage:**

L'objectif principal du filtrage est de réduire les variations d'intensité au sein de chaque région de l'image tout en respectant l'intégrité des scènes: les transitions entre régions homogènes, les éléments significatifs de l'image doivent être préservés au mieux.

Différentes méthodes de filtrage ont été développées suivant le type et l'intensité du bruit, ou les applications auxquelles on destine l'image. Les premières et les plus simples de ces méthodes sont basées sur le filtrage linéaire stationnaire (invariant par translations), mais

les limitations de ces techniques (en particulier leur mauvaise conservation des transitions) a conduit au développement des filtres "non-linéaire".

A partir de l'image obtenue dans la première étape, on a appliqué le filtre non-linéaire médian, car il n'influence pas sur le temps de réponse de système.

L'idée de ce filtre est de remplacer chaque entrée par la valeur médiane de son voisinage, par exemple, si on considère ces neuf pixels, dont une valeur aberrante (ici 111):

5	6	7
6	111	8
7	8	9

Le filtre médian va considérer les valeurs du voisinage par valeurs croissantes:

5	6	6	7	7	8	8	9	111
---	---	---	---	---	---	---	---	-----

Et prendre la valeur médiane, ici la valeur 7. La sortie du filtre donnera donc:

5	6	7
6	7	8
7	8	9

Ce qui a permis de remplacer la valeur aberrante par une valeur "de consensus" entre les valeurs voisines.

L'image suivante montre l'image originale (à gauche) et l'image avec le filtre médian (à droite) qui réduit le bruit et améliore les résultats du traitement comme la détection de bords. Ainsi, il permet, sous certaines conditions, de réduire le bruit tout en conservant les contours de l'image. Le résultat est illustré dans la (figure 3.5): l'image originale est à gauche, l'image en RGB au milieu, et l'image à droite sur laquelle on a appliqué le filtre médian:

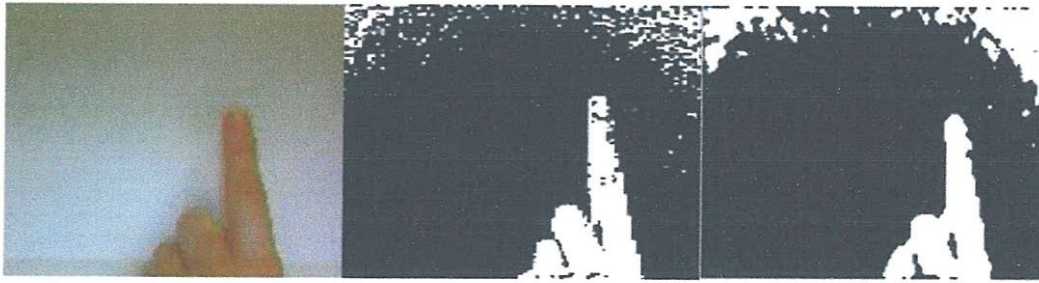


Figure 3. 5: le filtre médian

- **étiquetage:**

Cette étape est nécessaire pour la distinction entre les différentes composantes. Le principe de cette phase est d'affecter à chaque zone une étiquette correspondant à une couleur.

Tous les pixels d'une même région doivent avoir le même numéro (étiquette).

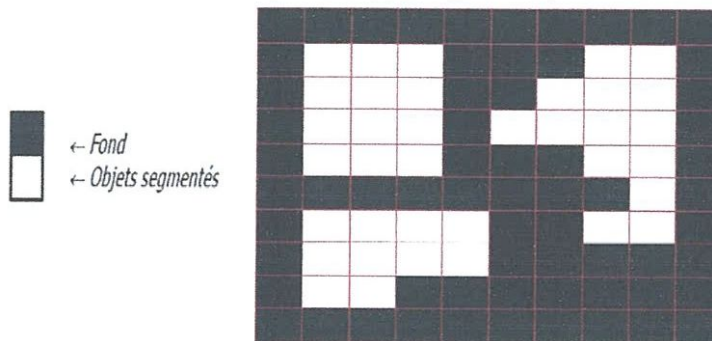


Figure 3. 6: Représentation des régions dans une image binaire

Nous faisons deux parcours de l'image:

Premier parcours de l'image:

Pour chaque pixel d'une région, nous lui affectons:

- Soit la plus petite étiquette parmi ses voisins haut et gauche.
- Soit une nouvelle étiquette.

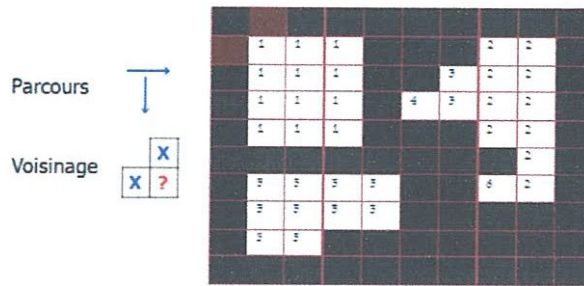


Figure 3. 7: Premier balayage

Deuxième parcours de l'image:

Pour chaque pixel d'une région, on lui affecte la plus petite étiquette parmi la sienne et celles de ses voisins bas et droite.

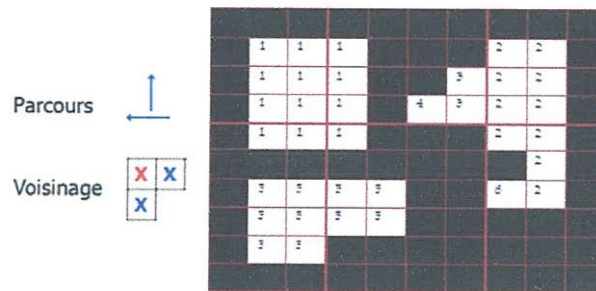


Figure 3. 8: Deuxième balayage

Après avoir terminé ces deux parcours d'étiquetage, nous distinguons plusieurs zones et c'est la plus grande parmi elles qui nous intéresse. La figure 3.9 illustre l'image filtrée (à gauche), l'image étiquetée (au milieu) et l'image contenant seulement la zone de plus grande surface (à droite).



Figure 3. 9: Image avant et après l'étiquetage

Donc dans notre image traitée, la main doit occuper la zone de plus grande surface, en évitant son chevauchement avec d'autres régions.

Pour bien comprendre notre traitement d'étiquetage nous avons défini le graphe suivant (figure 3.10):

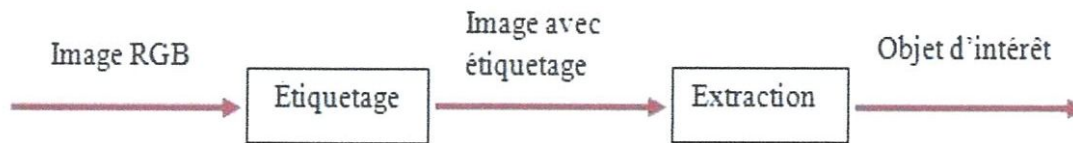


Figure 3. 10: le processus d'étiquetage de l'objet

(réf)

4.2. Module de reconnaissance:

À la fin de l'étape précédente, nous avons obtenu l'objet d'intérêt. Dans cette étape de reconnaissance, il faut reconnaître le bout du doigt.

→ localiser

Dans notre application, on a travaillé avec le geste déictique d'après la taxonomie de Quek (cf. 5-2).

Les gestes déictiques sont des gestes de pointage, pour indiquer quelque chose par un geste de direction. La direction peut être indiquée par l'index ou le pouce.

La cible du geste déictique peut être présente *in situ* (geste « déictique concret » ou « définition ostensive ») ou représentée de façon virtuelle (geste « déictique abstrait »). Les gestes déictiques s'appliquent aussi à la temporalité dans le discours (comme dans *hier* ou *avant*, exprimé souvent par un geste derrière). Le pointage pour énumération est inclus dans la catégorie du geste déictique.

Sont en revanche exclus de la catégorie les gestes ci-après: la main ouverte indiquant une direction (comme, par exemple, le geste qui désigne poliment l'interlocuteur en disant *vous*), puisqu'un tel geste est difficile à distinguer d'un geste ayant la même forme mais à caractère discursif ; un geste dit « pointage discursif » (l'index orienté vers le haut, par exemple) dont la destination du pointage n'est pas précisée.

