EIDGENÖSSISCHE TECHNISCHE HOCHSCHULE LAUSANNE
POLITECNICO FEDERALE LOSANNA
SWISS FEDERAL INSTITUTE OF TECHNOLOGY

ÉCOLE POLYTECHNIQUE
FÉDÉRALE DE LAUSANNE

Projet de 7ème Semestre 2002/2003

Interface utilisateur basée sur les gestes visuels pour chirurgie

Professeur: Reymond Clavel
Assistants: Sébastien Grange, Terry Fong
Candidat: Chauncey Graetzel
Section: Microtechnique

Virtual Reality and Active Interfaces Group

1 Introduction

1.1 Contexte : La vision en général

La vision par ordinateur est devenue un sujet très actuel. En effet, la mise au point de caméras bon marché et l'augmentation continuelle de la puissance de calcul des ordinateurs rend la vision artificielle abordable et réalisable.
Les moyens classiques d'interagir avec un ordinateur (HCI : Human - Computer Interaction) sont la souris, le clavier et les systèmes périphériques tels qu'un joystick. Ces moyens ont une chose en commun : elles ont été développées spécifiquement pour l'interaction avec les machines. L'utilisateur a dû apprendre à les utiliser convenablement. L'exemple le plus évident étant le clavier, chacun de nous ayant passé un certain nombre d'heures à apprendre où se trouvaient les touches avant de taper efficacement.
Le principe innovateur de la vision est d'aller dans l'autre sens. Les gestes visuelles nous sont intuitifs, nous les utilisons depuis notre plus jeune âge afin de communiquer avec d'autres personnes, et ceci de façon naturelle. Le geste est donc un candidat idéal pour les HCI. Pourtant les défis sont nombreux dans ce domaine, raison pour laquelle les gestes ne supplanteront pas totalement nos traditionnelles souris de si tôt :

- La reconnaissance doit être robuste : le taux de détection doit être élevé, même si les gestes sont effectués dans des conditions variées : gestes plus ou moins précis, obstruction d'une partie de l'objet détecté, condition de luminosités différentes, plusieurs utilisateurs, etc.
- Les gestes doivent être facilement exécutables pour l'utilisateur. Le fait de pouvoir répéter un geste plusieurs fois à la suite joue un rôle important. En effet, un geste banal comme lever son bras face à un écran peut devenir très fatigant s'il doit être répété plusieurs fois à la suite.
- La vision est gourmande en temps de calcul. Les applications qui doivent tourner en parallèle en souffrent. La nécessité que les calculs se fassent en temps réel rend le problème plus difficile encore, puisque les délais de reconnaissance doivent être courts.

Ces problèmes, qui sont d'ailleurs très similaires à ceux rencontrés dans la reconnaissance de paroles, ont donnés naissances à une multitude de type de solutions, adaptées au besoin spécifique des applications visées. Une recherche sur ces solutions a été faite au chapitre 4.

1.2 Présentation du problème

1.2.1 Description du projet

Ce projet vise une application médicale. Il consiste à développer une interface utilisateur basée sur la reconnaissance de gestes dynamiques, dans le but de permettre aux chirurgiens de bénéficier de l'aide d'ordinateurs pendant les interventions, sans avoir à utiliser un clavier ou une souris.
L'acquisition de l'image se base sur une caméra stéréo couleur, qui permet d'avoir une information sur la distance séparant un objet de la caméra. L'association de l'information

couleur, qui est rapide en ce qui concerne le temps de calcul mais dépendante de l'intensité lumineuse, et l'information stéréo, dont la qualité est liée à la texture et qui est plus lente mais indépendante de l'intensité, est une partie intégrante de ce projet.
Le développement se base sur plusieurs logiciels de recherche déjà existants(voir §1.4.1).
Les types d'interface, de gestes, de volume de travail, de système de reconnaissance ont été laissés à définir.

La première question que l'on pourrait se poser est « pourquoi avoir choisi le médical ?». Les 2 paragraphes suivants expliquent brièvement la motivation derrière ce choix.

1.2.2 Pourquoi le médical ? : Utilités du côté médical.

L'application de la vision dans la salle d'opération est très intéressante pour le chirurgien.Voici une présentation des points les plus importants.

- Le chirurgien se trouve dans un environnement stérile. Il passe environ 20 minutes à se stériliser les mains avant de pouvoir commencer l'opération. L'utilisation d'un clavier ou d'une souris est donc quasiment inconcevable. La vision n'a pas ce problème, puisqu'elle est sans contact.
- Les systèmes actuels d'interaction utilisent des pédales. Ces pédales sont peu souples quand à leur utilisation. Un autre moyen d'interaction, en parallèle, serait souhaitable.
- La parole pourrait être utilisée pour l'interaction. Le problème étant que beaucoup de communications se font déjà par la parole entre le chirurgien et ces assistants. Cette solution reste tout à fait possible, l'idéal étant une association de la parole et des gestes, augmentant ainsi la robustesse du système général.
- L'ordinateur peut agir sur une quantité d'outils très variés

1.2.3 Pourquoi le médical ? : Avantages du côté du développement

L'application médicale de la vision dans la salle d'opération présente aussi des avantages importants du côté du développement, car la salle d'opération, avec un environnement complètement défini et contrôlé dans le temps, évite certains problèmes qui sont traditionnellement associés à la vision :

- La luminosité est constante.
- Le volume de travail est donné.
- Les objets à détecter sont bien définis, tant par leur emplacement que par leur couleur.

Par contre, il ne faut pas oublier que dans le domaine médical, les erreurs, même dans la commande d'instruments annexes, ne sont pas tolérables. Une étude plus poussée de la demande médicale est présentée au chapitre 3. On peut aussi se référer à [1]

1.3 Objectifs du projet

Le projet peut se diviser dans 5 objectifs distincts.

1. Rechercher les solutions existantes dans le domaine de la reconnaissance de gestes visuels. (Chapitre 4)
2. Etudier rigoureusement le système de vision utilisé. (Chapitre 2)
3. Analyser la demande médicale, définir la zone de travail. (Chapitre 3)
4. Développer et tester un système de reconnaissance dynamique (Chapitre 5 et 6).
5. Développer et tester une interface médicale utilisant le système de reconnaissance dynamique (Chapitre 5 et 6).

Le but final étant de créer un système de reconnaissance montrant les possibilités techniques, qui puisse servir à mieux savoir dans quels domaines il est possible du point de vue de la technique, mais aussi intéressant du point de vue médical, de poursuivre le travail.

1.4 Matériel utilisé

Cette section décrit le matériel, hardware et software, utilisé dans le cadre de ce projet. Ceci permet une comparaison des résultats avec des travaux similaires effectués avec du matériel différent ou aussi une reproduction éventuelle des expériences.

1.4.1 Software

L'environnement de programmation utilisé est le Visual C++ 6.0 sous Windows 2000 NT.

Les fonctions de base du traitement d'image (prise d'image, affichage, filtres morphologiques, etc) proviennent de la bibliothèque TLib, développée par S.Grange du VRAI Group, EPFL.

Le calcul des images stéréoscopiques est entièrement basé sur un système de vision développé au SRI et dénommé Small Vision System[2] et[3]. Cet environnement software permet la calibration du système de vision stéréoscopiques et le traitement de ces images (calcul de la distance de chaque pixel).

1.4.2 Hardware

L'ordinateur utilisé est un Pentium 4 à 1.8 GHz, ayant 512 Mb de RAM.
L'acquisition de l'image est réalisée par un système de vision stéréoscopique digital, le STH-MD1 [3].Ce système est composé de 2 caméras couleurs CMOS digitales, de 1.3 megapixel chacune, correspondant à une image de 1288x1032 pixels. Ces caméras sont montées sur un bâti rigide qui contient un module d'interface IEEE 1394. Ce bâti permet aussi de monter les caméras sur un trépied.
Le transfert des images se fait via une interface IEEE 1394 FireWire, une liaison haut-débit, permettant de contrôler les paramètres de la caméra depuis un ordinateur. On peut contrôler en outre le niveau de réglage du gain, le temps d'exposition et la résolution

effective des caméras (la résolution de l'image peut être baissée afin d'augmenter la fréquence de transfert ou de baisser le temps de calcul).

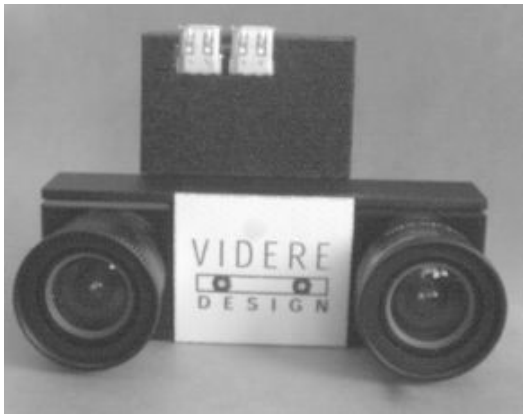


Figure 1.1: La caméra STH-MD1 utilisée dans ce projet.

2 Analyse du système de vision

L'analyse du système de vision nous permet de connaître précisément les caractéristiques de l'image utilisée. Cela nous permet de fixer les limites techniques, nous donnant ainsi une bonne base de travail pour des applications se basant sur ce système de vision.
Nous avons focalisé notre analyse sur trois domaines principaux de l'image :

- La résolution (dans le plan normal à l'axe focal et en distance caméra-objet)
- Le bruit de l'image stéréo.
- L'information couleur de l'image mono.

L'étude de la résolution de l'image s'est divisée en une étude théorique et une étude pratique, suivi d'une comparaison des résultats.
Les deux autres domaines se sont contentés d'une brève analyse théorique suivi d'expériences pratiques plus poussées.

2.1 Etude théorique de la résolution

2.1.1 Résolution de l'image :

La prise d'image est en fait une projection en perspective, décrite ici par le modèle pinhole, qui lie les coordonnées 3D des objets aux coordonnées 2D de la caméra. Le plan image est placé entre l'objet et la caméra. Il se trouve à une distance f égale à la focale de la lentille (voir figure 2.1).
La relation entre les coordonnées sur le plan image **m**(xi, yi) et les coordonnées spatiales **M**(X,Y,Z) est simplement :

$$\left\{\begin{array}{l} x_i = (X/Z)f \\ y_i = (Y/Z)f \end{array}\right\} \qquad (1)$$

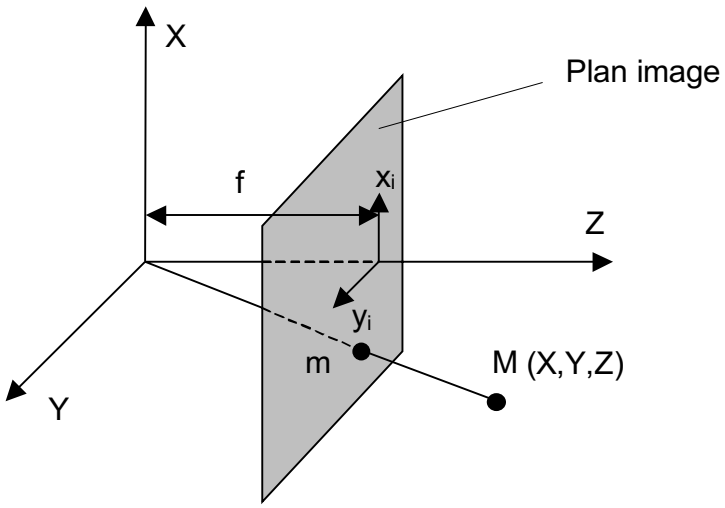


Figure 2.1 Projection en perspective de l'objet M(X,Y,Z) sur le plan image , cette dernière étant située à une distance f de la caméra placée en [0,0,0]. m (xi,yi) est le résultat de cette projection.

Nous voulons calculer la résolution en fonction de la distance de l'objet. La résolution est dans tout les cas limitée par la discrétisation.
Nous dénotons $\boldsymbol{\delta} = (\delta_x, \delta_y)^T$ la taille d'un pixel sur le plan image.

La taille du pixel dépend du champ de vision (« Field of view » ou FOV en anglais) de la caméra. Les valeurs des champs de vision horizontaux et verticaux en fonction de la focale f sont données par la documentation de la caméra [CAMERA DOC]. Celle-ci utilise le modèle pinhole décrit plus haut. Pour la lentille utilisée qui a une focale de 7,5mm, l'application numérique donne :

$HFOV_{théorique} = 2 \arctan(4,8/f) = 65.2°$ (2a)
$VFOV_{théorique} = 2 \arctan(3,8/f) = 53.7°$ (2b)

Ces valeurs ont été vérifiées expérimentalement en plaçant un objet à la limite de l'image et en mesurant sa position par rapport à la caméra pour en déduire l'angle :

$HFOV_{experimental} = 71.7°$
$VFOV_{experimental} = 51.4°$

Ces valeurs sont proches des angles calculés théoriquement. Voici un tableau présentant les valeurs théoriques des champs de vision en fonction des 2 focales utilisées au laboratoire :

Longueur focale [mm]	Champ de vision horizontal théorique	Champ de vision vertical théorique
4.8	90°	73°
7.5	65°	54°

Tableau 2.1 Valeurs théoriques du champ de vision en fonction de la focale.

La taille du pixel se déduit facilement de la figure 2.2 et des équations (2a) et (2b) ;

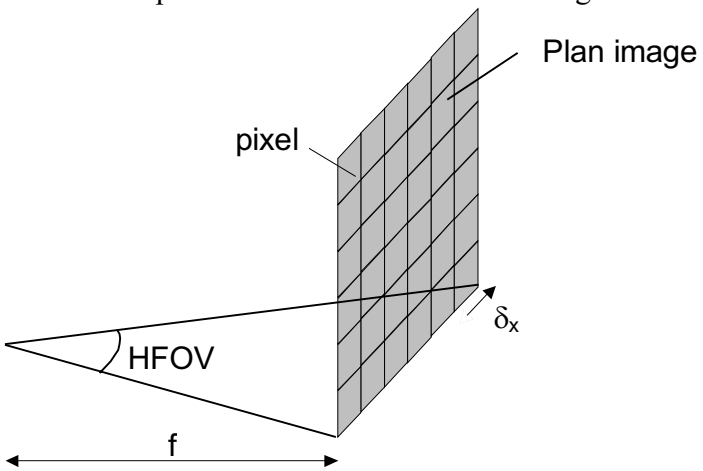


Figure 2.2 : Relation entre le champ de vision et de la résolution de l'image.

Nous utilisons une résolution d'image de 320*240, ce qui donne :

$$\delta = \begin{pmatrix} 9.6/320 \\ 7.6/240 \end{pmatrix} = \begin{pmatrix} 0.03 \\ 0.0316 \end{pmatrix} mm \qquad (3)$$

Maintenant que nous connaissons la taille du pixel, nous pouvons calculer la résolution dans le plan en fonction de la profondeur.

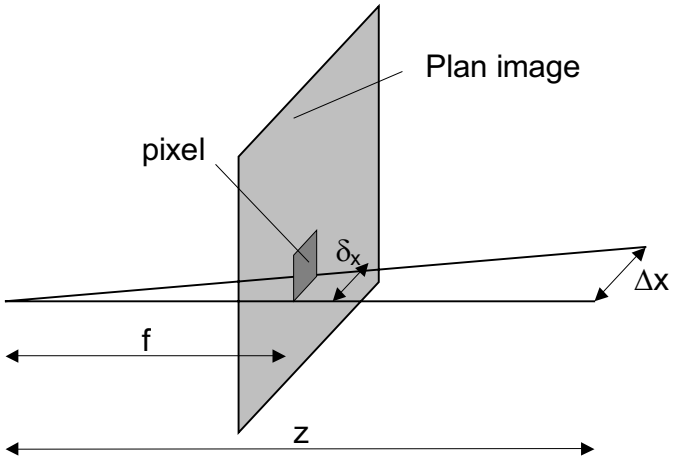


Figure 2.3 : Relation géométrique entre la résolution Δx,
la distance z , et les paramètres de la caméra δ et f.

Nous déduisons de la figure 2.3 :

$$\begin{pmatrix}\Delta x\\ \Delta y\end{pmatrix} = \frac{1}{f}\begin{pmatrix}\delta x\\ \delta y\end{pmatrix} z \qquad (4)$$

Il est intéressant de noter que la résolution est inversement proportionnelle à la longueur focale. Voici un tableau présentant quelques valeurs pour f =7.5mm

	Z [m]	1	1.5	2	2.5	3
f=7.5mm	Δx [mm]	4	6	8	10	12
f=4.8mm	Δx [mm	6.3	9.4	12.5	15.6	18.8

Tableau 2.2 : Résolution Δx *en fonction de la distance Z pour différentes longueurs focales*

Remarques : La largeur d'un doigt ne représente que 1-2 pixels à 3m. Ces valeurs ont été rapidement vérifiées en plaçant un objet d'une taille de 2.5cm à une distance de 3m. Celui-ci mesurant 2-3 pixels, nous avons conclu que les valeurs théoriques étaient du même ordre de grandeur que les mesures.

2.1.2 Résolution en profondeur en fonction de la distance.

2.1.2.1 Introduction

L'utilisation d'une caméra stéréo permet de connaître la distance d'un objet par rapport à celle-ci. Nous caractérisons ici la résolution théorique de cette mesure.
La caméra stéréo ne fournit pas directement la distance d'un objet. La caméra retourne une valeur qui s'appelle la disparité de l'objet.
La distance d'un objet est fondamentalement reliée à sa disparité dans l'image stéréo. La disparité est une mesure du décalage de l'image d'un objet (les coordonnées x_i et y_i de la figure 2.1) entre le plan image venant de la caméra à gauche et du plan image de droite.
L'algorithme du système stéréo cherche dans chaque image des points en commun, et ensuite calcule la disparité pour en déduire la distance de l'image. Idéalement, cette recherche se fait sur des lignes horizontales sur le plan image.

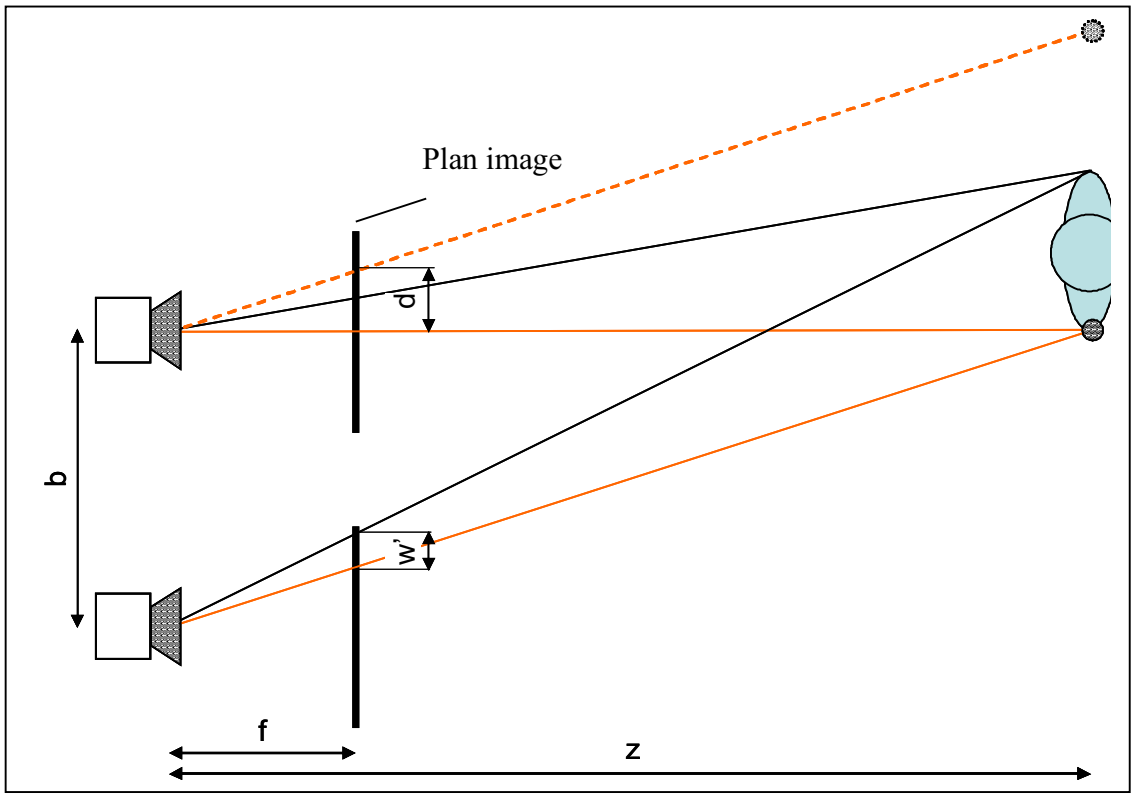


Figure 2.4 : Représentation de la disparité d. L'objet n'est pas vu au même endroit par les 2 caméras, une étant décalée par rapport à l'autre. Le trait tillé montre l'angle sous lequel l'épaule est vue par l'autre caméra.

Relation entre disparité d et profondeur z :
A partir de la figure, nous avons :

$$\frac{d}{f} = \frac{b}{z} \Rightarrow z = \frac{bf}{d} \qquad (5)$$

$$d \propto \frac{1}{z}$$

Nous voulons maintenant caractériser la précision de la mesure de la profondeur. Les erreurs sur le calcul de la profondeur se divisent en 4 catégories :

- Erreurs de discrétisation
- Erreurs de traitement d'images
- Erreurs de calibration
- Aberrations optiques

2.1.2.2 Erreurs de discrétisation

Du au caractère discret de l'image, les coordonnées images (x_i,y_i) ont une résolution de ±½ pixel à chaque point. Le calcul de la disparité se fait donc sur des positions estimées à ½ pixels près , ce qui introduit des erreurs sur le calcul de la profondeur.
Le calcul des erreurs de discrétisations se fait comme suit :

Les rayons lumineux venant de P croisent le plan image (voir figure 2.5). Les coordonnées de l'intersection avec le plan sont discrétisées. Suivant quelles valeurs elles prennent, la position estimée de P pourra prendre les valeurs P+ et P-. La résolution vaut Δz.

