

**Titre:** Interface humain-machine de détection visuelle de la posture de la  
Title: main

**Auteur:** Fabienne Lathuilière  
Author:

**Date:** 1999

**Type:** Mémoire ou thèse / Dissertation or Thesis

**Référence:** Lathuilière, F. (1999). Interface humain-machine de détection visuelle de la  
Citation: posture de la main [Mémoire de maîtrise, École Polytechnique de Montréal].  
PolyPublie. <https://publications.polymtl.ca/8663/>

 **Document en libre accès dans PolyPublie**  
Open Access document in PolyPublie

**URL de PolyPublie:** <https://publications.polymtl.ca/8663/>  
PolyPublie URL:

**Directeurs de  
recherche:** Jean-Yves Hervé  
Advisors:

**Programme:** Non spécifié  
Program:

UNIVERSITÉ DE MONTRÉAL

INTERFACE HUMAIN-MACHINE DE DÉTECTION  
VISUELLE DE LA POSTURE DE LA MAIN

FABIENNE LATHUILLIÈRE

DÉPARTEMENT DE GÉNIE ÉLECTRIQUE ET DE GÉNIE INFORMATIQUE  
ÉCOLE POLYTECHNIQUE DE MONTRÉAL

MÉMOIRE PRÉSENTÉ EN VUE DE L'OBTENTION  
DU DIPLÔME DE MAÎTRISE ÈS SCIENCES APPLIQUÉES (M.Sc.A.)  
(GÉNIE ÉLECTRIQUE)

Décembre 1999



National Library  
of Canada

Acquisitions and  
Bibliographic Services

395 Wellington Street  
Ottawa ON K1A 0N4  
Canada

Bibliothèque nationale  
du Canada

Acquisitions et  
services bibliographiques

395, rue Wellington  
Ottawa ON K1A 0N4  
Canada

*Your file* *Votre référence*

*Our file* *Notre référence*

The author has granted a non-exclusive licence allowing the National Library of Canada to reproduce, loan, distribute or sell copies of this thesis in microform, paper or electronic formats.

The author retains ownership of the copyright in this thesis. Neither the thesis nor substantial extracts from it may be printed or otherwise reproduced without the author's permission.

L'auteur a accordé une licence non exclusive permettant à la Bibliothèque nationale du Canada de reproduire, prêter, distribuer ou vendre des copies de cette thèse sous la forme de microfiche/film, de reproduction sur papier ou sur format électronique.

L'auteur conserve la propriété du droit d'auteur qui protège cette thèse. Ni la thèse ni des extraits substantiels de celle-ci ne doivent être imprimés ou autrement reproduits sans son autorisation.

0-612-53586-X

Canada

UNIVERSITÉ DE MONTRÉAL

ÉCOLE POLYTECHNIQUE DE MONTRÉAL

Ce mémoire intitulé:

INTERFACE HUMAIN-MACHINE DE DÉTECTION  
VISUELLE DE LA POSTURE DE LA MAIN

présenté par: LATHUILIÈRE Fabienne,

en vue de l'obtention du diplôme de: Maîtrise ès sciences appliquées

a été dûment accepté par le jury d'examen constitué de:

M. THISTLE John, Ph.D., président

M. HERVÉ Jean-Yves, Ph.D., membre et directeur de recherche

M. DUDEK Gregory, Ph.D., membre

Il lui écrivait des vers sur la neige,  
Hélas, les neiges fondent.  
Mais alors la neige tombait encore  
Et on était libre d'écrire sur la neige.

*Vladimir Vissotski*

# Remerciements

Je tiens en premier lieu à remercier chaleureusement mon directeur de recherche Jean-Yves Hervé qui a su me guider et me motiver tout au long de mon projet et sans la confiance ni l'énergie de qui le résultat n'aurait été le même. Je tiens également à remercier le CRSNG pour son soutien financier.

J'adresse pour terminer une pensée amicale à tous les autres membres du Groupe de Recherche en Perception et Robotique de l'École Polytechnique de Montréal, étudiants, assistants de recherche et professeurs pour leurs conseils, leurs idées et les fructueuses discussions qui s'en sont suivies.

## Résumé

Ce projet s'inscrit dans le domaine de l'Interface Humain-Machine et aborde plus particulièrement la reconnaissance de la posture de la main en mouvement dans le but d'exploiter l'information de dextérité gestuelle propre à l'être humain. L'approche choisie privilégie un système de détection visuelle de la main, robuste et capable de fonctionner en temps réel.

Une unique caméra couleur dont les caractéristiques intrinsèques sont connues est utilisée pour filmer le mouvement de la main pendant les phases d'approche et de saisie d'un objet par un opérateur. La main est revêtue d'un gant muni d'indices de couleur caractéristique placés aux extrémités des doigts ainsi que sur la paume dorsale. La cinématique de la main est modélisée à l'aide d'un modèle squelette à 26 degrés de liberté incluant la position et l'orientation de la paume ainsi que quatre degrés de liberté pour chaque doigt (trois degrés en flexion et un degré en abduction). L'extraction des positions des indices de couleur dans l'image permet de calculer tout d'abord la position et l'orientation de la paume dans le référentiel de la caméra puis de reconstruire la valeur des différentes variables d'articulation des doigts par com-

binaison d'heuristiques et de techniques de cinématique inverse. Il peut arriver que l'extrémité d'un doigt ne soit pas détectée par la caméra si elle est masquée par la paume, un autre doigt ou bien encore par les propres phalanges du doigt. Dans ce cas, un algorithme de prise en compte des occlusions est appliqué, basé sur la prédiction de la position des doigts non visibles et sur la validation des occlusions à l'aide d'une méthode de visibilité géométrique. Une fois la configuration de la main obtenue, une réplique graphique de la main développée à l'aide de la librairie graphique OpenGL peut être affichée afin d'apprécier visuellement le degré de fidélité du modèle reconstruit par rapport à la main originale. Les résultats obtenus sur des séquences de synthèse puis sur des séquences réelles de gestes ont permis de valider la méthode de reconstruction de la posture de la main, et ce, même en présence d'occlusions.

Le processus de détection des indices et de reconstruction des variables d'articulation de la main a été ensuite optimisé afin de constituer une interface humain-machine permettant à l'utilisateur d'effectuer un geste tout en visualisant en temps réel la réplique graphique de la main à l'écran. Ce dispositif présente l'avantage de permettre d'enregistrer plusieurs séquences de gestes afin d'en faire une étude détaillée. Nous nous sommes alors intéressés à l'étude de gestes de préhension. Dans le cadre d'une saisie d'objets par l'utilisateur, plusieurs types de prises ont ainsi été analysées et des caractéristiques propres à chacune ont été extraites, offrant par conséquent des critères de distinction des différentes prises.

La suite de nos travaux propose l'exploitation de l'information issue de la réalisation



du geste par l'être humain pour faire effectuer à une pince robotique la saisie d'un objet similaire par téléopération. Le contrôle de la pince est basé sur un asservissement de type maître-esclave où le robot, situé dans un espace de travail éloigné, est soumis à la réalisation de la tâche par la main. Le contrôle du robot bénéficie alors de l'information sur la trajectoire sélectionnée ainsi que sur la configuration des doigts de la main lors de la saisie de l'objet. On suppose calibrée la caméra filmant la main de l'opérateur et connues la position et l'orientation de l'objet à saisir par rapport à la caméra, de sorte qu'à chaque instant, les postures de la main sont déterminées par rapport à l'objet. Il est ensuite possible de contrôler le manipulateur en position en lui donnant comme commandes les positions successives de la main par rapport à l'objet que, lui, doit saisir. Le geste effectué par l'opérateur peut être reconnu grâce aux critères de reconnaissance des différentes prises, ce qui permet d'adapter la configuration de la pince du robot au type de geste sollicité. Des expériences en simulation ont prouvé l'intérêt du système de suivi de la posture de la main pour contrôler un robot manipulateur.

# Abstract

Hand gestures constitute a powerful means of interaction encompassing manipulative gestures as well as communicative ones. Consequently, several techniques have been developed in order to capture hand gestures, in particular non-intrusive ones such as vision-based hand detection systems. We propose a real-time visual hand tracking and posture estimation system to guide a robotic arm in gripping gestures. We have opted for an easy-to-use and robust human-computer interface. Our approach has been chosen so as to track hand posture in an image sequence in real time with a single camera detecting color cues on the hand. We present an original twenty-six degree-of-freedom kinematic model of the hand, for which forward and inverse kinematics formulations have been developed. The location of the hand is given by the wrist's middle point and its orientation is given by that of the palm. Each finger has four degrees of freedom, namely one in abduction/adduction and three in flexion/extension. The hand is covered with a dark glove marked with colored cues on the upper palm and on each fingertip. The position of the color markers in the image is used to reconstruct first the pose of the palm and then all the joint angles of the

fingers. Tip occlusions (of a fingertip by the palm, by another finger, or by its own phalanxes) are handled by predicting the finger positions, which are then validated by testing 3D geometric visibility conditions. The overall good performance of the hand posture reconstruction is validated on a graphic hand model developed with the OpenGL graphic library. Experiments carried out on graphical and real image sequences have led to meaningful 3D hand configurations.

The detection system and the reconstruction procedure have been optimized in order to react in real time to the operator's gestures while displaying the graphic hand model. This interface allows different gripping gestures to be recorded in order to analyze their characteristics. We have studied several different grasps and we have extracted meaningful features among the hand's joint angles in order to classify the gestures. It is thus possible to recognize one particular gesture performed by the operator among several alternatives.

In the context of telerobotics, we have been interested so far in hand-based interfaces, that is, the replication by a robotic system of all or part of the hand posture changes performed by an operator. The analysis of the gripping of an object can give a precise description of the human grasp trajectory. This information can be used to teleoperate a robot gripper in charge of grabbing a similar object in a remote workspace. The gripper's control replicates the operator's hand and finger motion with respect to the object to be grasped. The camera tracking the hand is calibrated and the position and the orientation of a reference frame attached to the object is assumed known

with respect to the camera 's reference frame. Therefore, successive hand postures can be computed with respect to the object. The robot gripper can thus be guided by position-based control using the corresponding configurations of the hand with respect to the robot's object. The gesture recognition stage enables us to provide the relationship between the hand and the gripper so that the gripper configuration can be controlled with greater accuracy to grasp the object. We have implemented this application in a graphical simulation and proved the efficiency of this master/slave approach allowing us to take advantage of human grasping skills rather than relying on automatic grasp generation computations.

# Table des matières

Dédicace . . . . .	iv
Remerciements . . . . .	v
Résumé . . . . .	vi
Abstract . . . . .	ix
Table des matières . . . . .	xii
Liste des tableaux . . . . .	xvii
Liste des figures . . . . .	xix
1 Introduction . . . . .	1
2 Revue bibliographique sur l'interprétation visuelle des gestes . . .	7
2.1 Définition du geste . . . . .	7
2.1.1 Classification . . . . .	7

2.1.2	Caractéristiques temporelles du geste . . . . .	9
2.1.3	Motivation du geste . . . . .	10
2.2	Analyse du mouvement de la main . . . . .	11
2.2.1	Le geste considéré comme statique . . . . .	12
2.2.2	Le geste considéré comme une séquence d'images de la main .	14
2.2.3	Le geste modélisé par un segment temporel . . . . .	17
2.2.4	Le geste analysé par rapport à l'enchaînement temporel . . . .	20
2.3	Comment nous avons envisagé le problème . . . . .	22
<b>3</b>	<b>Modélisation 3D de la main . . . . .</b>	<b>24</b>
3.1	Modèle cinématique de la main . . . . .	24
3.2	Modèle infographique de la main . . . . .	28
3.3	Cinématique des doigts II à V . . . . .	29
3.3.1	Cinématique directe . . . . .	29
3.3.2	Cinématique inverse . . . . .	32
3.4	Cinématique du pouce . . . . .	34
3.4.1	Cinématique directe . . . . .	34
3.4.2	Cinématique inverse . . . . .	37
<b>4</b>	<b>Détection et reconstruction de la main . . . . .</b>	<b>41</b>
4.1	Choix des indices . . . . .	42
4.2	Recouvrement de la position et de l'orientation de la paume . . . . .	45

4.3	Recouvrement de la position des doigts . . . . .	47
4.3.1	Cas des doigts II à V . . . . .	47
4.3.2	Cas du pouce . . . . .	49
4.4	Lissage temporel des résultats . . . . .	54
4.4.1	Filtre de Kalman . . . . .	54
4.4.2	Lissage envisagé . . . . .	58
4.5	Tests et validation . . . . .	58
4.5.1	Résultats obtenus sur des séquences de synthèse . . . . .	58
4.5.2	Résultats obtenus sur une séquence réelle . . . . .	69
<b>5</b>	<b>Traitement des occlusions . . . . .</b>	<b>81</b>
5.1	Conditions d'existence d'une occlusion . . . . .	82
5.1.1	Occultations pendant la phase d'approche . . . . .	82
5.1.2	Occultations pendant la phase de saisie . . . . .	84
5.2	Expressions mathématiques des conditions d'occlusion . . . . .	85
5.2.1	Occlusion d'une sphère par une sphère . . . . .	86
5.2.2	Occlusion d'une sphère par un cylindre . . . . .	88
5.3	Reconstruction en cas d'occlusions . . . . .	92
5.4	Résultats . . . . .	93
5.4.1	Validation de l'algorithme du traitement des occlusions . . . . .	93
<b>6</b>	<b>Étude de gestes . . . . .</b>	<b>100</b>

6.1	Interface Humain-Machine de détection de la posture de la main en temps réel . . . . .	100
6.1.1	Plateforme utilisée . . . . .	101
6.1.2	Détails de l'algorithme de recouvrement de la posture de la main	102
6.1.3	Améliorations apportées à l'étape de détection des taches . . .	106
6.2	Étude de plusieurs gestes de préhension . . . . .	108
6.2.1	Mise à l'échelle temporelle . . . . .	108
6.2.2	Reconnaissance de gestes . . . . .	110
6.2.3	Résultats obtenus . . . . .	112
6.2.4	Conclusion . . . . .	114
<b>7</b>	<b>Application au contrôle d'une pince robotique . . . . .</b>	<b>116</b>
7.1	Revue bibliographique sur l'asservissement visuel . . . . .	117
7.1.1	Contrôle basé sur l'image . . . . .	119
7.1.2	Contrôle basé sur la position . . . . .	121
7.2	Modèle du manipulateur . . . . .	122
7.2.1	Cinématique directe . . . . .	123
7.2.2	Cinématique inverse . . . . .	125
7.3	Modèle de la pince . . . . .	126
7.4	Contrôle de la pince par suivi de la main . . . . .	128
7.4.1	Principes . . . . .	128
7.4.2	Illustration . . . . .	131



7.4.3	Précision de la prise . . . . .	132
7.4.4	Conclusion . . . . .	135
8	<b>Conclusion . . . . .</b>	<b>143</b>
	<b>Bibliographie . . . . .</b>	<b>148</b>

## Liste des tableaux

4.1	Pouce . . . . .	60
4.2	Index . . . . .	60
4.3	Majeur . . . . .	60
4.4	Annulaire . . . . .	61
4.5	Auriculaire . . . . .	61
4.6	Position du poignet . . . . .	61
4.7	Orientation du plan de la paume . . . . .	62
4.8	Longueur des phalanges en cm, caméra 1 . . . . .	71
4.9	Longueur des phalanges en cm, caméra 2 . . . . .	72
4.10	Angles en degrés de la main fermée reconstruite . . . . .	72
4.11	Angles en degrés de la main ouverte reconstruite . . . . .	73
4.12	Pouce . . . . .	78
4.13	Index . . . . .	79
4.14	Majeur . . . . .	79
4.15	Annulaire . . . . .	80

4.16	Auriculaire . . . . .	80
5.1	Pouce . . . . .	95
5.2	Index . . . . .	95
5.3	Majeur . . . . .	95
5.4	Annulaire . . . . .	96
5.5	Auriculaire . . . . .	96
6.1	Verdict des distances $d_p$ et $d_i$ obtenues pour chaque geste de test . . .	112
6.2	Distance de ressemblance obtenue pour chaque geste de test . . . . .	113
7.1	Paramètres de Denavit-Hartenberg . . . . .	123
7.2	Écart de la position et de l'orientation de la pince par rapport à la configuration idéale lors de la fermeture sphérique . . . . .	133
7.3	Écart de la position et de l'orientation de la pince par rapport à la configuration idéale lors de la fermeture cylindrique . . . . .	134

# Liste des figures

1.1	Étapes de reconstruction de la posture de la main . . . . .	5
2.1	Taxonomie des gestes . . . . .	8
2.2	Le geste en tant que processus stochastique . . . . .	17
3.1	Os de la main (face palmaire) (tiré de l'Encyclopédia Universalis) . .	25
3.2	Exemples de modèles cinématiques de la main . . . . .	26
3.3	Modèle squelette de la main . . . . .	27
3.4	Modèle infographique de la main . . . . .	28
3.5	Modèle cinématique des doigts . . . . .	30
3.6	Cinématique directe des doigts II à V . . . . .	31
3.7	Notations utilisées pour la cinématique du doigt . . . . .	32
3.8	Modèle cinématique du pouce . . . . .	35
3.9	Cinématique du pouce . . . . .	36
3.10	Surface de cinématique du pouce en 3D pour $q_0$ fixé (cm) . . . . .	38
3.11	Surface de cinématique du pouce projetée dans le plan $(x, y)$ (cm) . .	38

4.1	Cône hexagonal HSV . . . . .	43
4.2	Les deux solutions possibles au problème de reconstruction de la paume . . . . .	46
4.3	Reconstruction de la position 3D des extrémités des doigts . . . . .	47
4.4	Solutions multiples au problème de cinématique inverse du doigt . . . . .	48
4.5	Sélection de la configuration la plus éloignée des configurations limites . . . . .	49
4.6	Positions de l'extrémité du pouce pour différentes valeurs de $q_0$ et $q_1$ . . . . .	52
4.7	Extrémité du pouce solution, TIP, par rapport à la surface de mouvement du pouce . . . . .	53
4.8	Projection dans le plan image de la solution . . . . .	53
4.9	Portions des séquences de synthèse originale et reconstruite . . . . .	64
4.10	Angle d'abduction de chaque doigt . . . . .	65
4.11	Reconstruction du pouce pour la séquence étudiée . . . . .	65
4.12	Reconstruction de l'index pour la séquence étudiée . . . . .	66
4.13	Reconstruction du majeur pour la séquence étudiée . . . . .	66
4.14	Reconstruction de l'annulaire pour la séquence étudiée . . . . .	67
4.15	Reconstruction de l'auriculaire pour la séquence étudiée . . . . .	67
4.16	Cas d'une des plus mauvaises reconstructions . . . . .	69
4.17	Main ouverte et fermée prise de deux caméras . . . . .	70
4.18	Reconstruction de la main ouverte vue de chaque caméra . . . . .	74
4.19	Reconstruction de la main fermée vue de chaque caméra . . . . .	74
4.20	Angle d'abduction de chaque doigt . . . . .	75