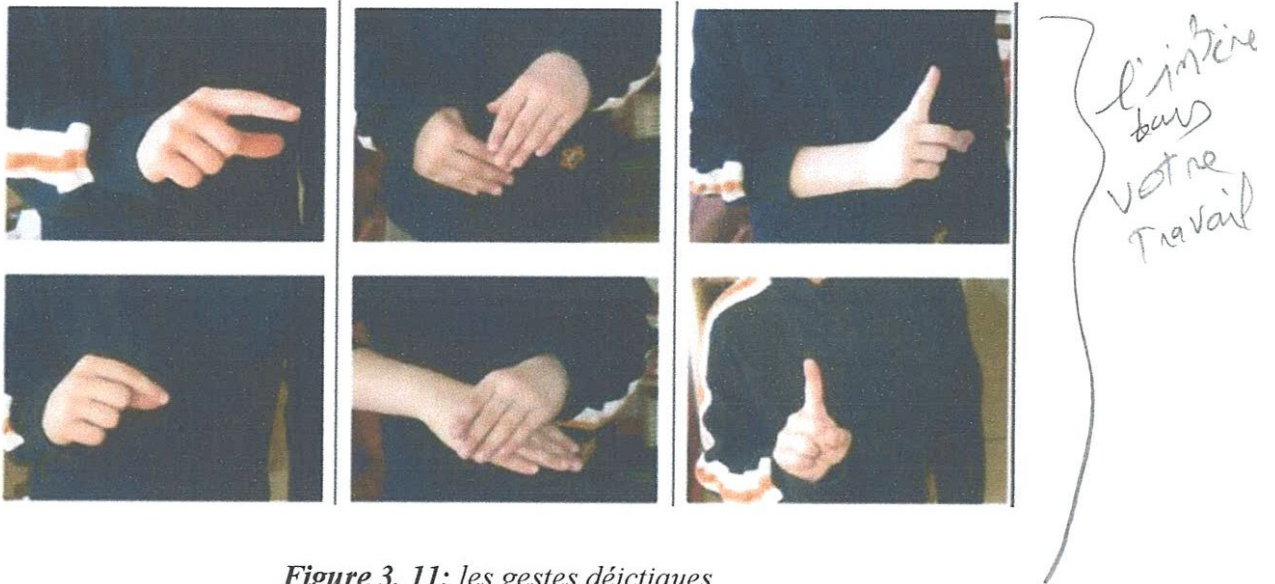


Figure 3. 11: les gestes déictiques

✓ **Reconnaitre le bout du doigt:**

Les bouts des doigts sont des points caractéristiques très utilisés pour le suivi de la main et la reconnaissance de gestes. Ils ont l'avantage d'être facilement détectables, avec une complexité de calcul relativement limitée.

Après le filtrage et l'étiquetage on aura uniquement la main en blanc, tout le reste en noir.

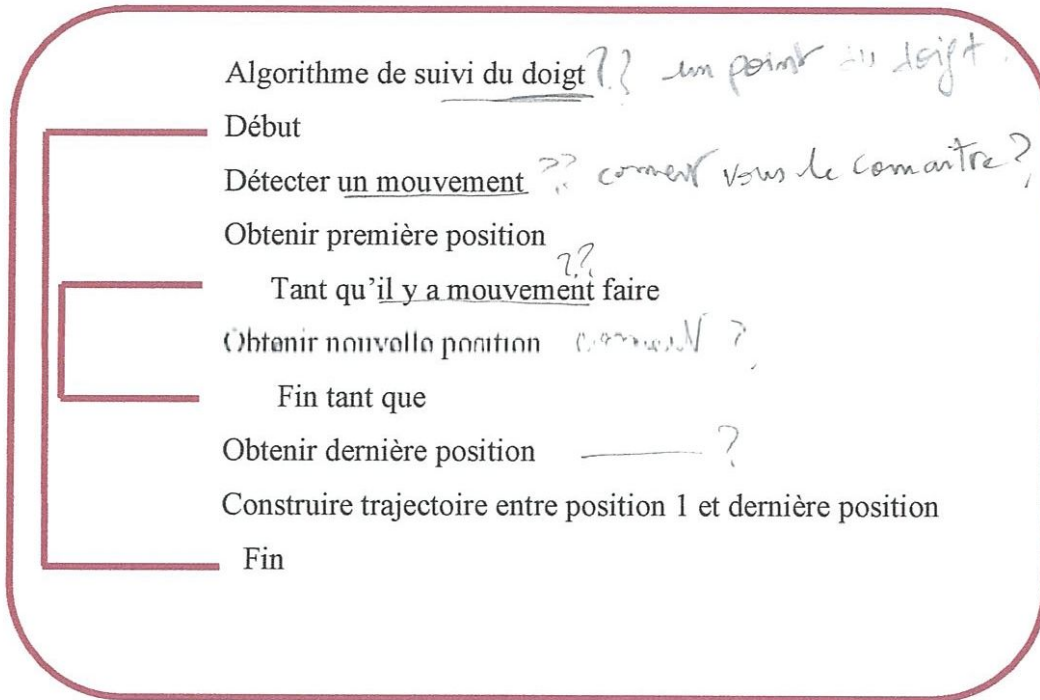
✓ Pour localiser le bout du doigt, on fait un simple parcours de l'image, du haut vers le bas ainsi le premier pixel blanc est bien notre bout de doigt. Parce que notre geste comporte par principe l'index comme seul doigt tendu (figure 3.12).



Figure 3. 12: la forme du doigt pour lancer l'action

✓ **Suivi de doigt:**

L'algorithme ci-dessous montre les étapes de calcul de la trajectoire du bout du doigt:



*Il faut expliquer
chaque instanc
de la procédure
comment vous l'avez
implémenté!*

A chaque frame, on calcule la position du bout du doigt:

On fait un parcours ligne par ligne, et colonne par colonne, de l'image, puis lorsque le doigt est trouvé, on prend les coordonnées de sa position dans l'image. A chaque image dans le flux vidéo, on calcule la nouvelle position et on la sauvegarde dans un tableau.

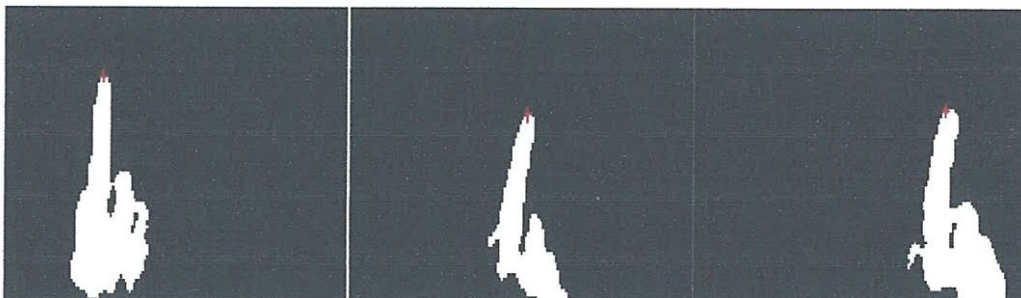


Figure 3. 13: le suivi du doigt

4.3. Module d'interaction:

Notre système consiste en une interface gestuelle qui prend en charge de quelques opérations nécessaires pour la manipulation de texte écrit ou imprimé sur une feuille de papier. Trois opérations sont prises en charge: la sélection de texte, la sauvegarde du texte sélectionné et l'impression du texte. Ces opérations sont commandées avec le doigt de l'utilisateur.

Le schéma suivant représente les étapes d'exécution de ces opérations (figure 3.14):

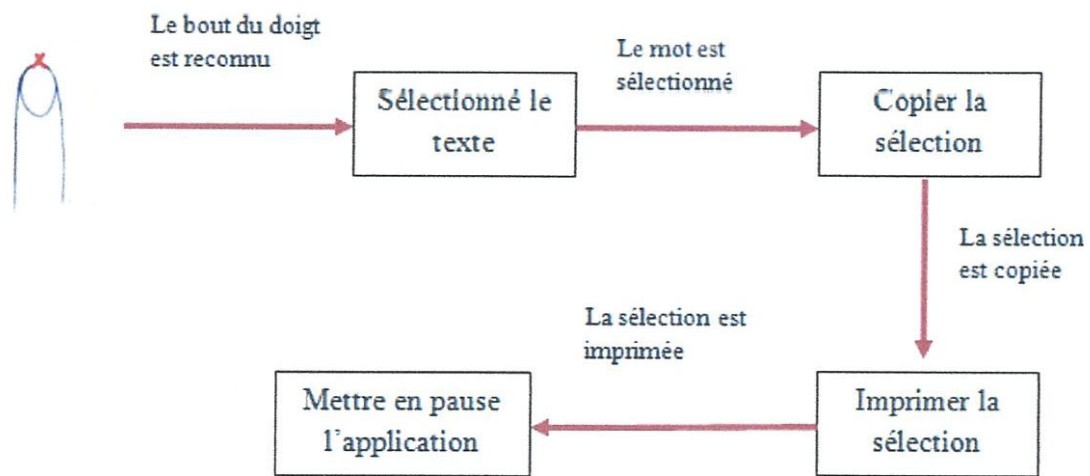
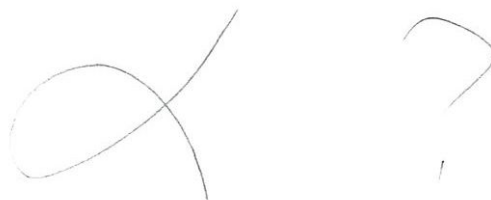


Figure 3. 14: les commandes d'interaction.