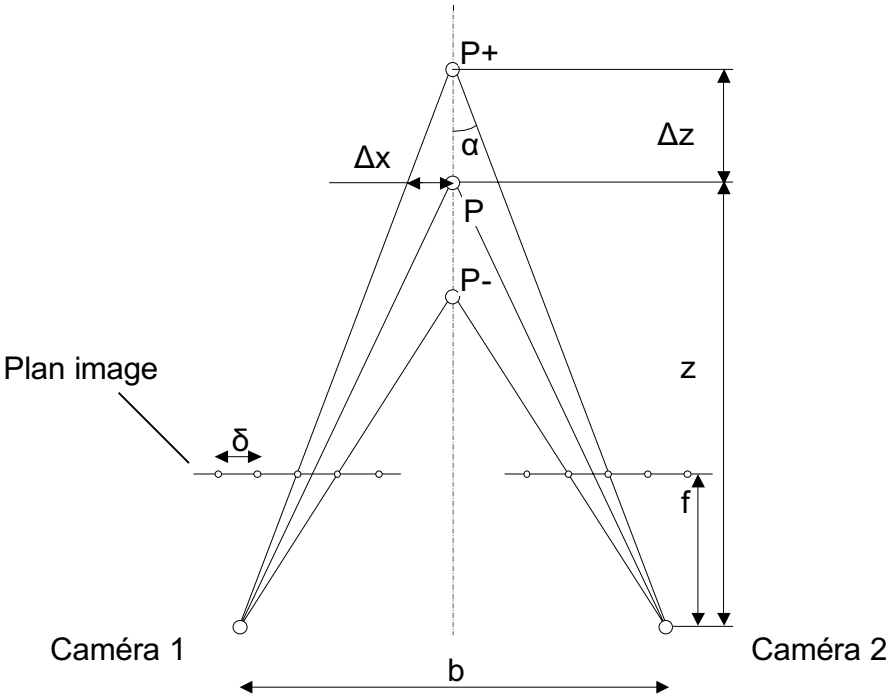


Figure 2.5 Représentation géométrique de la résolution en profondeur Δz

$$\alpha = \tan^{-1}\left(\frac{b/2}{z+\Delta z}\right)$$

$$\Delta x = \frac{z}{f}\delta$$

$$\tan\alpha = \frac{\Delta x}{\Delta z} \qquad (6)$$

$$\Rightarrow \Delta z = \Delta x \frac{z+\Delta z}{b/2} \cong \frac{\delta_x}{bf} z^2$$

2.1.2.3 Erreurs de traitement d'images

Les algorithmes qui travaillent sur l'image stéréographique introduisent des erreurs sur le calcul de profondeur. Ces erreurs sont souvent dues aux méthodes utilisées pour trouver des pixels dans chaque image correspondant au même objet. Les algorithmes ont des méthodes différentes. Calculer théoriquement ces erreurs dépasse le cadre de cette recherche.

2.1.2.4 Erreurs de calibration

Des erreurs de calibration influencent aussi la résolution en profondeur. La principale source d'erreur lors de la calibration reste les erreurs de discrétisations. Il est aussi difficile de calculer théoriquement l'amplitude de ces erreurs.

2.1.2.5 Aberrations géométriques

Les aberrations géométriques sont dues au type de lentilles utilisées. Ces erreurs sont mesurables et prévisibles. Des algorithmes sont disponibles afin de calculer la distance d'un objet en tenant compte des aberrations.

2.1.2.6 Remarques

Notons qu'une caméra «idéale » aura toujours des erreurs de discrétisation. La limite inférieure de la résolution est fixée par celles-ci. Pour diminuer cette erreur, il y a 2 solutions :

- augmenter la résolution du système. La caméra peut fonctionner à des résolutions plus hautes.Cette solution fait diminuer la valeur de δ. Son inconvénient réside dans l'augmentation de calcul nécessaire pour le traitement d'image.

- augmenter la longueur focale. Plus la longueur focale est grande et plus Δz est petit, mais il faut aussi être conscient que le champ de vision diminue aussi (voir équations n° 2). C'est donc un compromis.

La solution choisie pour ce projet de semestre est présentée au §5.2.

2.2 Etude pratique : Mesure de la résolution de la profondeur en fonction de la distance

2.2.1 But

Le but de cette expérience était de calculer empiriquement la résolution en profondeur, et ensuite de comparer les valeurs trouvées avec celles calculées théoriquement. Nous avons surtout cherché à connaître un ordre de grandeur de la résolution. Cette étude doit nous aider à définir les gestes possibles pour la reconnaissance. Ces gestes seront de toute façon exécutés plus ou moins précisément par les utilisateurs.

2.2.2 Déroulement

L'expérience s'est déroulée dans des conditions de luminosités intérieures normales. L'objet mesuré était un rectangle en carton, texturé à la main afin d'améliorer sa reconnaissance en stéréo. Il était placé à un angle de 45° par rapport à la caméra afin de pouvoir mesurer la différence de profondeur calculée par le système de vision stéréo entre le bord le plus proche et le plus lointain (voir figure 2.6). Ces bordures étaient facilement visibles sur l'image de référence en mettant le carton devant un fond noir. La disparité était mesurée sur l'image même. La différence entre la disparité maximale et la disparité minimale correspondant respectivement au bord le plus proche et le plus lointain du carton, donne le nombre discret de divisions en profondeurs. Les dimensions du rectangle étant connu, la résolution en profondeur Δz était calculée.
La résolution de l'image stéréo était de 320*240 pixels.L'expérience a été répétée pour 2 longueurs focales.

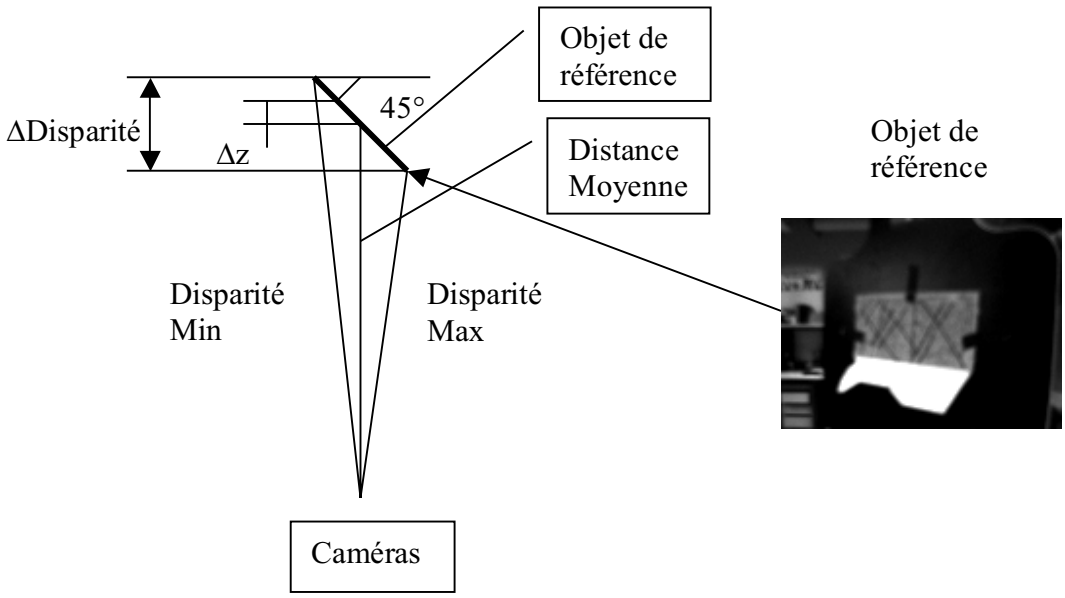


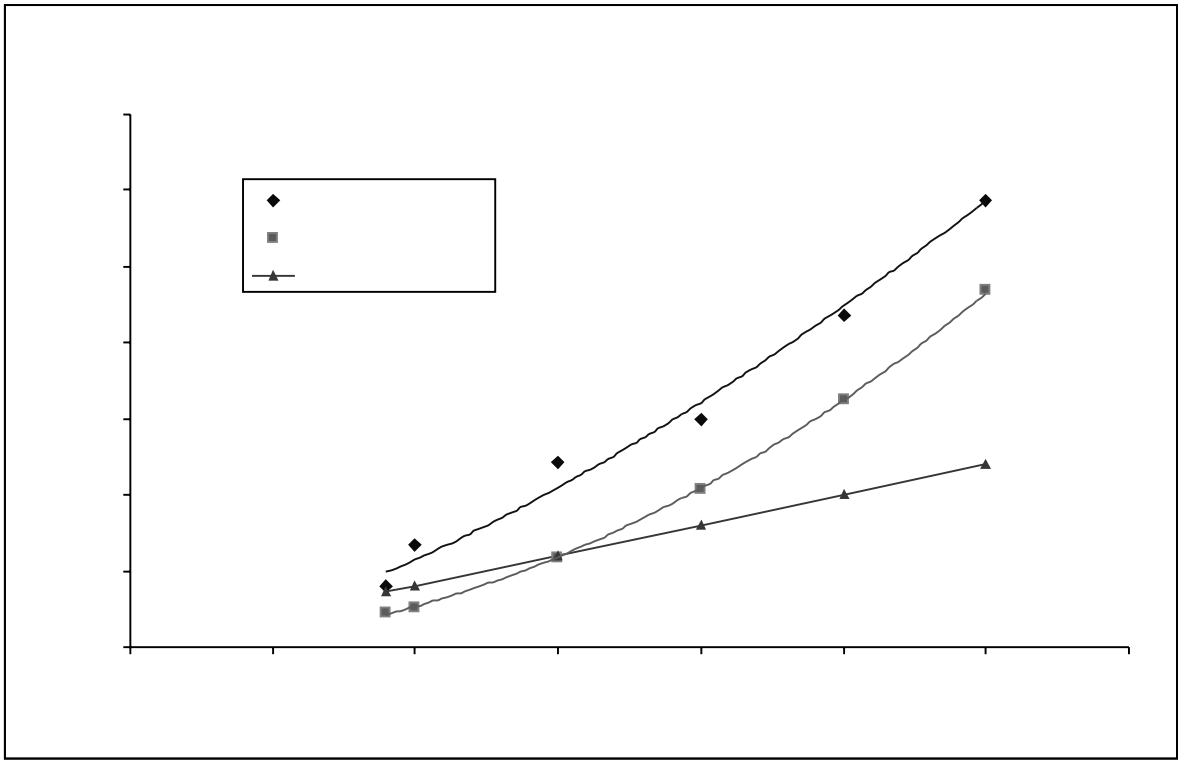
Figure 2.6 : Schéma de l'expérience, vue du dessus. La disparité, qui est inversement proportionnelle à la distance, est mesuré aux 2 extrémités. La taille de l'objet (photo) étant connue, la résolution en profondeur Δz est déduite

2.2.3 Résultats

Les résultats des disparités correspondent à la moyenne de 5 valeurs mesurées sur les bordures. Pour comparaison, nous avons calculé les valeurs théoriques de la résolution en profondeur (Δz-théorique) ainsi que la résolution dans le plan (Δx).

Focale	Distance [m]	Disparité min	Disparité max	ΔD [pixel]	Δz [mm]	Δz-théorique [mm]	Δx [mm]
7.5 mm							
	0.9	194	236.2	42.2	4.0	2.3	3.6
	1	165.4	190.6	25.2	6.7	2.6	4
	1.5	119.2	133.2	14	12.1	5.9	6
	2	87.4	98.8	11.4	14.9	10.4	8
	2.5	74.6	82.4	7.8	21.8	16.3	10
	3	62.2	68	5.8	29.3	23.4	12
4.8mm							
	0.45	220	160	60	2.8	0.9	2.8
	0.8	129	106	23	7.4	2.8	5
	1	104	88	16	10.6	4.3	6.3
	1.5	70	62	9	18.9	9.8	9.4
	2	50	45	6	28.3	17.4	12.5
	2.5	39	36	4	42.5	27.1	15.6
	3	31	29	3	56.7	39.1	18.8

Tableau 2.3: Valeurs de la résolution en profondeur Δz , la résolution en profondeur théorique Δz-théorique, et la résolution perpendiculaire à l'axe focal Δx, en fonction de la distance, pour une longueur focale de 7.5mm.



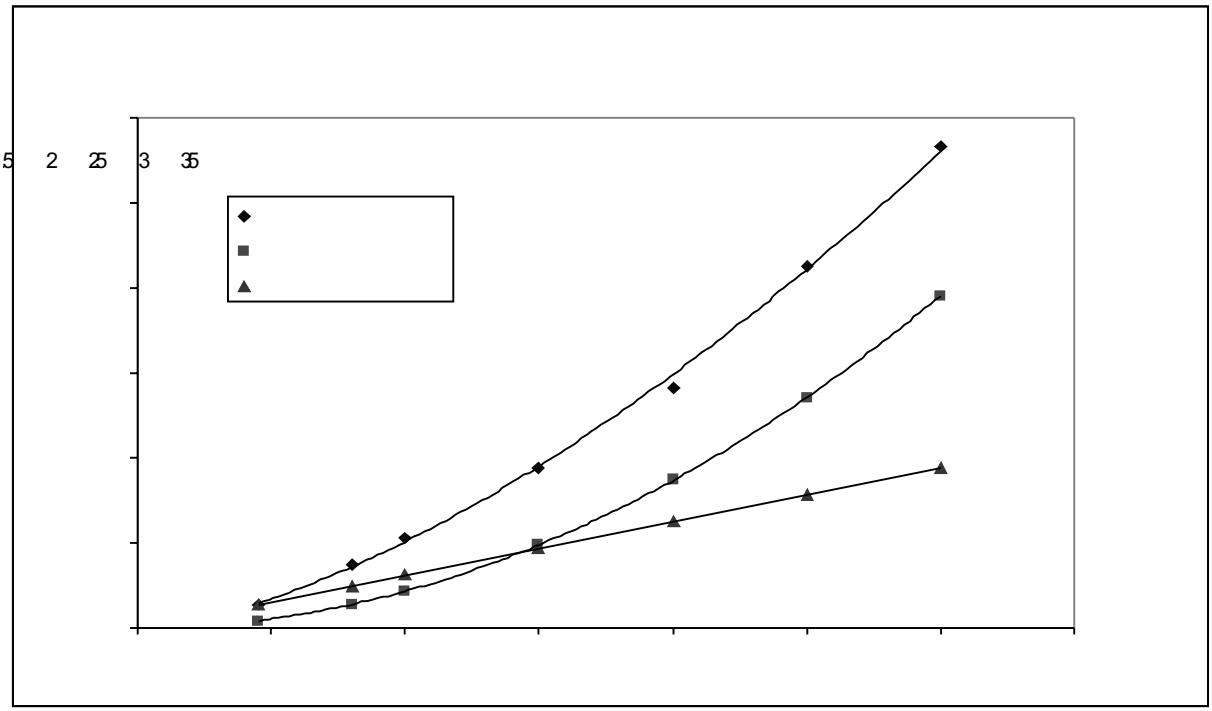


Figure 2.7 : Graphiques de la résolution en profondeur Δz, comparée avec la résolution minimale théorique Δz-théorique, et la résolution dans le plan normal à l'axe focal de la caméra Δx, en fonction de la distance séparant la caméra de l'objet.

2.2.4 Analyse

-A une distance d'environ 80cm, le système de vision n'arrive plus à calculer la disparité. Il est donc impossible de tirer une information de profondeur sur un objet placé plus proche.
-La valeur théorique pour la résolution en profondeur ne tient compte que de l'erreur de discrétisation. La différence avec la courbe trouvée empiriquement est égale à la somme des erreurs de calibration , du traitement d'image et des erreurs de mesures.
-Nous n'avons pas mesuré la précision de l'information de distance elle-même. La valeur de la distance pourrait être décalée, mais cela ne nous importe pas dans ce projet car c'est plutôt l'amplitude des mouvements et non pas leur éloignement par rapport à la caméra qui nous intéresse vraiment. Ceci se comprend intuitivement : lorsqu'on nous cherchons à interpréter un geste, nous nous soucions du mouvement relatif des bras par rapport à eux-mêmes. La distance à laquelle ses gestes sont exécutés ne vient pas en compte, à condition bien sûr que ces gestes soient effectués à une distance depuis laquelle nous arrivions encore à les distinguer. Cette approche est exactement la même pour un système de reconnaissance dynamique : la distance, en absolu, n'est pas importante, tant que celle-ci n'est pas trop grande.

2.3 Etude du bruit dans l'image stéréoscopique

La résolution de l'image stéréo est limitée par la résolution des deux images qui la composent. En pratique, les algorithmes stéréo introduisent une quantité importante de bruit. En effet, deux images d'un environnement statique prises à des intervalles de temps très proches montreront des différences significatives dans leur image stéréo. Ceci est du au fait que des petites variations de lumières rendent certaines textures sur l'image reconnaissables (ou pas) par les algorithmes de traitement. Typiquement, sur un objet avec peu de textures, l'image stéréo variera beaucoup, suivant si l'algorithme a réussi a faire le lien entre les deux images ou pas.

L'expérience consistait a prendre 2 images stéréo à la suite et à les soustraire l'une à l'autre. Une étude statistique était ensuite faite sur les pixels ayant changés, pour savoir la moyenne et l'écart type des différences avec l'image initiale. La stabilité temporelle de l'image stéréo en fonction de l'intervalle de temps entre la prise des images a ainsi pu être déterminée.

L'image était prise dans des conditions de luminosités intérieures. L'environnement était statique et représentait des objets à une multitude de distances. Les premières statistiques ont été prises sur des fenêtres de différentes dimensions. Ensuite, le temps entre les prises d'images a été allongé intentionnellement, en gardant la taille de la fenêtre fixe. Les résultats sont dans le tableau 2.4.

taille de la fenêtre	**temps**	**pixels différents**	**écart-type**
[pixels]	[ms]	[%]	[niveau de gris 8bits]
20*20	80	3.25	3
50*50	70	1.16	16
75*75	80	1.13	17
100*100	80	1.15	20
150*150	70	1.03	33
200*200	80	1.32	45
320*240	80	1.13	54
100*100	100	1.76	38
100*100	150	1.98	42
100*100	200	2.23	39
100*100	250	2.43	41
100*100	300	2.47	42

Tableau 2.4: Résultats de l'analyse du bruit stéréo

Remarques :

- Les valeurs indiquées sont des valeurs moyennes pour 5 mesures.
- Etonnamment, le temps de calcul de la disparité d'une portion d'image est indépendante de la taille cette portion. Cela porte à croire que l'algorithme calcule la disparité pour toute l'image chaque fois et ne garde que la portion désirée.

- Les écarts-types augmentent avec la taille de la fenêtre. Ceci s'explique par le fait que l'image complète comporte des zones proches et lointaines. Lorsque nous prenons une petite fenêtre, les valeurs de disparités risquent d'être proches puisqu'elles sont plus probables d'appartenir à un même objet, induisant un faible écart-type.
- L'augmentation du temps entre chaque prise d'image augmente l'erreur de façon presque linéaire. La valeur pour la plus petite fenêtre est considérée comme aberrante.

2.4 Analyse de l'information couleur

L'information couleur permet de segmenter l'image source très rapidement. Avant d'utiliser la couleur, il faut caractériser la confiance que l'on peut avoir dans les mesures, afin de distinguer ce qui est significatif du reste.
Le but de cette analyse est de déterminer expérimentalement les variations de couleurs dans le domaine temporel et de la position sur l'image. De plus, les caractéristiques de la couleur de la peau ont été étudiées.

2.4.1 L'acquisition de l'information couleur sur la caméra STH-MD1-C :

Les capteurs CMOS sont monochromatiques. On obtient l'information couleur en utilisant des filtres afin de mesurer les valeurs des trois couleurs de base. Les capteurs sont disposés selon le « Bayer Pattern », qui alterne sur une ligne les capteurs à filtres verts et bleus, et sur la ligne suivante les capteurs à filtre verts et rouges [5]. La moitié des capteurs sont donc verts. Une conclusion importante est qu'une caméra noir/blanc, avec le même nombre de capteurs, a une résolution quatre fois supérieure à une caméra couleur.
La caméra mesure une couleur sur 8bits/pixel. Des algorithmes sont appliqués afin de créer le format RGB, et la couleur corrigée est ensuite interpolée. Le résultat est donné sur 32bits/pixels.

2.4.2 Relation entre la couleur et la position sur l'image :

Une étude statistique a été faite sur la couleur RGB en fonction de la position de l'image, afin de déterminer si l'information couleur est plus stable temporellement au centre de l'image qu'au bord.
L'étude n'a pas cherchée à vérifier la précision de l'information couleur. En effet, la couleur RGB varie énormément en fonction de l'illumination, et il n'est pas possible de créer expérimentalement des conditions de luminosités parfaitement uniformes.

L'expérience s'est déroulée de la façon suivante :
Les mesures ont été faites sur 9 points différents de l'image. A chaque point, 10 images ont été prises de l'objet de référence, un classeur de couleur bleu. Ces images étaient prises à la suite alors que l'objet de référence ne bougeait pas.
Ensuite, les valeurs RGB d'un point sur chaque image ont été mesurées, afin de faire une comparaison. Les écarts-types sont présentés dans la figure 2.8.