4.21 Reconstruction du pouce pour les deux séquences . . . . .	76
4.22 Reconstruction de l'index pour les deux séquences . . . . .	76
4.23 Reconstruction du majeur pour les deux séquences . . . . .	77
4.24 Reconstruction de l'annulaire pour les deux séquences . . . . .	77
4.25 Reconstruction de l'auriculaire pour les deux séquences . . . . .	78
5.1 La main tourne légèrement autour du poignet . . . . .	83
5.2 Cas où le doigt s'auto-occulte . . . . .	84
5.3 Le poing se ferme . . . . .	85
5.4 Cas où la main enveloppe l'objet . . . . .	85
5.5 Modélisation du doigt et de son extrémité . . . . .	86
5.6 Occlusion d'une sphère par une sphère . . . . .	87
5.7 Première condition d'occlusion d'une sphère par un cylindre . . . . .	89
5.8 Seconde condition d'occlusion d'une sphère par un cylindre . . . . .	90
5.9 Condition d'auto-occlusion d'un doigt . . . . .	91
5.10 Algorithme de décision en cas d'occlusions . . . . .	94
5.11 Cas d'occlusions lors de la fermeture du poing . . . . .	97
5.12 Cas d'occlusions lors de la rotation du poignet . . . . .	97
5.13 Cas d'auto-occlusions . . . . .	98
6.1 Diagramme des différents états dans lesquels peut se trouver la main	104
6.2 Organigramme de l'interface de reconstruction de la posture de la main	106
6.3 Étalement dans l'image des taches à détecter . . . . .	107

6.4	Estimation du vrai centroïde de la tache à l'intersection des axes . . .	107
6.5	Différentes prises . . . . .	108
6.6	Mise à l'échelle temporelle de toutes les séquences . . . . .	109
6.7	Moyennes des angles pour les trois types de prise . . . . .	111
7.1	Structure du contrôleur dans le cas d'asservissement visuel . . . . .	118
7.2	Définition des articulations du robot CRS A-460 (tiré de la documen- tation technique du robot CRS A-460) . . . . .	123
7.3	Modèle cinématique du robot . . . . .	124
7.4	Modèle de la pince Barrett (tiré de la documentation technique de Barrett Technology Inc.) . . . . .	127
7.5	Différentes configurations de la pince graphique . . . . .	127
7.6	Repères de la grille de calibration et de l'objet . . . . .	129
7.7	Repère de l'objet par rapport à la base du robot . . . . .	129
7.8	Mouvement de la main et de la pince pour une prise sphérique . . . .	137
7.9	Mouvement de la main et de la pince pour une prise cylindrique . . .	138
7.10	Exemple de configuration de la prise sphérique . . . . .	139
7.11	Trajet de la pince pour différentes séquences d'une prise sphérique . .	140
7.12	Exemple de configuration de la prise cylindrique . . . . .	141
7.13	Trajet de la pince pour différentes séquences d'une prise cylindrique .	142

# Chapitre 1

## Introduction

Le domaine de l'Interface Humain-Machine connaît actuellement un essor important et doit son émergence au développement d'applications 3D de plus en plus complexes. Les jeux vidéo, les programmes de conception assistée par ordinateur, de modélisation et d'animation graphique en particulier ont accru les besoins en périphériques d'entrée d'un genre nouveau. Obsolètes maintenant les souris ou joysticks traditionnels, limités en vitesse ou en naturel d'utilisation! Spaceballs, gants dotés de capteurs d'angles, suiveurs magnétiques, casques munis de capteurs à ultra-sons se sont multipliés (Hand, 1994), donnant naissance plus récemment à des manettes d'un type nouveau comme les périphériques dotés de propriétés haptiques (Burdea, 1996), qui donnent à l'utilisateur un retour de force et ainsi un réalisme accru. En effet, face à l'arrivée massive des applications de réalité virtuelle, de manipulation d'objets virtuels, de modélisation d'objets en 3D, de jeux aux contrôles multipliés, les inter-



faces usagers visent dès lors à accroître le nombre de degrés de contrôle et à offrir à l'utilisateur plus de naturel d'utilisation et d'intuition.

Un des axes de recherche dans ce domaine concerne en particulier le développement d'applications dans lesquelles l'utilisateur interagit directement avec la machine, usant de moyens de communications naturels comme la voix ou les expressions du visage. Les gestes constituent en ce sens un excellent moyen d'interaction, d'une part comme compléments illustratifs au discours et d'autre part comme outils de manipulation d'objets. Plusieurs catégories d'interfaces basées sur l'interprétation des gestes ont ainsi vu le jour, aussi variées qu'une télécommande basée sur la détection de la direction pointée par l'index ou qu'une souris 3D basée sur la reconnaissance de certains gestes iconiques, incluant bien sûr les tentatives de reconnaissance des langages des signes. Les techniques couramment utilisées comprennent l'exploitation de gants munis de senseurs capables de donner l'angle de flexion des doigts, complétés des données de suiveurs magnétiques afin de préciser la position et l'orientation du poignet par exemple. Ces dispositifs sont en général coûteux et encore imprécis. D'autre part, ils présentent l'inconvénient de restreindre les mouvements de l'utilisateur à cause de la présence de câbles. D'autres applications leur préfèrent l'usage de la vision comme senseur non immersif. Le nombre croissant d'ordinateurs personnels munis d'une caméra disposée sur le moniteur vient de surcroît encourager l'usage de ces techniques. Bien sûr, plusieurs solutions sont apparues, faisant appel à une ou plusieurs caméras afin de bénéficier dans ce dernier cas de l'information en stéréo-

scopie. D'autre part, l'analyse du geste peut s'effectuer directement dans l'image ou bien par l'intermédiaire d'un passage en 3D. De plus amples détails sur ces techniques sont donnés au chapitre 2.

L'approche envisagée dans ce mémoire consiste à développer un module efficace de reconstruction du mouvement de la main afin de l'exploiter dans une étude des gestes de saisie. L'interface de suivi de la main peut également être exploitée pour des usages parallèles incluant des périphériques d'entrée 3D, des outils de contrôle à multi degrés de liberté et un support à l'étude de mouvements de lancers rapides. Les exigences de ce dispositif sont les suivantes : détecter en temps réel la posture de la main en 3D tout en laissant l'utilisateur libre de ses mouvements. En conséquence, les options suivantes ont été retenues afin de mettre au point un système robuste de détection de la posture de la main : détection visuelle à l'aide d'une seule caméra d'indices de couleurs sur la main et reconstruction basée sur des techniques de cinématique inverse relativement peu coûteuses en terme de temps de calcul.

La méthodologie suivante a été appliquée afin de mener à bien cette tâche. Un modèle cinématique de la main a été conçu afin de pouvoir définir les variables d'articulation de la main et de pouvoir modéliser son mouvement. Il est présenté au chapitre 3 qui contient également les équations de cinématique directe et inverse du modèle. Une sélection d'indices de couleurs à disposer sur la main a été décidée afin de posséder suffisamment d'information visuelle pour reconstruire la posture de la main dans sa totalité. Le chapitre 4 présente le système de détection visuelle des indices de couleurs

suivi du module de reconstruction des variables d'articulation de la main à l'aide de techniques d'optimisation et des équations de cinématique inverse. Dès lors, la validation de la qualité de la reconstruction de la posture de la main peut s'appuyer sur un modèle graphique de la main développé à l'aide de la librairie graphique OpenGL. Il est alors possible d'afficher ce modèle à partir de différents points de vues afin de souligner les détails de la configuration étudiée. La possibilité de non-détection des indices de couleur due à des occlusions a été envisagée au chapitre 5 et traitée à l'aide de méthodes heuristiques de prédiction des configurations de la main. Ces prédictions sont validées au moyen de techniques de détermination de surfaces visibles. Cette première partie termine la mise en œuvre de l'interface Humain-Machine capable de restituer en temps réel la posture de la main en mouvement filmée à l'aide d'une caméra, et ce, malgré la présence possible d'occultations des indices de couleur sur la main. Le processus complet est illustré à la figure 1.1. La figure 1.1(a) dépeint le banc d'essai expérimental, les figures 1.1(b) et 1.1(c) respectivement la main éclairée et la main dans les conditions d'opération. Enfin, la figure 1.1(d) affiche le modèle de la main reconstruite.

La seconde partie de ce mémoire concerne les applications développées à l'aide de ce système. Cette interface permet en particulier d'enregistrer les variables d'articulation de la main filmée sur plusieurs séquences de gestes. Le chapitre 6 présente l'interface dans sa version optimisée permettant l'acquisition en temps réel de séquences d'images. Plusieurs gestes de préhension ont ainsi été étudiés afin d'extraire des variables

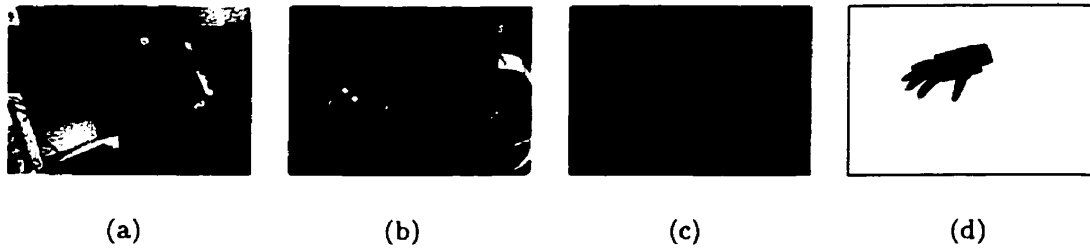


Figure 1.1 : Étapes de reconstruction de la posture de la main

d'articulation du modèle d'état de la main des informations caractéristiques de ces types de gestes. Des critères de reconnaissance des prises sphériques et cylindriques entre autres ont ainsi été mis en lumière permettant de caractériser de façon unique certaines saisies d'objets. Le chapitre 7 présente l'exploitation de l'information de mouvement en vue de contrôler une pince robotique à partir du mouvement guide de l'humain. Une revue des techniques de contrôle robotique par asservissement visuel fait l'objet du chapitre précédent. La caméra attachée à l'opérateur est calibrée par rapport à l'objet saisi par la main de sorte que les configurations successives de la main en mouvement puissent s'exprimer par rapport à l'objet. Considérons maintenant un manipulateur robotique dans un espace de travail à distance. Supposons connues la position et l'orientation d'un objet similaire par rapport au repère de base du robot. Il est possible de faire suivre au robot la trajectoire qu'avait suivie initialement la main, mais cette fois par rapport à l'objet que lui doit saisir par contrôle en position. Le robot est alors contrôlé de façon à suivre les configurations de passage des doigts de la main transformés dans son propre référentiel. Le gain d'information

tiré du geste effectué par l'être humain permet de prévoir le type de saisie administré au robot et de lui appliquer une configuration de la pince plus adaptée à l'objet à saisir. Ce système constitue en définitive un moyen de génération automatique de prise robotique.

## Chapitre 2

# Revue bibliographique sur l'interprétation visuelle des gestes

### 2.1 Définition du geste

#### 2.1.1 Classification

La nature à la fois intentionnelle et involontaire des mouvements humains rend délicate la définition de geste. Beaucoup d'auteurs s'accordent sur la définition suivante : un geste est défini par un mouvement du corps ou d'une partie du corps (comme le bras ou la main) qui véhicule ou non de l'information, qui exprime une pensée ou un sentiment. Afin de classer plus précisément les différentes catégories de gestes, plusieurs taxonomies ont été proposées (Kendon, 1986; Nespoulous et Lecours, 1986; Quek, 1994; Pavlovic, Sharma et Huang, 1997). Une version synthétique est proposée à la

figure 2.1.

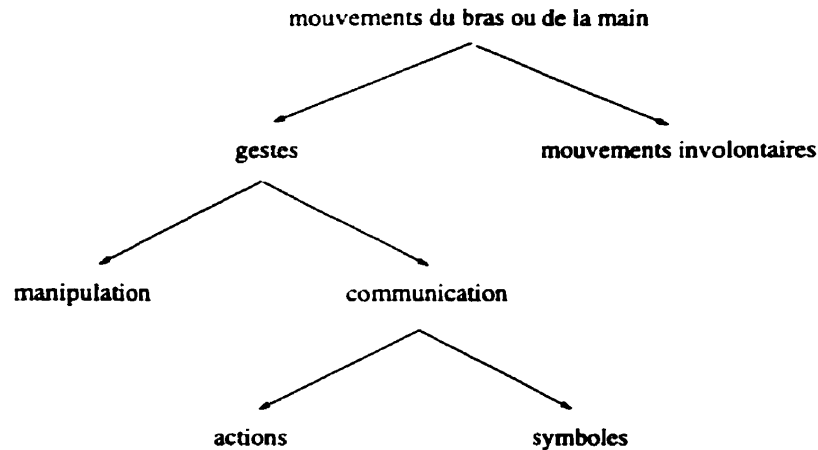


Figure 2.1 : Taxonomie des gestes

Une première distinction est faite entre les gestes et les mouvements involontaires qui ne véhiculent pas d'information. Ensuite, les gestes sont séparés en deux catégories : les gestes de manipulation et les gestes de communication. Une manipulation est une action sur un objet et n'est pas destinée à un récepteur, tandis qu'un geste communicatif transmet un message et possède des caractéristiques propres reconnaissables immédiatement. Dans cette dernière catégorie, on distingue les gestes symboliques des actions. Les symboles ont une signification linguistique et sont arbitraires par nature, comme les éléments du langage des signes. Les actions sont généralement divisées en gestes mimétiques (qui reproduisent une action) et en gestes déictiques (qui pointent quelque chose).

Dans le cadre de notre recherche, nous nous intéresserons davantage aux gestes de manipulation et dans la section suivante le mot geste désignera un mouvement de

préhension.

### 2.1.2 Caractéristiques temporelles du geste

Cette section s'attarde sur les caractéristiques cinématiques et temporelles des mouvements balistiques. Tout d'abord, en ce qui concerne les profils de vitesse, plusieurs auteurs s'accordent sur le modèle de la courbe en cloche (Jeannerod, 1984; Plamondon, 1995). La phase initiale à vitesse élevée constitue le mouvement d'approche qui finit par l'ouverture maximale des doigts suivie par la phase à vitesse faible qui conduit à la fermeture des doigts autour de l'objet. Selon Jeannerod (1984), la seconde phase débute à approximativement 75% de la durée totale du mouvement, ce qui lui laisse supposer que la coordination entre les phases de transport et de manipulation s'effectue dans le domaine temporel. De plus, le *feedback* visuel n'interviendrait que dans la précision de la prise et non avant. Plamondon et Alimi (1997) proposent comme schéma de vitesse une réponse en loi Delta log-normale pour expliquer la formation du mouvement par différence des contributions des muscles agonistes et antagonistes. Ils expliquent l'allure du mouvement par la cinématique uniquement. D'autres invariants temporels, notamment en ce qui concerne les fréquences de mouvement de manipulation de la main ont été étudiés (Kunesch, Binkofski et Freund, 1989). Parallèlement à ces recherches, des travaux portant plus précisément sur la durée et le contrôle des mouvements humains ont été menés. La première grande contribution concernant la relation entre la durée et la précision d'un mouvement



revient à Fitts (1954). Le mouvement qu'il étudie consiste à diriger rapidement la main vers une cible. La relation entre la durée du mouvement  $MT$ , la distance du point de départ à la cible  $A$  et la précision requise ( $W$  largeur de la cible) est alors modélisée par la loi logarithmique suivante :

$$MT = a + b \log_2 \left( \frac{2A}{W} \right).$$

Cette formule de régression largement testée s'est avérée un modèle robuste du comportement psychomoteur humain. Depuis, cette relation a entre autre été appliquée à des périphériques d'entrée dans des interfaces Humain-Machine (Card, English et Burr, 1978) puis a conduit à d'autres modélisations plus fidèles du compromis vitesse/précision du mouvement, comme les lois quadratique ou de puissance (Plamondon, 1995).

### 2.1.3 Motivation du geste

La tâche de saisie d'objet ne peut pas s'expliquer uniquement par son aspect moteur. Elle est le résultat de la volonté d'interagir avec l'objet. La nature du processus qui enclenche l'action à partir de l'information visuelle sur l'objet ne fait pas l'unanimité. Gibson (1977) propose la théorie des "affordances" qui confère à l'objet certaines propriétés intrinsèques et invariantes. Un stylo, par exemple, présente la qualité de servir à l'écriture. Cette propriété est toujours là pour être perçue même si je n'ai pas besoin d'écrire au moment où je vois le stylo. L'objet offre ce qu'il fait parce

que c'est ce qu'il est. Par opposition, les travaux de l'école de psychologie Gestalt (Koffka, 1935) affirment que l'objet s'offre avec un certain usage mais qui est perçu uniquement si l'utilisateur a besoin de l'objet. La valeur d'un objet, en ce sens, varie en fonction des besoins de l'observateur. Plus récemment, en accord avec cette dernière théorie, Jeannerod (1994) soutient que la façon de saisir un objet provient d'une représentation interne que l'on se fait de l'objet. C'est l'action intentée qui détermine la nature de la prise et non le *feedback* visuel. Par exemple un stylo ne sera pas saisi de la même manière selon qu'il s'agisse de le déplacer ou d'écrire avec. Le mouvement et le positionnement de la main sont déterminés au niveau de la représentation. Selon lui, la formation et l'exécution de la prise d'un objet reposent sur une représentation de l'objet de plus haut niveau que l'information acquise par la vision. Les mécanismes qui déterminent la formation de la prise ne sont donc pas encore tout à fait unifiés.

## 2.2 Analyse du mouvement de la main

Dans le cadre des interfaces Humain-Machine, les applications de reconnaissance de gestes ou de suivi de la main exigent plusieurs étapes indispensables à leur élaboration (Pavlovic et al., 1997). Tout d'abord, l'étape de modélisation du geste, qui intègre les caractéristiques spatiales et/ou temporelles du geste. Vient ensuite l'analyse du geste qui sert à déterminer les paramètres du modèle choisi à partir de l'information dans l'image. Ces paramètres donnent une certaine description de la pose de la main

ou de sa trajectoire et reflètent le modèle choisi précédemment. La dernière étape consiste à reconnaître le geste à partir des paramètres issus de l'analyse et interprétés à la lumière du modèle choisi et possiblement à travers les règles d'une grammaire. La mise en place d'une telle application exploitant l'information gestuelle est largement dominée par l'étape initiale de modélisation du geste, étape qui conditionne la suite des opérations. Nous nous intéresserons donc tout particulièrement aux différentes approches envisagées concernant la modélisation du geste et en particulier sur l'**interprétation temporelle** sous-jacente, qui est de grande importance dans le cadre de notre recherche : le geste est-il considéré comme statique ou bien dynamique? Et dans ce dernier cas, comment le temps est-il intégré dans la modélisation du geste?

### 2.2.1 Le geste considéré comme statique

Cette section fait état des modèles dans lesquels le temps n'est pas pris en compte. Il s'agit des gestes interprétés comme des postures statiques, soit dans l'image, soit dans l'espace.

#### 2.2.1.1 Étude des postures en 2D

Le modèle est dans ce cas basé sur l'**apparence dans l'image**. Les applications touchées concernent la reconnaissance de gestes iconiques, par exemple certains éléments du langage des signes. La main est statique et sa posture seule contient toute l'information à reconnaître. Dans cette approche, l'analyse du geste s'inspire des

techniques classiques de reconnaissance de forme, faisant appel entre autre aux images des mains, aux moments géométriques, contours, silhouettes, arêtes, vecteurs propres. Ahmad, Taylor, Lanitis et Cootes (1997) choisissent le contour de la main qu'ils modélisent à l'aide des composantes principales.

Ces méthodes se révèlent dans l'ensemble plutôt efficaces en terme de coût de calcul mais elles peuvent difficilement se généraliser étant donné qu'elles se bornent à des applications très spécifiques dans des environnements contraints.

#### **2.2.1.2 Étude des postures en 3D**

Le modèle repose dans ce cas sur un modèle 3D de la main, le plus souvent un modèle squelette. L'analyse de la posture de la main fait appel à une étape de cinématique inverse pour calculer la configuration spatiale de la main à partir d'indices caractéristiques sur une ou plusieurs images.