Pour commander une des opérations possibles (sélection, copie et impression), la feuille de papier est divisée en quatre zones, présentées ci-dessous:



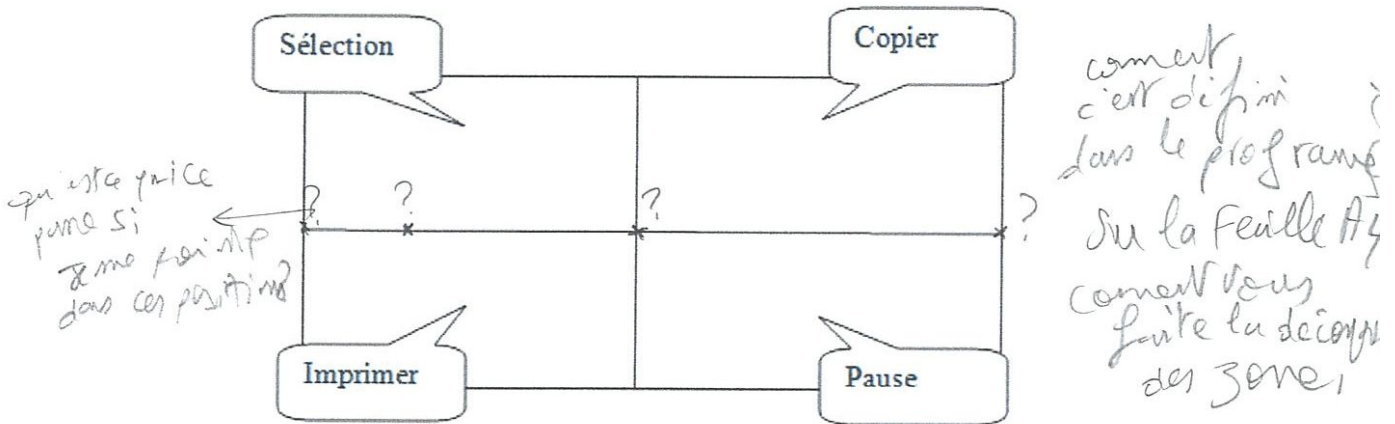


Figure 3. 15: Zones de commande.

Selon la position du doigt, dans une de ces zones, l'opération correspondant à la zone est exécutée. Pour sélectionner un texte sur la feuille de papier, il suffit de le souligner avec le bout du doigt. Il sera alors possible de le copier ou de l'imprimer.

La figure (3.16) montre un exemple pour la sélection d'une phrase, l'image à gauche illustre la première position de doigt, l'image au milieu illustre la dernière position de doigt et l'image qui sélectionné à droit.

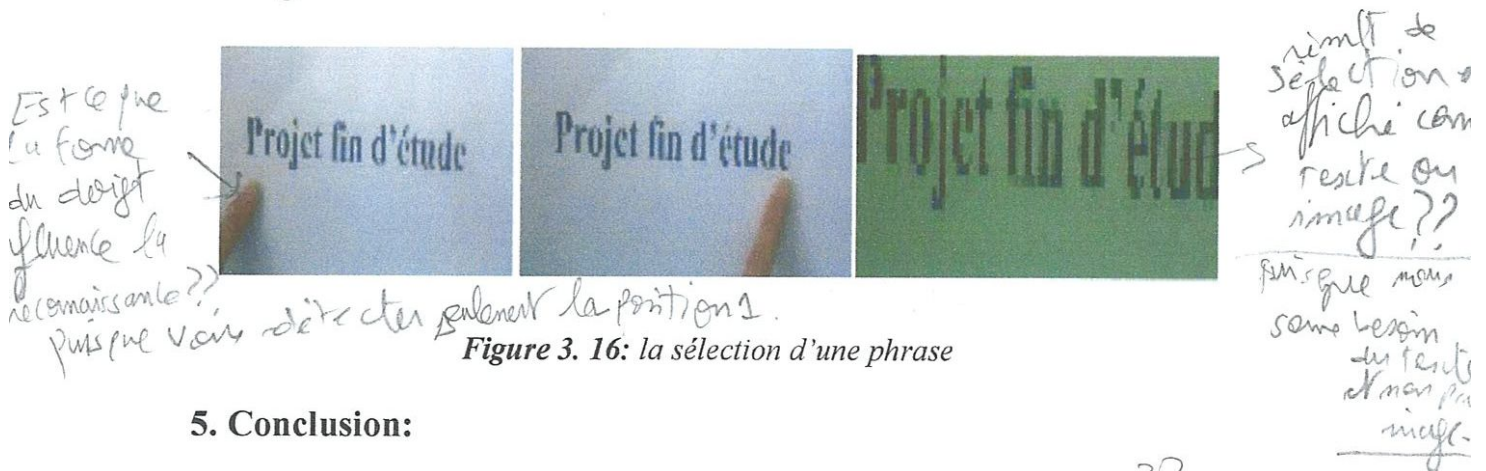


Figure 3. 16: la sélection d'une phrase

5. Conclusion:

Dans ce chapitre, nous avons présenté la conception un système de suivi de doigt pour une interaction par geste.

Différentes étapes sont suivies sur notre système pour exécuter une opération. Premièrement, l'étape de détection basée sur la couleur de la peau, le filtrage et l'étiquetage, qui permettent d'obtenir l'objet d'intérêt. Deuxièmement, l'étape de reconnaissance basée sur le bout du doigt, qui permet de suivre la trajectoire du doigt sur la feuille. Et enfin,

l'étape de l'interaction proprement dite, par l'exécution d'une des opérations (sélectionner, copier, imprimer) commandée par l'utilisateur.

Chapitre IV

Implémentation et résultats expérimentaux

1. Introduction:

Ce chapitre est consacré à présenter les différents outils de développement de notre système, et les différents composants matériels et logiciels nécessaires à son bon fonctionnement, ainsi que l'interface de notre système.

Nous allons tout d'abord introduire l'environnement de développement matériel (caractéristiques matérielles de l'ordinateur) puis l'environnement de développement logiciel (langage de programmation) et enfin, nous avons discuté les performances de notre système.

2. Environnement de développement matériel et logiciel:

Pour la réalisation de notre application, nous avons utilisé les environnements suivants:

2.1. Environnement de développement matériel:

Nous avons utilisé un ordinateur qui possède les caractéristiques décrites dans le tableau ci-dessous (*Tableau 4.1*):

L'environnement matériel	Caractéristique
Processeur:	Intel®Core™i3-380M CPU2.53GHz.
Mémoire installée (RAM):	4GB
Camera:	Résolution 160*120 pixels.

Tableau 4. 1: les caractéristiques de l'environnement matériel

2.2. Environnement de développement logiciel:

Le développement des systèmes de reconnaissance de geste en temps réel, peuvent être développé avec plusieurs langages de programmation: JAVA, C++, Matlab. Nous avons choisi Matlab pour le développement de notre système.

Matlab est un langage de programmation de quatrième génération émulé par un environnement de développement du même nom. Il est utilisé pour le calcul numérique.

? ? ?
 from Ai mi
 pas Java
 C++ pour
 le temps
 réel

Développé par la société The Math Works, MATLAB permet de manipuler des matrices, d'afficher des courbes et des données, de mettre en œuvre des algorithmes, et de créer des interfaces utilisateurs.

Matlab offre les avantages suivants [35]:

- **Sa richesse:** l'environnement Matlab comporte une palette d'outils s'adressant à de très nombreux domaines. Par exemple, l'intégration dans un seul environnement homogène de l'ensemble des outils permettant la conception fiable d'un système de contrôle, jusqu'à son prototypage en temps réel, constitue un atout déterminant.

- **La finition:** l'offre Matlab constitue un ensemble cohérent, soigné et fiable. Ce sont des qualités plus difficiles à obtenir lorsque les développements sont issus d'une communauté décentralisée et moins organisée, comme souvent dans le domaine du logiciel libre.

- **C'est un standard de fait,** enseigné dans la plupart des écoles et utilisé dans la plupart des grandes entreprises, ce qui facilite les échanges entre utilisateurs. Les développements sont assurés par une équipe de plusieurs centaines d'informaticiens, qui assurent une évolution permanente du logiciel et la sortie régulière de nouvelles versions comportant chaque fois davantage de fonctionnalités.

- **La documentation** livrée avec Matlab est complète et professionnelle. La qualité de la documentation est un des points faibles couramment reprochés à Scilab.

Matlab fonctionne sur Windows, Mac et une variété de systèmes.

Les caractéristiques ci-dessus, nous ont encouragées à choisir Matlab comme outils de développement de notre système.

La figure 4.1 montre l'interface principale de l'environnement de développement de Matlab:

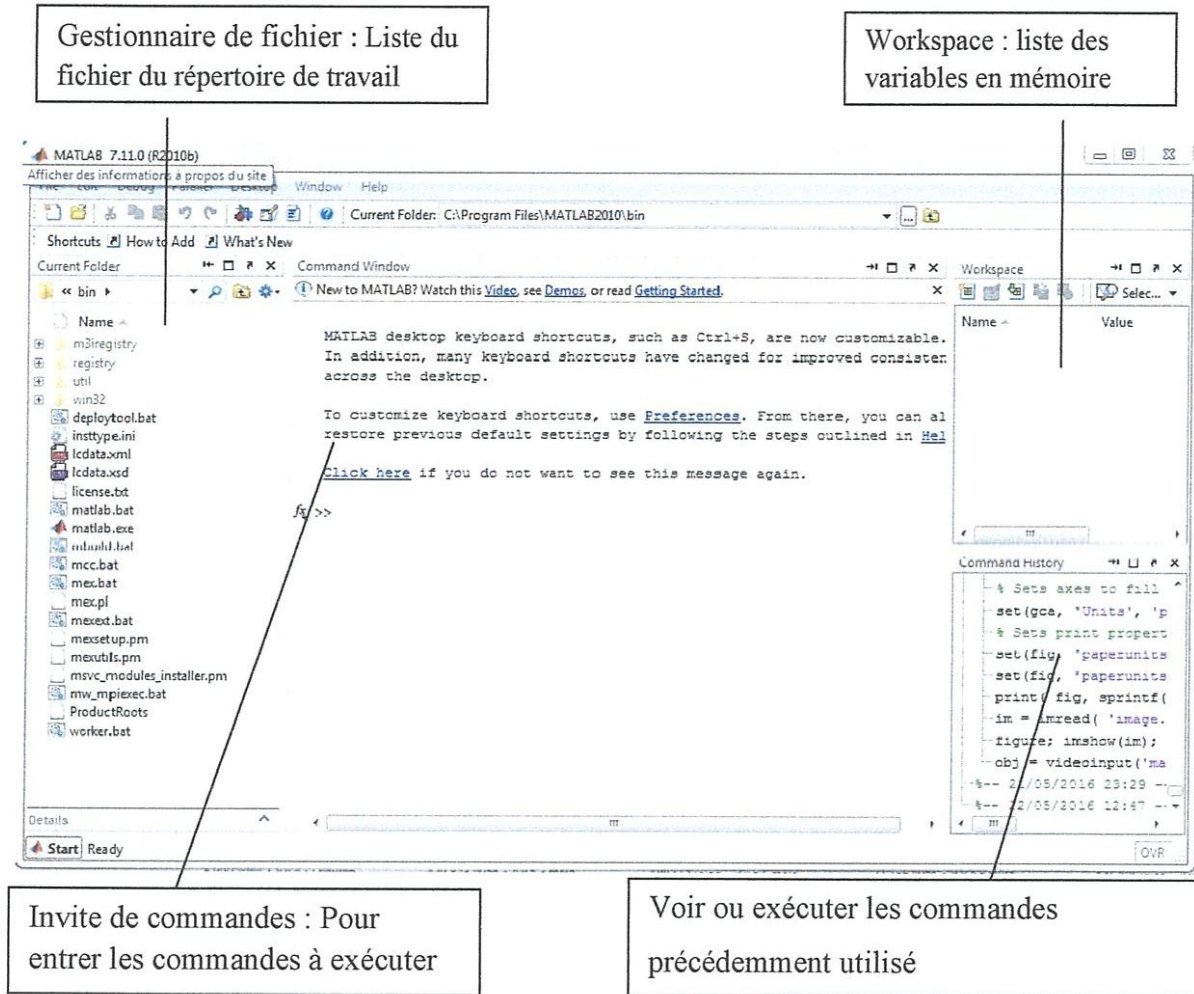


Figure 4. 1 l'interface principale de l'environnement de Matlab.

3. Description de l'application:

3.1. Les composants de l'interface:

Dans notre application, nous avons réalisé une interface facile à manipuler pour l'utilisateur. La figure suivante illustre notre interface (Figure 4. 2):

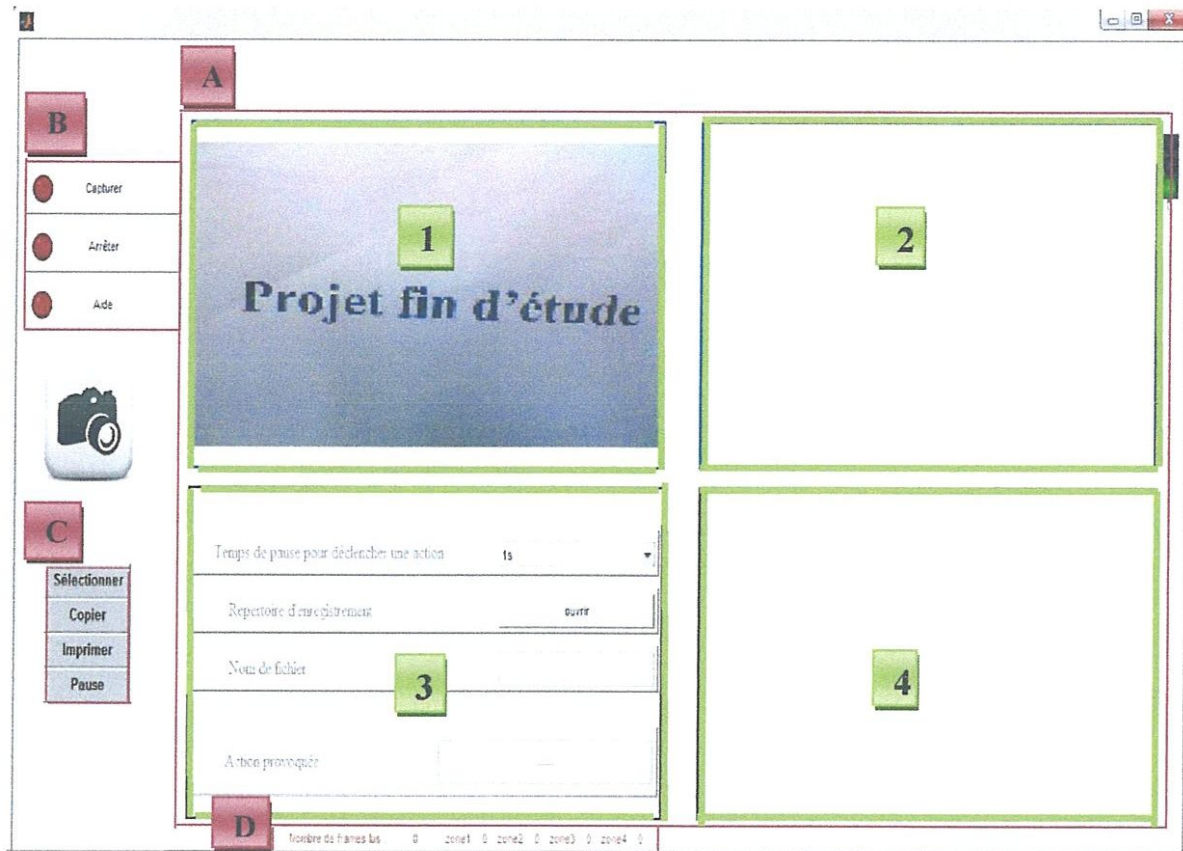


Figure 4. 2: l'interface de notre application.

Cette interface contient: quatre groupes: A, B, C, D

➤ Le groupe (A) contient quatre axes, le premier (1) pour l'affichage de la vidéo capturée en temps réels. Le deuxième (2) pour l'affichage du traitement appliqué aux frames de la vidéo. Le troisième (3) contenant les paramètres configurables de notre système. Et dans le quatrième (4), l'affichage de la sélection de l'utilisateur.

➤ Le groupe (B) contient trois boutons:

- Le premier bouton: « Capturer», pour lancer la capture de la vidéo
- Le deuxième bouton: « arrêter», pour arrêter le traitement.
- Le troisième bouton: «Aide» pour afficher une aide sur notre système.

➤ Le groupe (C) présente l'opération commandée par l'utilisateur, lorsqu'elle est détectée. Dans ce cas, un des textes statiques change de couleur.

➤ Le groupe (D) contient des informations techniques sur l'exécution de notre système, telles que le nombre total de frame acquis et le nombre de frames dans lesquels le doigt est détecté dans chaque zone de commande.

Lorsque l'utilisateur clique sur le bouton « Capturer », le système commence le traitement et affiche un flux vidéo en temps réel dans l'axe (1) et la vidéo traitée dans l'axe (2) (Figure 4. 3)

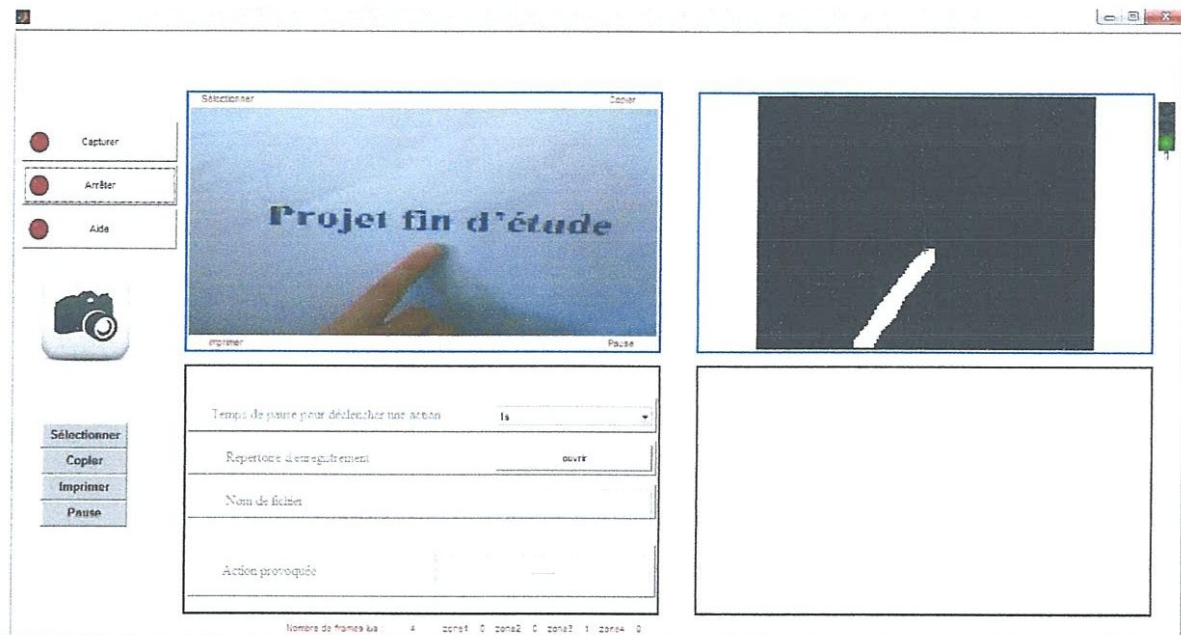


Figure 4. 3: fenêtre principale lors du traitement d'une commande gestuelle.