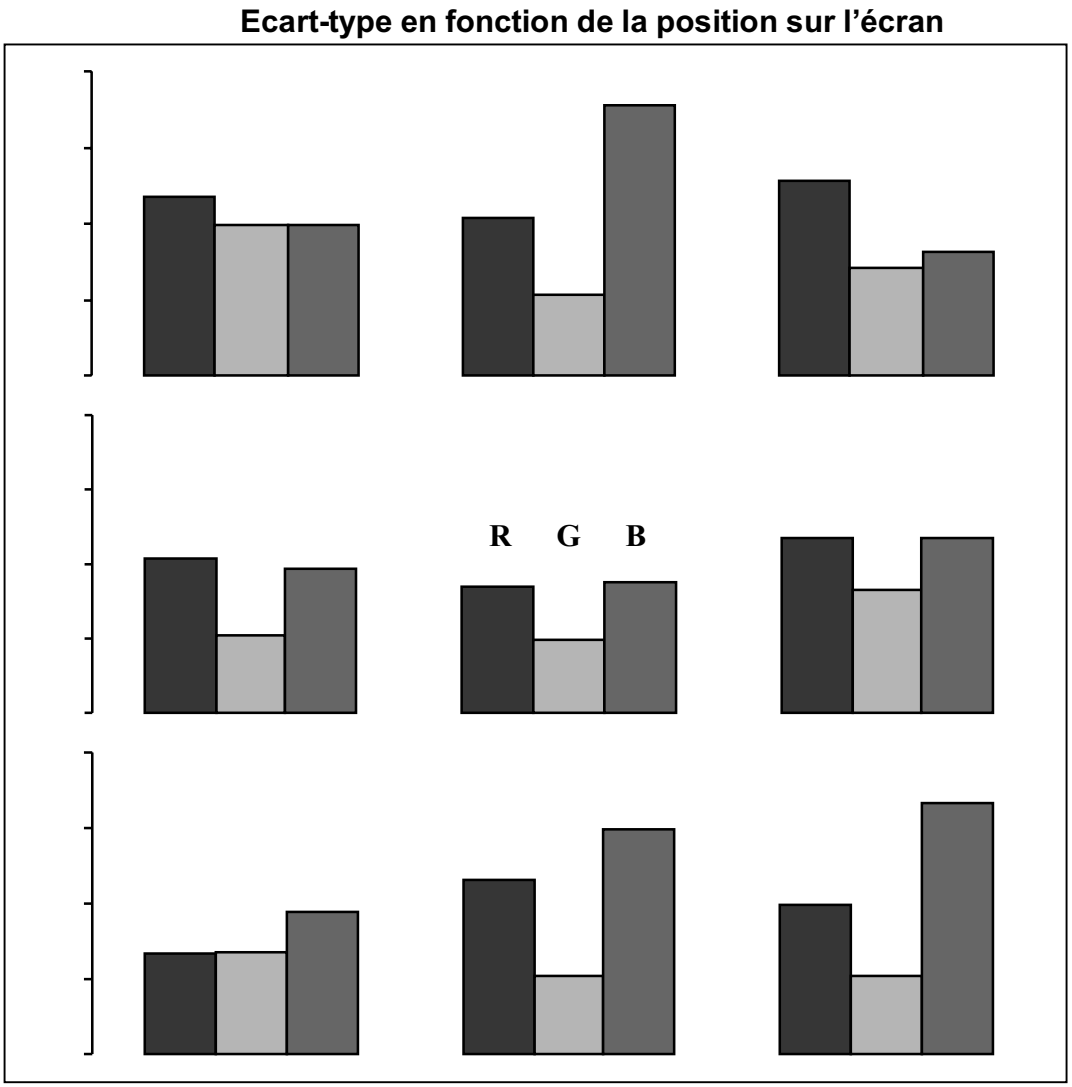


Figure 2.8: Représentation des écarts types du rouge, du vert et du bleu (respectivement RGB) en fonction de la position sur l'écran (position des histogrammes)

Discussion :

- Les histogrammes montrent une légère diminution de la variation de la couleur au centre de l'image. Cette anisotropie peut avoir différentes causes :
  -Les mesures de chaque composant de la couleur sont théoriquement indépendantes de l'intensité lumineuse, mais en pratique il s'avère que

non. Les capteurs ont des sensibilités qui varient un peu en fonction de l'intensité.

- La valeur de ces écarts-types reste relativement faible. La valeur de la couleur devrait être considérée à ±3 près. Tous les 8 bits mesurés par la caméra sont significatifs.
- La faible variation de la couleur verte est due au fait qu'il y a 2 fois plus de capteurs verts (Bayer Pattern). L'interpolation accumule donc moins d'erreurs.
- La fluctuation plus grande de la couleur bleue s'explique par le fait que les photodiodes des capteurs CMOS sont moins sensibles quand on s'approche de l'ultraviolet [5].

2.4.3 Etude de la couleur de la peau

La peau a une couleur bien distincte, qui peut être utilisée afin de reconnaître des parties humaines. L'étude de la couleur de la peau est donc indispensable. Il faut en particulier tenir compte de la variation des caractéristiques quand on passe d'une personne à une autre. Il est intéressant de noter que le bronzage de la peau affecte principalement l'intensité lumineuse. C'est en fait le sang sous la peau et un pigment, la mélanine, qui sont responsables pour la couleur caractéristique. Ainsi, en prenant un format de couleur normalisé, une peau plus foncée diffère peu d'une peau claire [5].
L'expérience consistait à mesurer la couleur de la peau sur 6 personnes différentes. La couleur était mesurée 30 fois pour chaque personne. Les surfaces mesurées se trouvaient soit sur la main, l'avant-bras ou le visage. Le format de couleur qui a été utilisé était le NRG (Normalized Red Green), format qui donne les valeurs de rouge et de vert indépendamment de l'intensité.

Les valeurs trouvées expérimentalement, codées sur 8 bits, sont:

	Espérance	Ecart-type
Rouge normalisé	112.8	8.1
Vert normalisé	70.5	3.6

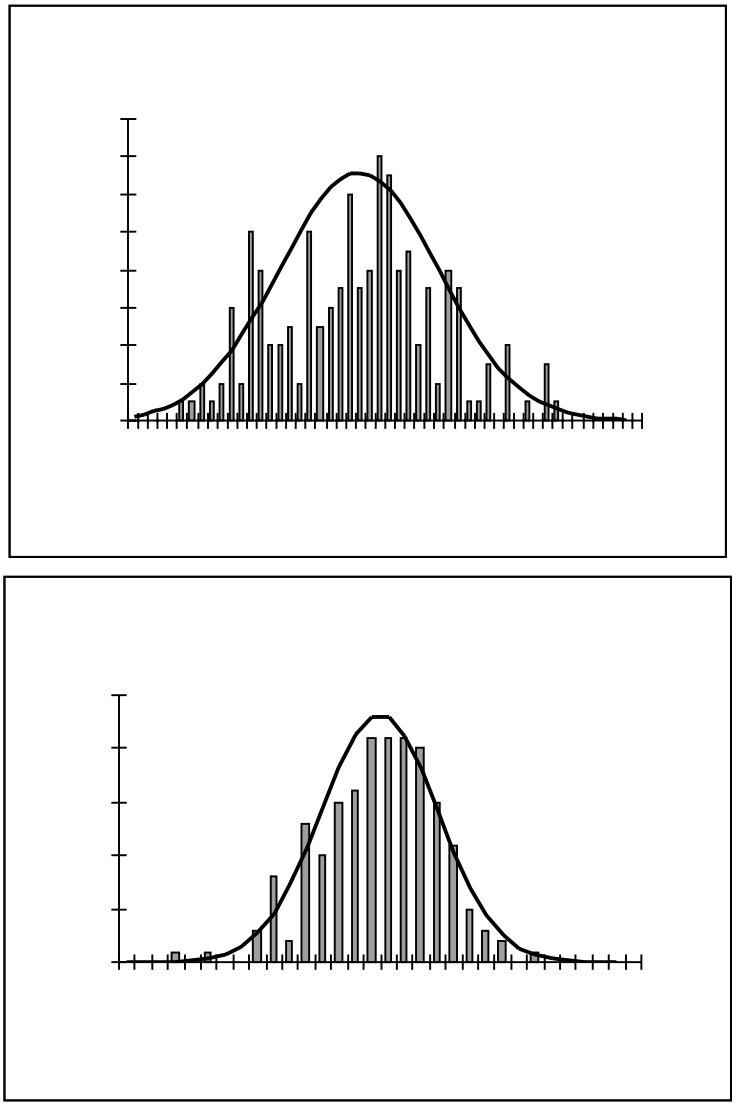


Figure 2.9: Répartition de la couleur de la peau humaine.

Remarques :

- Les valeurs mesurées peuvent être modelés assez précisément comme une distribution gaussienne théorique ayant les mêmes écart-type et variance, on notera néanmoins une légère dissymétrie pour les valeurs de rouge, qui présentent plusieurs pics secondaires.

2.4.4 Variation temporelle de la couleur

Il s'agit de déterminer la variation au cours du temps de la couleur d'une série d'images fixes successives prises dans les mêmes conditions.
Pour ce faire, nous avons étudié la variation des couleurs normalisées rouge et verte pour un même pixel, dans une série de 40 images.

Les images étant prises à une fréquence d'environ 25Hz, la durée totale de l'échantillonnage est dans notre cas de 1,6 s.

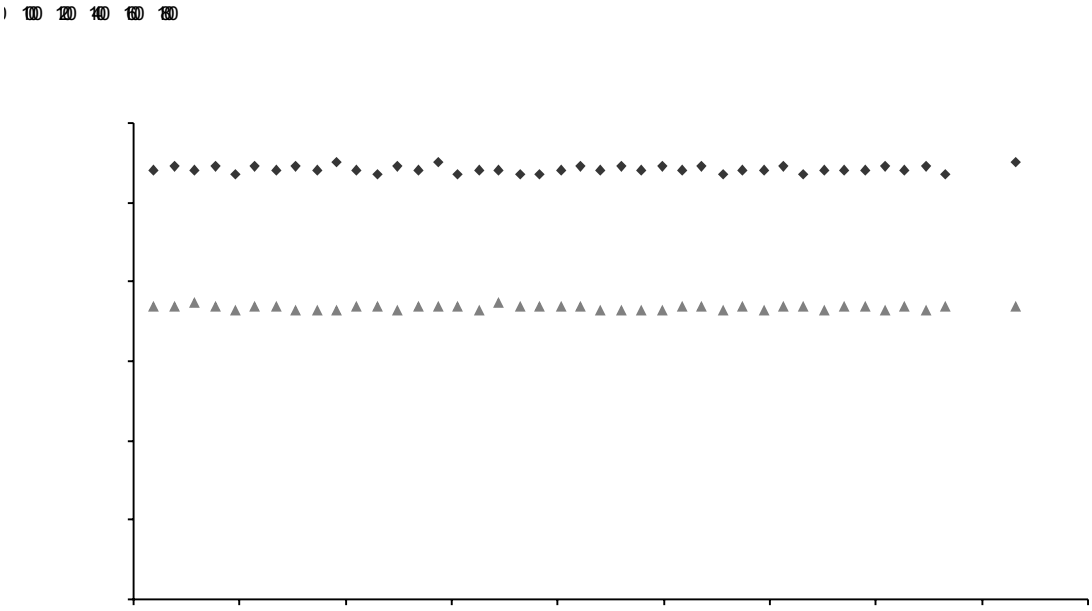


Figure 2.10: Stabilité temporelle de la couleur.

Cette mesure nous donne une idée du bruit statique de la caméra, car toutes les images sont prises dans les mêmes conditions. Ce bruit est notamment dû à la variation lumineuse ambiante, à l'optique de la caméra, et aux différents algorithmes d'acquisition et de traitement de l'image.

Néanmoins, on constate que ce bruit est faible, on obtient les écarts-types suivants :

0.82 pour le rouge
0.57 pour le vert

Nous considérerons donc cette variation comme négligeable dans nos algorithmes futurs.

3 Analyse de la demande médicale

3.1 Introduction

Tout développement recherche la qualité. La qualité peut être représentée comme étant l'intersection des trois cercles de la figure 3.1.

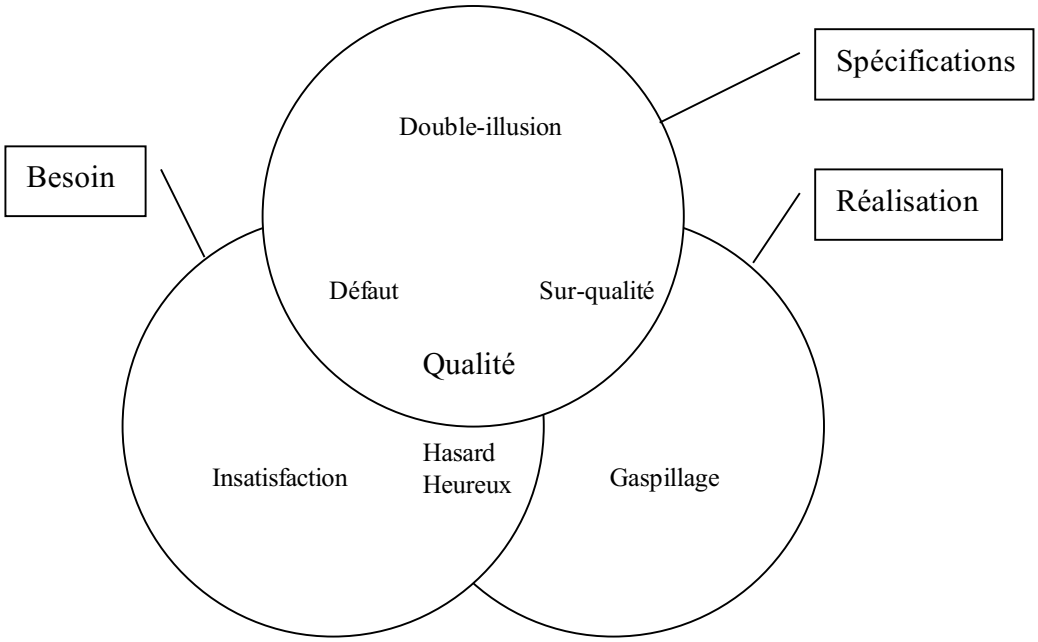


Figure 3.1 : Illustration du concept de qualité. Extrait de [6].

Au début d'un projet, ces cercles ne sont pas définis. C'est justement le but de ce chapitre de mieux connaître l'intérêt médical c'est-à-dire, dans la figure 3.1, le besoin. La difficulté vient du fait que le côté médical veut très souvent savoir ce qui est faisable du point de vue technique, et ainsi juger si cela est intéressant pour eux. De plus, les chirurgiens ne sont pas des ingénieurs. Ils ne voient pas tous d'un bon œil une innovation technique, ayant craintes que cela ne leur complique encore plus la tâche.

Une étude technologique préalable doit donc être faite. Sur ces bases, les points critiques essentiels peuvent en être tirés. Un questionnaire a été élaboré en fonction de ces points critiques, ce qui structure la discussion.

L'analyse de la demande ne se fait pas simplement une fois. Ce chapitre représente une étude initiale. Un échange permanent au cours du développement doit se faire afin de cibler au mieux les besoins.

3.2 Etablissement des points essentiels

Les points essentiels, sur lesquels la discussion devait porter, ont été établis en début de projet et ont suivi plusieurs modifications.Le questionnaire qui a été envoyé aux

médecins et qui a servi comme base lors des interviews effectuées se trouve en annexe. On trouve içi une liste des thèmes discutés.

- Types de systèmes qui pourraient être contrôlés par l'interface
  - Genre de contrôle nécessaire pour leur commande (digitale, analogique, analogique sur plusieurs dimensions, autre)
- Emplacement de la caméra dans la salle d'opération
- Gestuelle :
  - Libération des mains pour l'interaction.
  - Zone de travail (où s'effectuent les gestes)
    - Position
    - taille
    - fréquence d'occlusions
    - quantités de gestes « parasites »
  - Type de gestes souhaités
- Gants :
  - Différence de couleur avec le reste de l'habillage
  - Changement de couleur en cours d'opération
  - Gants « adaptés » (texture, marqueurs couleurs)

3.3 Prise de contact avec les médecins

Comme dit précédemment, un questionnaire a été formulé sur la base des points essentiels du §1.2. Celui-ci a été transmis à environ une dizaine de chirurgiens en tout. X réponses ont été recueillies.
Deux interviews de chirurgiens ont été effectuées :

- Dr Zambelli du HOSR,
  Effectuée à l'EPFL, le lundi 25 novembre 2002.

- Dr Caversaccio, Hôpital de l'Isle, Berne
  Effectuée le jeudi 16 janvier 2003.

Finalement, une assistance à une opération réelle a été réalisée le jeudi 16 janvier 2003, à l'hôpital de l'Isle, Berne.

3.4 Rapport des Résultats

Le premier résultat qui peut être établi concerne la grande variation d'avis parmi les médecins!
Une synthèse a été réalisée, se basant sur les points où les chirurgiens étaient d'accord entre eux, donnant ainsi une idée générale de la demande.
Une des causes de ces différents avis vient en partie du fait que les chirurgiens questionnés travaillent dans différents domaines de la chirurgie. Ce projet visant à long terme une application dans la chirurgie minimale invasive, c'est dans cette optique que

les réponses ont été recueillies. Ce type de chirurgie représente 80% des opérations effectuées aujourd'hui. De plus, la tendance est à la hausse.

3.4.1 Synthèses des réponses aux interviews et aux questionnaires

La méthode la plus simple pour présenter les résultats est de reprendre les points essentiels du §3.2.

- **Types de systèmes qui pourraient être contrôlés par l'interface**

R :Intérêt général dans le système afin de commander des appareils, de diminuer les ordres donnés au personnel et l'utilisation des pédales, ainsi que pour la recherche de données informatiques (par exemple informations sur les patients).
Les appareils cités par les médecins comprennent :

- lumières
- table d'opération
- fluoroscope
- bipolar diathermy
- caméras vidéo
- electrocoagulation
- Rayons X intra-opératoire

Les chirurgiens ont montré un intérêt particulier dans un système permettant de contrôler un curseur sur un écran, donnant ainsi une grande flexibilité quant aux applications médicales envisageables.

  - **Genre de contrôle nécessaire pour leur commande (tout ou rien, analogique, analogique sur plusieurs dimensions, autre)**

    R : Dans tout les cas, un contrôle digital et analogique est nécessaire. Le contrôle analogique sur 2 dimensions est souhaitable, surtout si l'on tient compte de l'intérêt pour le maniement d'un curseur. L'utilité d'un contrôle sur 3 dimensions (ou plus) n'est pas encore bien défini et nécessiterait un démonstrateur afin de connaître les possibilités techniques.

- **Emplacement de la caméra dans la salle d'opération**

R: La réponse était sans équivoque : à côté de l'écran de contrôle. C'est la où les occlusions sont les plus rares (dans l'endoscopie, le chirurgien doit toujours voir l'écran). L'écran est situé en face du chirurgien, de l'autre côté de la table d'opération, à une distance d'environ 1m à 1m50.

- **Gestuelle :**

C'est dans la gestuelle que les avis différaient le plus.

  - **Libération des mains pour l'interaction**

R : On distingue trois cas principaux :

-le chirurgien est en train d'opérer avec les mains prises par des outils.
Dans le cas d'une opération endoscopique, l'interaction devrait alors se faire par les pouces, ceux-ci étant les seuls doigts encore libres.
Les chirurgiens ont montré une sceptitude à interagir avec l'ordinateur en utilisant des gestes dans ce cas là. En effet, la concentration requise dans ces moments est importante. Les commandes données ne peuvent être que très simples. L'utilisation de pédales ou d'un bouton placé sur l'instrument est préféré, de part leur excellente fiabilité et leur facilité d'utilisation.

-le chirurgien n'est pas actif dans l'opération, mais doit maintenir ses outils en place.
Là aussi, les chirurgiens étaient sceptiques à l'idée d'utiliser des gestes pour effectuer des commandes. En effet, différents doigts peuvent être bloqués et la réalisation de gestes par les doigts peut être difficile, surtout si les outils doivent rester très précisément en place. De plus, ces gestes sont assez limités et l'utilisation de pédales ou d'autres interrupteurs sont des solutions plus fiables.

-le chirurgien peut libérer au moins une de ses mains.
C'est dans ce cas que les chirurgiens ont montré le plus d'intérêt, malgré la contrainte supplémentaire. Dans ce cas, l'utilisation de gestes pour interagir avec un ordinateur est plus naturelle et rapide que des pédales ou des interrupteurs. La possibilité d'interagir sans avoir à poser ces outils (mais en pouvant bouger le bras) était souhaitée.

- **Zone de travail**

  - **Position**

    R : La zone de travail semblant être la plus naturelle se situait environ 30cm au-dessus du patient, donc à une distance d'environ 1m de la caméra.

  - **taille**

    R : Les chirurgiens s'accordaient sur un parallélépipède d'une hauteur de 40cm, une longueur en direction de la caméra de 30cm et d'une largeur de 45cm.

  - **fréquence d'occlusions**

    R : Les occlusions sont très rares dans la zone de travail, car elle est très contrôlée. En effet, un flux de gaz souffle sur cette zone pour des raisons de stérilités. Peu de personnes ont le droit de s'aventurer dans cette zone.