Rijpkema et Girard (1991) ont développé un modèle très précis de la cinématique de la main dans le but de synthétiser des mouvements de saisie d'objets en infographie. La cinématique inverse est résolue dans ce cas par la méthode du jacobien inverse, ou "resolved motion rate control method". Ces travaux ont été exploités par Lee et Kunii (1995) pour reconstruire la posture de la main à partir de points de couleur et d'une méthode de reconstruction itérative basée sur la minimisation du couple exercé sur l'ensemble des doigts. D'autres travaux ont été conduits depuis afin de minimiser le temps de calcul de la cinématique inverse. Lien et Huang (1998) font appel à un

algorithme génétique, Millar et Crawford (1996) à des heuristiques appliquées à un modèle simplifié et Shimada, Shirai, Kuno et Miura (1998) à un filtre de Kalman avec des contraintes d'inégalités. Plusieurs caméras sont utilisées par Gavrilu et Davis (1995), mais pour détecter cette fois les mouvements du corps humain, comme un danseur en mouvement. Il s'agit de superposer le modèle volumétrique 3D du corps aux images prises de divers points de vue et de reconstruire la configuration des différents membres. Kakadiaris et Metaxas (1995) reconstruisent également le modèle 3D d'un corps humain par intégration des contours apparents des différents membres observés de trois points de vue orthogonaux.

Ces approches conduisent à des modèles plus élaborés du corps humain en mouvement ou plus précisément de la main (dont les angles des doigts), mais il n'est pas prouvé qu'une telle précision soit nécessaire à la compréhension du geste. De plus, les coûts de calculs sont souvent prohibitifs pour des applications en temps réel.

### **2.2.2 Le geste considéré comme une séquence d'images de la main**

Cette section concerne les gestes modélisés par une séquence d'images analysées les unes à la suite des autres dans un procédé en temps réel. Les applications sont surtout celles de **suiti** de la main. Dans l'étape d'analyse des paramètres du modèle, la valeur des variables d'état considérées dans une image permet alors de les prédire dans l'image suivante et donc de lisser ou filtrer ces grandeurs. La vitesse

d'exécution du mouvement est alors prise en compte implicitement. Il ne s'agit pas à proprement parler de reconnaissance de geste dans ce cas, mais plutôt de "tracking" de la main ou des doigts dans une application de périphérique d'entrée en 3D par exemple. Une première catégorie de méthodes utilisées ici sont basées sur le principe suivant : un modèle squelette 3D d'articulation de la main est proposé puis un modèle géométrique est apposé par-dessus, cylindrique ou bien volumétrique pour une plus grande précision. Les variables d'articulation du modèle 3D de la main sont modifiées légèrement jusqu'à ce que la projection du modèle 3D se superpose avec l'image de la main réelle. Kuch et Huang (1995) utilisent un modèle volumétrique de la main et appliquent l'opérateur ou exclusif entre l'image de la main réelle binarisée et l'image prédite pour définir l'erreur de suivi. Un filtre de Kalman est utilisé par Goncalves, Bernardo, Ursella et Perona (1995) dans le suivi du bras humain. L'algorithme de Gauss/Newton minimisant l'erreur de position des doigts en 3D par détection de la main nue est exploité dans un processus de prédictions et corrections d'états par Rehg et Kanade (1993).

La seconde approche repose directement sur l'information de vitesse du mouvement de la main. Les propriétés dynamiques du mouvement de la main sont prises en compte par Kjeldsen et Kender (1997) dans une application de souris 3D. Ils détectent la position sur l'écran pointée par l'index de l'utilisateur et modélisent le curseur par un ressort non linéaire qui répond à une force  $F = f(D)$  où  $D$  représente la distance entre le curseur et la main réelle.  $f$  est une fonction sigmoïde afin de borner le dé-

placement du curseur. Les caractéristiques de flot de mouvement dans l'image sont également une source d'information possible. L'équation du flux optique dans l'espace des joints est exploitée par Yamamoto et Koshikawa (1991) afin de mettre à jour les valeurs des angles des joints du bras humain en mouvement modélisé par un manipulateur robotique. Cipolla, Okamoto et Kuno (1993) font usage des propriétés de parallaxe de mouvement pour interpréter les vitesses de 4 points de la main en termes de changements d'échelle (zoom), rotation et changement d'orientation de la surface afin d'animer un objet graphique sur l'écran. Exploitant les travaux précédents et les propriétés d'invariance affine sous projection orthographique avec mise à l'échelle, Kuno, Sakamoto, Sakata et Shirai (1994) calculent la position et l'orientation de la main dans un repère fixe centré sur l'utilisateur. Ces méthodes sont surtout destinées à appréhender le mouvement relatif de la main en vue d'interagir avec un objet sur l'écran.

Pour le moment, les méthodes basées sur la détection d'indices visuels requièrent des contraintes sur l'arrière-plan de la scène pour éviter le cas des occlusions. La prise en compte des occlusions est envisagée dans (Rehg et Kanade, 1995). D'autre part, le traitement n'est pas toujours en temps réel en ce qui concerne le suivi mais cette approche paraît prometteuse pour appréhender la finesse du mouvement.

### 2.2.3 Le geste modélisé par un segment temporel

Dans cette section, l'aspect dynamique du geste est maintenant pris en considération. Un geste est considéré ici comme une action dynamique et implique que le mouvement de la main véhicule autant d'information que sa posture. Le geste est vu comme un procédé stochastique au cours du temps. Il doit être reconnu quels que soient l'instant de son exécution et sa durée. Il est alors modélisé par une trajectoire aléatoire  $\mathcal{M}_T$  dans l'espace des paramètres du modèle sur l'intervalle  $\mathcal{I}$ , voir figure 2.2 où l'espace illustré correspond au cas particulier de l'espace cartésien  $[x, y, z]$ . Un geste est considéré comme une réalisation particulière appartenant à l'espace des possibles. C'est

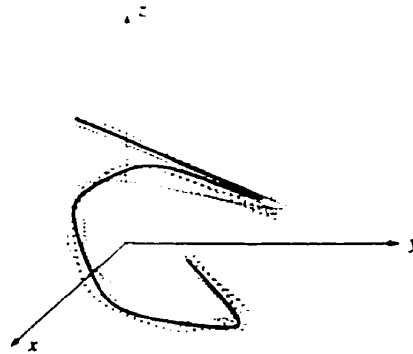


Figure 2.2 : Le geste en tant que processus stochastique

en général l'aspect de la main dans l'image qui est exploité. L'analyse de l'image a lieu après que toute la séquence a été enregistrée.

La première technique consiste à éliminer la composante temporelle du flot d'images par la technique de "Time collapsing" (Bobick et Wilson, 1995). Une séquence gestuelle est alors définie comme un vecteur d'états dans un espace de configura-



tions, où l'état caractérise la variance du mouvement à un instant donné par rapport à un prototype du geste issu d'un ensemble de tests. Le prototype du geste est calculé à partir d'un ensemble de trajectoires après élimination de la dimension temporelle de la courbe tout en conservant l'ordre dans lequel les états sont atteints successivement.

Une seconde technique vise à ramener la séquence à un nombre fixe d'échantillons par la technique de contraction/dilatation temporelle, dite "Dynamic Time Warping" (Darrell et Pentland, 1993). Tous les gestes sont ainsi définis par le même nombre d'images correspondant à la séquence la plus longue. Une corrélation normalisée est ensuite appliquée entre chaque image et un ensemble de vues du modèle obtenues par apprentissage. Le pic de corrélation détermine l'image de référence qui correspond le plus fidèlement à l'image traitée. La propriété de "Time Warping" se retrouve également implicitement dans les techniques de reconnaissance par modèles de Markov cachés (Hidden Markov Models) : la variabilité de la durée des échantillons de tests est prise en considération dans le modèle. Les modèles de Markov cachés constituent un modèle stochastique double caractérisé par une chaîne d'états possédant chacun une probabilité de transition d'un état à un autre. Ils ont été appliqués avec succès dans le domaine de la reconnaissance de la parole (Rabiner, 1989). Ici, le but est uniquement de reconnaître l'action humaine en exploitant l'apparence dans l'image (Yamato, Ohya et Ishii, 1992; Yang, Xu et Chen, 1994; Nechyba et Xu, 1996; Yang, Xu et Chen, 1997). Les HMM sont caractérisés par leur caractère d'apprentissage et

l'invariance par rapport à l'échelle temporelle. Une séquence d'images est alors transformée par quantification vectorielle en un vecteur d'indices qui est ensuite comparé aux modèles du geste modélisés par HMM.

Dans d'autres travaux enfin, le temps est directement exploité au moyen de l'analyse des vitesses des images (Rigoll et Kosmala, 1997), où les éléments de modélisation du geste sont extraits du film des vitesses, constitué de la séquence des différences entre l'image et celle qui la précède. Chaque nouvelle image caractérise ainsi ce qui a bougé par rapport à l'image d'avant. Les auteurs calculent ce qu'ils appellent le "centre de mouvement" de l'image et les écarts types moyens mesurant l'amplitude du mouvement en  $x$  et  $y$  afin de quantifier chaque geste.

L'exploitation du mouvement dans l'image conduit par ailleurs à expliciter la localisation et la "fraîcheur" du mouvement dans l'image. Davis et Bobick (1997) cherchent à décrire où se passe le mouvement dans l'image et comment il se déplace. Ils construisent pour cela l'image binaire de l'énergie de mouvement appelée MEI ("Motion-Energy Image") d'une part, et l'image scalaire de l'historique du mouvement appelée MHI ("Motion-History Image") d'autre part. La première représente la localisation du mouvement dans une séquence d'images tandis que dans la seconde, l'intensité d'un pixel indique si le mouvement en ce point est récent ou non.

Soit  $I(x, y, t)$  une séquence d'images et  $D(x, y, t)$  la séquence d'images binaires indiquant les régions en mouvement.  $D$  est en général obtenue par différentiation.

Alors l'image binaire MEI est définie de la façon suivante :

$$E_{\tau}(x, y, t) = \bigcup_{i=0}^{\tau-1} D(x, y, t - i),$$

où  $\tau$  désigne la durée totale de l'action. L'image MHI se construit, elle, de la façon suivante :

$$H_{\tau}(x, y, t) = \begin{cases} \tau & \text{si } D(x, y, t) = 1 \\ \max(0, H_{\tau}(x, y, t - 1) - 1) & \text{sinon.} \end{cases}$$

Ces techniques tentent de saisir la façon dont l'être humain effectue un geste dans le temps et sont tout à fait valides pour la reconnaissance de gestes de communication. Cependant leur application semble limitée pour la reconnaissance de gestes de manipulation car l'information spatiale précise fait défaut.

#### 2.2.4 Le geste analysé par rapport à l'enchaînement temporel

Cette section concerne les gestes qui sont reconnus s'ils apparaissent à un certain moment dans un contexte temporel. Le geste est considéré comme une attitude particulière (ou une occurrence) ayant lieu à un moment donné. Il s'agit de segmenter le geste des autres mouvements non-intentionnels.

Kendon (1986) décompose un geste symbolique en trois phases : préparation, pic et rétraction, sur un intervalle qu'il nomme "la phrase gestuelle". La phase de préparation met la main en mouvement à partir d'une position de repos, le pic caractérise la pose statique du geste et se produit à vitesse faible et enfin la phase de rétraction

caractérise le retour de la main à la position de repos initiale ou bien son repositionnement pour une nouvelle phase gestuelle. Les travaux de Quek (1994) tirent parti de cette définition pour déceler à quel moment le geste est statique dans une séquence d'images. Pendant le mouvement grossier du pic, le changement de configuration de la main est négligé. L'auteur adopte une méthode basée sur le flux optique pour interpréter le geste. Il calcule un ensemble de vecteurs spatio-temporels qui suivent le contour de la main en mouvement. ceci en trois étapes :

- Détection de la position des contours en mouvement par combinaison des dérivées spatiales et temporelles de l'image.
- Calcul du flot de mouvement sur certains points sélectionnés le long du contour. La correspondance des points de contours sur les différentes images est obtenue par un processus de corrélation.
- Lissage du flot de mouvement par contrainte de variance.

Appliqué sur une séquence de mouvement, le processus conduit à un ensemble de vecteurs de flot de mouvement dans le temps. L'intervalle où la vitesse est la plus faible correspond à l'occurrence de la posture statique de la main qui peut ensuite être reconnue dans le dictionnaire de gestes.

La modélisation du geste dans un contexte permet de définir le geste par rapport au passif temporel de la séquence. Dans (Davis et Shah, 1994), les différentes étapes du geste sont reconnues au fur et à mesure par une machine à états finis, de la po-

sition de départ spécifique (main à plat signifiant “hello”) à une position d’arrivée caractéristique du geste à reconnaître. ceci quelle que soit la durée des mouvements intermédiaires. D’autres auteurs envisagent le geste comme une transition d’un état à un autre. Jo, Kuno et Shirai (1998) établissent un diagramme de transitions d’états entre différentes actions comme saisir, relacher, bouger ou pointer. Un geste particulier est reconnu à l’aide de la définition de la transition par rapport au geste précédent.

Cette modélisation a le mérite d’intégrer le geste à un contexte, ce qui s’avère utile dans le cas d’interfaces de reconnaissance de gestes (quand mettre le processus de reconnaissance de geste en route par exemple). Mais elle se cantonne à une description de haut niveau du geste (et sans doute arbitraire) et ne donne pas plus d’information sur le geste en lui-même.

## 2.3 Comment nous avons envisagé le problème

À la lumière de ces travaux, nous allons maintenant détailler les motivations de l’approche que nous avons choisie dans le développement de notre projet. Nous désirons concevoir une interface capable de suivre la position de la main en mouvement tout en détectant sa posture de façon relativement précise. L’analyse de la posture de la main dans l’espace est donc préférable pour capter plus de détails et éviter les configurations multiples conduisant à une même projection dans le plan image si l’analyse dans l’image était envisagée. D’autre part, le dispositif doit réagir

en temps réel pour suivre la main en mouvement. C'est pour cela que la solution de reconstruire la posture de la main à partir d'indices disposés sur les doigts plutôt qu'à partir d'une détection de contours est préférée. C'est aussi cette motivation qui justifie l'usage d'une seule caméra plutôt que deux afin d'éviter un traitement en stéréoscopie. Dans notre étude, le geste est envisagé aussi bien en termes de poses statiques lorsqu'il s'agit de détecter les indices de couleur sur la main et de reconstruire sa posture qu'en termes d'enchaînement temporel dans l'étape de suivi de la paume, de lissage des variables d'articulation du modèle reconstruit et de prédiction de la posture de la main en cas d'occlusion. La force de notre système est d'intégrer à la fois un suivi de la main et une détection robuste de sa posture en temps réel, contrairement aux systèmes qui imposent au poignet de rester immobile ou bien qui nécessitent de lourdes charges calculatoires.

## Chapitre 3

# Modélisation 3D de la main

### 3.1 Modèle cinématique de la main

La main est composée de vingt-sept os qui forment trois groupes distincts (figure 3.1) :

- le carpe;
- les métacarpiens. Ils constituent le squelette de la paume et s'unissent aux phalanges au niveau des articulations métacarpo-phalangiennes.
- les phalanges. Chaque doigt possède trois phalanges (première, deuxième et troisième) sauf le pouce qui n'en possède que deux. Elles s'articulent autour des articulations interphalangiennes proximales (I.P.P.) et interphalangiennes distales (I.P.D.).

La main se compose de cinq rayons reliés au carpe, formant une voûte fixe assurant la stabilité de la préhension. Le pouce a l'avantage de posséder au niveau de

la jonction du métacarpien et du trapèze une articulation permettant le mouvement d'opposition à chacun des quatre autres doigts, qui eux ne sont mobiles que dans un seul plan. Afin de modéliser l'attitude de la main et les différents mouvements possi-

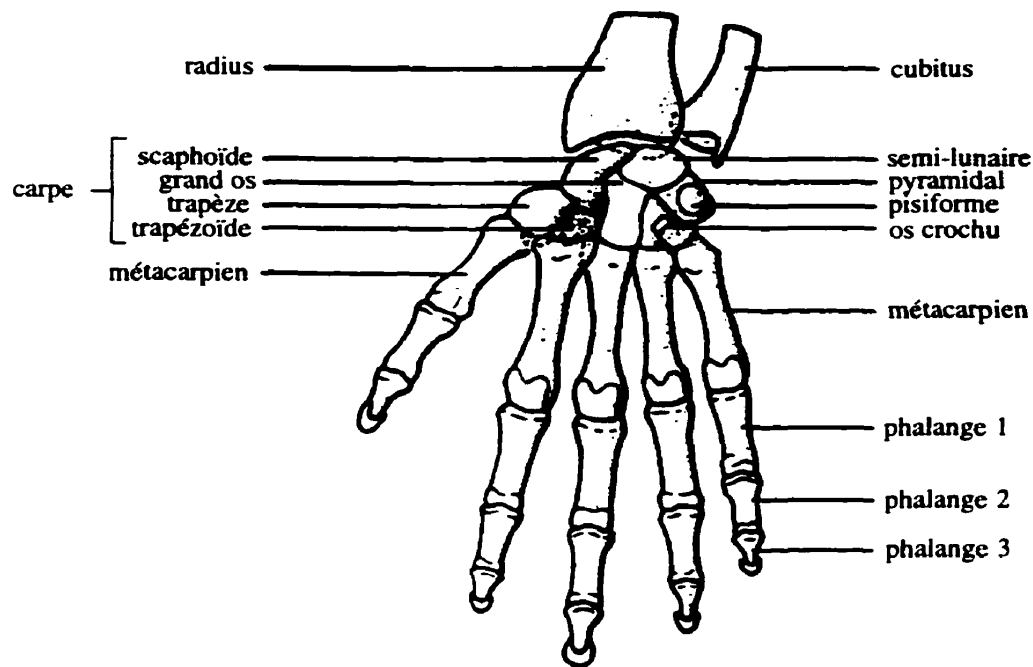


Figure 3.1 : Os de la main (face palmaire) (tiré de l'Encyclopédia Universalis)

bles des doigts, un modèle cinématique qui prend en compte l'anatomie fonctionnelle de la main est nécessaire. Dans le domaine de la modélisation de l'articulation de la main, plusieurs modèles cinématiques ont été proposés dans la littérature scientifique (Rehg et Kanade, 1993; Rijpkema et Girard, 1991; Lee et Kunii, 1995; Millar et Crawford, 1996). La plupart d'entre eux comportent vingt-sept degrés de liberté et se distinguent surtout par la localisation du point d'ancrage de la chaîne de chaque doigt (figure 3.2(a)) et par la cinématique du pouce (figure 3.2(b)).