3.2. La description des différentes commandes:

3.2.1. La commande de sélection:

Avant de sélectionner une zone sur la feuille de papier filmé, l'utilisateur doit d'abord commander une sélection. Pour cela, une doit positionner son doigt sur le quart haut droit de la feuille de papier, pendant une seconde. Lorsque le doigt est détecté, la commande est alors lancée (le texte statique **Sélectionner** devient rouge, voir figure 4.4).

Ensuite, pour sélectionner la zone, l'utilisateur doit montrer du doigt le début de la zone à sélectionner, puis le déplacer jusqu'à la fin de la zone. Lorsqu'il met son doigt sur le début et la fin de la zone à sélectionner, l'utilisateur doit laisser son doigt en pause pendant un bref moment avant de le déplacer. La zone sélectionnée est alors affichée dans l'axe (4). La figure (4.4) suivante illustre le résultat:

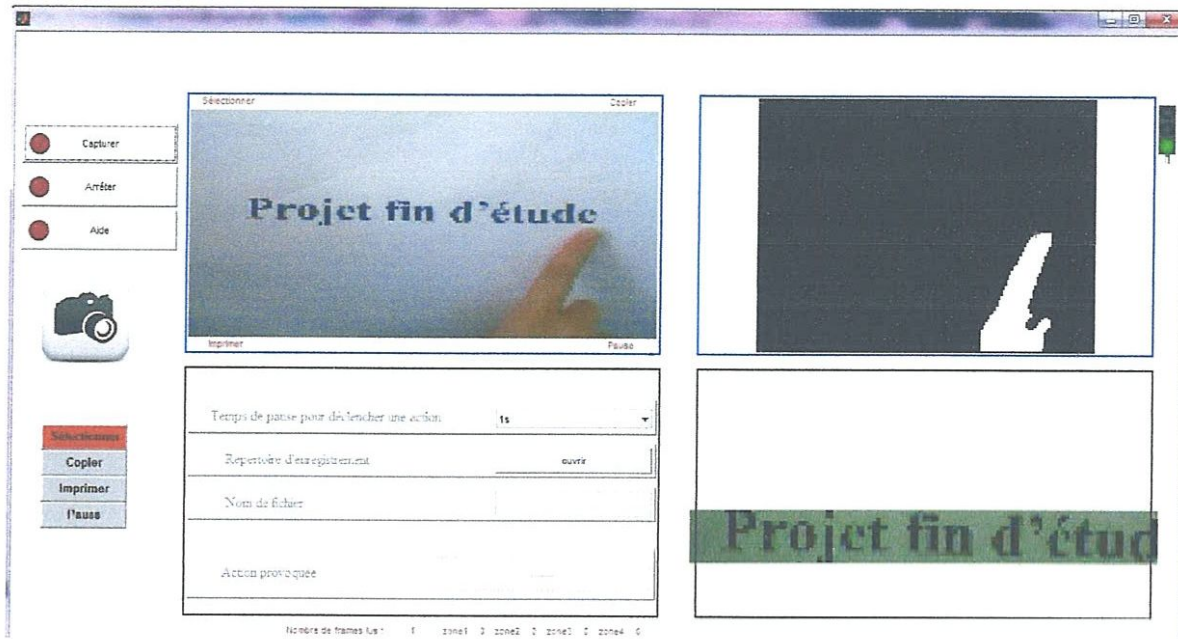


Figure 4. 4: résultat de la commande de sélection.

La zone sélectionnée peut être alors copiée ou imprimée.

3.2.2. La commande de copie:

Dans notre application, copier une zone sélectionnée permet de l'enregistrer sur disque dur. Pour exécuter cette commande, l'utilisateur doit positionner son doigt dans le quart haut gauche de la feuille de papier. La figure 4.5 illustre le résultat:

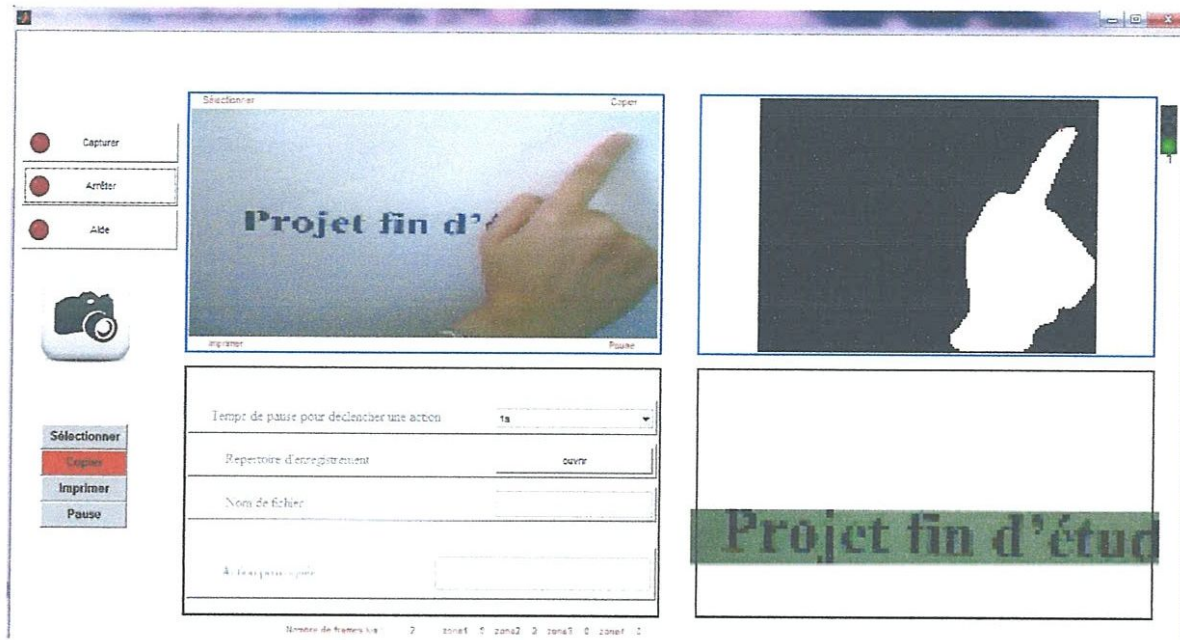


Figure 4. 5: résultat de la commande Copier.

3.2.3. La commande d'impression:

Même principe que précédemment, la commande imprimer est exécutée lorsque le doigt est positionné dans le quart bas droit de la feuille de papier. La zone sélectionnée est alors envoyée vers l'imprimante par défaut de l'ordinateur.

La figure 4.6 illustre le résultat de l'impression:

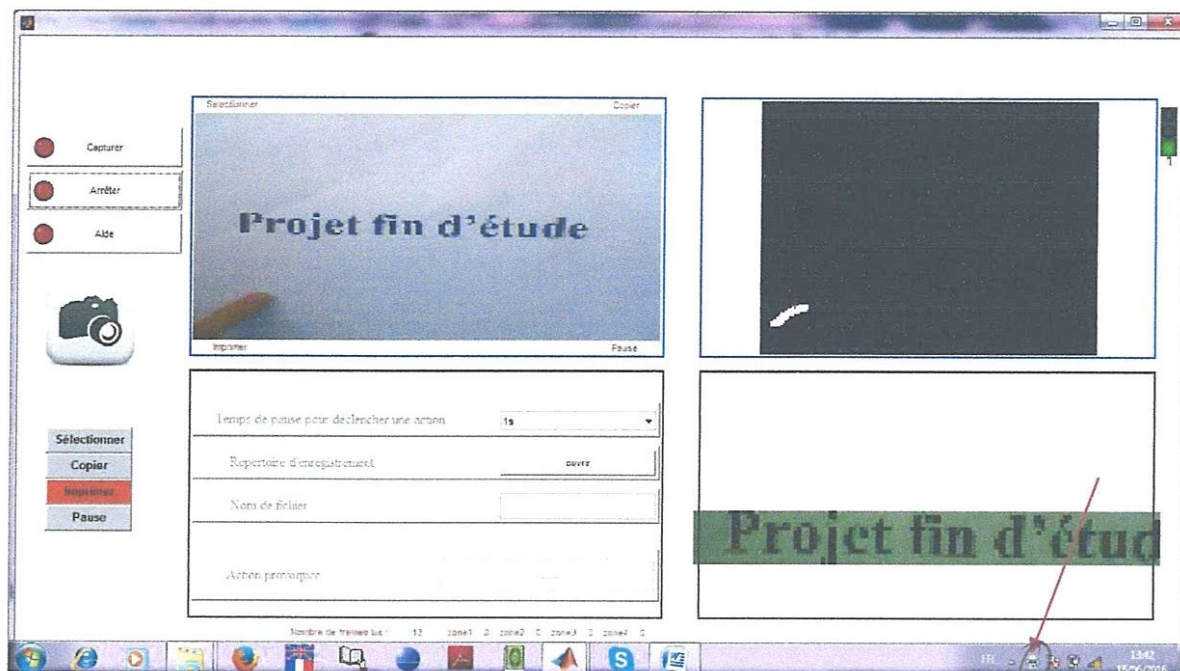


Figure 4. 6: résultat de la commande Imprimer.

3.2.4. La commande de pause:

Cette commande est utilisée lorsque l'utilisateur a besoin d'arrêter l'acquisition de la vidéo et le traitement des gestes.