  - **quantités de gestes « parasites »**

R : Comme cela a été dit précédemment, les gestes effectués dans la zone de travail ont chaque fois un but précis. Il est néanmoins difficile de quantifier ses gestes en général, puisqu'ils dépendent de l'opération spécifique.

- **Type de gestes souhaités**

R : Les chirurgiens veulent un système qui soit le plus simple possible. Un système pour lequel le chirurgien doit apprendre tout un vocabulaire de gestes et les exécuter d'une façon précise n'est pas envisageable. Les gestes déictiques (les gestes où l'on pointe vers quelque chose) semblent être les plus naturels. Quelques gestes simples représentant une action, par exemple une croix faite dans l'air pourrait signifier « annulation », sont envisageables, mais en nombres limités.

- **Gants :**
  - Différence de couleur avec le reste de l'habillage

R : Les gants sont d'une couleur bien distincte et différente des couleurs du reste de l'habillage du chirurgien.

- **Changement de couleur en cours d'opération**

R : La différence entre la chirurgie « ouverte » et la chirurgie « minimalement invasive » est ici marquante. Dans la chirurgie ouverte, presque la moitié du gant peut être recouvert par du sang lors d'une opération (ce chiffre a été donné de manière informative, évidemment la quantité de sang est dépendante du type d'opération). Alors qu'en chirurgie endoscopique, la couleur des gants reste presque toujours intacte. Notons qu'il est toujours possible d'utiliser la couleur du gant et celle du sang pour l'analyse de l'image, en élaborant un système où les caractéristiques de la main sont mesurées à nouveau en cours d'opération.

- **Gants « adaptés » (texture, marqueurs couleurs)**

R : L'utilisation de gants plus texturés que normaux (afin de rendre le calcul de l'image stéréo plus efficace) ou avec des petits marqueurs de couleurs, n'a pas posé de problèmes particuliers pour les chirurgiens.

3.4.2 Assistance à une opération réelle

Description

L'opération s'est effectuée le jeudi 16 janvier 2003, à l'hôpital de l'Ile située à Berne. Celle-ci était de type endonasale.

Mr Grange et moi-même avons pu assister à l'opération en salle même, nous donnant un point de vue idéal, comme nous le voyons sur la figure 3.2

Figure 3.2 : Dr Caversaccio en pleine opération

La salle d'opération avait la particularité d'être muni d'un système de navigation permettant de connaître la position des outils endoscopiques par rapport au crâne du patient. Le système consistait d'un détecteur placé sur un piédestal à 2 m de la table d'opération (figure 3.3), d'émetteurs placés sur les outils et d'un écran de contrôle sur lequel la position de l'outil était représentée. La vue depuis la caméra endoscopique était aussi représentée sur cet écran (figure 3.4). Pour plus d'informations sur ce système de navigation, se référer à [7].

Figure 3.3 : Les 3 détecteurs du système de navigation sont visibles en haut à gauche. Ils permettent la localisation des outils par rapport au patient.

Figure 3.4 : L'écran de contrôle du système de navigation, situé à environ 1.5 à 2m du chirurgien, qui voit en temps réel le positionnement de ses outils ainsi que la vue depuis la caméra endoscopique (en bas à droite).

Déroulement de l'opération

L'opération a duré environ 2 heures. Celle-ci consistait à enlever un foyer d'infection situé dans l'arrière de la cavité nasale. La patiente était évidemment sous anesthésie complète. Dr Caversaccio avait à disposition deux écrans pour visualiser l'image prise depuis la caméra endoscopique. Ces écrans devaient être impérativement visible pour le chirurgien. Durant toute la durée de l'opération, pas une occlusion de ces écrans ne s'est produite.
L'opération a débutée avec une phase de calibration du système de navigation. Durant cette phase, le chirurgien devait placer des outils de pointage sur le patient afin de connaître le positionnement relatif de ces outils par rapport au patient.
Durant cette phase, le chirurgien avait à sa disposition une petite commande (le "virtual keyboard") lui permettant de choisir diverses options sur le menu à l'écran. La validation se faisait via une pédale située sous la table d'opération. Le chirurgien demandait aussi à un assistant de changer des options, ce que ce dernier accomplissait en utilisant le clavier et la souris à disposition. A un moment, l'assistant s'est trompé, appuyant sur le faux bouton de la souris. La confusion qui s'en est suivi a nécessité l'intervention du chirurgien, qui a dû enlever ses gants, utiliser l'ordinateur de façon classique, et se stériliser les mains à nouveau. Cet épisode a duré 7 à 8 minutes, montrant clairement qu'un système d'interaction plus adapté serait utile. La manière d'interagir avec l'ordinateur était ici rustique, nécessitant un assistant supplémentaire qui n'avait aucun autre rôle dans l'opération et créant une perte de temps importante pour une opération à priori simple.
Une fois cette phase de calibration terminée, l'opération même a débuté. A partir de ce moment, plus aucune interaction avec l'ordinateur n'a été nécessaire, les outils nécessaires à l'intervention étant contrôlés par des boutons montés sur les instruments. Le chirurgien tenait de sa main gauche la caméra et de sa main droite l'outil. Les changements d'outils étaient assez fréquents (environ toutes les minutes). Au cours de cette opération, les gants n'ont pas eu de taches de sang. Il était intéressant de noter que la zone au-dessus du patient était réservée pour le chirurgien. Même l'assistante posait les outils nécessaires au chirurgien sur le côté, s'approchant rarement à moins de 30 cm du patient. Ces observations sont toutes positives pour l'introduction d'un système de reconnaissance dans la salle d'opération.
Un problème se pose néanmoins pour l'utilisation de caméras : la salle d'opération a été assombrie au cours de cette opération endoscopique afin de rendre l'écran de contrôle plus visible. La luminosité était trop faible pour espérer utiliser une caméra normale.Ce problème devra être étudié plus en détail dans un travail futur.

Figure 3.5 : Dr Caversaccio en train d'utiliser l'outil de navigation au cours de l'opération. Un des détecteurs est visible en haut à gauche de l'image.

L'opération s'est terminée de façon tout à fait normale.

3.5 Conclusion

Cette recherche a permis de mieux cibler les buts à atteindre du système de reconnaissance dynamique.
Voici les spécifications souhaitables que nous pouvons tirer de l'analyse de la demande médicale.

- Le système devra pouvoir pointer et cliquer dans une fenêtre Windows en utilisant des gestes déictiques, employant des postures de main les plus générales que possible afin de diminuer le travail d'apprentissage, et afin de rendre le système utilisable même si un outil est encore tenu par la main. Les bras sont considérés comme étant libres. Le contrôle doit être analogique sur 2 dimensions au moins. La robustesse du système est primordiale.
- La caméra est placée à une distance de 1m à 1m50 de la zone de travail.
- La zone de travail sera celle donnée au §3.4.1
- Une méthode doit être développée afin de distinguer les gestes ayant comme intention l'interaction avec le système des autres gestes fait dans la zone de travail.
- Le système doit pouvoir continuer à suivre la main du chirurgien même si d'autres mains passent dans la zone de travail.
- L'utilisation de l'interface utilisateur médicale doit être intuitive et permettre de régler certains paramètres du système de reconnaissance tels ceux qui sont spécifiques au gant employé.

- Le système doit pouvoir être contrôlable en même temps par des moyens très fiables (souris, clavier). Ceci peut être utile au cas où la reconnaissance ne fonctionnerait pas correctement, donnant la possibilité de reprendre le contrôle par des moyens classiques.

4 Etude de solutions

Ce chapitre décrit les différentes méthodes utilisables pour aborder le problème. C'est un résumé de l'état d'art actuel du domaine de la reconnaissance de gestes dynamiques.
Le problème initial est trop vaste pour être abordé tel quel.
La plupart des applications visuelles séparent le travail en trois.

1) Segmentation
2) Détection ou tracking
3) Interprétation

La segmentation vise à séparer l'information étant potentiellement utile du reste. Cette sélection initiale permet de concentrer le travail qui suit la segmentation seulement sur les parties relevantes de l'image.
La détection ou le tracking suivent la segmentation. Dans cette partie, l'on veut chercher à connaître les caractéristiques de l'objet cherché. La détection se distingue du tracking par le fait que la détection ne se soucie pas de ce qui a été calculé dans les images précédentes. La détection cherche sur toute la zone de travail, décidant quel est l'objet cherché parmi plusieurs hypothèses, alors que le tracking se concentre sur une zone particulière, où l'on pense que l'objet précédemment détecté se trouve.
Une fois que les caractéristiques actuelles de l'objet cherché sont trouvées, il s'agit d'interpréter leur signification, prenant en compte aussi les caractéristiques calculées précédemment.
La segmentation n'est pas très différente de la détection et du tracking. Habituellement, la détection et le tracking se font à un niveau plus élevé de discernement. L'étude de solutions s'est divisé en 2 grandes parties : Reconnaissance élémentaire et Interprétation de gestes dynamiques. La première se concentrant sur les techniques d'extraction d'information relevante d'une image, et la deuxième sur leur interprétation. Des exemples complets de reconnaissance visuelle et d'interaction médicale sont présentés à la fin de ce chapitre.

4.1 Reconnaissance élémentaire

Une liste est présentée des principales méthodes employées dans la sélection d'éléments importants.

4.1.1 Information couleur

La couleur est très souvent utilisée, de part sa rapidité de calcul. Evidemment, la couleur est efficace seulement si l'objet recherché a une couleur spécifique. Si d'autres objets dans le champ de vision ont une couleur similaire, il faut utiliser d'autres méthodes en parallèle afin d'affiner la sélection sur l'objet recherché. Un autre problème lié à la couleur est la variation de l'intensité lumineuse. Dans le format de couleur classique, le RGB, l'intensité lumineuse fait varier énormément la couleur mesurée. Des formats différents, moins dépendants de l'intensité lumineuse, ont été développé. Les deux plus connus sont le NRG (Normalised Red Green) et le HSI (Hue Saturation Intensity). Le NRG calcule la valeur d'une couleur en format RGB en la normalisant par son intensité :

$$\begin{cases} NR = \dfrac{R}{R+G+B} \\ NG = \dfrac{G}{R+G+B} \end{cases}$$

Le format HSI se calcule de façon un peu plus complexe, pour plus de détails, se référer à [8]. Le HSI donne une valeur de teinte (hue) et une valeur correspondant à la saturation. Le choix entre les différents formats de couleur dépend du type d'objet à observer.
Le filtre de la couleur peut être dynamique, c'est-à-dire qu'une fois qu'il a détecté l'objet recherché, les caractéristiques couleurs de cet objet sont enregistrées et utilisées dans le filtre suivant. Il faut néanmoins faire attention à l'instabilité temporelle d'un tel filtre, liée au fait que le filtre pourrait après un certain temps, ne plus chercher la bonne couleur du tout.
Le filtrage couleur est un choix intéressant pour ce projet, car l'étude médicale a montré que les gants étaient toujours d'une couleur spécifique.

4.1.2 Filtres morphologiques

L'utilisation de filtres morphologiques est très répandue. Les filtres morphologiques de base sont l'érosion et la dilatation. La dilatation permet, sur une image binaire, de connecter des groupements de pixels. Une analyse de ces groupements de pixels, ou blobs en anglais, permet d'extraire des objets. L'érosion est l'opérateur inverse de la dilatation. Les contours des groupements de pixels sont diminués. L'érosion peut être utilisée afin d'éliminer les pixels isolés, souvent associés au bruit. Ces filtres induisent une perte d'information, il s'agit donc de trouver, le plus souvent expérimentalement, le meilleur compromis.

4.1.3 Mouvement

Le mouvement est bien sûr réservé pour les applications où l'objet recherché bouge, avec néanmoins la contrainte que le reste bouge le moins possible. L'application d'un filtre basé sur le mouvement devient beaucoup moins efficace si la caméra bouge (par exemple si elle est fixée sur un robot où si elle tourne sur elle-même), du fait que tout l'environnement change aussi.

4.1.4 Elimination du décor

Cette technique est très similaire à la précédente. La différence vient du fait que dans ce cas, une image du décor est prise à un moment précis, par exemple au lancement du programme. Cette image est ensuite soustraite chaque fois à l'image actuelle, ce qui ne laisse que le changement visible. On diminue ainsi considérablement la quantité d'information à traiter. Cette technique marche bien lorsque la caméra reste fixe, et lorsque l'arrière-plan ne change pas constamment.

4.1.5 Contours

Les contours d'une image sont assez facilement calculable par des algorithmes existants. Le contour de l'objet recherché doit présenter des éléments caractéristiques. Les

extrémités des doigts sont souvent détectées par recherche sur les contours, car leur courbure est très significative [9].

4.1.6 Stéréoscopie

La stéréoscopie fournit une information sur la distance séparant la caméra des éléments de l'image. La stéréoscopie nécessite l'utilisation de deux caméras, séparées l'une de l'autre. Les deux images fournies par les caméras sont ensuite analysées par des algorithmes, qui cherchent à faire correspondre des éléments des deux images. La stéréoscopie est donc dépendante de la texture des objets. En effet, il est très difficile de faire correspondre correctement les images d'une surface lisse et de couleur uniforme. Une fois que la correspondance a été réalisée, l'évaluation de la distance est effectuée (voir §2.1.2).
La stéréoscopie est gourmande en temps de calcul, mais a l'avantage d'être peu sensible à l'illumination.
Cette méthode peut être utilisée soit pour filtrer les objets hors d'un intervalle de distance, soit pour estimer la taille réelle d'objets observés.

4.1.7 Auto-correlation

L'auto-correlation cherche à faire correspondre au mieux un modèle avec une certaine partie de l'image. Cette convolution peut devenir assez lente si elle doit s'effectuer sur toute l'image et s'il y a plusieurs modèles à chercher.

4.1.8 Estimation de paramètres

Lorsque du tracking est effectué, il est utile de pouvoir estimer la position la plus probable de l'objet détecté préalablement, afin de centrer la fenêtre de recherche de la position actuelle. Ceci se fait le plus souvent à l'aide de filtre de Kalman, ou de Markov. La théorie sur ces 2 méthodes n'est pas explicitée plus en détail ici, mais le filtre de Kalman est souvent préféré à celui de Markov pour sa rapidité de calcul. En contrepartie, le filtre de Kalman a le défaut d'être à hypothèse unique (suivi d'une seule position probable), le rendant plus instable.

4.1.9 Utilisation de Marqueurs

Les marqueurs sont très souvent utilisés dans la vision. Evidemment, la possibilité d'avoir des marqueurs dépend entièrement de l'application voulue. Les marqueurs passifs peuvent être des patchs, placés par exemple sur chaque extrémité de doigts ou sur un gant. Leur utilisation est motivée habituellement par une nécessité de robustesse accrue. Une autre classe de marqueurs est formée par les marqueurs actifs. Leur utilisation est beaucoup plus encombrante que les marqueurs passifs, nécessitant souvent des câbles ou des capteurs assez lourds. Les marqueurs actifs simplifient la détection et le tracking. Ils permettent d'avoir des mesures très précises de la position. Ils étaient beaucoup plus utilisés au début de la reconnaissance de gestes, lorsque la vision en temps réel n'était pas encore envisageable dû au temps de calcul.

4.1.10 Autres techniques

Cette liste ne présente que les techniques principales. D'autres tels que le Elastic Graph matching ou les modèles de Bayes (voir [10]).
Toutes les techniques présentées ici sont spécifiques à l'apparence 2D de l'objet sur l'image. Si nous prenons l'exemple qui nous intéresse le plus, la main, ces méthodes visent à connaître des paramètres « externes » tels que la position du centre de la main, le nombre de doigts visibles, leur position relative, etc.
Il existe une autre approche, où le but visé est de reconstruire un modèle 3D de la main. Beaucoup moins utilisé, le modèle 3D de la main nécessite, si l'on veut travailler en temps réel, de porter des capteurs. En effet, la reconnaissance devient plus complexe. Un exemple de modèle 3D peut être trouvé en [11]

4.2 Interprétation de gestes

L'interprétation de gestes est un sujet complexe. La première difficulté pour une interprétation de gestes réside dans la distinction entre les mouvements involontaires et les gestes eux-mêmes.
Le geste est ensuite décomposé en 3 parties principales : la préparation, le geste en lui-même et la rétraction.
Une taxonomie possible des gestes est représentée dans la figure suivante:

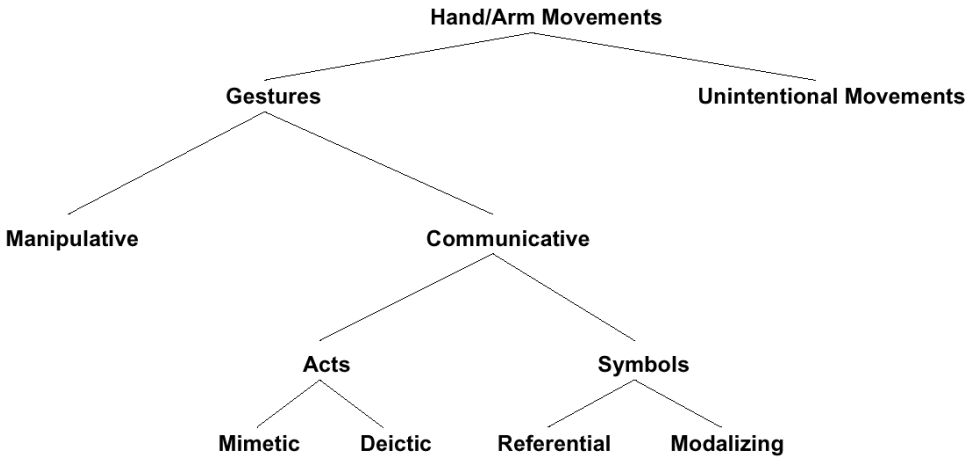


Figure 4.1: Une des taxonomies de gestes. Extrait de [14].

Le système d'interprétation doit être :

- Indépendant du moment où le geste débute
- Le plus indépendant possible de la vitesse et de la qualité d'exécution du geste.

Les méthodes d'interprétation sont souvent héritées de la reconnaissance vocale, qui fait face à des problèmes similaires.

4.2.1 Hidden Markov Model

Le Hidden Markov Model ou HMM est une méthode stochastique se basant sur un apprentissage préalable des gestes à reconnaître.
Tout comme dans le chaînes de Markov, les HMM sont composés d'observations O, issus de la détection ou du tracking, d'états ou states S, qui sont internes au HMM, et d'une probabilité que la suite d'observations appartiennent au HMM : $P(O \mid \lambda)$ où λ représente les caractéristiques spécifiques du HMM.
Un HMM est défini pour chaque séquence d'observations que nous voulons reconnaître. Ceci peut être des gestes en entiers ou des fragments de gestes, tout comme en reconnaissance vocale, on peut utiliser des HMM pour des mots en entiers ou pour des phonèmes.
La méthode générale d'utilisation des HMM est la suivante : après avoir entraîné chacun des HMM spécifiques, en lui donnant les observations liés au geste à reconnaître, nous obtenant λ. Ensuite, les HMM sont prêts à être employé pour de la reconnaissance. Les observations O sont calculées, et ensuite chaque HMM calcule la probabilité $P(O \mid \lambda)$ que ces observations soient causées par le geste pour lequel il a été entraîné. En comparant ces probabilités, on peut estimer si un geste a été effectué.
Une conclusion importante de cette explication est que plus il y a de HMM, plus le temps de calcul sera élevé.

Pour voir une étude plus poussée des HMM, se référer à [12]

Il y a avantages à utiliser les HMM lorsque les gestes à détecter sont compliqués (langage des signes, par exemple). Pour des gestes simples, leur implémentation n'est pas justifiée.

4.2.2 Représentation en états des caractéristiques dominants (PCA)

Cette méthode, dont la théorie est développée en détail dans [13], consiste à distinguer les caractéristiques ayant le plus de signification, puis ensuite d'analyser ces caractéristiques afin reconnaître le type de geste.
La première sélection analyse les caractéristiques des observations. Certaines propriétés des observations sont spécifiques à un ou plusieurs des gestes recherchés. Imaginons que nous voulons développer un système permettant de reconnaître les mouvements suivant : {Horizontaux gauches et droites, verticaux gauches et droites, rotation dans les sens positif et négatif}. Des grandes amplitudes dans la composante y et de petites amplitudes dans la composante x des observations sont forcément dus à un mouvement vertical. Il s'agit dans un deuxième temps de distinguer auquel des deux mouvements verticaux ce geste appartient. Par contre, si l'amplitude des mouvements est grande en x et en y, un mouvement de rotation a du être produit. La deuxième phase consistera alors à distinguer dans quel sens ce mouvement de rotation a été effectué.
Cet exemple est relativement simpliste. Un système doit aussi pouvoir détecter quand aucun des gestes recherchés n'a été effectué. Mais la méthode générale est applicable dans des cas plus complexes. Cette technique s'applique néanmoins principalement à des gestes simples. L'implémentation de cette technique est beaucoup moins lourde que les HMM.

4.2.3 Autre techniques d'interprétation

Par rapport à la reconnaissance élémentaire, les techniques d'interprétation sont encore plus dépendants du type d'application visé. L'interprétation doit simuler notre capacité à comprendre, à distinguer, l'intention. L'ordinateur étant loin d'avoir notre intelligence globale, il s'agit de créer un modèle pour l'application spécifique, souvent en prenant certaine conventions (tout comme les langages entre hommes sont basés sur certaines règles). Les méthodes d'interprétation sont trop nombreuses pour être listées dans cette étude. Pour plus d'informations, se référer à [14] et à [15].