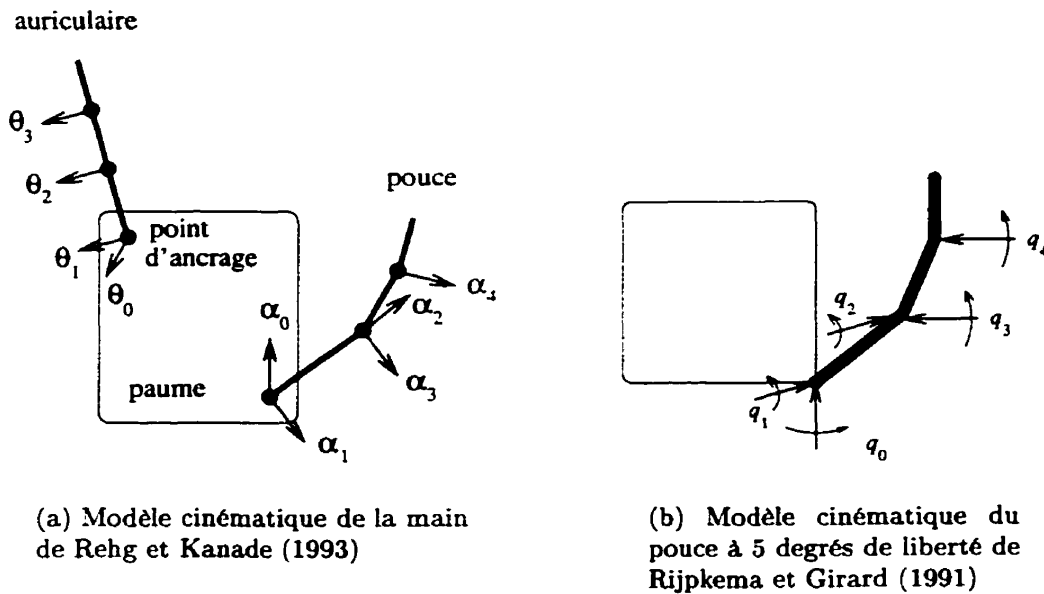


Figure 3.2 : Exemples de modèles cinématiques de la main

La suite de cette section présente le modèle cinématique de la main que nous avons développé et analyse le domaine de validité du modèle, compte-tenu des mouvements possibles de la main humaine. Le modèle est conçu de sorte qu'il respecte la plupart des degrés de liberté des doigts tout en restant suffisamment simple pour ne pas alourdir le problème de cinématique inverse. La main est assimilée à un ensemble de chaînes cinématiques articulées représentées par un modèle squelette à vingt-six degrés de liberté. Sa position et son orientation sont données respectivement par la position du centre du poignet,  $C$ , et l'orientation de la paume, sous la forme des trois angles d'Euler : lacet, tangage et roulis. Ce choix minimise le nombre de degrés de liberté en position et orientation de la main à six. La position des doigts est calculée

par rapport au plan de la paume qui sert de référence. D'après les travaux de Tubiana (1981), on fait la supposition que les doigts forment cinq raies concourantes au centre du poignet,  $C$ , lors de la fermeture du poing (figure 3.3). Ceci procure l'avantage de pouvoir exprimer entièrement la position d'un doigt par rapport à  $C$ , fixe, plutôt que par rapport aux jointures métacarpo-phalangiennes mobiles par exemple.

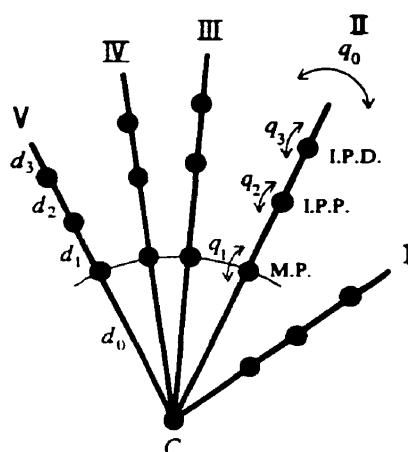


Figure 3.3 : Modèle squelette de la main

Les doigts sont énumérés de I à V du pouce à l'auriculaire. Chaque doigt est représenté par un manipulateur composé de quatre segments rigides à quatre degrés de liberté,  $q_0, q_1, q_2$  et  $q_3$ . Tout d'abord, l'angle d'abduction/adduction  $q_0$  définit l'écartement du doigt dans le plan de la paume. Viennent ensuite les trois angles de flexion/extension  $q_1, q_2$ , et  $q_3$ , qui déterminent la pliure du doigt dans le plan défini par  $q_0$  perpendiculaire à la paume. Ils correspondent respectivement aux articulations métacarpo-phalangienne (M.P.), interphalangienne proximale (I.P.P.) et interphalangienne distale (I.P.D.). Une fois l'angle d'abduction déterminé, les doigts II à V sont assimilés

à des manipulateurs planaires. Le pouce est également modélisé par un manipulateur à quatre degrés de liberté, mais non planaire. Son modèle est simplifié, en ce sens qu'il ne permet pas l'abduction au niveau l'articulation interphalangienne proximale (I.P.P.), ceci dans l'optique de ne pas surcharger l'étape de cinématique inverse.

### 3.2 Modèle infographique de la main

Afin de simuler la cinématique choisie, un modèle graphique de la main a été construit puis animé à l'aide de la librairie graphique OpenGL. La paume est représentée par un parallélépipède et chaque doigt est construit à l'aide de trois cylindres. Les dimensions ont été calibrées à partir de mesures prises sur la main de l'auteur. La cinématique est fidèle au modèle squelette de la main (figure 3.4).

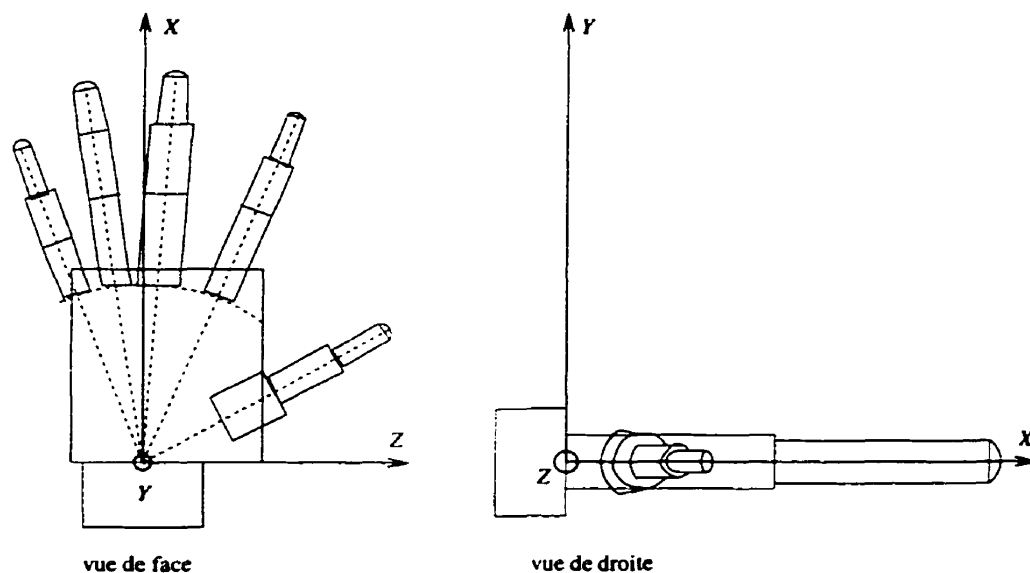


Figure 3.4 : Modèle infographique de la main

### 3.3 Cinématique des doigts II à V

#### 3.3.1 Cinématique directe

Nous munissons le modèle de la main de 3 repères afin de représenter la cinématique des doigts :

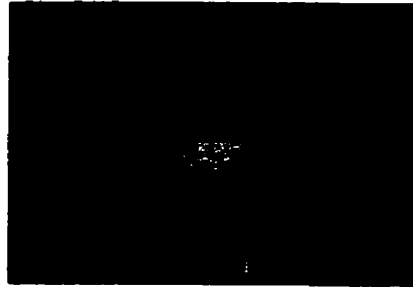
- $R_c = [C, X, Y, Z]$  est le repère de référence de la main dont le centre est le point milieu du poignet,  $C$ , et dont les axes sont définis par l'orientation de la paume, comme illustré à la figure 3.5.
- $R_p = [O, x, y, z]$  est le repère de référence du doigt dont l'origine se situe sur le premier joint de flexion/extension et dont l'axe  $x$  est la direction du doigt dans la plan de la paume, comme illustré à la figure 3.5.
- $R_f = [M, x_f, y_f, z_f]$  est le repère final du doigt, où  $M$  est l'extrémité du doigt et les axes  $x_f$  et  $y_f$  se situent dans le plan de mouvement, comme illustré à la figure 3.5. Le troisième axe,  $z_f$ , est alors parallèle à l'axe  $z$  du repère de référence du doigt.

Chaque doigt est assimilé à un manipulateur plan à quatre degrés de liberté. L'abduction/adduction  $q_0$  représente l'angle formé par le plan contenant le doigt et l'axe  $X$  du repère de référence de la main. Les angles  $q_1$ ,  $q_2$  et  $q_3$  représentent les angles de flexion/extension, figure 3.5.

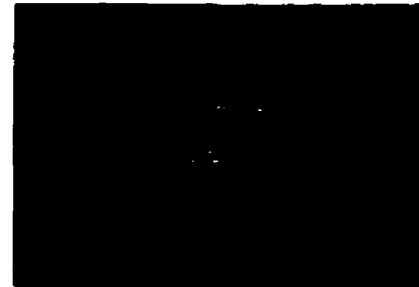
La cinématique directe du doigt s'exprime au moyen de la matrice de transformation homogène suivante, définissant la position et l'orientation du repère final,  $R_f$ ,



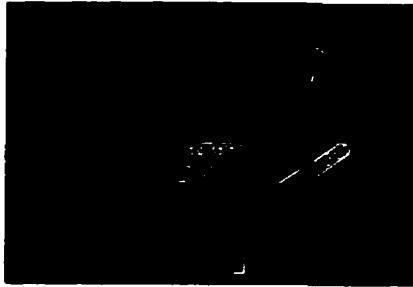
main à la figure 3.6.



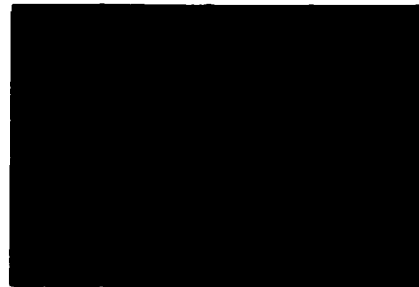
(a) Position initiale de la main



(b) Après flexion



(c) Après abduction



(d) Après abduction et flexion

Figure 3.6 : Cinématique directe des doigts II à V

Dans le repère  $[0, X, Y, Z]$ , les coordonnées  $[X_M, Y_M, Z_M]$  de l'extrémité  $M$  du doigt sont données par les relations suivantes :

$$\begin{cases} X_M = \cos(q_0)[d_0 + d_1 \cos(q_1) + d_2 \cos(q_1 + q_2) + d_3 \cos(q_1 + q_2 + q_3)] \\ Y_M = d_1 \sin(q_1) + d_2 \sin(q_1 + q_2) + d_3 \sin(q_1 + q_2 + q_3) \\ Z_M = \sin(q_0)[d_0 + d_1 \cos(q_1) + d_2 \cos(q_1 + q_2) + d_3 \cos(q_1 + q_2 + q_3)] \end{cases}$$

### 3.3.2 Cinématique inverse

Le formalisme utilisé pour décrire la cinématique directe des doigts assimilés à des manipulateurs planaires rappelle celui de la robotique. Il est intéressant d'en tirer parti pour résoudre l'étape de cinématique inverse par raisonnement géométrique plutôt que par des méthodes itératives comme c'est souvent le cas dans le domaine de reconnaissance de gestes ou de l'animation graphique par ordinateur (Lee et Kunii, 1995; Welman, 1993).

Supposons le point  $[M]_{R_p}$  donné par ses trois coordonnées  $[X_M, Y_M, Z_M]$ . On appelle  $l_0$  la distance  $OM$ , et  $l_1$  et  $l_2$  les longueurs définies comme indiqué sur la figure 3.7. On

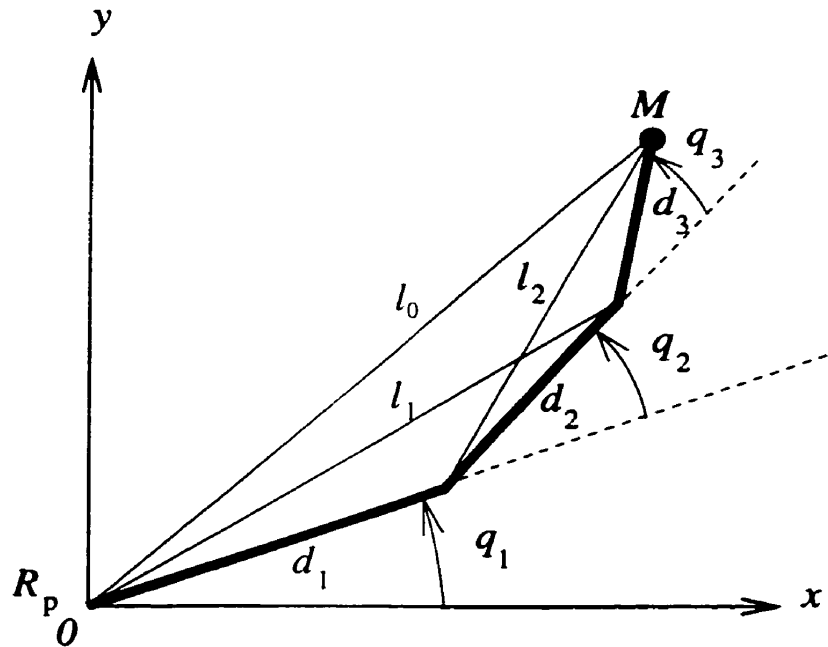


Figure 3.7 : Notations utilisées pour la cinématique du doigt

applique alors la formule de Al-Kashi (règle du cosinus qui lie les longueurs des côtés

d'un triangle au cosinus d'un de ses sommets) aux triangles suivants :  $\{l_1, d_1, d_2\}$ ,  $\{l_1, l_0, d_3\}$ ,  $\{l_2, d_2, d_3\}$  et  $\{l_2, l_0, d_1\}$ . On obtient les résultats suivants :

$$\left\{ \begin{array}{l} l_0^2 = X_M^2 + Y_M^2, \\ l_1^2 = d_1^2 + d_2^2 + 2d_1d_2 \cos q_2 \\ \quad = l_0^2 + d_3^2 - 2l_0d_3 \cos \left[ \text{atan2}(X_M, Y_M) - \left( \frac{\pi}{2} - q_1 - q_2 - q_3 \right) \right], \\ l_2^2 = d_2^2 + d_3^2 + 2d_2d_3 \cos q_3 \\ \quad = l_0^2 + d_1^2 - 2l_0d_1 \cos [\text{atan2}(Y_M, X_M) - q_1]. \end{array} \right.$$

D'où

$$\left\{ \begin{array}{l} q_1 = \text{atan2}(Y_M, X_M) - \arccos \left( \frac{l_0^2 + d_1^2 - d_2^2 - d_3^2 - 2d_2d_3 \cos(q_3)}{2l_0d_1} \right), \\ d_1^2 + d_2^2 + 2d_1d_2 \cos q_2 \\ \quad = l_0^2 + d_3^2 - 2l_0d_3 \cos \left[ q_2 + q_3 - \arccos \left( \frac{l_0^2 + d_1^2 - d_2^2 - d_3^2 - 2d_2d_3 \cos(q_3)}{2l_0d_1} \right) \right]. \end{array} \right.$$

Pour finir, les quatre joints sont donnés par les expressions suivantes :

$$\left\{ \begin{array}{l} q_0 = \text{atan2}(Z_M, X_M), \\ q_3 = 2q_2/3, \\ d_1^2 + d_2^2 - l_0^2 - d_3^2 + 2d_1d_2 \cos q_2 + \\ \quad 2l_0d_3 \cos \left[ 5q_2/3 - \arccos \left( \frac{l_0^2 + d_1^2 - d_2^2 - d_3^2 - 2d_2d_3 \cos(2q_2/3)}{2l_0d_1} \right) \right] = 0, \\ q_1 = \text{atan2}(Y_M, X_M) - \arccos \left( \frac{l_0^2 + d_1^2 - d_2^2 - d_3^2 - 2d_2d_3 \cos q_3}{2l_0d_1} \right). \end{array} \right.$$



On note que  $q_2$  s'obtient par résolution numérique en annulant la troisième équation. L'algorithme de Newton a été employé pour trouver le zéro de la fonction ainsi définie. Les autres angles s'obtiennent de façon directe.

## 3.4 Cinématique du pouce

### 3.4.1 Cinématique directe

Comme nous l'avons mentionné plus tôt, le pouce possède un degré de liberté supplémentaire en abduction/adduction au niveau de l'articulation interphalangienne proximale. Pour ne pas compliquer le modèle cinématique du pouce par rapport aux autres doigts, ni incrémenter la difficulté du problème de cinématique inverse, le pouce est approximé par un manipulateur non planaire à quatre degrés de liberté. Son modèle restreint le spectre des mouvements possibles du pouce tout en conservant le mouvement de la pince, fondamental pour notre étude. Son mouvement est composé d'une rotation autour de l'axe  $z$  suivie de deux rotations autour de l'axe  $y$ , le tout exprimé en transformations relatives (figure 3.8). Les transformations successives sont illustrées sur la main réelle pour plus de clarté, voir figure 3.9. La cinématique directe du pouce s'exprime par la matrice de transformation homogène suivante :

$$[M]_{R_c} = R[Y, q_0] T[x, d_0] R[z, q_1] T[x, d_1] R[y, q_2] T[x, d_2] R[y, q_3] T[x, d_3] [M]_{R_f}.$$

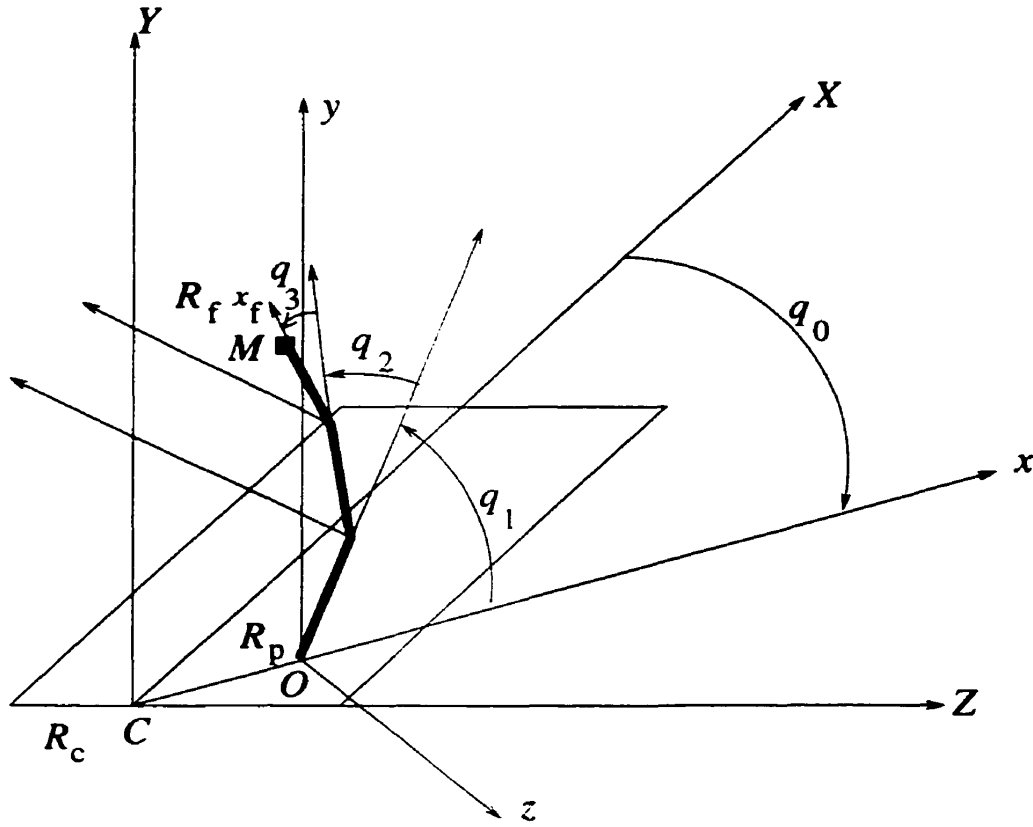


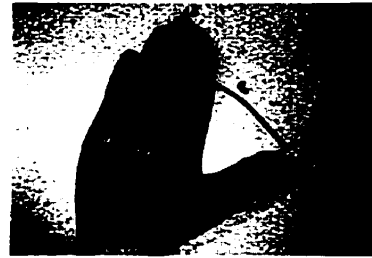
Figure 3.8 : Modèle cinématique du pouce

On choisit cette fois de lier les deux derniers joints selon la relation suivante :  $q_3 = q_2$ , qui s'avère par simulation proche de la réalité. Le calcul de cinématique inverse sera encore une fois simplifié. La position de l'extrémité du pouce  $M[X_M, Y_M, Z_M]$  exprimée dans le repère  $R_p = [0, x, y, z]$  (figure 3.8) s'écrit :

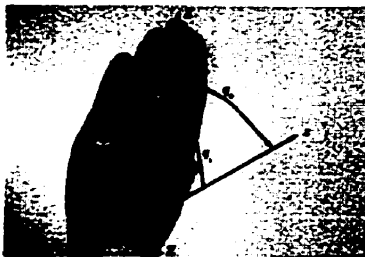
$$\begin{cases} X_M = \cos q_1 (d_1 - d_3 + d_2 \cos q_2 + 2d_3 \cos q_2^2), \\ Y_M = \sin q_1 (d_1 - d_3 + d_2 \cos q_2 + 2d_3 \cos q_2^2), \\ Z_M = -\sin q_2 (d_2 + 2d_3 \cos q_2). \end{cases} \quad (3.1)$$



(a) Pouce dans la direction de  $X$



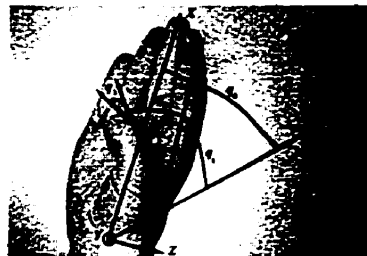
(b) Après abduction



(c) Après première flexion



(d) Après seconde flexion



(e) Après dernière flexion

Figure 3.9 : Cinématique du pouce

On remarque que  $Z_M$  ne dépend pas de l'angle  $q_1$ , ce qui facilite les calculs de cinématique inverse.