L'exécution de cette commande suit le même principe que les commandes précédentes. L'utilisateur doit positionner son doigt dans le quart bas gauche de la feuille de papier. La figure 4.7 illustre le résultat d'une sélection suivie d'une pause:

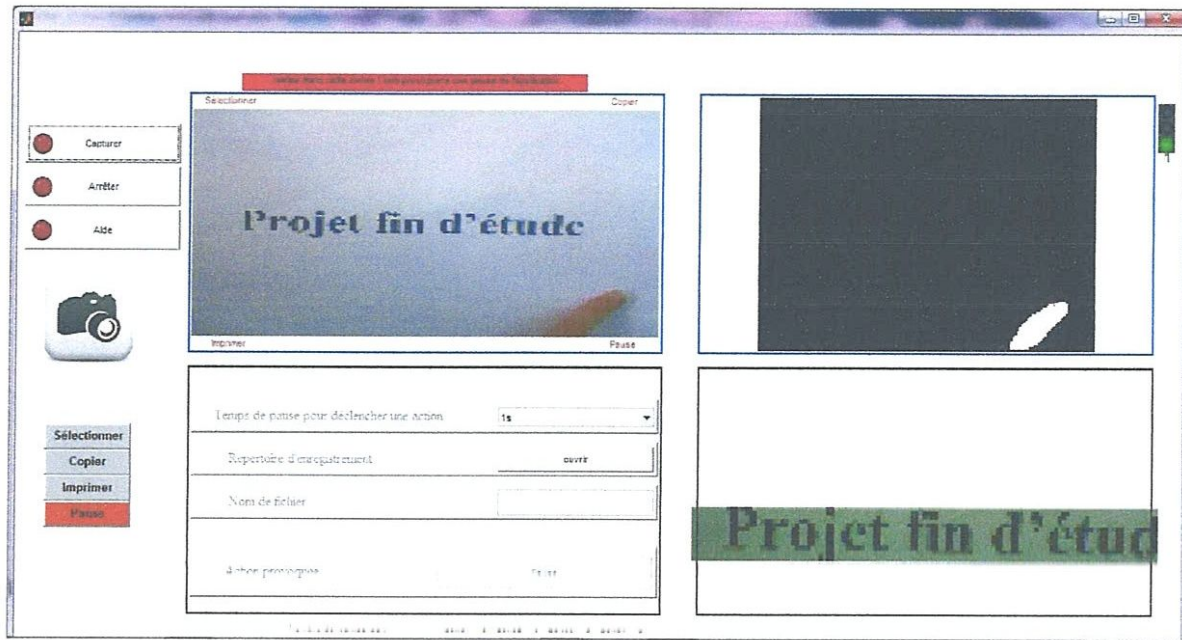


Figure 4. 7: le résultat de la commande Pause.

4. Test et Résultat:

Avec notre système, nous avons procédé à une série de tests portant sur des feuilles de papier contenant des textes, avec des polices, des tailles et des couleurs différentes. On a également testé avec des images, des tableaux et des graphes.

Pour ses tests, on a utilisé deux résolutions de webcam différentes, pour comparer les résultats.

Nous avons choisi comme police de texte à tester: Time New Roman, Arial, Cambria, et Magnéto. Ce dernier a été choisi à cause de sa ressemblance avec l'écrit manuscrit.

Nous avons également testé des textes de tailles différentes, de huit jusqu'à soixante-douze, et de trois couleurs (rouge, vert, bleu).??

Pour les tableaux, nous nous sommes intéressés à la taille de leurs bordures, et pour les graphes à la largeur des traits. La figure suivante montre le résultat de teste sur une image: ??

en réalité, nous nous dispenserons pas de ces simplifications??

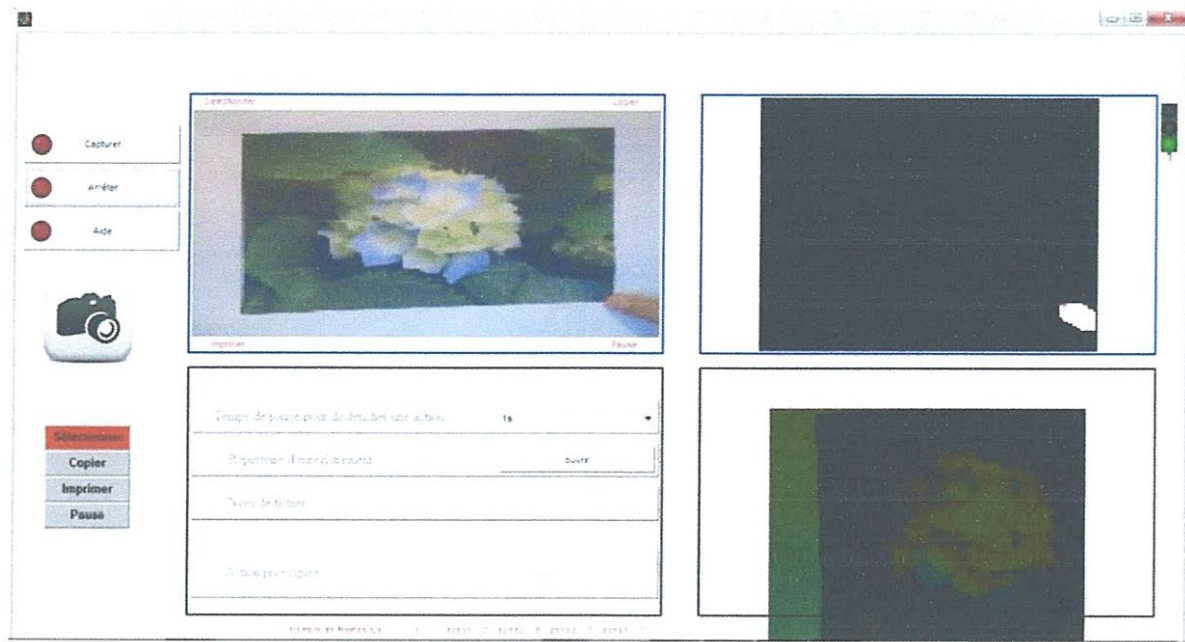


Figure 4. 8: résultat de teste sur une image

Et la figure 4.9 illustre le résultat de teste sur un graphe:

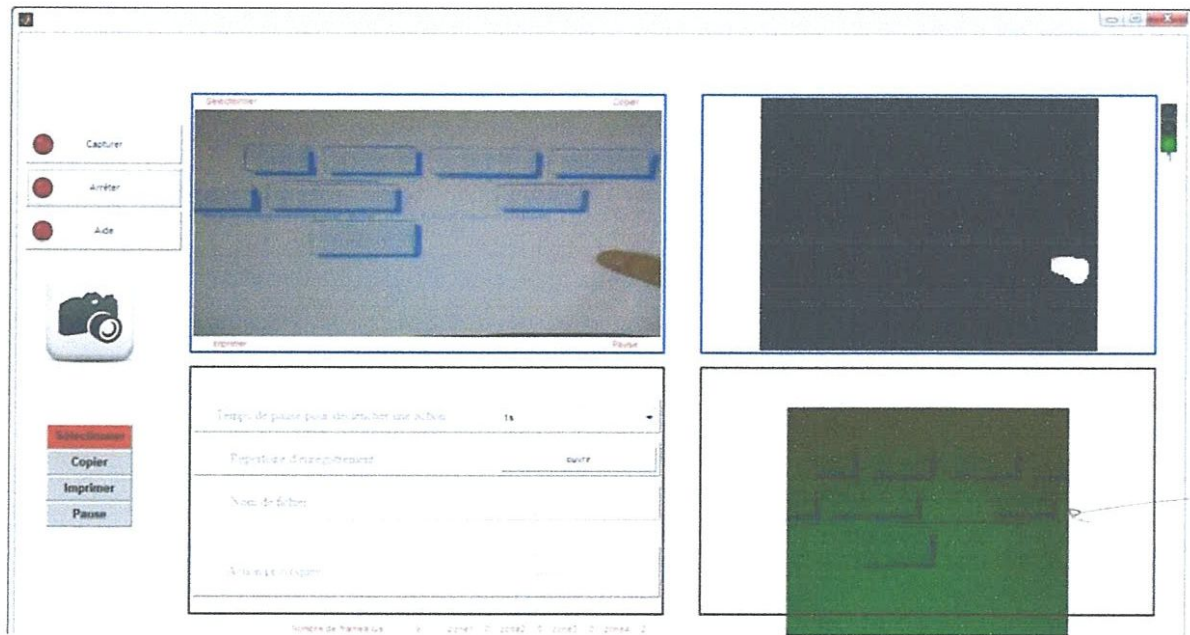


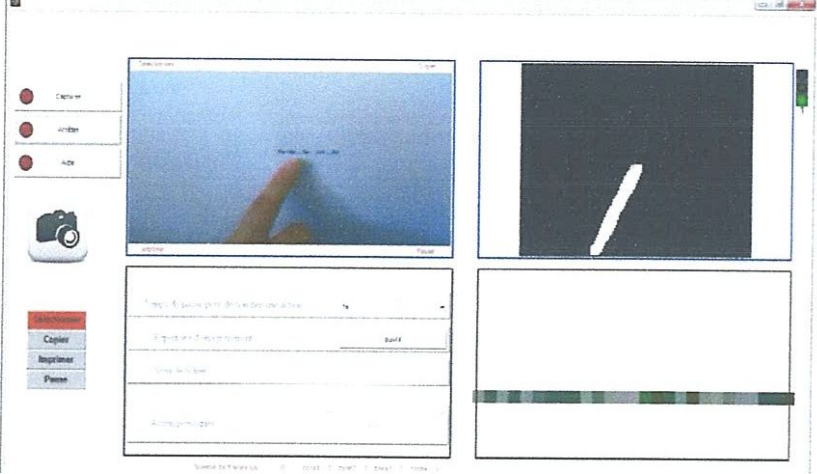
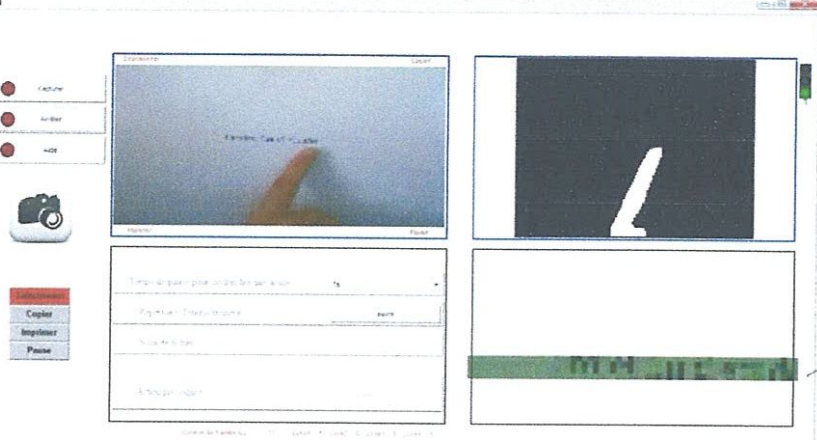
Figure 4. 9: le résultat de teste sur un graphe

Les tests ont été réalisés, dans un environnement de luminosité stable, sur une feuille de papier A4 contenant un texte-témoin: « Projet fin d'étude », dont nous avons varié la police, la taille des caractères et la couleur. Ensuite, nous avons effectué la commande Sélection.

à coloriser?

L'objectif des tests était de vérifier la lisibilité du texte sélectionné, pour une utilisation ultérieure, notamment pour un **OCR**.
Reconnaissance optique de caractères

Il s'agissait donc de voir si un texte sélectionné était illisible, peu clair ou lisible. Le tableau ci-dessous montre les résultats possibles:

L'état	Résultat
Illisible	
Peu clair	

*peu clair ??
non lisible*

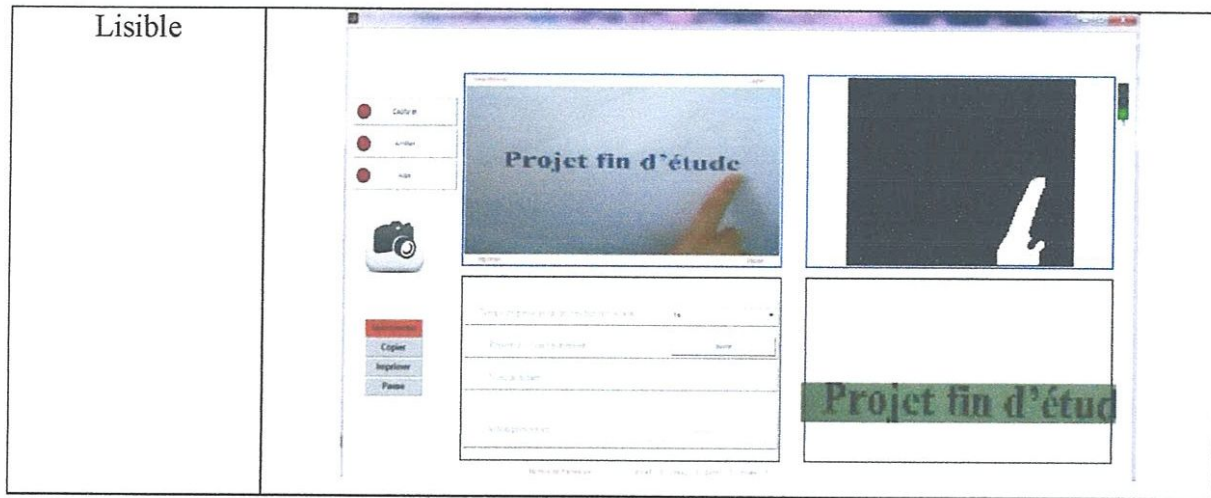


Tableau 4. 2: Résultats possibles de la sélection.

Le tableau ci-dessous montre les différents résultats obtenus avec l'ordinateur utilisé pour le développement (cf. 2.1):

Police \ Taille	Police	Time New Roman	Arial	Cambria	Magneto
8-18	8-18	Illisible	Illisible	Illisible	Illisible
18-20	18-20	Illisible	Illisible	Illisible	Illisible
20-24	20-24	Peu claire	Peu claire	Peu claire	Peu claire
24-26	24-26	Lisible	Lisible	Lisible	Lisible
26-72	26-72	Lisible	Lisible	Lisible	Lisible
Avec des couleurs		Lisible			
Image		Lisible			
Tableau		Lisible			
Graphe		Lisible			

pour que ce graphique

avec 24 c'est quoi? utilise des données vec bones valeurs ou faibles

pour quelle police??

m résultat pour toutes les polices, le mieux c'est de choisir une police avec des résultats différents.

Tableau 4. 3: les différents résultats du premier test.

Pour le deuxième test, nous avons utilisé un ordinateur ayant les caractéristiques suivantes:

L'environnement matériel	Caractéristique
Processeur:	Intel®Atom™N570 CPU1.66GHz.
Mémoire installée(RAM):	2GB
Camera:	Acer Crystal Eye Webcam Résolution: 640*480 pixels.

Tableau 4. 4: les caractéristiques de l'environnement de test

Le tableau ci-dessous montre les différents résultats du deuxième test:

Police \ Taille	Time New Roman	Arial	Cambria	Magneto
8-18	Illisible	Illisible	Illisible	Illisible
18-20	Peu claire	Peu claire	Peu claire	Peu claire
20-24	Lisible	Lisible	Lisible	Lisible
24-26	Lisible	Lisible	Lisible	Lisible
26-72	Lisible	Lisible	Lisible	Lisible
Avec des couleurs	Lisible			
Image	Lisible			
Tableau	Lisible			
Graphe	Lisible			

Tableau 4. 5: les différents résultats du deuxième test.

① Est ce que le matériel en matière de processeur influence le système

② le temps est un facteur déterminant dans les applications temps réel pour faire ce n'est pas pris en considération

③ utiliser d'autres phrases ou même écrit à la main pour l'évaluation très réduite mon satisfaisante pour évaluer les peu amonies

5. Discussion des résultats:

En comparant les deux tableaux ci-dessus, on observe que les résultats varient d'un test à l'autre:

- La lisibilité du texte dépend de sa taille, et non de la police utilisée: ??
 - Le texte de tailles entre 8 et 18 points est illisible dans les deux tests.
 - Le texte de tailles de 18 à 20 commence à devenir lisible avec une caméra d'une résolution de 640*480.
 - Avec une taille comprise entre 20 et 24, le texte commence à être lisible dans le premier test, alors qu'elle est lisible avec la deuxième résolution.
 - A partir d'une taille de 26 points, le texte est toujours lisible, avec les deux résolutions de caméra utilisées.

- Le changement dans la taille des bordures des tableaux n'a pas de conséquence sur leur lisibilité.

↳ rien n'est mis dans les tableaux de test.

- De même par rapport aux graphes. On observe que, dans nos tests, ces traits restent lisibles quel que soit leur taille.

??

- Les images restent lisibles dans les deux tests.

quelque soit leur taille ??

On peut considérer que la lisibilité d'un texte, sélectionné avec notre système, ne dépend pas de la police utilisée pour écrire le texte, en tout cas pour les polices les plus populaires (Times New Roman, Arial et Cambria). Par contre, elle varie avec la taille du texte. Plus la résolution de la caméra utilisée pour la capture est élevée, plus la taille à laquelle le texte reste visible diminue.

pourquoi pas la position de caméra, le calibrage ??

Pour les tableaux, les graphes et les images, la résolution des caméras utilisées, n'a pas l'air d'avoir des conséquences sur leur lisibilité, ceci devant être valable uniquement pour des résolutions acceptables.

6. Les difficultés et problèmes rencontrés:

Parmi les problèmes rencontrés lors de la réalisation de notre système, on peut en citer deux:

• **Problèmes liée à la luminosité et l'éclairage:** une mauvaise luminosité ou l'éclairage augmente le niveau de bruit dans l'image. Ce qui rend la détection et le suivi du doigt de l'utilisateur très difficile. *?? c'est pas une détection du doigt.*

• **Problème liée à la résolution de la caméra:** selon la résolution de la caméra les résultats de la sélection sont différent, surtout sur des textes de petites tailles, rendant le texte sélectionné illisible.

7. Conclusion:

Dans ce chapitre, nous avons présenté notre système à travers son interface. Nous avons également présentée l'environnement de développement matériel et logiciel utilisé. Puis, nous avons décrit les tests réalisés sur notre système, ainsi que les résultats obtenus. Nous avons conclu que la sélection d'une zone de texte sur la feuille de papier s'effectue correctement, mais la lisibilité du texte sélectionné dépend de la résolution de la caméra utilisée pour la capture et de la taille du texte.

Conclusion générale :

La vision par ordinateur est une source potentielle de nombreuses avancées en interaction Homme-machine. Dans certaines situations, la vision semble être la seule approche capable de fournir une solution au problème d'interaction. Dans d'autres situations, la vision apporte un avantage par rapport aux approches traditionnelles.