4.3 Systèmes complets de reconnaissance

Cette section décrit brièvement des systèmes fonctionnels ayant des buts similaires à celui développé dans le cadre de ce projet. Cette liste, non exhaustive, sert à donner une idée des principaux axes de recherche dans ce domaine. Il est intéressant de noter les points suivants:

- Les systèmes sont le plus souvent des démonstrateurs de performance technologique, visant à montrer les possibilités atteignables avec le minimum d'investissement en matériel (nombreux "gadgets" techniques utilisant une webcam bon marché).
- L'assistance aux handicapés est un des seuls domaines dans lequel l'interaction par vision entre homme et ordinateur est exploitée pratiquement. Des systèmes existent dans lesquels une personne ayant perdu la faculté de bouger ses bras arrive néanmoins à utiliser un ordinateur en mouvant sa tête. Un autre exemple d'application est décrit dans [17], où un ordinateur peut être utilisé pour interpréter et traduire un langage de signe.
- A notre connaissance, aucun système de vision actuel n'a été développé spécifiquement dans le but d'améliorer les interactions entre chirurgien et ordinateur durant une opération.

4.3.1 Dey S. "Système de reconnaissance de postures de mains" [10]

Objectif

Reconnaissance de 6 différentes postures statiques des mains.

Méthode utilisée

Segmentation couleur, puis technique dite de "histogram matching". Environ 6 histogrammes circulaires, centrés sur la main et de rayons croissants, sont établis à partir d'une image segmentée et binaire. L'analyse de ces histogrammes permet d'établir si un doigt est tendu ou pas. Le système arrive donc à déterminer le nombre de doigts et leurs positions relatives.

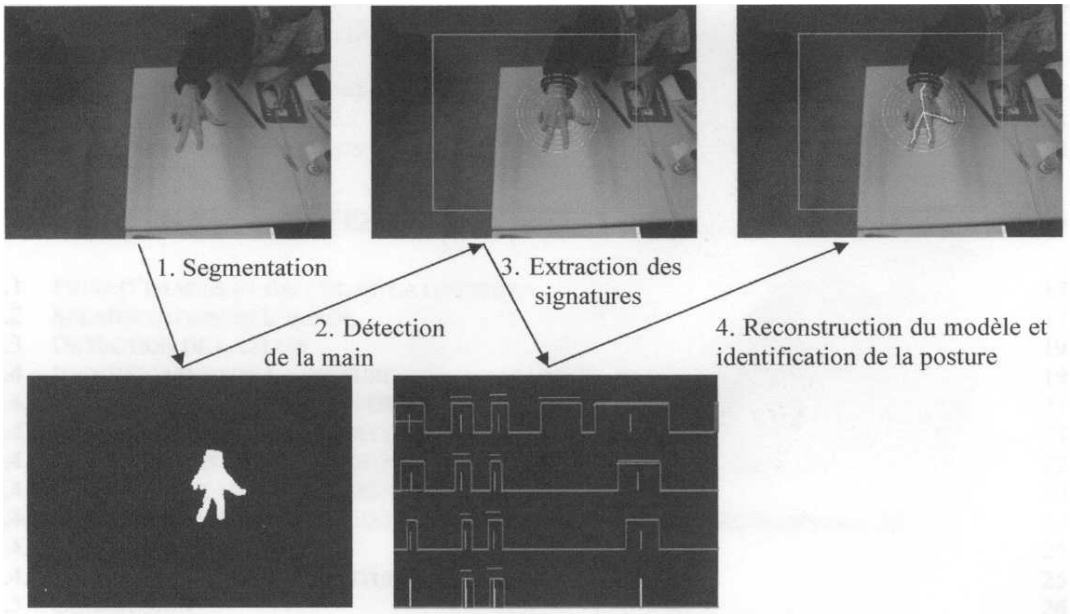


Figure 4.2: Processus d'identification de postures. Extrait de [10]

Remarques

Système statique. Caméra placée en dessus. Flexibilité réduite: nécessité de n'avoir qu'une seule main dans la zone de travail. Zone de travail limité. Bon taux de détection correct.

4.3.2 Von Hardenberg "Bare-Hand Human-Computer Interaction" [9]

Objectif

Contrôle d'un curseur avec le doigt

Méthode utilisée

Le décor est tout d'abord éliminé, puis les doigts sont recherchés à partir de la courbure significative des extrémités. La distance caméra-main est de l'ordre de 50cm.

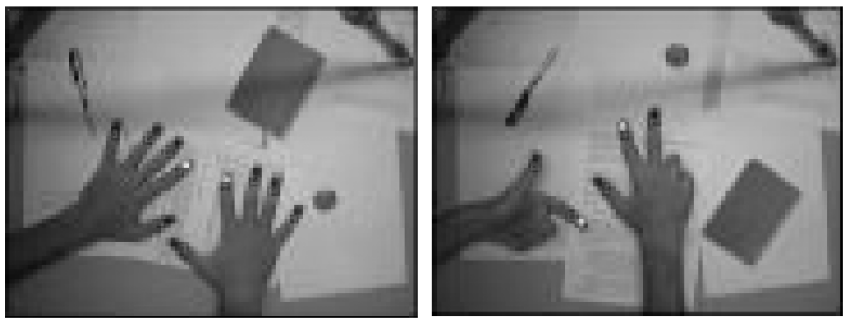


Figure 4.3: Extraction des extrémités des doigts.

Remarques

Trois applications visées: contrôle d'une présentation (p.ex. Powerpoint), dessin à main levée et déplacement d'icônes. Distances trop faibles pour application médicale. Technique néanmoins robuste.

4.3.3 Iannizzotto G. "Graylevel VisualGlove" [16]

Objectif

Contrôler un curseur avec la main sans contact en utilisant une webcam normale.

Méthode employée

Détection des extrémités des doigts de part leur contours. Travail en niveaux de gris. Distance caméra main < 50 cm. Click en faisant connectant l'index et le pouce.

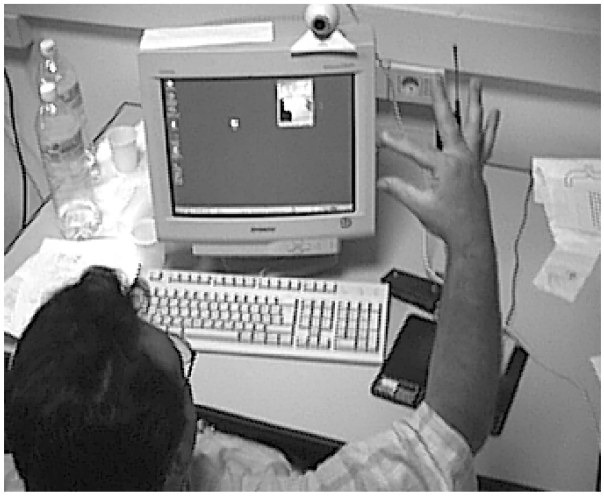


Figure 4.4: Utilisateur du Graylevel VisualGlove. Extrait de [16]

Remarques

Zone de travail pas compatible avec l'interface médicale. De plus, le pouce et l'index doivent être entièrement libres pour pouvoir utiliser le système.

4.3.4 Starner T."Virtual Recognition of American Sign Language Using Hidden Markov Models" [17]

Objectif

Reconnaissance d'un vocabulaire réduit du langage de signe (ASL) en utilisant un marqueur couleur.

Méthode employée

Une gant de couleur unique est employé pour faciliter la reconnaissance. Les caractéristiques analysées de la main sont la position, les axes d'inertie et l'excentricité de l'ellipse entourant la main.

Figure 4.5: Reconnaissance du langage des signes. Extrait de [17]

Remarques

Les HMM ont des excellents résultats pour ce genre d'applications. Le vocabulaire était ici faible, mais il est à souligner que cette recherche a été faite en 1995. Aujourd'hui, un système similaire aurait un vocabulaire sûrement beaucoup plus vaste.

4.4 Systèmes existants d'interaction entre chirurgien et ordinateur

Les ordinateurs sont déjà présents dans certaines salles d'opération. Le chirurgien doit dans la plupart des cas pouvoir contrôler ces ordinateurs. Une méthode d'interaction est donc nécessitée. Mis à part les solutions décrites dans les chapitres 1 et 3, dans lesquelles le chirurgien contrôle l'ordinateur au moyen d'un système de pédales et d'ordres donnés à un assistant, il existe actuellement un nombre restreint d'autres solutions plus adaptées.

4.4.1 Virtual Keyboard [18]

Le "Virtual Keyboard" fait partie du système de navigation utilisé durant l'opération que nous avons suivi ([7]). Le "Virtual Keyboard" est constitué d'une petite commande sur laquelle le chirurgien peut activer des boutons et calibrer les émetteurs placés sur les outils (voir figure 4.6). La commande est reliée à un ordinateur par un câble. Le chirurgien utilise les outils pour appuyer sur les boutons.
Ce système a une très bonne fiabilité. Il a le défaut d'être dédié à l'application pour laquelle il a été développé.

Figure 4.6: Le "Virtual Keyboard" et les outils qui peuvent le contrôler.

4.4.2 Ecrans tactiles (Touchscreen) Stériles

Les écrans tactiles sont une solution très intéressante pour le chirurgien. Différents fabricants proposent des modèles spécialement adapté pour leur intégration dans la salle d'opération (un exemple se trouve en [19]). Ces écrans résiste aux désinfectants médicaux et peuvent donc être nettoyé avant l'opération. Par contre, ils ne sont pas conçus pour soutenir les températures auxquelles on stérilise les outils. L'écran n'est donc pas "parfaitement" stérilisé, et le chirurgien ne peut pas toucher l'écran directement avec ses doigts. Par conséquent, les fabricants proposent des petits pointeurs en plastiques qui résistent à quelques stérilisations à haute température et qui sont ensuite remplacés (voir figure 4.7).
Cette solution allie rapidité et fiabilité, son seul inconvénient provient de la proximité nécessaire de l'écran tactile et donc d'un encombrement supplémentaire dans un environnement déjà chargé.

Figure 4.7: Ecran tactile et pointeur stérilisable pour application médicale

5 Description du système

Ce chapitre montre le choix des solutions et leur implémentation. Il se base sur le chapitre 3 pour les spécifications (présenté au §3.5) et sur le chapitre 4 pour le choix des possibilités.
Voici un organigramme représentant les parties principales de la boucle du système de reconnaissance et les données qui sont transmises entre chaque partie:

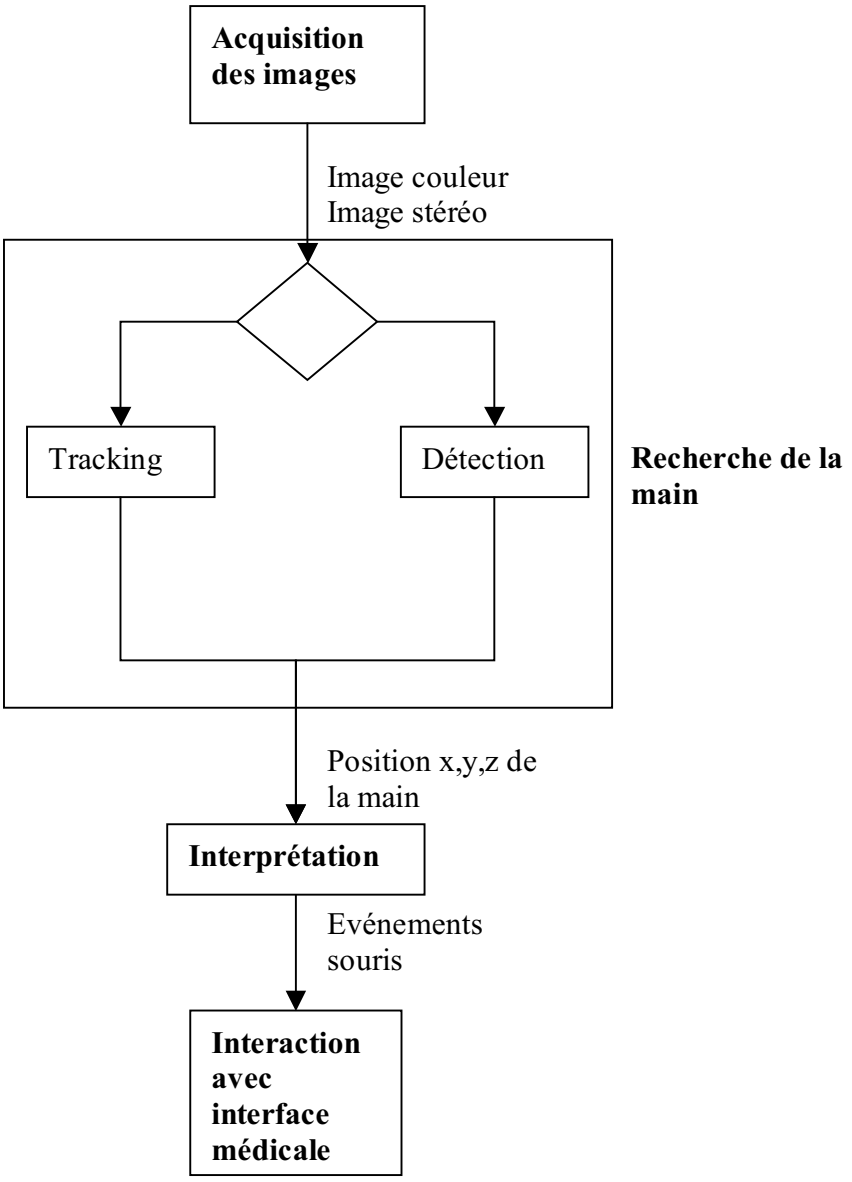


Figure 5.1 : Descriptif de la boucle principale du système développé.

Le système peut donc se diviser en 4 grandes parties correspondant à chacun des blocs principaux de l'organigramme.
Chaque partie a un rôle précis. Le système a été développé afin que le minimum d'informations soit passé entre chacun des blocs, les rendant les plus indépendants possibles, ce qui est avantageux du point de vue de la modularité et du développement futur.
Les paragraphes suivant décrivent les buts, le choix de solutions ainsi que le fonctionnement de chacune des parties du système de reconnaissance.

5.1 Initialisation du système de reconnaissance

A ce stade, une image de l'environnement est enregistrée. Cette image est considérée par la suite comme étant le décor. Celle-ci est utilisée pour l'élimination du décor. Le reste de l'initialisation est tout à fait classique.

5.2 Acquisition des images mono et stéréo

Les 2 images provenant de chacune des caméras sont transférées à l'ordinateur via le câble Firewire. Le calcul de l'image stéréographique se fait par des algorithmes déjà développés. L'image est d'une résolution de 320*240. Ce format a été choisi parce qu'il était le meilleur compromis entre temps de calcul et résolution.
Parmi les deux longueurs focales disponibles, 7.5 et 4.8 mm, le choix a été porté sur celui de 4.8 mm. En effet, la distance arrive à être calculée pour des objets plus proches avec cette lentille. Le champ de vision est grand, et la perte de résolution en profondeur ne pose pas de problème particulier pour cette application (voir § 2.1.2).
L'algorithme stéréo fournit une image dont le niveau de gris de chaque pixel est inversement proportionnelle à sa distance estimée.
L'acquisition des images mono et stéréo fournit donc deux images au bloc « recherche de la main »: une image couleur classique et une image en niveaux de gris représentant la distance.

Figure 5.2 : Les 2 images acquises par la caméra. Les pixels clairs sur l'image de gauche correspondent à des distances proches. Les trous noirs montrent les zones où la distance n'a pas réussi a être calculée.

5.3 Recherche de la main

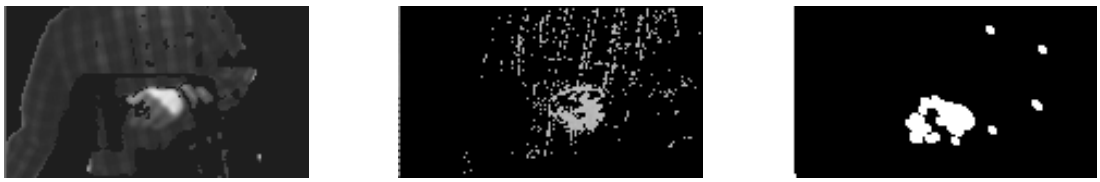
Description générale

Le but de la recherche de la main est de connaître la position (x,y,z) de la main, si une main a été trouvée.
La recherche de la main se base sur les 2 images acquises à chaque début de boucle.
L'utilisation en commun de l'information stéréo et l'information couleur a été choisie car ces deux techniques se complètent bien : la rapidité de l'information couleur est alliée à la robustesse (aux changements de luminosités) de l'information stéréo.
Le fonctionnement général est assez simple. Le système de recherche peut se trouver dans 2 modes : la détection ou le tracking. Comme cela a été dit précédemment, la détection effectue une recherche sur toute la zone de travail, alors que le tracking recherche dans une zone spécifique une main qui a été détectée précédemment.
Les techniques utilisées pour trouver la main sont différentes dans la détection et le tracking. Par contre, le traitement initial de l'image couleur est le même, la seule différence étant que le tracking effectue ce traitement d'image seulement dans la zone de recherche spécifique afin de gagner du temps de calcul.
Initialement, le système est en mode détection. Si la détection trouve un objet satisfaisant les critères de ressemblances avec la main, elle initialise un tracking sur cet objet. Le système passe donc en mode tracking. La qualité du tracking est évaluée à chaque boucle. Si celle-ci est jugée insuffisante, le système repasse en mode de détection.

Traitement d'image initial

Le traitement de l'image initiale utilise les techniques d'élimination du décor, de filtres couleur et morphologiques. L'information stéréo est utilisée plus loin.
Le décor est tout d'abord soustrait (voir figure 5.3 a), puis l'image est changée au format NRG (voir §4.1.1). Un filtre couleur est ensuite appliqué à l'image (figure 5.3 b), suivi d'une érosion et d'une dilatation. Le résultat de ce traitement (figure 5.3 c) est une image binaire dans laquelle les objets étant de la couleur recherchée, mais n'appartenant pas au décor, sont en blanc.

Une fois cette partie terminée, le système passe à des manipulations spécifiques au mode dans lequel il se trouve.



a) Soustraction du décor *b) Filtrage couleur* *c) Filtrage morphologique*

Figure 5.3 : Différentes étapes du traitement d'image initial.

5.3.1 Détection

La détection calcule plusieurs positions hypothétiques de la main. Une sélection est ensuite faite par rapport à la ressemblance de chacune de ces hypothèses avec un modèle prédéfini de la main. Un seuil de ressemblance est fixé. L'objet ayant le meilleur facteur de ressemblance, et étant au-dessus du seuil, sera considéré comme « la » main recherchée. Le système passera alors en mode tracking afin de suivre cet objet. Si aucun objet n'est au dessus du seuil, le système restera en mode détection.

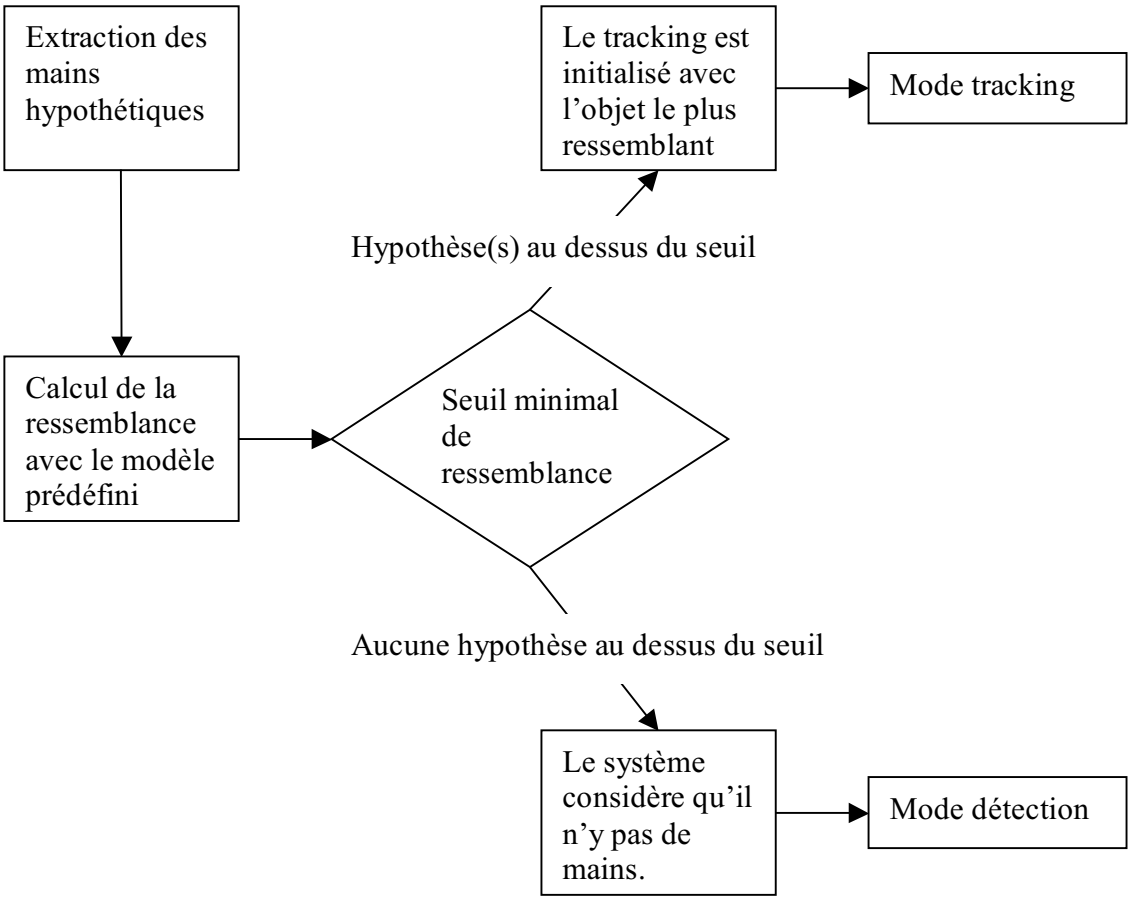


Figure 5.2 : Description du fonctionnement de la détection

L'extraction des mains hypothétiques se fait en classant les groupes de pixels qui se connectent. Une première sélection est déjà faite ici : le système ne recherche que les groupes de pixels étant plus grand qu'une certaine limite. Les blobs dus au bruit sont ainsi éliminés. Les blobs restants constituent donc les mains hypothétiques. Il s'agit alors de calculer leurs caractéristiques et de les comparer avec le modèle prédéfini de la main. Afin de calculer leurs caractéristiques, le système va calculer leur taille réelle. Celle-ci dépend de la taille en pixels sur l'écran et de leur distance les séparant de la caméra. Cette distance n'est pas obtenue de façon aussi triviale que l'on pourrait le croire :

Calcul de la distance séparant la caméra de la main

Un histogramme des valeurs de gris de l'image stéréo est construit à partir des pixels appartenant à la main. Une moyenne est calculée en utilisant les niveaux de gris correspondant à une distance de 0-3m et apparaissant fréquemment (voir figure 5.3) Cette méthode permet d'avoir une mesure stable de la distance. Elle permet de filtrer le bruit, car le bruit ne cause pas de grands pics dans l'histogramme, mais plutôt des valeurs faibles changeant aléatoirement d'emplacement. Elle permet aussi d'éviter les problèmes dus aux zones de l'image où l'algorithme stéréo n'a pas réussi à faire le lien entre les 2 images (zones sans textures).