### 3.4.2 Cinématique inverse

Supposons maintenant connues les coordonnées de l'extrémité du pouce  $M[X_M, Y_M, Z_M]$  dans le repère  $R_p = [0, x, y, z]$ . Notons que cette dernière hypothèse inclut implicitement la connaissance de l'angle d'abduction  $q_0$  de la posture du pouce. Les équations de cinématique inverse se résolvent alors aisément et donnent immédiatement les angles de flexion du pouce :

$$\begin{cases} q_1 = \text{atan2}(Y_M, X_M). \\ q_2 = q_3 = \text{acos} \frac{-d_2 + \sqrt{d_2^2 - 8d_3(d_1 - d_3 - X_M/\cos q_1)}}{4d_3}. \end{cases} \quad (3.2)$$

Comme on le note,  $X_M$  et  $Y_M$  suffisent à reconstruire  $q_1$ ,  $q_2$  et  $q_3$ . Cependant cela implique que  $Z_M$  vérifie l'équation 3.1.3. c'est-à-dire que le point  $P$  obtenu en 3D après estimation de la position de l'extrémité du pouce appartienne à la courbe de cinématique du pouce définie par les équations 3.1 pour  $(q_1, q_2) \in [0; \pi/2]^2$ , ce qui est fortement improbable en pratique. Pour pallier cette incertitude, on choisit de calculer les coordonnées du point sur la courbe de cinématique du pouce qui s'avère le plus proche du point obtenu en 3D.  $P$ . Les figures 3.10 et 3.11 illustrent la surface admissible que peut parcourir l'extrémité du pouce pour  $(q_1, q_2) \in [0; \pi/2]^2$  et  $q_0$  fixé. Sur ces figures, le plan de la paume est représenté par le triangle  $(A, B, C)$  où  $A$  est le centre du poignet. Le point  $A$  correspond en fait au point  $C$  de la figure 3.8.

On cherche donc à minimiser la distance entre  $P$ , l'extrémité du pouce reconstruite en 3D, et la courbe définie par les équations 3.1. On définit par  $[X_P, Y_P, Z_P]$  les

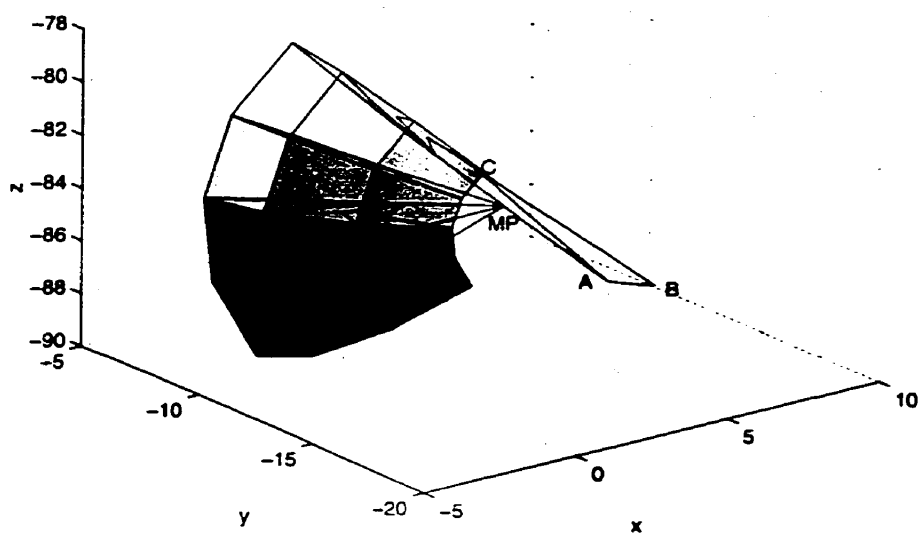


Figure 3.10 : Surface de cinématique du pouce en 3D pour  $q_0$  fixé (cm)

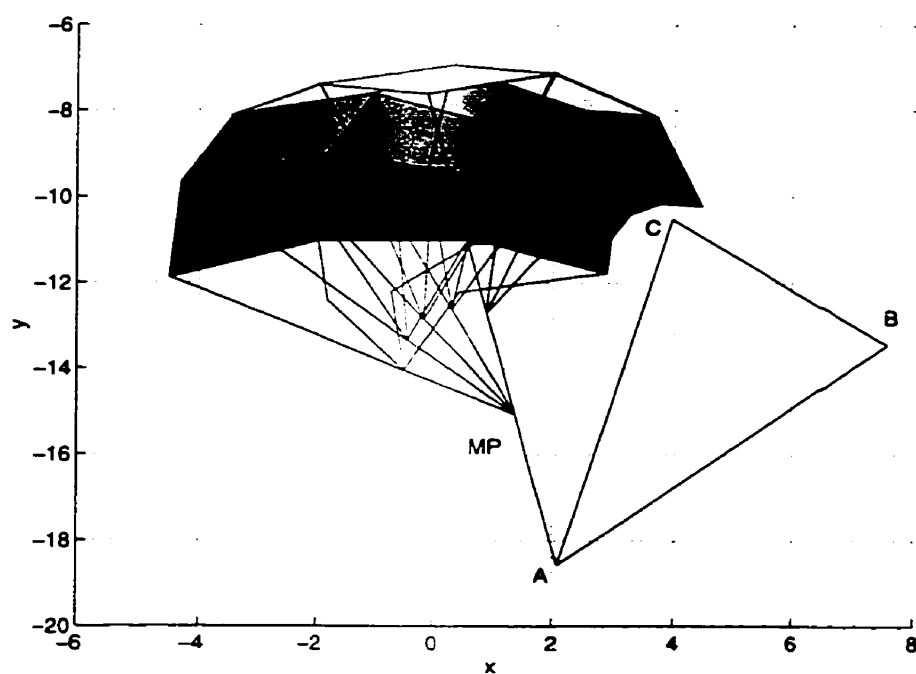


Figure 3.11 : Surface de cinématique du pouce projetée dans le plan  $(x, y)$  (cm)

coordonnées de l'extrémité du pouce  $P$  dans le référentiel  $R_p$  (figure 3.8). Encore une fois, l'angle d'abduction  $q_0$  est supposé déterminé. On cherche  $q_1$  et  $q_2$  qui minimisent le carré de la norme  $f(q_1, q_2)$  de la différence des vecteurs  $\vec{OP}$  et  $\vec{OM}$ , avec  $M$  appartenant à la courbe de cinématique du pouce.

$$\begin{aligned} f(q_1, q_2) = & \left[ X_P - \cos q_1 (d_1 - d_3 + d_2 \cos q_2 + 2d_3 \cos q_2^2) \right]^2 \\ & + \left[ Y_P - \sin q_1 (d_1 - d_3 + d_2 \cos q_2 + 2d_3 \cos q_2^2) \right]^2 \\ & + [Z_P + \sin q_2 (d_2 + 2d_3 \cos q_2)]^2. \end{aligned}$$

$f(q_1, q_2)$  est une fonction convexe et atteint son minimum si  $P$  appartient à la courbe caractéristique. Une condition nécessaire et suffisante pour minimiser  $f(q_1, q_2)$  est la suivante :

$$\begin{cases} \frac{\partial f}{\partial q_1} = 0, \\ \frac{\partial f}{\partial q_2} = 0, \end{cases}$$

ce qui conduit après calculs aux relations suivantes :

$$\left\{ \begin{array}{l} \sin q_1 X_P - \cos q_1 Y_P = 0. \\ [d_2 + 4d_3 \cos q_2][(X_P \cos q_1 + Y_P \sin q_1 - d_1) \sin q_2 + Z_P \cos q_2] \\ - 2d_3 Z_P - d_2 d_3 \sin q_2 = 0. \end{array} \right. \quad (3.3)$$

L'angle  $q_1$  est obtenu directement par l'équation précédente. On remarque qu'il correspond à celui issu du tout premier calcul de cinématique inverse. En fait, c'est uniquement la composante  $Z_P$  qui est modifiée par l'intermédiaire de  $q_2$  de sorte que le point  $P$  appartienne à la surface de cinématique du pouce. L'angle  $q_2$  se calcule par la méthode numérique de Newton pour calculer le zéro d'une fonction. Les deux angles  $q_1$  et  $q_2$  assurent la position de l'extrémité du pouce sur la surface de cinématique du pouce pour  $q_0$  déterminé tout en minimisant la distance avec le point  $P$  initialement reconstruit. Ainsi est résolue l'étape de cinématique inverse pour le cas du pouce.

## Chapitre 4

# Détection et reconstruction de la main

On désire implanter un système visuel capable de détecter en temps réel la posture de la main en mouvement. Maintenant que le modèle cinématique de la main et en particulier les équations de cinématique inverse ont été définis, il importe de connaître l'information minimale indispensable à la reconstruction de la posture de la main. Il s'agit d'orner la main d'un certain nombre d'indices de nature différente que l'on va détecter sur l'image prise par la caméra. Ces indices doivent permettre de reconstruire la posture de la main définie par son modèle cinématique. Ce chapitre présente donc les étapes de choix et de détection des indices sur le gant ainsi que le processus de reconstruction des variables d'articulation de la main basé sur l'information procurée par les indices.



## 4.1 Choix des indices

Le modèle de la main possédant vingt-six degrés de liberté, un nombre non négligeable d'indices est nécessaire afin de procéder au recouvrement de la posture de la main dans sa totalité. Le choix de détecter des motifs particuliers a ainsi été rejeté en raison du nombre insuffisant de formes possibles et de la complexité de calcul qui risque d'être mise en jeu pour les détecter. L'utilisation d'une caméra couleur ainsi que des indices de couleur distincte est apparue le meilleur compromis pour la rapidité de détection et pour le nombre suffisant d'indices différents (choix de couleurs). La position du poignet seule ne permet pas de reconstruire l'orientation de la paume. On s'intéresse surtout à la position du plan de la paume et non à l'inclinaison du poignet par rapport à l'avant-bras. En conséquence, on ne prend pas en compte la position de l'avant-bras par rapport à la main mais uniquement l'orientation de la paume. On choisit de marquer le dessus de la main de trois points formant un triangle de dimensions connues afin de reconstruire la position 3D de ces trois points, donnant de ce fait la position et l'orientation du triangle ornant la paume supérieure. D'après les équations de cinématique directe des doigts, en faisant la supposition que le plan du mouvement des doigts est connu, la position de l'extrémité du doigt suffit à calculer la valeur de l'angle de chaque articulation. On marque donc de couleur différente les extrémités de chaque doigt. Finalement, afin d'éviter que la couleur de la peau n'interfère lors de la segmentation couleur avec les indices caractéristiques de la main (les trois points de la paume ainsi que les cinq extrémités des doigts), on place des points de couleur vive

sur un gant noir. On choisit de transformer les composantes RGB des couleurs dans l'espace HSV (Hue, Saturation, Value qui correspondent respectivement à la teinte, la saturation et l'intensité lumineuse) afin de séparer au mieux les classes de couleur différente, figure 4.1. L'angle de la teinte (Hue) détermine la couleur, la saturation le degré de pureté de la couleur et la valeur correspond à l'intensité.

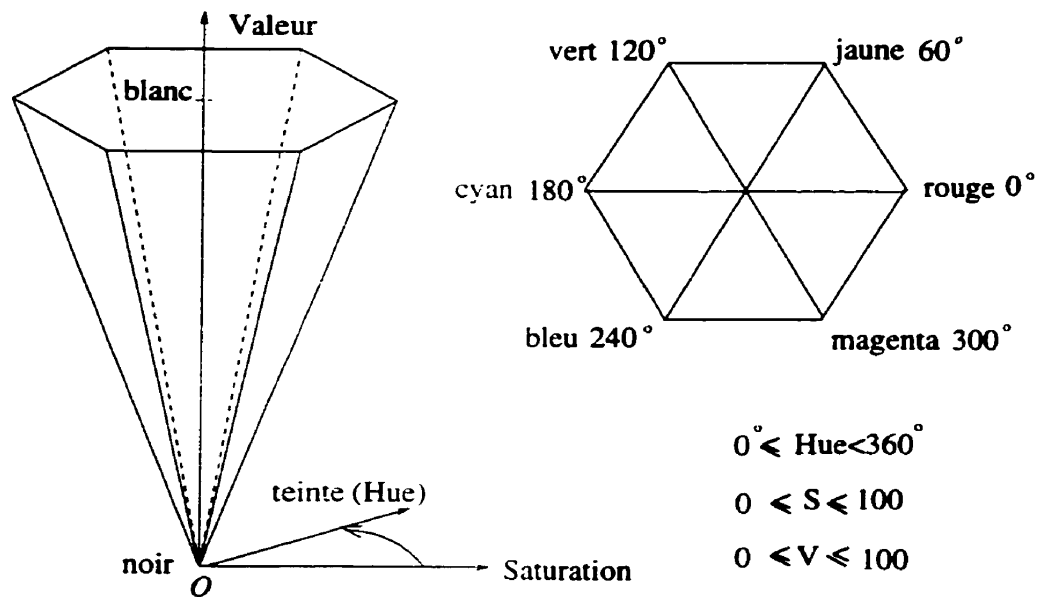


Figure 4.1 : Cône hexagonal HSV

Les composantes (H,S,V) sont obtenues à partir des composantes (R,G,B) à l'aide des équations suivantes (Foley, van Dam, Feiner et Hughes, 1996) où r, g et b, chacun dans l'intervalle  $[0,1]$  désignent respectivement les composantes R,G,B et h dans  $[0,360)$ , s et v dans  $[0,1]$  les composantes H.S.V :

$$\max = \text{maximum}(r, g, b)$$

$$\min = \text{minimum}(r, g, b)$$

```

v=max

si (max  $\neq$  0)

    s=(max-min)/max

si max==0

    s=0 et h est non défini

sinon delta=max-min

si (r==max)

    h=(g-b)/delta

sinon si (g==max)

    h=2.0+(b-r)/delta

sinon si (b==max)

    h=4.0+(r-g)/delta

h=h*60.0

si h<0.0

    h=h+360.0

```

On choisit ainsi 5 couleurs espacées de manière équidistantes dans l'espace des teintes (Hue). Les différents essais avec la caméra ont conduit au choix des teintes suivantes : blanc, magenta, orange, cyan, vert et jaune.

## 4.2 Recouvrement de la position et de l'orientation de la paume

Afin de détecter la position et l'orientation de la paume, on choisit d'orner la partie supérieure de la main de trois points  $A$ ,  $B$  et  $C$  formant un triangle dont le point de base  $A$  est centré au milieu de poignet et correspond au centre de rotation pour l'angle d'abduction des doigts. Les dimensions de ce triangle ainsi que son orientation définie dans le repère  $R_C$  de la figure 3.5 sont connues. Les trois points  $A$ ,  $B$  et  $C$  se projettent dans le plan image aux points respectifs  $p_A$ ,  $p_B$  et  $p_C$ . La caméra est placée de manière à ce que ces points d'intérêt soient toujours visibles. En pratique, la caméra est légèrement surélevée par rapport à la main. La détection dans l'image des trois points donne leurs coordonnées en 2D. Il existe deux solutions au problème de reconstruction du triangle (DeMenthon et Davis, 1990). Les auteurs proposent une alternative à la solution exacte qui, elle, ne s'exprime pas sous forme analytique. Ils utilisent l'approximation de la projection orthoperspective pour le modèle de la caméra, qui conduit à une expression simple des solutions. Là encore, il faut discriminer la solution exacte parmi les deux solutions possibles obtenues par symétrie miroir (figure 4.2). On considère le plan  $\pi$  contenant le point  $A$  orthogonal à la ligne de vue ( $Op_A$ ). Les points  $B_i$  et  $C_i$  avec  $i = 1, 2$  se projettent orthogonalement sur ce plan en  $H_B$  et  $H_C$ . Les points  $p_A$ ,  $p_B$  et  $p_C$  dans le plan image sont les images respectives de  $A$ ,  $H_B$  et  $H_C$  par projection perspective. Les triangles  $(A, B_1, C_1)$  et  $(A, B_2, C_2)$  constituent les

deux solutions au problème de reconstruction du triangle défini par les points image  $p_A$ ,  $p_B$  et  $p_C$ , sous l'hypothèse de projection orthoperspective.

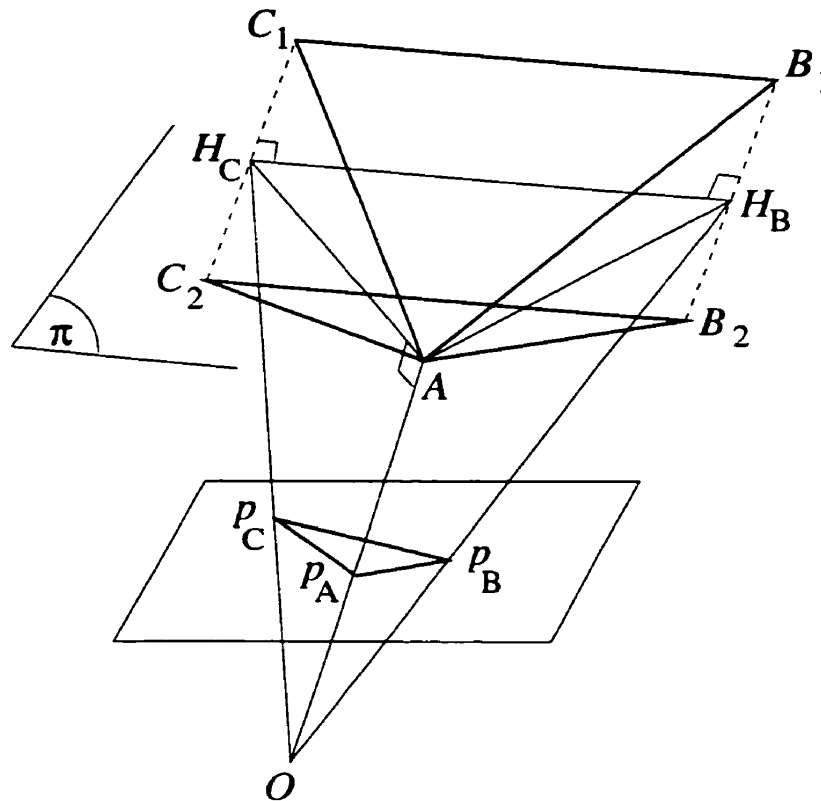


Figure 4.2 : Les deux solutions possibles au problème de reconstruction de la paume

Afin de discriminer la solution exacte, on effectue une prédiction du vecteur normal à la paume à partir des dérivées premières et secondes de ses coordonnées sur les images précédentes. On sélectionne parmi les normales des deux triangles candidats celle dont l'angle avec la normale prédite est minimal. Le triangle correspondant est alors choisi comme caractéristique de la position et orientation de la paume à l'image suivante. Il importe alors de connaître l'orientation initiale de la paume afin de lancer le processus de suivi.