Une des possibilités ^{3'} données par l'utilisation de la vision par ordinateur dans l'interaction homme machine est l'interaction à travers l'usage du corps, en particulier les gestes de la main de l'utilisateur. L'interaction gestuelle est déjà utilisée dans plusieurs domaines, tels que les jeux vidéo et les différents dispositifs de communication tactiles.

Nous avons réalisé dans ce projet un système de capture et transformation d'inscriptions en objets virtuels. Ce système permet à un utilisateur d'interagir avec une feuille de papier A4 à travers certains gestes de la main. Il s'agit d'utiliser des gestes de pointage pour sélectionner du texte sur une feuille de papier, filmée par une webcam, et d'exécuter quelques commandes sur le texte sélectionné.

Notre système comporte deux phases, une phase pour la détection de peau et la reconnaissance du bout de doigt de la main effectuant le geste de pointage, et une phase d'interaction par l'exécution de différentes commandes (sélectionner, copier, imprimer).

Nous avons testé notre système avec des caméras de résolution différentes, sur des feuilles contenant un texte imprimé à différentes polices et tailles de polices, ainsi que des images, des graphes et des tableaux utilisant plusieurs largeurs de trait.

Les tests ont montré que la détection et le suivi du doigt, et donc l'interaction avec la feuille de papier est toujours correcte, par contre, lors de la sélection d'un texte, la lisibilité du texte sélectionné dépend de la résolution de la webcam utilisée. Pour des textes dans la taille de police est trop basse, la sélection obtenue est soit peu claire, soit illisible, rendant toute utilisation de l'objet sélectionné (dans un OCR, par exemple) difficile, et parfois même impossible.

↓ petite

Perspectives :

Pour augmenter l'utilisabilité de notre système, nous proposons un certain nombre d'améliorations :

1. Amélioration de la segmentation par deux approches, logicielle et/ou matérielle :
 - Implémenter des méthodes adaptées au changement de l'éclairage, qui portent sur la normalisation des intensités des images,
 - Utiliser des capteurs HD pour élever la résolution des vidéos.
2. Améliorer la reconnaissance du geste d'interaction, par la modélisation 3D de la main, permettant d'augmenter le corpus de gestes reconnus.
3. Utiliser la reconnaissance des caractères par l'ajout d'un OCR pour récupérer le texte sélectionné.
4. Utiliser des méthodes ou des moyens pour calibrer la caméra utilisée.

- [1] Jérôme Martin, RECONNAISSANCE DE GESTES EN VISION PAR ORDINATEUR, thèse de doctorat, Grenoble INP, 2000.
- [2] A.BERRACHED, F.OUAMRI, Commande basée sur le geste, Mémoire de master, USTO, LAERSI, 2010.
- [3] Simon Conseil, Suivi tridimensionnel de la main et reconnaissance de gestes pour les Interfaces Homme Machine. Thèse de Doctorat, Université Paul Cézanne - Aix-Marseille III, mars 2008.
- [4] LOCKTON, Ray. Hand gesture recognition using computer vision. 4th Year Project Report, 2002, Oxford University.
- [5] Guray ERUS, Reconnaissance d'objets cartographiques dans les images satellitaires à haute résolution, thèse de doctorat, Université PARIS DESCARTES, 2008
- [6] Felipe de Jesús Trujillo Romero, Modélisation et Reconnaissance active d'objets 3D de forme libre par Vision en Robotique, thèse de doctorat, l'université de Toulouse, 2008
- [7] Jean-François Jego. Interaction basée sur des gestes définis par l'utilisateur : Application à la réalité virtuelle. Thèse de doctorat. Ecole Nationale Supérieure des Mines de Paris, 2013.
- [8] KHADHRAOUI, Taher, BENZARTI, Fouazi, et ABDULRAHMAN, Hamid AMIRI. Gesture Determination for Hand Recognition. LSTS Laboratory National School of Engineers of Tunis (ENIT), Tunisia, 2011.
- [9] Nallaperumal K, Ravi S, Babu CNK, Selvakumar RK, Fred AL, Christopher S, Vinsley SS. Skin Detection Using Color Pixel Classification with Application to Face Detection: A Comparative Study. Proceedings of the International Conference on Computational Intelligence and Multimedia Applications, 2007.
- [10] HUYNH-HUU, Hung, MEUNIER, Jean, SEQUEIRA, Jean, et al. La détection et le suivi de régions d'intérêt pour la vidéosurveillance de la prise de médicaments. In : XXIIe colloque GRETSI (traitement du signal et des images), Dijon (FRA), 8-11 septembre 2009.

- [11] MENSER, Bernd et WIEN, Mathias. Segmentation and tracking of facial regions in color image sequences. In: Visual Communications and Image Processing 2000. International Society for Optics and Photonics, 2000. p. 731-740.
- [12] Hajarbouirouga, Reconnaissance des scènes vidéo pour adulte, thèse de doctorat, Université Mohammed V-Agdal, 2012
- [13] BencherietChemesse-Ennehar, BouallegAbd El halim, Tebbikh Hicham, Segmentation de la Couleur de Peau par Seuillage Selon Différents Espaces de Couleur, conférence .LAIG, Université 8 Mai 45 de Guelma, 2007
- [14] Matilde Gonzalez Prooiado. Computer Vision Méthods for Unconstrained Gesture Recognition in the Context of Sign Language Annotation. Thèse de doctorat. Computation and Language [cs.CL], Université Paul Sabatier - Toulouse III, 2012
- [15] Brougui Manel et Boumaraf Nestine, Détection et suivi d'un objet suspect dans le contexte de vidéosurveillance, mémoire de master, Université Kasdi Merbah - Ouargla, 2015.
- [16] Aurélie Bugeau. Détection et suivi d'objets en mouvement dans des scènes complexes, application a la surveillance des conducteurs. Thèse de doctorat. Interface homme machine. Université Rennes 1, 2007.
- [17] SINGLA, Nishu. Motion Detection Based on Frame Difference Method. International Journal of Information & Computation Technology, 2014, vol. 4, no 15, p. 1559-1565.
- [18] Nicolas Verbeke, suivi d'objets en mouvement dans une séquence vidéo, thèse de doctorat, l'université paris descartes, 2007.
- [19] Lucas, B. D. et Kanade, T. An itérative image registration technique with an application to stereo vision, journal. Vancouver, Canada. (1981)
- [20] J. J. Kuch, and T. S. Huang, «Vision Based Hand Modeling and Tracking for Virtual Teleconferencing and Telecollaboration». In Proc. Of fifth International Conference on Computer Vision, June 20-23, 1995, MIT, Cambridge, Massachusetts, p. 666-671.
- [21] J. Lee and T. Kunii, «Model-Based Analysis of Hand Posture». IEEE Computer Graphics and Applications, September 1995, p. 77-86.

- [22] Ajara-Badagry, La Pertinence De L'acquisition De La Competence Gestuelle En Didactique De Fle, journal, Village Français du Nigeria, 2014.
- [23] THOMET, Julien. Une vue d'ensemble de la reconnaissance de gestes. Département d'informatique-Université de Fribourg, Séminaire «Gesture recognition», Semestre d'automne, 2009.
- [24] KARAM, Maria, et al. A taxonomy of gestures in human computer interactions, Technical Report, Faculty Of Physical Science And Engineering, University Of Southampton, 2005.
- [25] P. A. Harling et A. D. N. Edwards: Hand tension as a gesture segmentation cue. In Proc. Of Gesture Workshop, 1996.
- [26] Rudy Morin. Espace de conception et modèle d'interaction multi-tactile gestuel : un environnement de développement pour enrichir le modèle. Thèse de doctorat. Computer Science. Université Toulouse le Mirail - Toulouse II, 2011.
- [27] Mathias BAGLIONI, Nouvelles Interactions Physiques pour Dispositifs Mobiles, Thèse de doctorat, TELECOM ParisTech, 2012.
- [28] M. Bertrand Tornil. Nudine Baptiste-Jessol : Adaptations et Interactions gestuelles et haptiques, ciblées utilisateurs. Vers plus d'utilisabilité et d'accessibilité. Thèse de doctorat. Université paul sabatier - toulouse iii. 2006.
- [29] François Bérard. Vision par ordinateur pour l'interaction homme-machine fortement couplée. Thèse de doctorat. Interface homme-machine. Université Joseph-Fourier - Grenoble I, 1999. Français.
- [30] Clément Réverdy. Modèles de Markov Cachés (HMM) pour de la reconnaissance de gestes humains. Stage Master Recherche. Machine Learning. 2014.
- [31] JM Rehg, T Kanade, Visual tracking of high DOF articulated structures: An application to human hand tracking, In Proc. European Conference on Computer Vision, 1994
- [32] B. Stenger, Template based Hand Pose recognition using multiple cues, In Proc. 7th Asian Conference on Computer Vision, 2006, Japan.

[33] VEERIAH, Vivek et SWAMINATHAN, P. L. Robust hand gesture recognition algorithm for simple mouse control. *International Journal of Computer and Communication Engineering*, 2013, vol. 2, no 2, p. 219.

[34] Feng-Sheng Chen, Chih-Ming Fu, Chung-Lin Huang, Hand gesture recognition using a real-time tracking method and hidden Markov models, Institute of Electrical Engineering, National TsingHua University. *Image and Vision Computing* 21 (2003) 745–758.

[35] <http://www.acsysteme.com/fr/matlab-ou-scilab>.