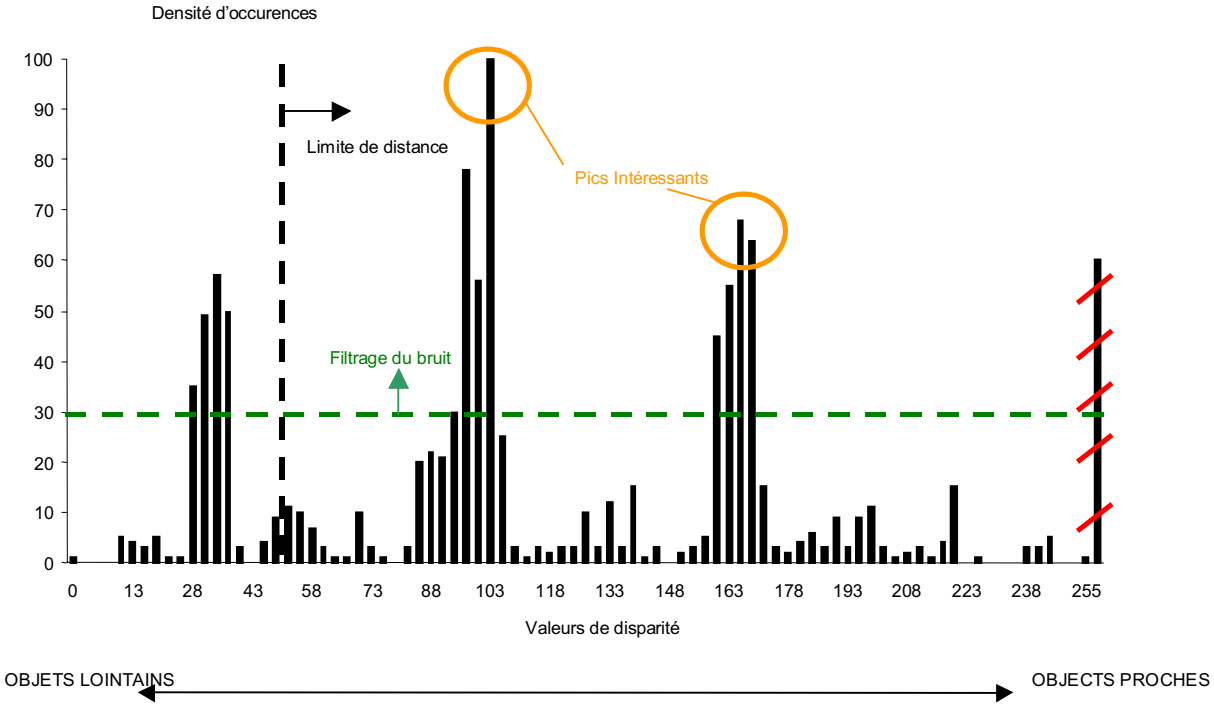


Figure 5.3 : Filtrage de l'image de disparité. Le calcul de distance se fait seulement sur le quadrant en haut à droit (par rapport aux traits tillés). Le pic en 255 n'est pas pris en compte car il représente les points où la disparité n'a pas pu être calculée.

Comparaison avec le modèle prédéfini

Une fois que la taille réelle de l'objet est calculée, une vérification est effectuée pour être sure que la main se trouve dans la zone de travail au niveau de la distance caméra - objet. Ensuite, il s'agit d'évaluer sa ressemblance avec le modèle de la main. La comparaison se fait sur 2 caractéristiques : le périmètre et l'aire. En calculant la différence avec les valeurs du modèle de la main, on obtient un facteur de ressemblance :

Ressemblance = Kp |Périmètre $_{\text{Modèle}}$ – Périmètre $_{\text{Objet}}$| + Ka |Aire $_{\text{Modèle}}$ – Aire $_{\text{Objet}}$|

Où Ka et Kp sont les pondérations de chacune des caractéristiques.

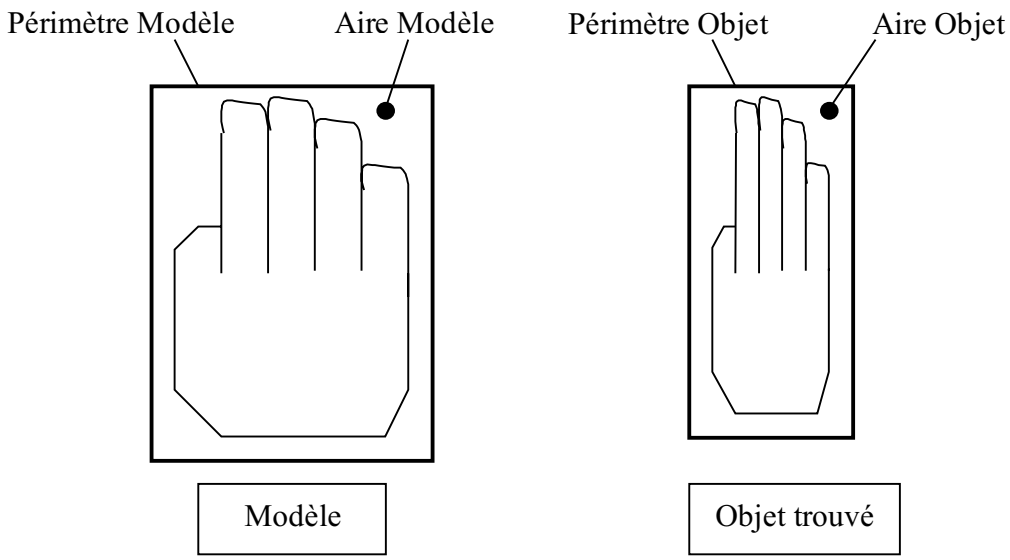


Figure 5.4 : Paramètres influençant le facteur de ressemblance

Plus l'objet sera similaire avec le modèle, plus le facteur de ressemblance sera petit. Un seuil est ensuite défini, au dessous duquel les objets sont considérées comme ayant une ressemblance acceptable avec une main. Si il y a plusieurs objets au-dessous de ce seuil, c'est celui qui a le plus petit facteur de ressemblance qui est choisi. Le résultat de la détection est présenté à la figure 5. .

Figure 5. : Résultat de la détection. La main sur la gauche de l'image (rectangle plus foncé) est considérée comme étant la main à suivre. Les autres rectangles sont des candidats potentiels.

5.3.2 Tracking

Fonctionnement général

Le tracking contient un modèle de l'objet à suivre qui a été défini lors de son initialisation. Une convolution est effectuée dans la fenêtre de recherche. L'emplacement où l'image du modèle a le plus de pixels correspondants avec l'image actuelle donne la nouvelle position de l'objet. Les positions de l'objet sont passées à un filtre de Kalman qui estime la futur position de l'objet, permettant ainsi de centrer la fenêtre de recherche suivante sur la position la plus probable de l'objet. La taille de cette fenêtre de recherche est déterminée empiriquement : trop petite et le tracking perdra souvent le suivi de la main. Trop grande et il pourrait y avoir confusion avec d'autres objets comme par exemple d'autres mains dans le champ de vision. Il ne faut pas oublier aussi qu'un des buts du tracking est de réduire le temps de calcul en travaillant sur une petite fenêtre.

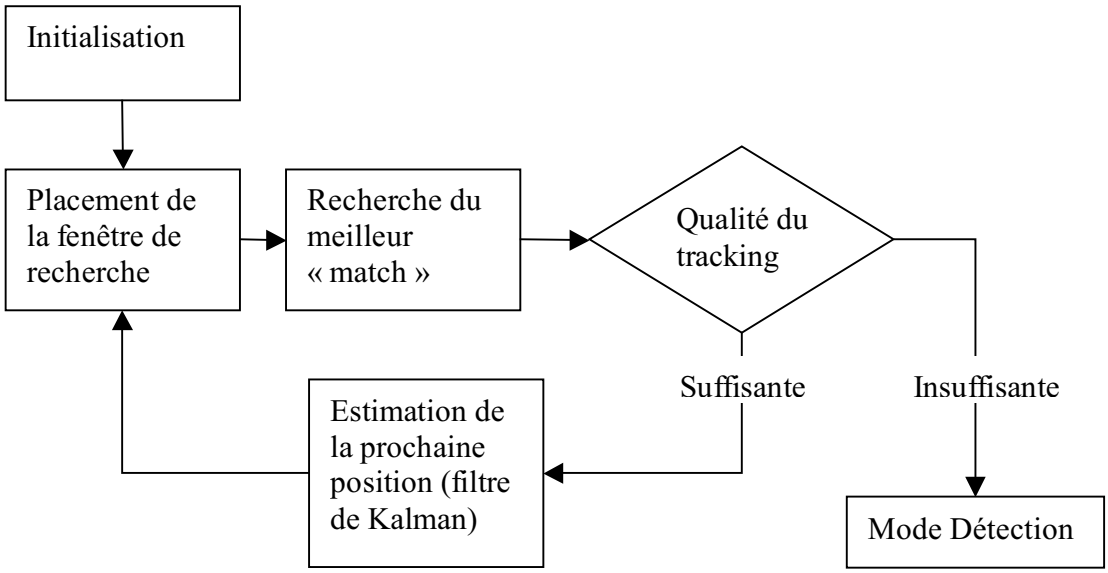


Figure 5.5 : Description du fonctionnement du tracking

Initialisation

C'est lorsque le système passe du mode de détection à celui du tracking que celui-ci est initialisé. Lors de son initialisation, le tracking enregistre l'image de l'objet à suivre. C'est cette dernière qui servira ensuite de modèle. L'initialisation définit aussi l'emplacement de départ de la fenêtre de recherche.

Evaluation de la qualité du tracking

Le défaut de ce tracking réside dans le fait qu'il n'a qu'une seule hypothèse sur la position de la main. S'il perd le suivi de la main et que celle-ci sort de la fenêtre de recherche, Le tracking ne va plus la suivre, mais il va suivre l'objet qui ressemble le plus à une main à l'intérieur de la fenêtre de recherche. Il s'agit donc de détecter quand le tracking a perdu le suivi de la main, et de refaire une détection. La main peut aussi avoir changé d'apparence. Il serait aussi utile de réinitialiser un tracking dans ces cas là.
En même temps, le fait que le tracking n'a qu'une seule hypothèse et une petite fenêtre de recherche rend le système insensible à d'autres objets dans l'environnement de travail (autres mains, têtes, etc).
Pour évaluer la qualité du tracking, on utilise le pourcentage de similitude entre le modèle et l'objet trouvé. Si ce pourcentage descend en dessous d'un seuil, un compteur est initialisé. Ce compteur continue tant que le pourcentage n'est pas repassé au-dessus du seuil. Cette méthode mesure donc le temps consécutif de mauvais tracking. Au-delà d'une limite de temps, le système repasse en mode détection. Cette technique permet de rendre le système assez flexible, donnant une chance au tracking de rattraper la main s'il l'a perdue. Le tracking continue donc même en cas d'occlusions à courtes durées.
Une évaluation similiaire est faite sur deux autres niveaux. Premièrement, un système mesure le temps depuis lequel le système a un pourcentage de similitude jugé « trop bon ». Un pourcentage élevé n'est pas toujours bon signe. Par exemple si le tracking a été initialisé avec une image de la main de profil, ce modèle arriverait, une fois la main tournée de face, à un pourcentage presque parfait de similitude. Le problème vient du fait

que ce modèle arrive à être placé à plusieurs endroits de la main vue de face, rendant le système peu stable. Ceci est détecté et le système initialise un nouveau tracking en passant en mode détection.
Deuxièmement, le temps consécutif écoulé depuis que l'objet suivi est sorti du volume de travail est mesuré. Le système continue à suivre la main pendant quelques secondes, mais si celle-ci ne revient pas dans la zone de travail, le tracking s'arrête. On évite ainsi une certaine frustration quand on travaille proches de bords du volume de travail.

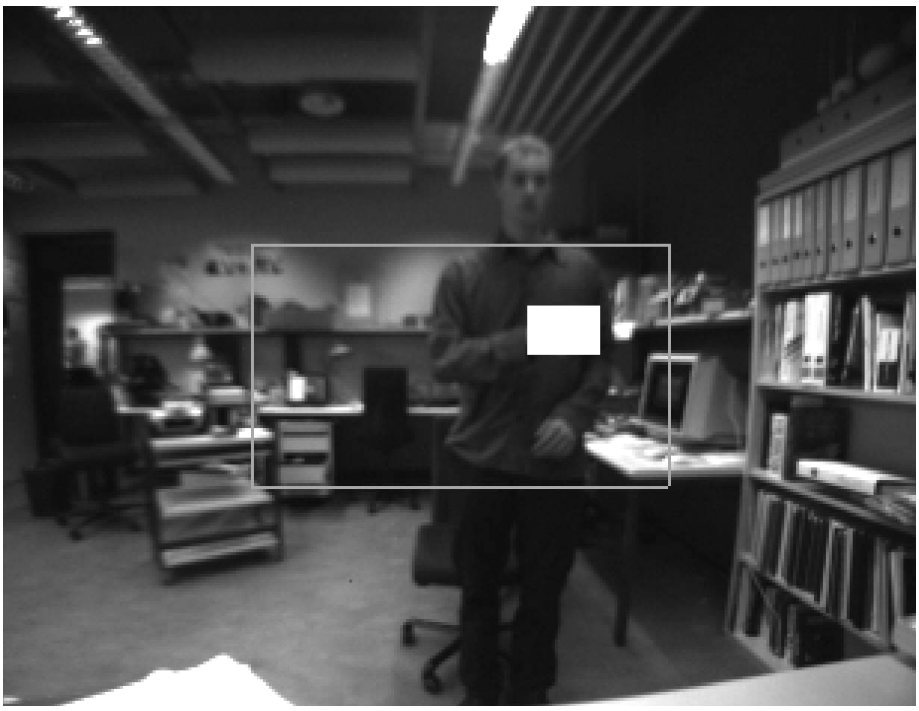


Figure 5. Résultat du tracking. On remarque que la deuxième main présente dans le volume de travail ne cause pas d'interférences.

.

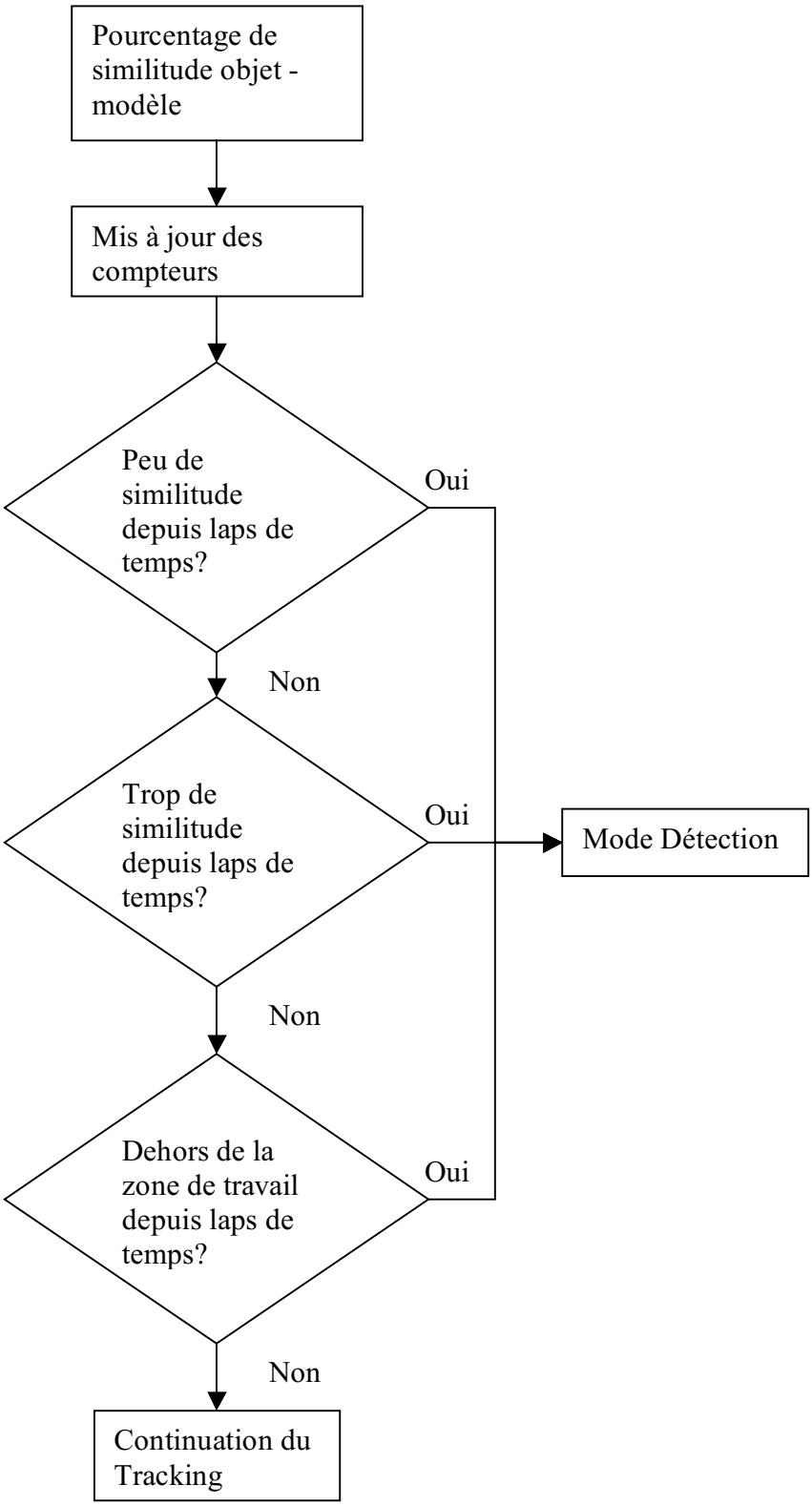

Pourcentage de similitude objet - modèle
Mis à jour des compteurs
Peu de similitude depuis laps de temps?
Oui
Non
Trop de similitude depuis laps de temps?
Oui
Mode Détection
Non
Dehors de la zone de travail depuis laps de temps?
Oui
Non
Continuation du Tracking

Figure 5.6 : Evaluation de la qualité du tracking

5.4 Interprétation

La seule information que le bloc « recherche de la main » passe au bloc « interprétation » est la position de la main. Si aucune main n'a été détectée, une valeur de référence est donnée.
Le but de l'interprétation est de donner une signification à la suite de positions qui lui ont été données.
Le choix a été fait de baser l'interprétation sur une représentation en différents états. Le système doit décider dans quel état il se trouve, et agir en conséquence sur l'interface utilisateur. Actuellement, le système a le choix entre 3 états:

- NO_HCI : L'utilisateur ne veut pas interagir avec le système.
- SCROLLING : Le système suit les positions de la main dans le volume de travail
- CLICK : L'utilisateur veut « cliquer ».