## 4.3 Recouvrement de la position des doigts

### 4.3.1 Cas des doigts II à V

La valeur de l'angle d'abduction  $q_0$  détermine le plan de mouvement du doigt considéré. À partir de la connaissance de ce plan, la position 3D de l'extrémité du doigt est connue par intersection avec la ligne de mire passant par l'indice dans le plan image (figure 4.3).

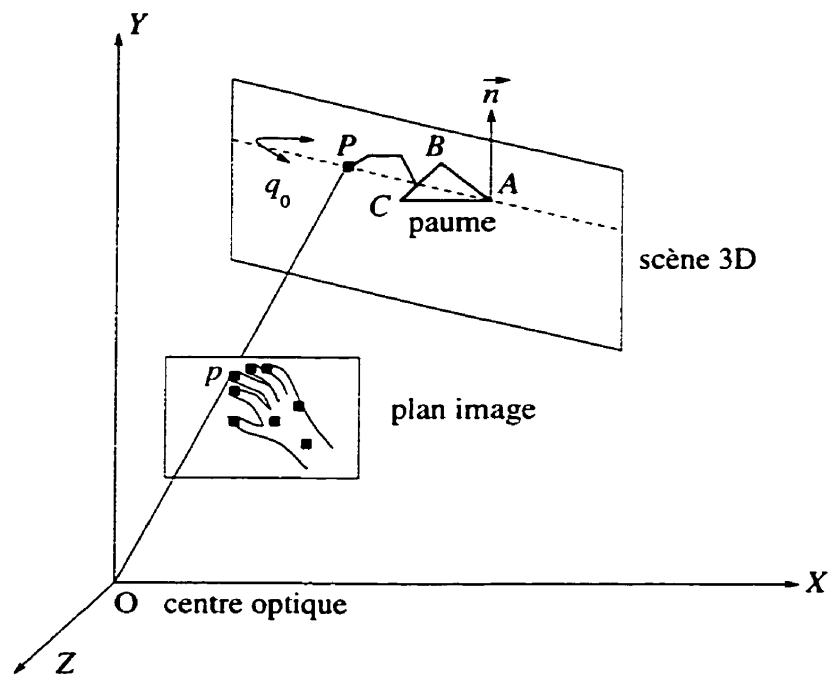


Figure 4.3 : Reconstruction de la position 3D des extrémités des doigts

Les équations de cinématique inverse donnent alors les angles de flexion  $q_1$ ,  $q_2$ , et  $q_3$ . Cependant, l'angle d'abduction  $q_0$ , donc la position du plan de mouvement, ne peut pas être déterminé de façon unique à partir de la seule connaissance de la position de

l'extrémité du doigt en 2D dans le plan image. Pour tout angle  $q_0$  donné, il existe en effet une solution unique pour les angles de flexion  $q_1$ ,  $q_2$ , et  $q_3$  qui conduit à la même projection de l'extrémité du doigt considéré dans le plan image, comme l'illustre la figure 4.4 pour le cas de l'index.

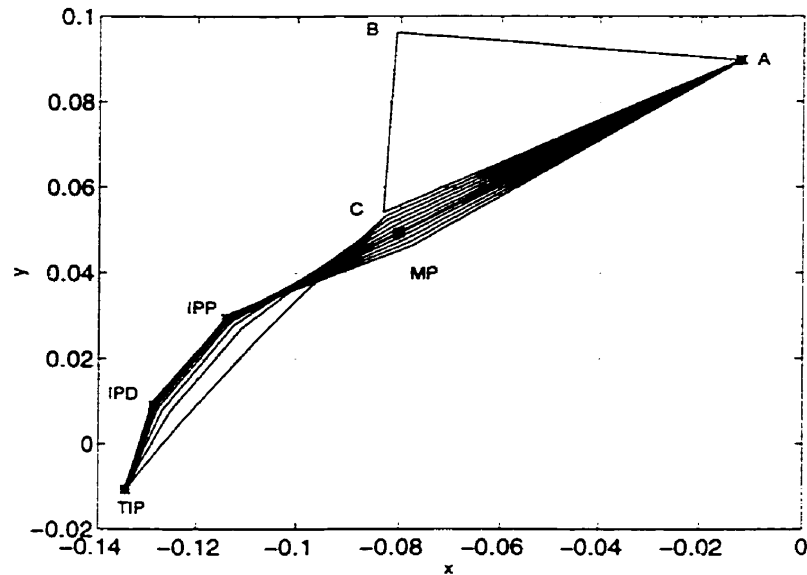


Figure 4.4 : Solutions multiples au problème de cinématique inverse du doigt

Sur cette figure, TIP désigne la position de l'indice de l'extrémité du doigt dans le plan image, et MP, IPP et IPD respectivement les articulations métacarpo-phalangienne, interphalangienne proximale et interphalangienne distale.

Il s'ensuit que l'angle d'abduction  $q_0$  est déterminé par un processus d'optimisation. On calcule toutes les solutions possibles sur l'intervalle des angles  $q_0$  admissibles physiologiquement. Parmi les angles  $q_0$  qui conduisent à des angles de flexion  $q_1$ ,  $q_2$ , et  $q_3$  admissibles, on sélectionne finalement l'angle  $q_0$  le plus éloigné des configurations

limites. La position finale est ainsi à égale distance entre la position tendue du doigt et la position fléchie à l'extrême. La figure 4.5 illustre dans l'espace la solution choisie mise en évidence à l'aide d'étoiles dans le cas de l'index.

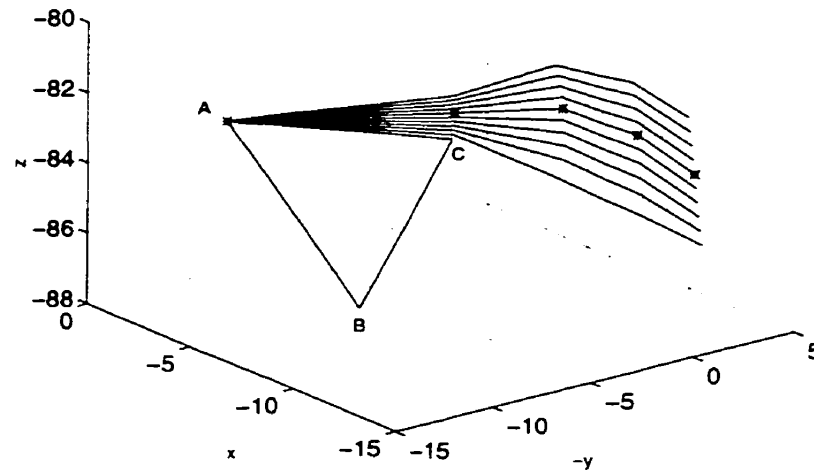


Figure 4.5 : Sélection de la configuration la plus éloignée des configurations limites

### 4.3.2 Cas du pouce

En ce qui concerne le pouce, son plan de mouvement est déterminé par les 2 angles  $q_0$  et  $q_1$ . L'intersection entre ce plan et la ligne de vue passant par l'indice du pouce dans le plan image donne les coordonnées 3D de la position  $P$  de l'extrémité du pouce pour des angles  $q_0$  et  $q_1$  particuliers. Comme précédemment dit, pour que ce point  $P$  soit une solution possible au problème de reconstruction du pouce, il importe que ce point reconstruit soit le plus proche possible de la surface de cinématique du pouce pour  $q_0$  et  $q_1$  donnés. Nous allons donc calculer, pour l'ensemble des valeurs  $q_0$  et  $q_1$



admissibles, la distance  $\delta_P$  entre le point  $P$  reconstruit et la surface de cinématique du pouce correspondante. Les résultats de cinématique inverse obtenus au chapitre 3 sont rappelés brièvement ici. Soit  $[X_P, Y_P, Z_P]$  les coordonnées de  $P$  dans le référentiel  $R_p$  de la figure 3.8. Les coordonnées du point  $M$  le plus proche de  $P$  appartenant à la surface de cinématique du pouce sont données par les équations 4.1 :

$$\begin{cases} X_M = \cos q_1(d_1 - d_3 + d_2 \cos q_2 + 2d_3 \cos q_2^2), \\ Y_M = \sin q_1(d_1 - d_3 + d_2 \cos q_2 + 2d_3 \cos q_2^2), \\ Z_M = -\sin q_2(d_2 + 2d_3 \cos q_2), \end{cases} \quad (4.1)$$

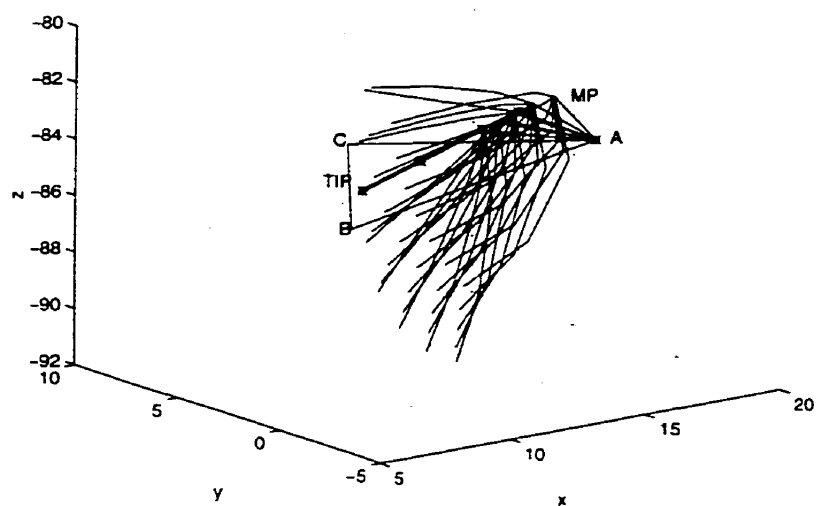
dans lesquelles  $q_1$  et  $q_2$  sont les résultats des équations de minimisation de la distance  $MP$  (4.2) :

$$\begin{cases} \sin q_1 X_P - \cos q_1 Y_P = 0, \\ [d_2 + 4d_3 \cos q_2][(X_P \cos q_1 + Y_P \sin q_1 - d_1) \sin q_2 + Z_P \cos q_2] \\ - 2d_3 Z_P - d_2 d_3 \sin q_2 = 0. \end{cases} \quad (4.2)$$

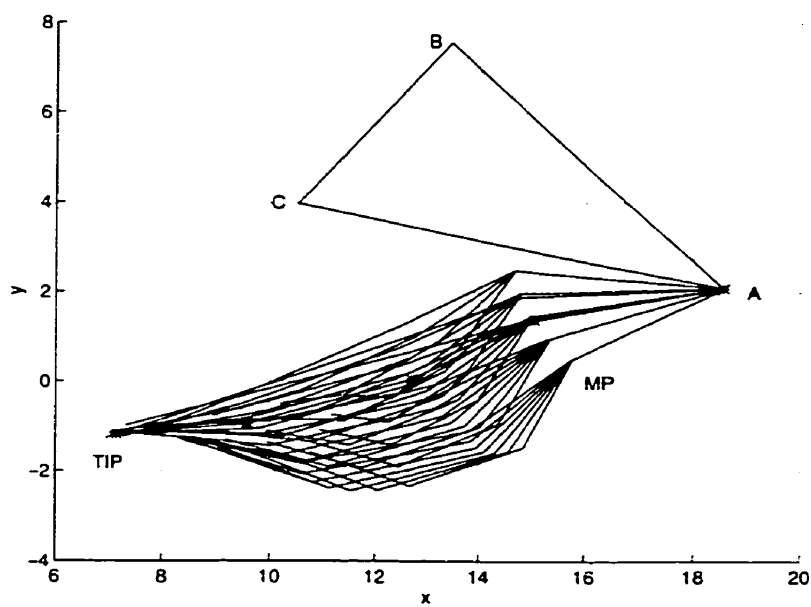
C'est ainsi que l'on obtient la distance  $\delta_P = MP$ .

Comme c'est le cas pour les autres doigts, il existe plusieurs solutions admissibles au problème de cinématique inverse du pouce, c'est-à-dire plusieurs configurations en  $q_0$  et  $q_1$  conduisant à la même projection ou à une projection très voisine de l'extrémité du pouce dans le plan image. Il ne suffit pas de minimiser la distance  $\delta_P$  pour trouver la solution au problème. En effet, cela signifierait seulement que le point  $P$

reconstruit se trouve sur la surface de cinématique du pouce. Cela n'implique par que les angles  $q_0$  et  $q_1$  calculés en conséquence sont les bons, étant donné qu'ils sont fortement dépendants du modèle cinématique choisi pour le pouce et non de la réalité. En conséquence, on sélectionne les points  $P$  tels que  $\delta_P$  soit inférieur à un certain seuil, ceci afin d'assurer que le point  $M$  choisi sur la surface de cinématique du pouce se projette en un point très proche de l'extrémité du pouce détectée. En pratique, ce seuil est choisi de l'ordre de 4 mm. Ensuite, par analogie avec le cas des autres doigts, on sélectionne parmi les angles  $q_0$  correspondants aux points  $P$  admissibles la valeur médiane afin de s'éloigner des configurations extrêmes. Enfin, pour cette valeur  $q_0$  sélectionnée, on choisit, parmi les valeurs de  $q_1$  validées, la valeur médiane en  $q_1$ . Les valeurs des angles  $q_2$  et  $q_3$  sont alors connues ( $q_2$  par l'équation 4.2 et  $q_3 = q_2$ ). La figure 4.6 illustre un exemple de reconstruction sur lequel la solution est représentée à l'aide d'étoiles. On remarque que plusieurs configurations en  $q_0$  et  $q_1$  nettement distinctes en 3D mènent à des projections relativement proches de la solution TIP dans le plan image. La figure 4.7 illustre la position de l'extrémité du pouce solution représentée par le label TIP par rapport à la surface de cinématique du pouce pour l'angle  $q_0$  solution. La figure 4.8 donne la projection dans le plan image de la surface précédente.



(a) Dans l'espace (cm)



(b) Projection dans le plan image (cm)

Figure 4.6 : Positions de l'extrémité du pouce pour différentes valeurs de  $q_0$  et  $q_1$

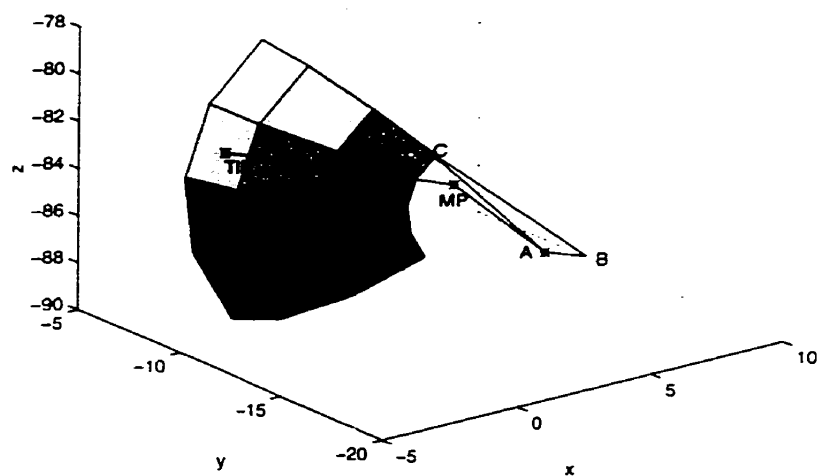


Figure 4.7 : Extrémité du pouce solution, TIP, par rapport à la surface de mouvement du pouce

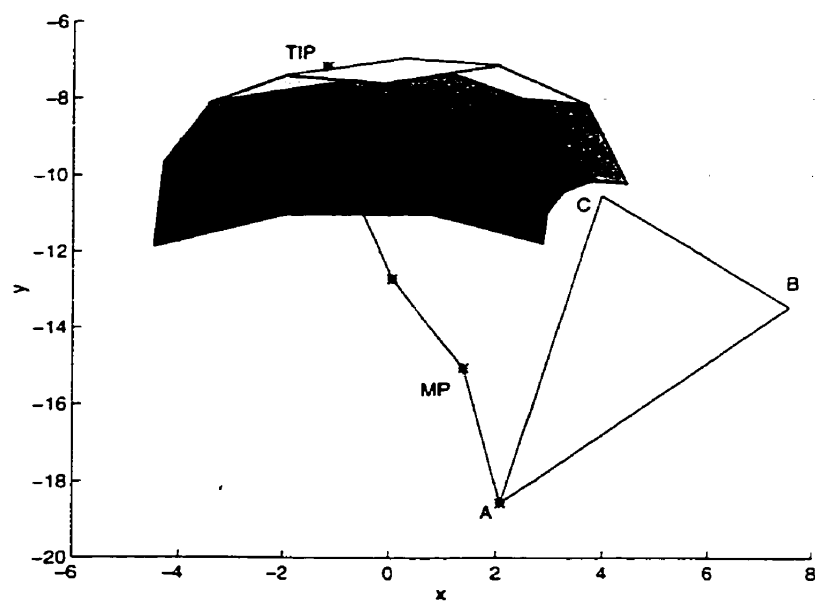


Figure 4.8 : Projection dans le plan image de la solution

## 4.4 Lissage temporel des résultats

### 4.4.1 Filtre de Kalman

Une méthode efficace souvent utilisée dans le cas de détection et suivi de cible est le filtre de Kalman. Il s'agit d'un algorithme de prédiction/correction de l'estimée du vecteur d'état d'un système, optimal au sens des moindres carrés parmi la classe des estimateurs linéaires sans biais de l'état d'un système linéaire avec bruits non corrélés. L'observation de l'état se fait à partir d'une sortie spécifiée par un vecteur de mesure et le gain  $G(k)$  de l'observateur doit minimiser au sens des moindres carrés l'erreur de l'estimée. L'utilisation de cette méthode est envisageable pour améliorer la qualité et le lissage des angles obtenus par reconstruction. Soit  $\underline{X} \in \mathbb{R}^n$  le vecteur d'état et  $\underline{V} \in \mathbb{R}^q$  le vecteur de mesure.

Le modèle d'état est régi par les équations suivantes :

$$\begin{cases} \underline{x}_{k+1} = \underline{x}_k + \Gamma_k \underline{\xi}_k \\ \underline{v}_k = g_k(\underline{x}_k) + \underline{\eta}_k \end{cases}$$

où  $\{\xi_k\}$  et  $\{\eta_k\}$  sont des séquences non corrélées et non corrélées avec l'état initial de bruit blanc gaussien centré telles que  $\text{Var}(\xi_k) = Q_k$  et  $\text{Var}(\eta_k) = R_k$ .  $\xi_k$  est l'erreur de modélisation et  $\eta_k$  le bruit de mesure. Ce modèle est ensuite linéarisé conduisant aux

équations suivantes :

$$\begin{cases} \underline{x}_{k+1} = \underline{x}_k + \Gamma_k \underline{\xi}_k \\ \underline{w}_k = C_k \underline{x}_k + \underline{\eta}_k \end{cases}$$

où

$$C_k = \left[ \frac{\partial g_k}{\partial x_k}(\hat{x}_{k|k-1}) \right].$$

On peut alors appliquer les formules du filtre de Kalman étendu :

$$\left\{ \begin{array}{l} P_{0,0} = \text{Var}(x_0), \quad \hat{x}_{0|0} = E(x_0). \\ \text{Pour } k = 1, 2, \dots, \\ \text{Phase de prédiction} \\ P_{k,k-1} = P_{k-1,k-1} + Q_{k-1} \\ \hat{x}_{k|k-1} = \hat{x}_{k-1|k-1} \\ G_k = P_{k,k-1} \left[ \frac{\partial g_k}{\partial x_k}(\hat{x}_{k|k-1}) \right]^T \cdot \left[ \left[ \frac{\partial g_k}{\partial x_k}(\hat{x}_{k|k-1}) \right] P_{k,k-1} \left[ \frac{\partial g_k}{\partial x_k}(\hat{x}_{k|k-1}) \right]^T + R_k \right]^{-1} \\ \text{Phase de correction} \\ P_{k,k} = \left[ I - G_k \left[ \frac{\partial g_k}{\partial x_k}(\hat{x}_{k|k-1}) \right] \right] P_{k,k-1} \\ \hat{x}_{k|k} = \hat{x}_{k|k-1} + G_k \left[ v_k - g_k(\hat{x}_{k|k-1}) \right] \end{array} \right.$$

Il est possible d'envisager plusieurs approches quant au choix du vecteur de mesure.

- Mesure des coordonnées en 3D

Le vecteur de mesure peut tout d'abord être construit avec les coordonnées 3D des extrémités des doigts, afin de simplifier les calculs de dérivées partielles

dans l'étape de linéarisation. Soit  $\underline{X} \in \mathbb{R}^{21}$  le vecteur d'état formé des variables d'articulation indépendantes de la main et  $\underline{V} \in \mathbb{R}^{21}$  le vecteur de mesure formé des coordonnées 3D des extrémités des doigts ainsi que de la position et orientation de la paume.

$$\begin{aligned}\underline{X} &= [\text{doigt}[0].q[0] \text{ doigt}[0].q[1] \text{ doigt}[0].q[2] \cdots \text{doigt}[4].q[0] \text{ doigt}[4].q[1] \\ &\quad \text{doigt}[4].q[2] \ X_P \ Y_P \ Z_P \ \alpha \ \beta \ \gamma]^T \\ \underline{V} &= [X[0] \ Y[0] \ Z[0] \cdots X[4] \ Y[4] \ Z[4] \ X_P \ Y_P \ Z_P \ \alpha \ \beta \ \gamma]^T\end{aligned}$$

où  $\text{doigt}[i]$  représente le  $i^{\text{e}}$  doigt,  $q[i]$  le  $i^{\text{e}}$  angle de flexion et  $[X[i] \ Y[i] \ Z[i]]$  les coordonnées 3D de l'extrémité du  $i^{\text{e}}$  doigt. Le point  $P[X_P \ Y_P \ Z_P]$  représente la position de référence de la paume et  $[\alpha \ \beta \ \gamma]$  son orientation en angles d'Euler. Si l'on analyse de plus près le vecteur de mesure, on se rend aisément compte que l'erreur de reconstruction 3D est implicitement contenue à l'intérieur du vecteur de mesure. En effet, en l'absence de connaissance du processus de reconstruction 3D, le filtre de Kalman ne pourrait rendre de façon fidèle la pose de la main telle qu'elle apparaît dans l'image. Cette approche n'est donc pas valide.

#### ● Mesure des coordonnées en 2D

Afin d'appliquer le filtre avec le maximum d'information sensorielle, le vecteur de mesure pourrait comprendre cette fois-ci les coordonnées  $[x[i], y[i]]$  des indices de couleur dans l'image. Il inclurait les cinq doigts ainsi que les projections des

3 points  $A, B$  et  $C$  de la paume d'où  $\underline{V} \in \mathbb{R}^{16}$ . La fonction  $g_k$  comprendrait alors l'information du modèle de projection de la caméra. Le vecteur d'état serait identique au cas précédent.

$$\underline{V} = [x[0] \ y[0] \ \cdots \ x[4] \ y[4] \ x_A \ y_A \ x_B \ y_B \ x_C \ y_C]^T$$

Cependant, là encore, les résultats ne seraient pas concluants à cause de l'absence de modèle dynamique du comportement de la main, et donc de l'imprécision du modèle d'état. Prédire la posture de la main uniquement à partir de la posture précédente additionnée de bruit est insuffisant. En outre, il n'est pas envisageable dans le cadre de ce projet de prétendre modéliser le comportement dynamique de la main. De plus, il est extrêmement difficile de modéliser les matrices de covariance du vecteur de mesure car des corrélations existent entre les différentes entrées à cause du processus de détection. Enfin, la contrainte d'une détection en temps réel risquerait de ne pas être vérifiée compte-tenu de l'étape d'inversion de la matrice apparaissant dans les équations du filtre de Kalman. Par conséquent, le filtre de Kalman a été laissé de côté au profit d'une méthode de lissage classique sans prédiction des variables d'articulation du modèle de la main.



### 4.4.2 Lissage envisagé

Le processus de reconstruction des variables d'articulation de la main est exécuté sans prédiction préalable. Un lissage temporel est appliqué à l'ensemble des grandeurs d'état du modèle cinématique de la main afin d'assurer un mouvement fluide dans le cas de la main en mouvement. Le lissage envisagé s'inspire d'un filtre avec perte de mémoire exponentielle dans lequel le poids  $\alpha \in [0, 1]$  accordé à la dernière valeur prédite constitue le paramètre à fixer. L'équation est la suivante :

$$X_p(t) = \alpha X_p(t-1) + (1 - \alpha) X_m(t)$$

où  $X_p(t)$  désigne la grandeur prédite (lissée) à l'instant  $t$  et  $X_m(t)$  la grandeur mesurée à l'instant  $t$ . Ce lissage permet de pallier l'absence de prédiction des angles de la main et donc de gommer les soubresauts dans les résultats obtenus.

## 4.5 Tests et validation

### 4.5.1 Résultats obtenus sur des séquences de synthèse

Afin de valider la pertinence de la méthode de reconstruction envisagée, la technique de détection des indices de couleur sur la paume suivie de la reconstruction des variables d'articulation de la main a été testée sur des séquences de synthèse de la main. Une centaine de séquences de la main en mouvement ont été produites

à l'aide de la librairie graphique OpenGL et du modèle infographique de la main (chapitre 3). On suppose ici connus les longueurs des phalanges de chaque doigt ainsi que les paramètres intrinsèques de la caméra (distance focale et ouverture). Chaque séquence est constituée de vingt-cinq images et toutes les variables de configuration de la main (vingt-six degrés de liberté) sont connues pour chacune d'entre elles. L'algorithme de détection des indices de couleur a été appliqué sur chaque séquence suivi de la reconstruction de la posture de la main avec lissage en utilisant la méthode précédemment décrite. Le paramètre  $\alpha$  du processus de lissage a été fixé à 0.55 après réglage sur une séquence de test. Pour chacune des images de chaque séquence ont été calculées les erreurs entre les variables d'articulation reconstruites et celles du modèle. Les moyennes des erreurs obtenues sur l'ensemble des séquences sont reportées dans les tableaux 4.1 à 4.5 pour le cas des doigts ainsi que dans les tableaux 4.6 et 4.7 pour la configuration de la paume. Les angles  $q_0$ ,  $q_1$  et  $q_2$  dénotent respectivement l'abduction, la flexion au niveau de la jointure métacarpo-phalangienne et la flexion au niveau de la jointure interphalangienne proximale. La position du poignet est donnée par le point du triangle de la paume situé au centre du poignet et l'orientation du plan de la paume est donnée par les trois angles d'Euler : lacet ( $\alpha$ ), tangage ( $\beta$ ) et roulis ( $\gamma$ ). Les grandeurs étudiées comprennent la moyenne des valeurs absolues des erreurs, le biais (valeur absolue de la différence entre la moyenne de l'angle modèle et la moyenne de l'angle reconstruit), la valeur absolue de l'erreur maximale et pour finir l'écart type de la valeur absolue de l'erreur.

Tableau 4.1 : Pouce

grandeur	$q_0$	$q_1$	$q_2$
erreur moyenne en degrés	4.2	11.0	9.0
biais en degrés	3.5	6.2	8.6
erreur maximale en degrés	7.4	24.5	15.3
écart type en degrés	1.9	7.0	4.0
vitesse moyenne en deg/s	0.0	56.0	-13.6

Tableau 4.2 : Index

grandeur	$q_0$	$q_1$	$q_2$
erreur moyenne en degrés	0.8	4.9	8.6
biais en degrés	0.6	3.3	5.0
erreur maximale en degrés	1.9	11.4	17.9
écart type en degrés	0.5	3.0	5.0
vitesse moyenne en deg/s	0.1	15.0	29.5

Étudions pour commencer l'erreur en abduction. On note la présence d'une erreur moyenne et d'un biais relativement semblables autour de 3-4 degrés entre les valeurs initiales et les valeurs reconstruites des angles d'abduction pour tous les doigts. Ceci s'explique par le processus de reconstruction qui sélectionne l'angle d'abduction médian parmi les candidats possibles. Il existe donc un écart avec la solution exacte. Ce biais n'est pas systématique, ni en valeur ni en signe, étant donné qu'il dépend du

Tableau 4.3 : Majeur

grandeur	$q_0$	$q_1$	$q_2$
erreur moyenne en degrés	3.3	8.7	4.8
biais en degrés	3.3	8.6	3.9
erreur maximale en degrés	4.6	13.7	10.0
écart type en degrés	0.7	2.5	2.7
vitesse moyenne en deg/s	-0.2	7.9	18.9

Tableau 4.4 : Annulaire

grandeur	$q_0$	$q_1$	$q_2$
erreur moyenne en degrés	4.5	11.4	7.2
biais en degrés	4.5	11.3	6.3
erreur maximale en degrés	6.1	17.4	13.7
écart type en degrés	0.8	3.4	3.4
vitesse moyenne en deg/s	-0.1	8.2	16.9

Tableau 4.5 : Auriculaire

grandeur	$q_0$	$q_1$	$q_2$
erreur moyenne en degrés	6.0	15.7	10.9
biais en degrés	6.0	15.7	10.4
erreur maximale en degrés	8.6	24.8	19.3
écart type en degrés	1.2	4.8	4.2
vitesse moyenne en deg/s	-0.2	16.1	18.4

point de vue de la main qui engendre plus ou moins de solutions possibles au problème de reconstruction de l'angle d'abduction. Néanmoins, il est très difficile de réduire cette erreur car l'indétermination au niveau de l'angle d'abduction est inhérente au choix de l'analyse *dans l'image* des indices de couleur. Naturellement, cette imprécision se répercute au niveau des angles de flexion de chaque doigt étant donné qu'ils sont calculés dans le plan de mouvement du doigt déterminé par l'angle d'abduction

Tableau 4.6 : Position du poignet

grandeur	$X$	$Y$	$Z$
erreur moyenne en cm	0.5	0.3	0.9
biais en cm	0.4	0.2	0.9
erreur maximale en cm	1.4	0.7	2.2
écart type en cm	0.5	0.2	0.6

Tableau 4.7 : Orientation du plan de la paume

grandeur	$\alpha$	$\beta$	$\gamma$
erreur moyenne en degrés	0.5	3.0	1.5
biais en degrés	0.4	2.9	1.4
erreur maximale en degrés	1.4	5.1	4.2
écart type en degrés	0.5	1.1	0.8

calculé précédemment. En ce qui concerne plus précisément les angles de flexion, on observe une erreur moyenne respectivement de l'ordre de 10 et 8 degrés pour  $q_1$  et  $q_2$ . Là encore, il s'agit d'un biais résultant cette fois de l'imprécision sur l'angle d'abduction. L'erreur maximale atteint en moyenne 18 degrés pour  $q_1$  et 15 degrés pour  $q_2$ , ce qui n'altère malgré tout pas de façon significative l'allure de la main reconstruite. D'autre part, l'erreur reste contrainte autour de sa valeur moyenne comme en témoignent les faibles valeurs de l'écart type autour de 4 degrés.

Avant d'illustrer les remarques précédentes d'images, un mot sur le recouvrement de la position et de l'orientation du triangle de la paume. L'erreur moyenne en position se situe autour de moins de 1 cm en profondeur dans l'espace et 4 mm en moyenne en X et Y, ce qui est très correct. De même, l'orientation est recouverte à quelque deux degrés près. Ainsi, la position de la paume est très bien détectée et reconstruite, tout comme son orientation. Les tests de reconstruction de la posture des doigts effectués avec les valeurs *exactes* de la position et de l'orientation de la paume démontrent que le gain est inférieur à un degré sur les angles des articulations des doigts. Donc, l'erreur obtenue dans la posture de la paume n'intervient pas dans la précision de

la reconstruction de la posture des doigts. Ce n'est qu'à partir d'une erreur de 4 pixels sur la position du centroïde des taches de la paume que l'erreur en angles croît de façon significative (4 degrés d'erreur de plus pour les angles de flexion). L'angle d'abduction, lui, n'est pas affecté, ce qui s'explique par le processus de sélection de la grandeur médiane.

Afin d'accompagner d'images l'explication des résultats obtenus, voici l'illustration de la reconstruction d'une séquence d'images avec les valeurs des angles à l'appui (figure 4.9). Il est à noter qu'un lissage a été effectué sur la position du poignet afin de fluidifier le mouvement général de la main, ce qui explique le léger décalage de la position de la réplique en hauteur. Cependant, ce lissage n'entre pas en compte pour la reconstruction des variables d'articulation des doigts. Pour la séquence illustrée, les valeurs des angles d'abduction initiales et reconstruites pour chaque doigt ont été reportées sur le même graphique, ce qui est rendu possible par l'absence de chevauchement des doigts (figure 4.10).

On observe une erreur d'environ 3-4 degrés au niveau de l'angle d'abduction (figure 4.10), qui est du même signe et du même ordre de grandeur pour tous les doigts sauf le pouce. C'est la configuration particulière de la prise de vue couplée à la sélection de la valeur médiane en abduction qui produit ce décalage répétable pour tous les doigts. Cette erreur est malgré tout acceptable visuellement. La différence est la plus marquée sur l'image initiale au niveau de l'auriculaire. L'erreur sur l'angle d'abduction se traduit par une flexion plus importante du petit doigt (figure 4.9).



(a) Séquence originale



(b) Séquence reconstruite

Figure 4.9 : Portions des séquences de synthèse originale et reconstruite

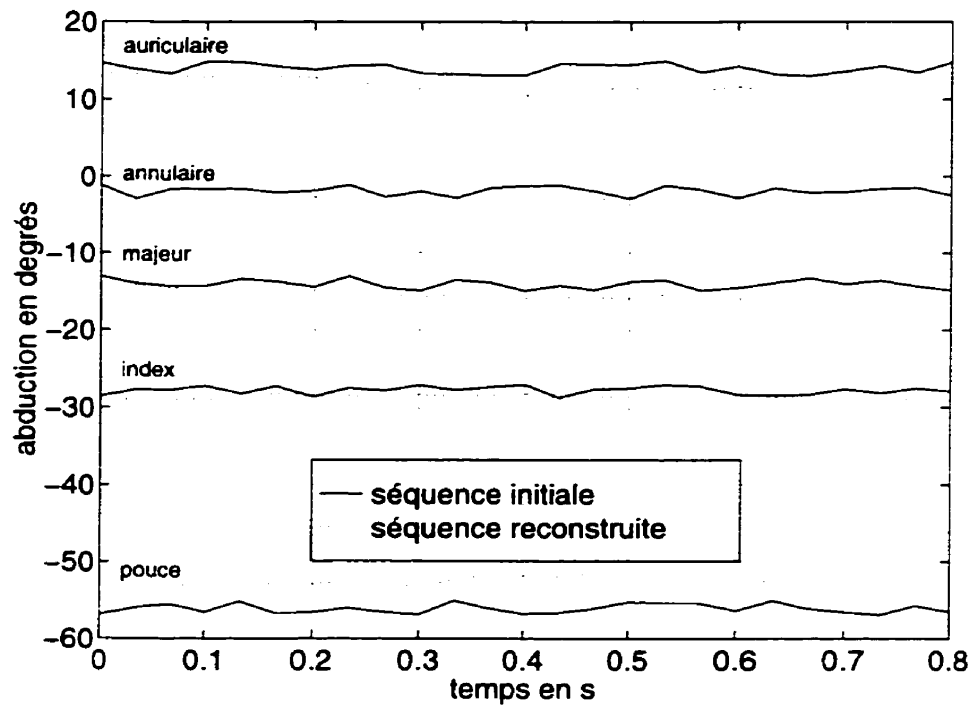
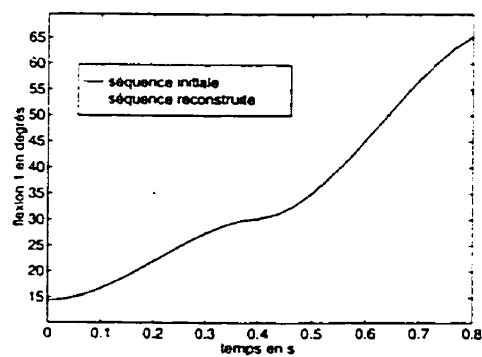
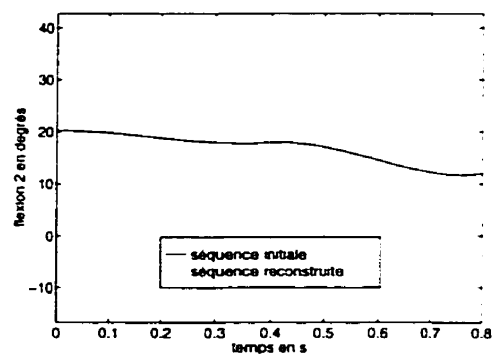