Un geste qui passe à travers le volume de travail n'est pas toujours intentionné à interagir avec l'ordinateur. Il s'agit de différencier ces gestes dits « parasites » des gestes d'interaction. Le même problème se pose lorsque l'utilisateur veut cliquer : il a positionné son curseur et veut valider cette position. La manière de valider doit pouvoir être distinguée des gestes de pointage.
Dans le problème qui nous occupe, ce moyen de faire comprendre à l'ordinateur l'intention de l'utilisateur est forcément convenu à l'avance. L'utilisateur et l'ordinateur doivent l'apprendre. Il est donc important de choisir une méthode qui soit intuitive pour l'utilisateur, qui nécessite le moins d'apprentissage possible et qui permette à l'ordinateur d'arriver à distinguer efficacement les types de gestes.
La méthode choisie consiste à garder sa main fixe dans le volume de travail durant un certain temps (actuellement une seconde). Cette méthode a été adoptée car elle permet de s'affranchir des problèmes cités ci-dessus : les gestes « parasites » sont rarement statiques, et lorsque l'on veut cliquer, on ne risque pas de perdre la position du curseur en gardant la main fixe. Evidemment, le défaut de cette technique réside dans le risque de cliquer sans le vouloir.
Un curseur spécial a été développé pour ce projet. Celui-ci est plus grand qu'un curseur classique, ce qui facilite son repérage lorsque l'on travaille à une certaine distance de l'écran.
Un feedback visuel permet à l'utilisateur de connaître l'état actuel du système (Curseur bleu : NO_HCI, rouge : SCROLLING, vert : CLICK) et aussi de voir l'état d'évolution d'un état à un autre : curseur bleu-rouge devenant plus rouge à mesure que la transition au SCROLLING s'effectue. Du rouge au vert pour la transition au CLICK (voir figure 5.7).
Lorsque l'utilisateur veut utiliser le système, il place sa main dans la zone de travail et la laisse fixe en attendant que le curseur, initialement bleu, devienne entièrement rouge. Le curseur rouge signifie que le système suit ses mouvements et dirige la souris en conséquence. S'il laisse sa main à nouveau fixe, le curseur va commencer à changer de couleur, passant du rouge au vert. Lorsque ce dernier sera entièrement vert, un click sera effectué à la zone pointée.

Une autre technique de click a été implémentée. Elle consiste à avancer son bras en direction de la caméra lorsque l'on veut cliquer. Cette technique a l'avantage de diminuer la quantité de clicks non voulus. Par contre, il est plus difficile de cliquer précisément sur une zone, car en avançant le bras, il y a le risque de bouger le curseur.

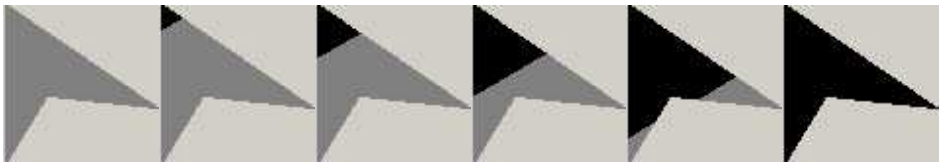


Figure 5.7 : Illustration du curseur développé et de son feedback visuel. Le curseur change progressivement de couleur, signifiant que le changement d'état approche.

La position spatiale de la main est donnée en distance pour z, et en coordonnées à l'écran pour x et y. L'interpréteur travaille dans un autre système de coordonnées. Il s'agit de faire la transformation inverse que celle vue au §2.1.1. En effet, les coordonnées à l'écran correspondent à une projection dans le plan image, mais nous cherchons la position spatiale de la main.

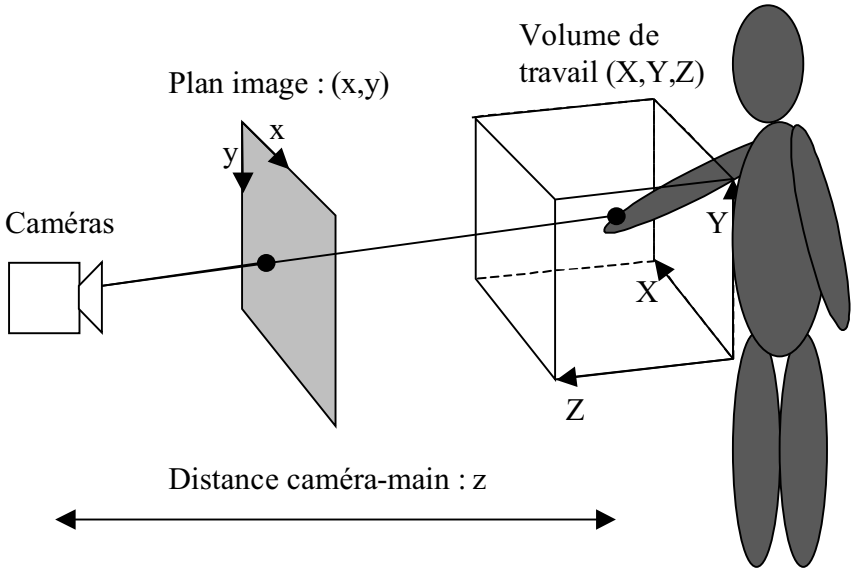


Figure 5.7 : Illustration du changement de coordonnées : (x,y,z) => (X ,Y,Z)

La position de la main est sujette à du bruit. Un filtre passe-bas réduit ce bruit, mais introduit un phénomène d'inertie. Le curseur réagit avec un certain retard aux mouvements rapides. La figure suivante décrit la manière dont le système d'interprétation choisit l'état du système :

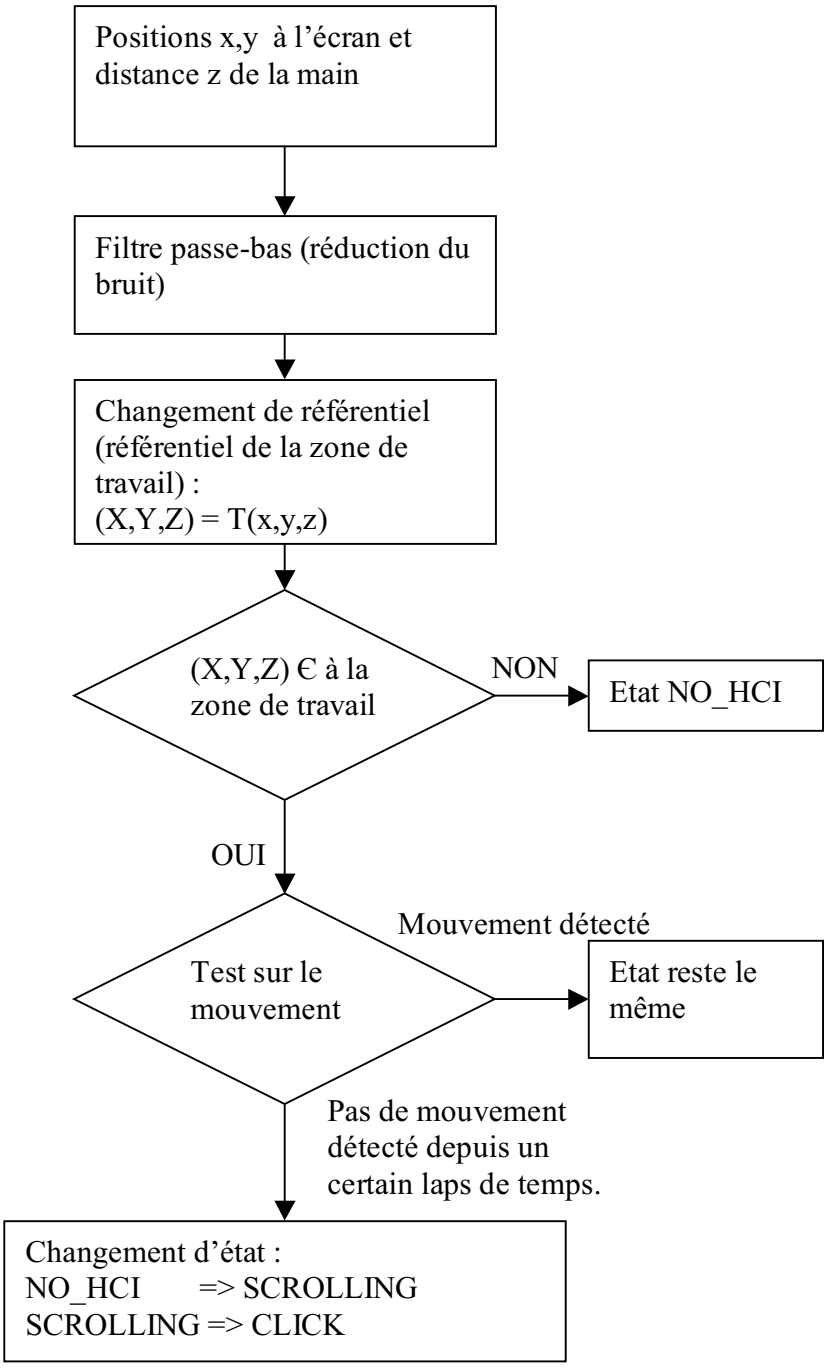


Figure 5.8 : Description du choix des états.

Une fois l'état choisi, le système d'interprétation gère l'interaction avec l'interface :

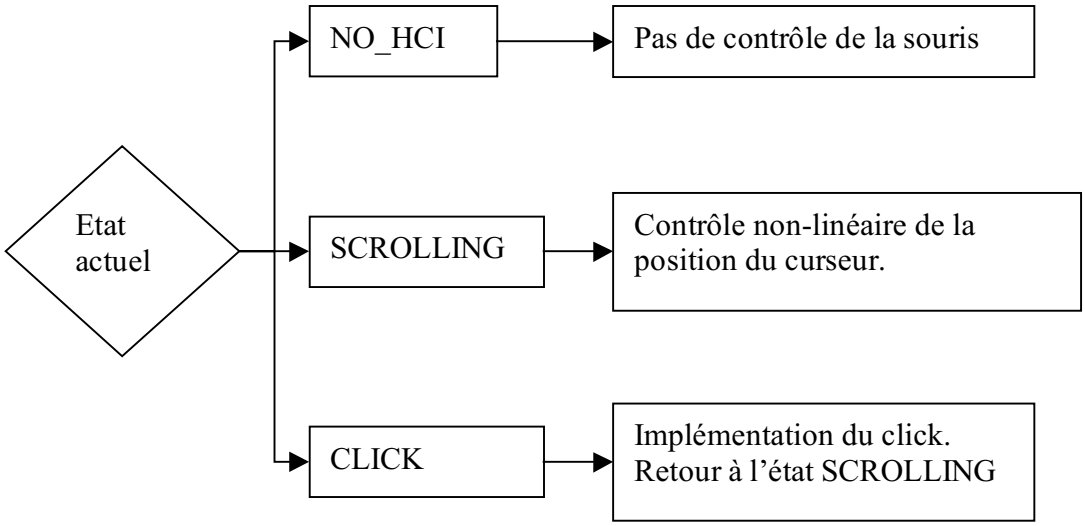


Figure 5.9 : Gestion des états

Le contrôle de la position du curseur à l'écran se fait de manière non-linéaire. Le gain entre le déplacement de la main et le déplacement du curseur à l'écran est dépendant de la vitesse de la main (voir figure 5.9). Ceci permet d'accéder à toutes les portions de l'écran, toute en restant précis dans les petits mouvements, malgré le fait que la résolution de détection de la main soit bien inférieure à la résolution du curseur à l'écran.

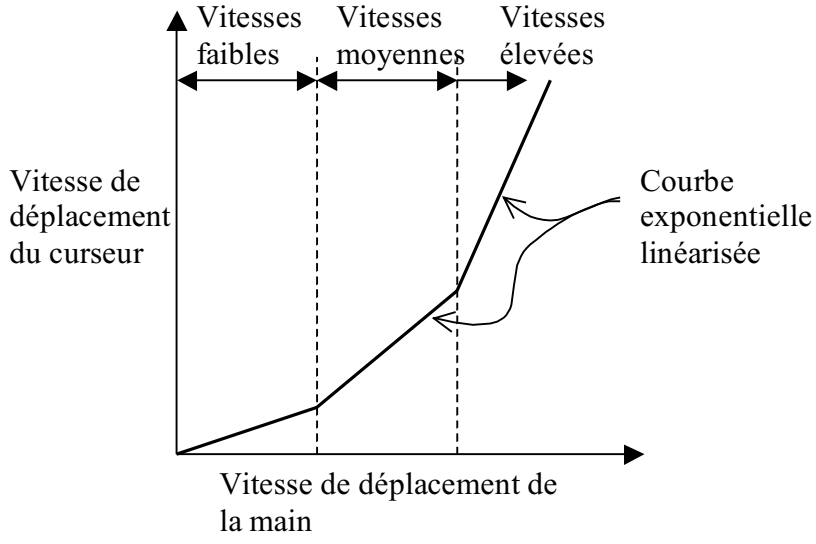


Figure 5.10 : Trois gains différents sont appliqués au déplacement du curseur, dépendant dans quelle plage de vitesses (faibles, moyennes ou élevées) la vitesse de la main se trouve.

5.5 Interface utilisateur médicale

L'interface doit permettre au médecin de régler des paramètres, tant au niveau des divers instruments cités dans le chapitre 3 qu'au niveau du système d'interaction.
L'interface qui a été développée vise une grande simplicité d'utilisation et une certaine efficacité. La navigation au sein de l'interface est basée sur des menus. Le démonstrateur montre les possibilités de contrôle offertes par le système de reconnaissance :

- Accès rapide à des données
- Utilisation d'un interrupteur digital
- Réglage analogique sur 1 dimension
- Réglage analogique sur 2 dimensions
- Réglage des paramètres du système de reconnaissance

Les figures suivantes montrent les différents menus de l'interface développée.

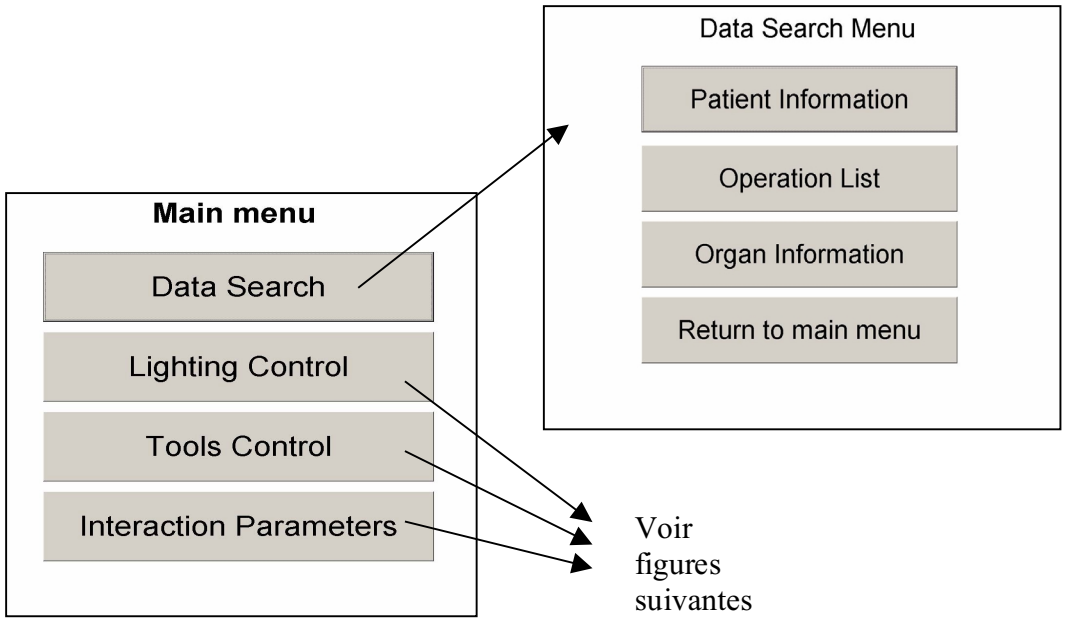


Figure 5. : Menu principal de l'interface. Les boutons sont de grande taille afin de faciliter leur pointage. Dans le menu de Data Search, des informations concernant l'opération en cours peuvent être accédées.

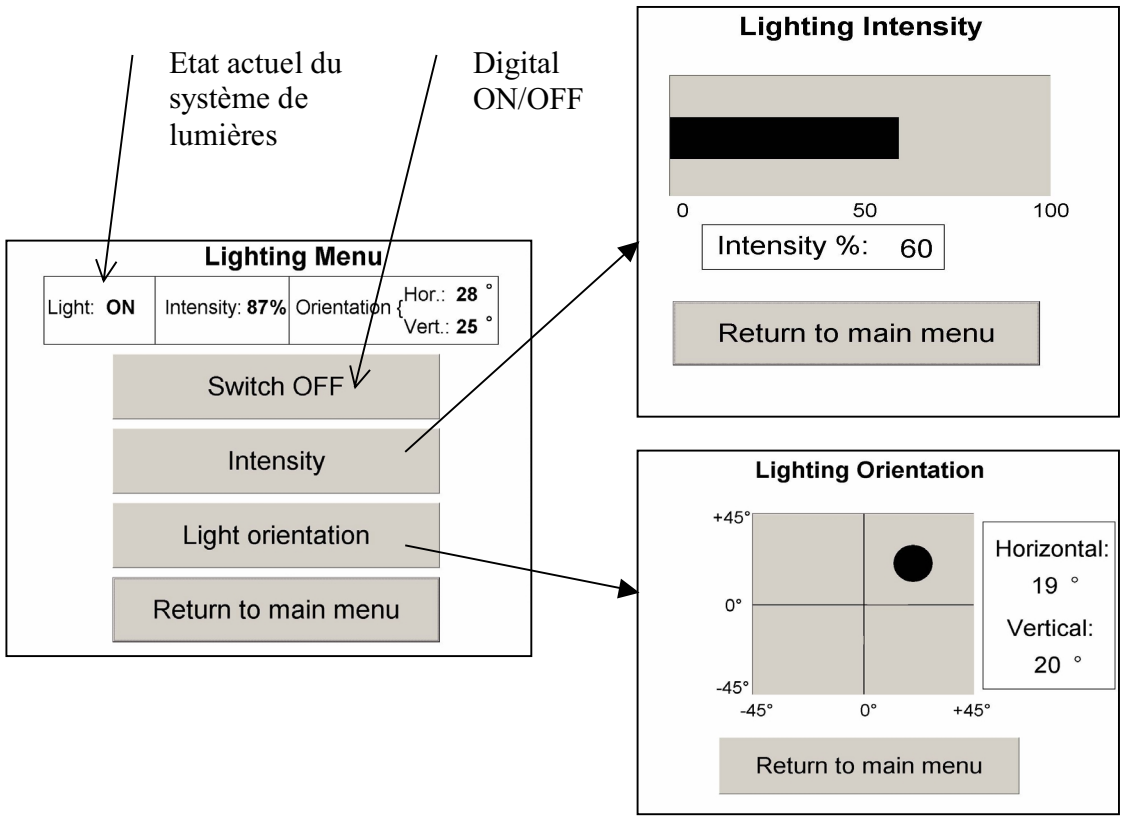


Figure 5. : Contrôle des paramètres de la luminosité. Les possibilités de contrôle digitales (switch), analogique sur 1 dimension (Intensity, 100 valeurs discrètes différentes) et analogique sur 2 dimensions (Orientation, 90 valeurs discrètes dans chaque direction) sont présentées ici. Le contrôle analogique se fait en plaçant le curseur dans la zone grise.

Les réglages des paramètres de l'interface médicale permettent dans l'état actuel du projet de:

- Reprendre la photo du décor (si la caméra a bougé).
- Changer la couleur sur laquelle le système se base pour trouver la main, en plaçant le nouvel objet de référence dans une zone précise.
- Redéfinir la zone de travail.

6 Résultats

Le système développé a été testé dans trois domaines principaux : le temps de cycle, la robustesse et la prise en main. Les deux derniers domaines ont été effectués dans le cadre d'un test utilisateur (user-study), durant lequel différentes personnes ont employé le démonstrateur afin d'effectuer des tâches prédéfinies et ont ensuite donné leur avis dessus.

6.1 Temps de cycle

Le temps de cycle n'est pas absolument constant. Les causes de sa variation sont les suivantes:

- Le mode dans lequel le système se trouve influence la vitesse d'exécution. Si le système est en mode détection, sa vitesse est plus lente qu'en mode tracking. En mode détection, la vitesse est inversement proportionnel au nombre de mains hypothétiques qui ont été trouvé dans la fenêtre de travail, alors qu'en mode tracking, le temps de cycle reste plus constant.
- Des interruptions du système d'exploitation peuvent intervenir pour diverses raisons. Ces interruptions augmentent le temps de cycle.
- La résolution minimale du système de mesure de temps est de 10ms, induisant des imprécisions.

Un échantillon de mesures de 35 valeurs a été pris afin de donner une idée sur la répartition des valeurs de temps de cycle. Les résultats sont présentés dans le tableau 6.1 et la figure 6.1.
Le temps de cycle moyen mesuré : **60.13 ms**
Ce qui donne une fréquence respective de : **16.6 Hz**
Rappelons que ces mesures ont été prises sur un Pentium 4 à 1.8 GHz, ayant 512 Mb de RAM.

	Acquisition images couleurs	Calcul Stéréo	Recherche de la main	Interprétation	Total
Unités	[ms]	[ms]	[ms]	[ms]	[ms]
Moyenne	25.03	26.78	6.81	1.50	60.13
Ecart-type	17.77	7.15	7.86	4.74	17.67
Médiane	31	31	0	0	57
Maximum	47	32	16	16	94
Minimum	0	15	0	0	31

Tableau 6.1 : Analyse statistique du temps de cycle réparti dans les 3 blocs principaux du système de reconnaissance.

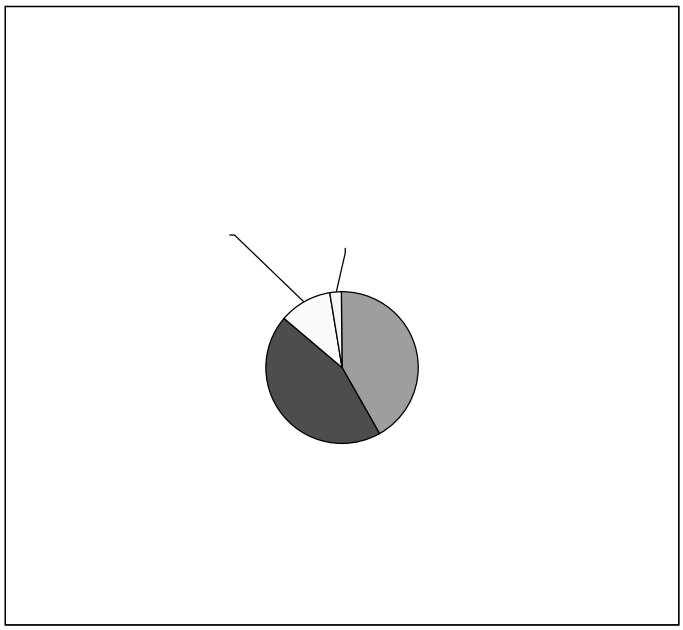


Figure 6.1 : Représentation du temps passé en moyenne dans chaque partie du système de reconnaissance durant un cycle entier.

Ces résultats montrent que le système développé fonctionne en temps réel. La grande partie du temps de calcul est utilisée par l'acquisition des images (ce qui comprend le calcul de l'image stéréo). La faible « consommation » du bloc interprétation est encourageante. Elle permet d'envisager de travailler plus en détail ce domaine sans pour autant perdre une quantité significative de temps.

6.2 Tests utilisateurs

Le système a été présenté à quinze personnes ne l'ayant jamais utilisé avant. L'expérience consistait en une petite phase de prise en main, et ensuite d'une phase durant laquelle la personne utilisait le système tout seul. Pour finir, les utilisateurs était invité à réaliser un test durant lequel ils devaient accomplir 4 opérations précises correspondant chacune à une des possibilités principales offertes par le système développé (accès d'informations, interrupteur on/off, réglage analogique sur 1 dimension et réglage analogique sur 2 dimensions.
Pour caractériser l'efficacité du système, le temps pris pour chacune de ces opérations a été mesuré. En effet, il est dur de mesurer un pourcentage de réussite dans un système dynamique, car dans la grande majorité du temps l'utilisateur arrivera finalement à accomplir la tâche requise. Mais si le système a perdu plusieurs fois le suivi, ou si l'utilisateur n'a pas réussi à cliquer au bon endroit, le temps mis pour accomplir l'opération aura augmenté en conséquence.
On pourrait donc croire que le temps est une mesure indirecte de l'efficacité du système, mais on serait en train d'oublier ce que l'utilisateur veut vraiment. Celui-ci n'est pas du

tout intéressé par le fait que le système arrive à suivre sa main avec un certain pourcentage de réussite. Ce qu'il cherche avant tout, c'est une interaction avec l'ordinateur qui soit intuitive, sans frustrations et rapide. Les deux premiers points sont abordés dans le questionnaire et la rapidité a pu être évaluée en chronométrant les personnes ayant essayé le système.

6.2.1 Résultats quantitatifs

Le test se divisait en 4 opérations:

1) Accéder la page d'informations sur le patient.
2) Allumer la lampe virtuelle.
3) Régler l'intensité de la lampe virtuelle précisément à 80%.
4) Régler l'orientation de la lampe virtuelle sur une valeur choisie à l'avance par l'utilisateur.

L'expérience a porté sur quinze personnes. Les temps ont été mesurés 2-3 minutes après la prise en main initiale des utilisateurs.

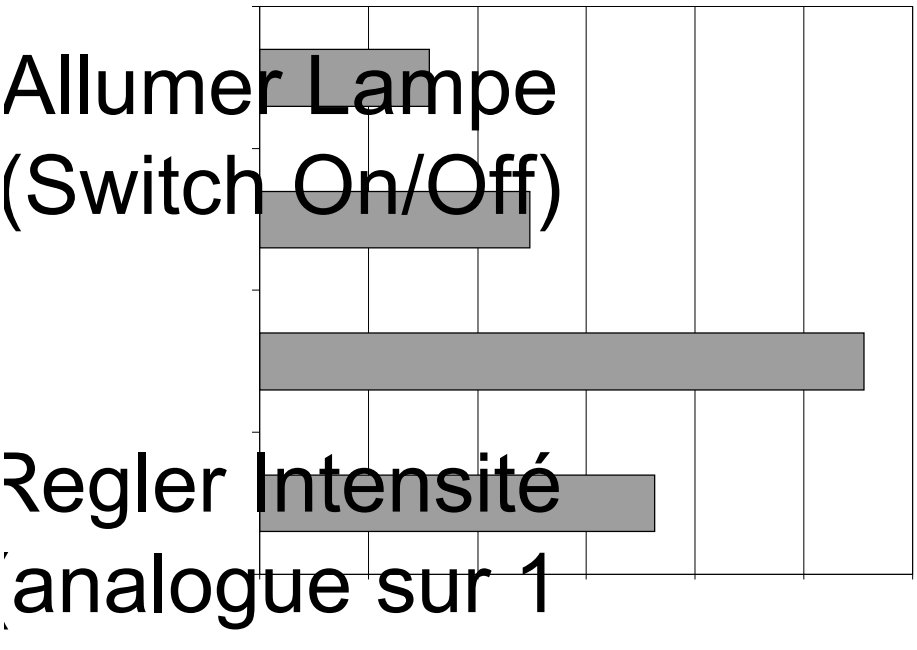


Figure 6.2: Temps moyen mis par les utilisateurs pour effectuer les 4 opérations.

Remarques:

- La navigation est relativement rapide, ainsi que l'interrupteur On/Off. Par contre, les réglages analogiques nécessitent plus de temps. Ceci est du en partie par le fait

que ces réglages nécessitent deux validations supplémentaires (une pour entrer dans la fenêtre de contrôle, une pour confirmer la valeur trouvée).

- Le temps de réglage analogique sur deux dimensions est sensiblement plus court au réglage analogique sur une dimension car la précision de réglage n'était pas aussi enforcée.
- L'expérience des utilisateurs était ici nulle. Avec plus d'expériences, ces temps deviennent sensiblement plus courts. Les temps sont approximativement 50% plus court lorsque j'utilise le système.

Remarques sur la fiabilité générale du système

Le système dans son ensemble s'est montré assez fiable, sans pour autant être proche de la fiabilité des périphériques d'ordinateurs classiques.
Il arrive parfois que l'utilisateur n'arrive pas précisément à faire ce qu'il veut (typiquement dans les réglages analogiques), ou pire, que le système prenne des décisions que l'utilisateur n'a pas du tout ordonné. Ces erreurs sont néanmoins rares si le système est utilisé correctement.
En effet, les erreurs proviennent principalement du fait que le système initialise le suivi avec autre chose que la main. Les objets que le système détecte faussement sont habituellement la tête, ou encore une autre main présente dans le volume de travail (notons que dans ce cas, ce n'est pas entièrement de la faute au système de reconnaissance puisqu'il a été conçu pour détecter une main). Il arrive parfois aussi que l'habillage de l'utilisateur soit de couleur similaire à la peau humaine. Dans ce cas, le système a beaucoup plus de peine à suivre correctement la main. Les essais avec une boule de couleur unique et un filtrage adapté montre que ces erreurs disparaissent presque entièrement. Ces erreurs sont d'ailleurs principalement du à un mauvais réglage du volume de travail (les mains d'autres personnes ou la tête de l'utilisateur sont présentes dans le volume de travail, etc.).
Dans le cas où le système a détecté l'objet correctement, le système devient alors très stable, surtout au niveau de la navigation (grands boutons). Une étude sur la précision de pointage au niveau des interactions analogiques n'a pas été faite. En effet, celle-ci aurait été assez complexe puisqu'il aurait fallut la rendre le plus indépendant possible des différentes conditions dans lesquelles les tests ont été effectués. Cette étude aurait donc pris trop de temps par rapport au reste, mais sa réalisation n'est pas à exclure des travaux futurs (voir §7). Notons néanmoins que 80% des utilisateurs arrivait à régler l'intensité lumineuse à environ 3 pourcent près lors de leur premier essai.
Une remarque doit être faite sur la technique de cliquage. Comme cela a été décrit au §5.4, deux techniques de click ont été implémentées, une consistant à laisser sa main fixe dans le volume de travail durant une seconde, l'autre consistant à avancer son bras en direction de la caméra.
La technique de l'attente a l'avantage d'être plus précise, le pointeur ne risquant pas de bouger comme c'est le cas lorsque le bras est avancé en direction de la caméra. Par contre, la quantité de click non voulus est beaucoup plus élevée. En effet, si le système commence son suivi sur un autre objet que la main de l'utilisateur, il y a une grande chance que cet objet reste fixe. Il causera donc une série de click non voulu, induisant une certaine confusion chez l'utilisateur.

L'autre technique n'a pas ce problème, mais un plus grand temps d'adaptation est nécessaire pour l'utiliser.

6.2.2 Résultats qualitatifs

Au terme de leurs essais, les utilisateurs étaient invités à évaluer le système de reconnaissance. Le questionnaire (voir annexe) était constitué de quatre questions, suivi d'une place pour les remarques éventuelles.
Les résultats sont présentés dans la figure 6.3.

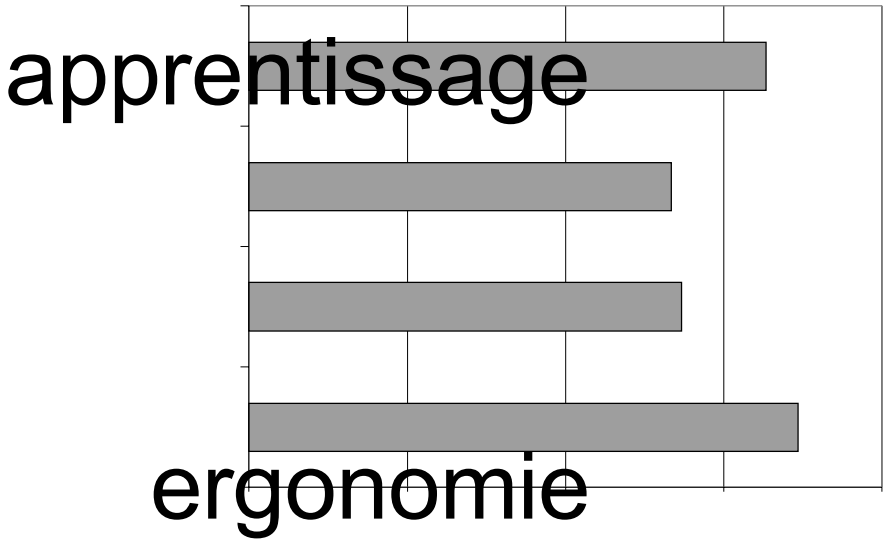


Figure 6.3: Résultats du sondage effectué.

Synthèses des remarques faites par les utilisateurs:

- Le curseur devrait être plus stable
- Le système devrait mieux réagir aux mouvements rapides
- Il manque un feedback visuel pour connaître l'emplacement de sa main dans la zone de travail. L'utilisateur a tendance à penser que l'emplacement de la souris à l'écran correspond à la position de la main dans la zone de travail, ce qui porte à confusion.

Remarques sur l'évaluation:

- On voit que le point faible du système réside dans son efficacité à changer les paramètres de façon rapide. Le problème vient principalement du click et des limites du volume de travail. La source d'insatisfaction avec l'ergonomie provenait aussi de l'absence d'un feedback visuel.
- Les personnes ayant testé le système ne sont pas tous des spécialistes du monde médical. La qualification du potentiel médical n'indique donc pas l'avis de chirurgiens, mais plutôt un avis général.

7 Travaux Futurs

Le système actuel peut bien sûr être amélioré. L'étude médicale qui a été menée au cours de ce projet n'était qu'une recherche initiale, et nous sommes encore loin d'avoir un système prêt à être implémenté en salle d'opération. Une étude plus approfondie est nécessaire. Grâce au démonstrateur développé, les médecins comprendront mieux les possibilités et aussi certaines limites que présente un système d'interaction visuelle. En effet, lors des discussions, les chirurgiens, afin de répondre à une question, nous demandaient souvent ce que nous avions à leur proposer, quelles étaient les possibilités réelles.
Une recherche plus poussée sur l'environnement médical (la salle d'opération) doit aussi être établie. Pour commencer, l'analyse de photos numériques prises durant l'opération serait utile. Ensuite, une opération pourrait être filmée avec la caméra stéréoscopique utilisée. Finalement, un prototype pourrait être testé avec des chirurgiens en salle d'opération.
Ces travaux devraient permettre de mieux cibler la poursuite du travail. Dans son état actuel, le système analyse les mouvements de façon très simple. La partie interprétation du système peut certainement être rendue plus complète en fonction des besoins. Les HMM, candidats idéaux pour une interprétation générale de gestes, n'ont malheureusement pas pu être employé, faute de temps. Ceux-ci ont néanmoins un grand potentiel, et leur implémentation est une suite logique du projet actuel.
En parallèle à ces développements, le système peut aussi être amélioré plus spécifiquement sur les points suivants si nécessaire:

- Recherche, à partir de la disparité, d'un corps humain. Le volume de travail pourrait bouger avec le chirurgien. Cette technique a le potentiel d'augmenter la robustesse du système puisqu'elle permet de vérifier si la main trouvée appartient vraiment au chirurgien.
- Système de navigation utilisant l'information de profondeur pour avancer dans les menus. L'utilisateur naviguerait en quelque sorte dans un menu en 3 dimensions ou le passage de portes ou fenêtres accéderait à des nouvelles possibilités. Cette méthode serait plus rapide qu'un click.
- L'emploi de gestes des doigts pour contrôle le système doit être étudiée. L'utilisation des doigts n'est pas compatible avec les buts fixés au §3.5, puisque le système ne pourrait plus être utilisé avec des outils dans la main. Pourtant la quantité d'informations supplémentaires qui pourraient en être tiré doit être évaluée, surtout si l'application changerait d'objectif.
- Divers caractéristiques du système peuvent être améliorées: stabilité du positionnement, insensibilité à la luminosité, etc.
- Utilisation d'informations externes: fusion de donnée (état de l'interface, position des outils, etc)
- Feedback sonore.

8 Conclusion

Ce projet a montré qu'il existe véritablement un intérêt médical pour un moyen d'interaction entre le chirurgien et l'ordinateur. Il est difficile d'imaginer comment cet intérêt diminuerait dans le futur, puisque les salles d'opérations deviendront, comme le reste de notre environnement, de plus en plus informatisées.
Le moyen utilisé dans ce projet, la vision, n'est qu'une des possibilités. D'autres études sont en courts sur des touch-screen stériles, ou encore l'utilisation de la reconnaissance vocale. Ces solutions ont des avantages les unes par rapport aux autre qui pourraient être mis en commun dans un système d'interaction robuste aux applications pas nécessairement médicales.
L'analyse du système de vision a permis de se donner une idée précise sur les performances de la caméra. Cette étude a été utile au cours de ce projet, et le restera pour tout autre travail utilisant ce système de vision.
Le système développé présente des résultats encourageants: fonctionnement en "temps réel", facilité d'apprentissage et d'utilisation, robustesse élevée. Les faiblesses du système, telles qu'une efficacité pas toujours optimale, peuvent encore être amélioré. De plus, la technique utilisée consistant à reconstruire la position spatiale des objets détectés permet d'avoir une zone de travail précisément définie dans l'espace. Ceci convient bien aux salles d'opérations où certains espaces sont strictement restreints au chirurgien.
Ce projet en explorant en détail la solution de la vision stéréoscopique, a permis le développement d'une bonne base de discussion pour des études futures dans le domaine de l'interaction chirurgien-ordinateur. Une comparaison médicale directe des techniques à disposition peut maintenant être effectuée afin d'en faire ressortir les solutions qui seront maintenues pour l'élaboration d'un système fonctionnel en salle d'opération.

Remerciements:
Je tiens à remercier Sébastien Grange et Terry Fong du VRAI Group pour leur aide précieuse dans le développement du système de reconnaissance, Dr Rippstein, Dr Zambelli et Dr Caversaccio pour leur grande coopération dans la recherche médicale, ainsi que toutes les personnes ayant pris part dans les diverses phases de ce projet.

Etudiant:

Chauncey Graetzel

Ecublens, le 10 février 2003

9 Références

[1] Grange S., "Vision-based Human Computer Interaction for Medical Applications", PhD Proposal, EPFL, November 2002.

[2] Konolige K., "Small Vision Systems: Hardware and Implementation", Artificial Intelligence Center, SRI International, 1997

[3] Videre Design, "STH-MD1/-C Stereo Head User's Manual", 2001

[4] Bayer Pattern description, <*www.ise.stanford.edu/class/psych221/99/dixiedog/vision.htm*>

[5] Siegwart R. "Autonomous Mobile Robots", cours EPFL, 2002

[6] Ryser P. "L'ingénieur dans la R&D industriel", cours EPFL, 2002

[7] Caversaccio M. et al., "The Bernese Frameless Optical Computer Surgery System", Clinical paper, University of Bern, 1999

[8] J. Kender, "Saturation, Hue, and Normalized Color: Calculation, Digitization Effects, and Use", Technical Report, Carnegie Mellon University, 1976

[9] Von Hardenberg C. et al, "Bare-Hand Human-Computer Interaction", TU Berlin, 2001

[10] Dey S., "Système de reconnaissance de postures de main" Technical Report, EPFL, June 2002.

[11] Turk M., Jojic N.and Huang T. "Tracking Self-Occluding Articulated Objects in Dense Disparity Maps" University of Illinois, Microsoft Research, 1999.

[12] Rabiner L., "A Tutorial on Hidden Markov Models and Selected Applications in Speech Recognition" Proceedings of the IEEE, Vol. 77, NO.2, February 1989.

[13] James P. Mammen, Subhasis Chaudhuri et Tushar Agrawal, "A Two-stage Scheme for Dynamic Hand Gesture Recognition", Indian Institute of Technology, Bombay, 2002

[14] Pavlovic V, Sharma R. and Huang T. "Visual Interpretation of Hand Gestures for Human-Computer Interaction: A Review", IEEE Transactions on Pattern Analysis and Machine Intelligence, Volume 19, No 7, pages 677-695, July 1997

[15] Shamaie A. and Sutherland A., "Graph-Based Matching of Occluded Hand Gestures", Dublin City University, 2001

[16] Iannizzotto G, Villari M, Vita L, "Hand-tracking for human-computer interaction with Graylevel VisualGlove: Turning bakc to the simple way", University of Messina, 2001

[17] Starner T. and Pentland A.,"Virtual Recognition of American Sign Language Using Hidden Markov Models", Massachusetts Institute of Technology, 1995

[18] Visarius H et al. "Man-Machine Interfaces in Computer Assisted Surgery" Journal of Computer Aided Surgery, Vol2, No. 2, pp. 102-107, 1997

[19] Medical Touchscreens, <*http://www.ftgdata.com/touchscreens/touchscreen-medical-market.html*>

10 Annexes

10.1 Questionnaire envoyé aux chirurgiens

I am working on a project that consists of developping a dynamic hand gesture recognition system that is applicable in a medical environment. It should allow the surgeon to interact with a computer during an operation without the use of a keyboard or a mouse.
The system is based on a camera capturing in real time the surgeons hand movements inside a predefined workspace, analyzing his gestures and interpreting them as a specific command. The camera is stereoscopic, giving it the capability to perceive depth.

Here are a few questions that are of interest for me:
I don't expect you to answer all the questions, but any kind of information would help me. Thank you.

1) What systems would be interesting to control (or are already controlled) by a computer (for example a tool, a camera or the lighting) ?
Note that the computer could replace both orders given to personnel and some of the pedals used to interact.

2) Among these systems, what kind of control would you need:

-digital (on/off)
-analog (multiple levels), what approximate resolution (how many levels)?
-analog on several dimensions (for example x,y and z positionning)
-other

3) The camera must be placed as close as possible (1-3m) to the workspace for the system to work correctly.
Where would it be the less intrusive? On the ceiling, on a pedestral, etc.

4) Types of gesture:

4a) Are you usually holding tools in your hands when you need to interact with the systems listed in question 1 ?
Could you free one hand when you want to make the gestures?

4b) Is it possible to have a particular zone, apart from the operation area itself, where you could perform the interacting gestures?

4c) In approximately what volume is it possible for you to perform the gestures?

4d) Knowing that the vision system, placed at a distance of 3m, cannot detect

changes on your hand position of about 3cm in depth and 1 cm sidewise, what gestures would be preferable for commanding the systems?

-linear: up/down, left/right, turn clockwise, counterclockwise
-mimetic: e.g. making a cross in the air would mean "cancel", opening your hand would mean "light on"
-deictic: following the path of your fingertip (mouse like command)
-small movements, like shaking a finger.
-other

4f) Is the color of your gloves usually significantly different from the color of the rest of your clothing?

4e) Would it be problematic if you had to wear gloves that would have a small color patch on the inside and the outside of the palm?
Would this patch get occluded during operation by blood, for example?

10.2 Questionnaire fourni pour le user-study

Evaluation du système de reconnaissance

	Insuffisante		Suffisante		Excellente
1) La phase d'apprentissage d'utilisation du système doit être rapide. Qualifier la facilité d'apprentissage.	1	2	3	4	5
2) Le système doit pouvoir être utilisé de façon intuitive, sans frustrations. Qualifier l'ergonomie.	1	2	3	4	5
3) Le système doit pouvoir accéder et changer rapidement les paramètres médicaux. Qualifier l'efficacité.	1	2	3	4	5
4) Le but à long terme est d'implémenter cette application en salle d'opération réelle. Qualifier le potentiel médical.	1	2	3	4	5

5) Toutes remarques sont les bienvenues :