Figure 4.10 : Angle d'abduction de chaque doigt



(a) Flexion MP



(b) Flexion IPP

Figure 4.11 : Reconstruction du pouce pour la séquence étudiée



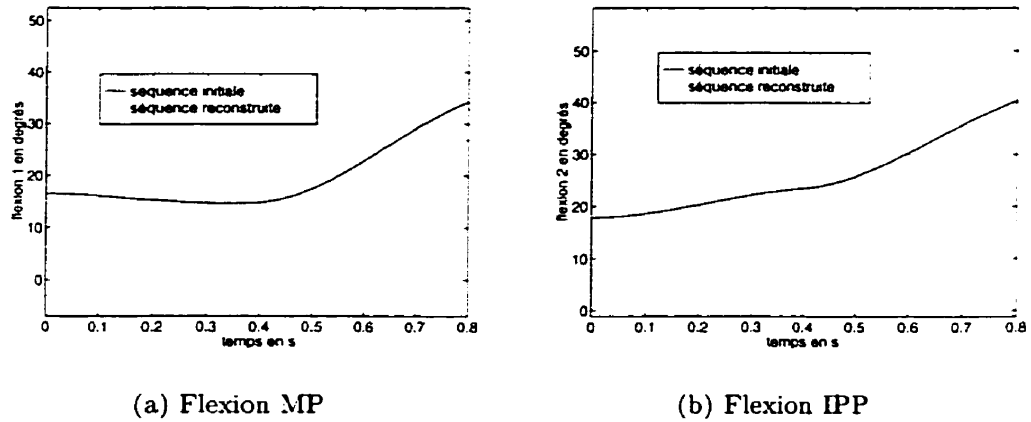


Figure 4.12 : Reconstruction de l'index pour la séquence étudiée

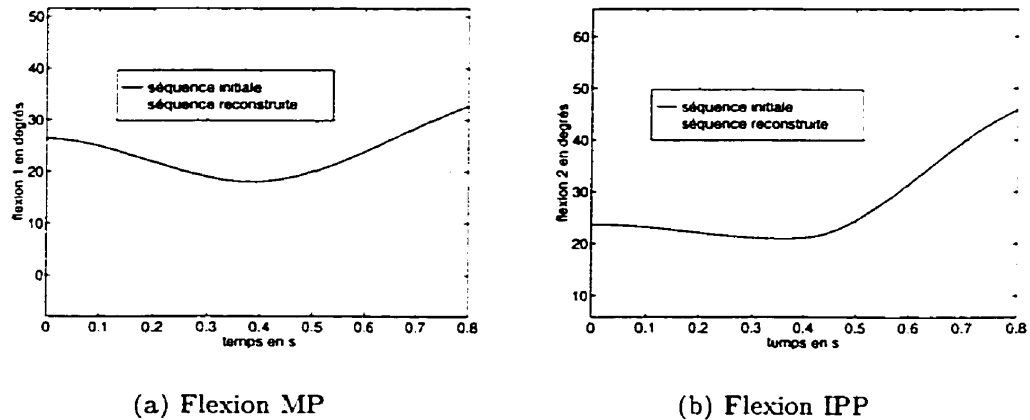
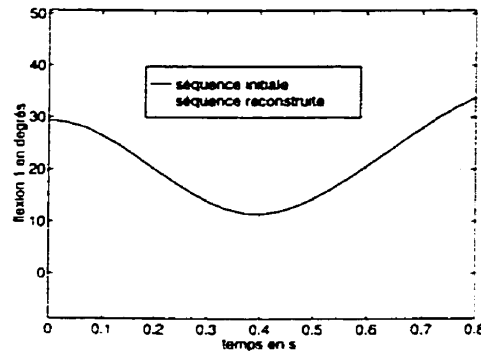
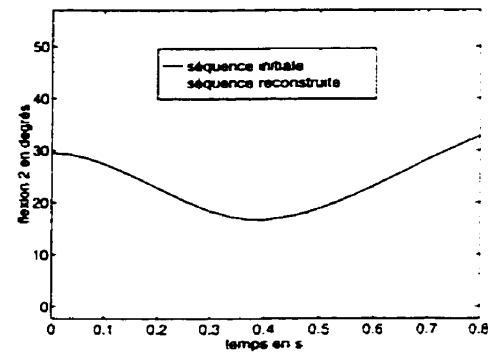


Figure 4.13 : Reconstruction du majeur pour la séquence étudiée

Les angles de flexion sont reportés aux figures 4.11 à 4.15 sur lesquelles MP désigne l'articulation métacarpo-phalangienne et IPP l'articulation interphalangienne proximale. La flexion au niveau de l'articulation interphalangienne distale IPD n'est pas représentée ici étant donné qu'elle est liée linéairement à l'articulation précédente par le choix initial du modèle cinématique de la main (chapitre 3). On note que l'erreur en

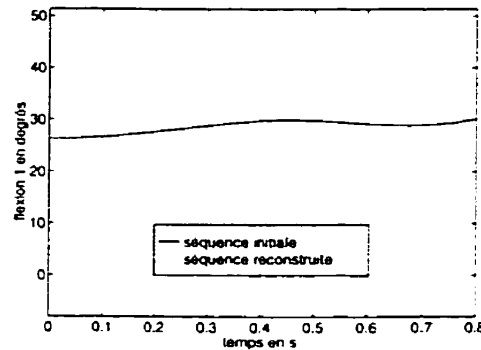


(a) Flexion MP

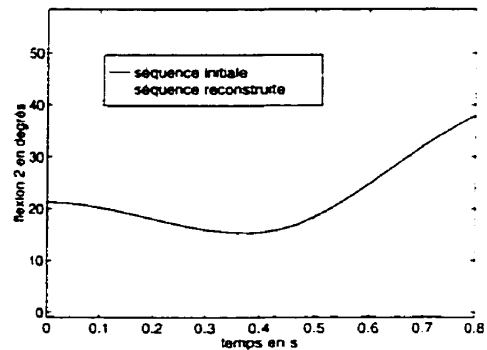


(b) Flexion IPP

Figure 4.14 : Reconstruction de l'annulaire pour la séquence étudiée



(a) Flexion MP



(b) Flexion IPP

Figure 4.15 : Reconstruction de l'auriculaire pour la séquence étudiée

flexion est légèrement plus importante puisqu'elle avoisine les 9 degrés. Les courbes reconstruites apparaissent translatées par rapport aux courbes de référence. Ceci s'explique par l'erreur introduite par l'angle d'abduction. Sur la première partie de la trajectoire, l'angle de flexion MP est légèrement inférieur tandis que l'angle de flexion

IPP est supérieur afin de compenser le décalage en abduction. Sur la seconde partie de la trajectoire, le phénomène s'inverse pour l'angle de flexion IPP. On remarque que le pouce présente des erreurs du même ordre de grandeur que celles des autres doigts, cependant l'allure des courbes est moins fidèle à celle des courbes d'origine. Ceci s'explique par le processus de reconstruction du pouce qui lève deux indéterminations à savoir l'angle d'abduction  $q_0$  et le premier angle de flexion  $q_1$  qui déterminent à eux deux le plan de mouvement final du pouce. Rappelons que le pouce n'est pas modélisé, lui, par un manipulateur planaire, ce qui complique la reconstruction de ses variables d'articulation. De plus, l'approche choisie pour résoudre sa reconstruction consiste à *forcer* la main réelle à adopter le modèle de cinématique du pouce (processus de minimisation de la distance de l'extrémité du pouce à la surface de cinématique du pouce) à chaque instant, ce qui se traduit par des données plus chaotiques que dans le cas des autres doigts.

Il résulte que le modèle de la main reste très cohérent avec la configuration initiale. La figure 4.16 illustre le cas d'une reconstruction parmi les plus mauvaises en termes d'erreurs d'angles dans les séquences analysées. Plusieurs vues de la main sont présentées afin de clarifier la pose dans l'espace. Il est clair que le décalage en abduction apparaît ainsi que l'erreur en abduction et en flexion du pouce. Cependant, la pose reste malgré tout cohérente avec l'originale si l'on envisage une application de reconnaissance de gestes clairement distincts ou une application de robotique nécessitant des propriétés de saisie.

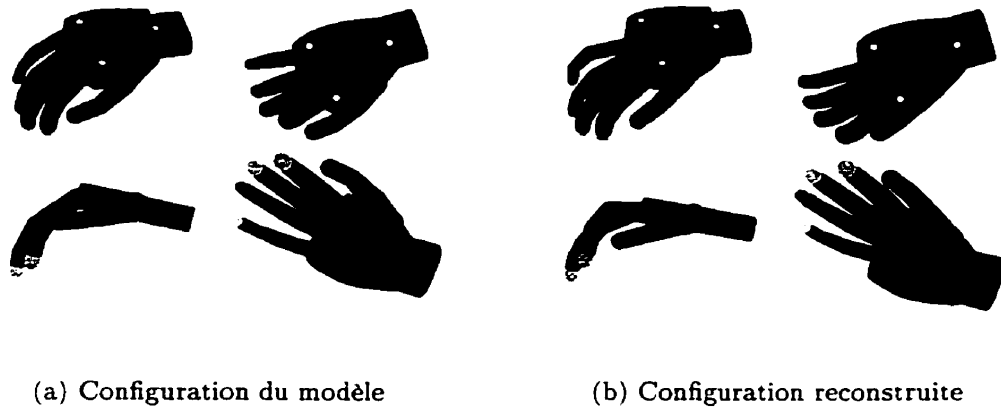


Figure 4.16 : Cas d'une des plus mauvaises reconstructions

#### 4.5.2 Résultats obtenus sur une séquence réelle

La main est filmée dans un mouvement laissant visibles tous les indices de couleur afin de valider la partie de reconstruction des variables d'articulation de la main. Deux caméras disposées en vis-à-vis filment la main, ceci permettant de bénéficier de deux points de vue d'une même posture de la main et de valider ou non l'invariance du processus de reconstruction. La première étape consiste à calibrer les longueurs des phalanges ainsi que les angles limites en abduction. Pour ce faire, on reconstruit la posture de la main à plat, les doigts complètement serrés d'une part et les doigts ouverts au maximum d'autre part.

#### 4.5.2.1 Calibration des phalanges et des angles d'abduction limites

Les longueurs des phalanges sont calibrées afin de conduire à des angles de flexion minimaux sur chacune des images ouverte et fermée. L'étape de reconstruction donne ensuite les valeurs des angles limites en abduction, à savoir pour la main complètement ouverte ou totalement fermée (figure 4.17).

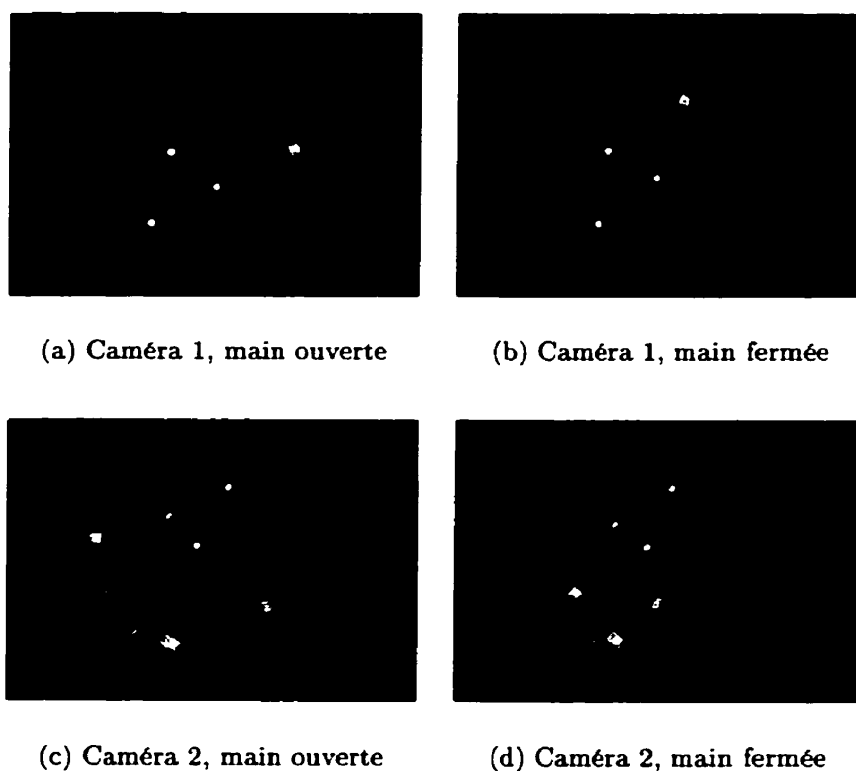


Figure 4.17 : Main ouverte et fermée prise de deux caméras

La calibration des longueurs des phalanges se fait à partir d'un fichier de référence obtenu par mesure directe sur le gant. Ce fichier contient les longueurs réelles des trois phalanges de chaque doigt. Cependant, l'utilisation de ces données brutes con-

duit souvent à des reconstructions erronées à cause des erreurs sur la localisation du centroïde des taches. Par exemple, en cas de doigt tendu, il se peut qu'aucune solution n'existe au problème de reconstruction car les longueurs fixées des phalanges sont trop *courtes* pour atteindre le centroïde de l'extrémité du doigt. Afin de résoudre cette difficulté, les deux images de la main ouverte et fermée sont utilisées pour calibrer les longueurs des phalanges afin d'assurer une solution au problème de cinématique inverse. Pour ce faire, on multiplie par un rapport croissant allant de 0.7 à 1.2 les longueurs des doigts du fichier de référence jusqu'à ce qu'une solution de reconstruction existe pour les deux images simultanément. On conserve alors la longueur minimale obtenue pour chaque doigt qui devient la nouvelle longueur de référence. Les résultats de la calibration des longueurs des phalanges pour chacune des caméras sont reportés aux tableaux 4.8 et 4.9.

Tableau 4.8 : Longueur des phalanges en cm, caméra 1

doigt	phalange 1	phalange 2	phalange 3
pouce	2.48	3.03	2.30
index	3.88	2.62	2.43
majeur	5.15	3.22	2.14
annulaire	4.38	3.46	2.88
auriculaire	4.02	3.05	2.91

L'erreur totale est comprise entre 0.17 cm pour l'index et 0.97 cm pour l'auriculaire. Cet écart est en partie dû au fait que la caméra 1 est pratiquement perpendiculaire au plan de la paume. Or, le processus de reconstruction de l'angle d'abduction fait intervenir l'intersection entre le plan de mouvement du doigt et la ligne de vue passant

Tableau 4.9 : Longueur des phalanges en cm, caméra 2

doigt	phalange 1	phalange 2	phalange 3
pouce	2.48	3.03	2.30
index	3.96	2.67	2.47
majeur	5.04	3.15	2.10
annulaire	4.10	3.24	2.70
auriculaire	3.63	2.75	2.63

par la position dans l'image de l'extrémité de couleur du doigt. Dans ce cas, les deux sont presque parallèles, introduisant ainsi une plus grande erreur de positionnement de l'extrémité du doigt en 3D. Mis à part le cas de l'auriculaire, les variations sont minimales. On conserve donc le fichier de calibration correspondant à chaque caméra, pour ainsi prendre en considération les variations induites par les orientations respectives des deux caméras.

Pour chaque caméra, ces longueurs conduisent à des angles minimaux en flexion afin d'assurer la pose à plat dans les deux configurations (main ouverte et main fermée).

Les angles résultant de la reconstruction sont détaillés dans les tableaux 4.10 et 4.11.

Tableau 4.10 : Angles en degrés de la main fermée reconstruite

doigt	Caméra 1			Caméra 2		
	$q_0$	$q_1$	$q_2$	$q_0$	$q_1$	$q_2$
pouce	-43.0	5.5	9.1	-44.5	14.5	7.4
index	-24.0	-7.5	18.7	-26.8	-10.1	24.0
majeur	-16.7	-6.7	20.2	-18.0	-9.6	22.7
annulaire	-7.9	-4.6	22.5	-6.6	-6.2	22.3
auriculaire	2.8	-2.5	25.9	6.2	-4.1	25.1

La différence moyenne en abduction est de 2.1 degrés entre les deux caméras pour

Tableau 4.11 : Angles en degrés de la main ouverte reconstruite

doigt	Caméra 1			Caméra 2		
	$q_0$	$q_1$	$q_2$	$q_0$	$q_1$	$q_2$
pouce	-70.0	4.0	6.4	-71.0	20.0	1.7
index	-34.0	-11.4	4.9	-37.7	-8.4	5.8
majeur	-21.2	-10.6	4.5	-23.3	-12.7	6.2
annulaire	-4.4	-9.3	5.9	-4.0	-13.1	4.8
auriculaire	17.1	-1.6	5.7	21.1	-8.1	1.4

une même pose, de 4.9 degrés en ce qui concerne la première flexion et 2.3 degrés pour la seconde flexion. La reconstruction en abduction reste proche d'une caméra à l'autre, ce qui assure le bon positionnement du plan de mouvement de chaque doigt. L'erreur en flexion est surtout due au cas du pouce avec des écarts de 9 et 16 degrés. En effet, à cause de sa non-planarité, il est difficile de calibrer à la fois la longueur du pouce et son premier angle de flexion, car plus la longueur du pouce est grande et plus il plie pour respecter la projection de son extrémité (cas de la main ouverte). Pour les autres doigts, la moyenne de l'erreur est de 3 degrés, ce qui est satisfaisant. Afin de mieux apprécier visuellement la différence entre les deux postures reconstruites et de juger de l'invariance du processus de reconstruction de la main, les figures 4.18 et 4.19 illustrent plusieurs points de vue de la main pour chaque caméra. Pour une pose donnée, les mains ont sensiblement la même allure quelle que soit la caméra. La calibration est donc bien réussie.



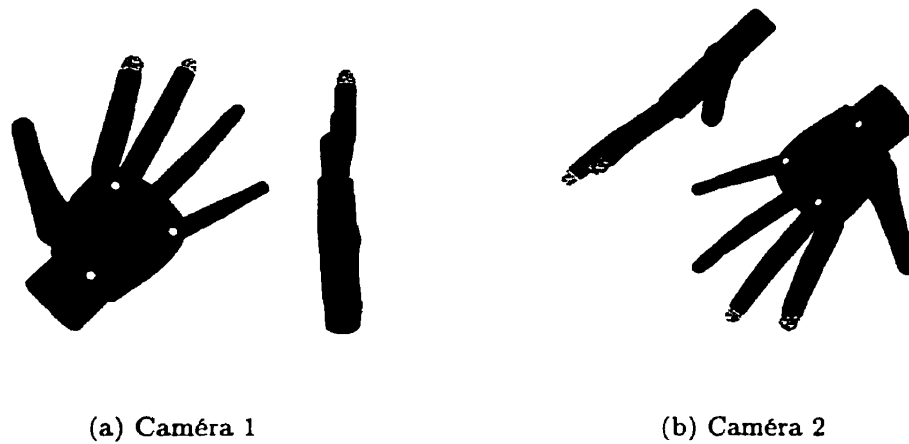


Figure 4.18 : Reconstruction de la main ouverte vue de chaque caméra

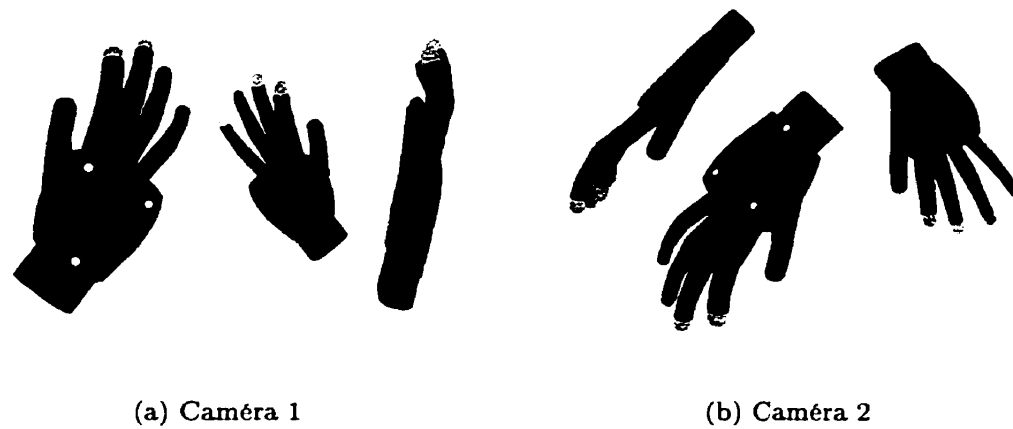


Figure 4.19 : Reconstruction de la main fermée vue de chaque caméra

#### 4.5.2.2 Analyse et comparaison d'une séquence prise des deux caméras

La main décrit cette fois-ci un mouvement lent, visible des deux caméras et sa posture est reconstruite en utilisant les calibrations précédentes en longueurs de phalanges et angles d'abduction limites. La posture de la main a été reconstruite pour chacune

des deux séquences sur 65 images, soit sur une durée de 2s environ. Les différences entre chaque variable d'articulation pour les deux séquences sont illustrées sur les figures 4.20 à 4.25.

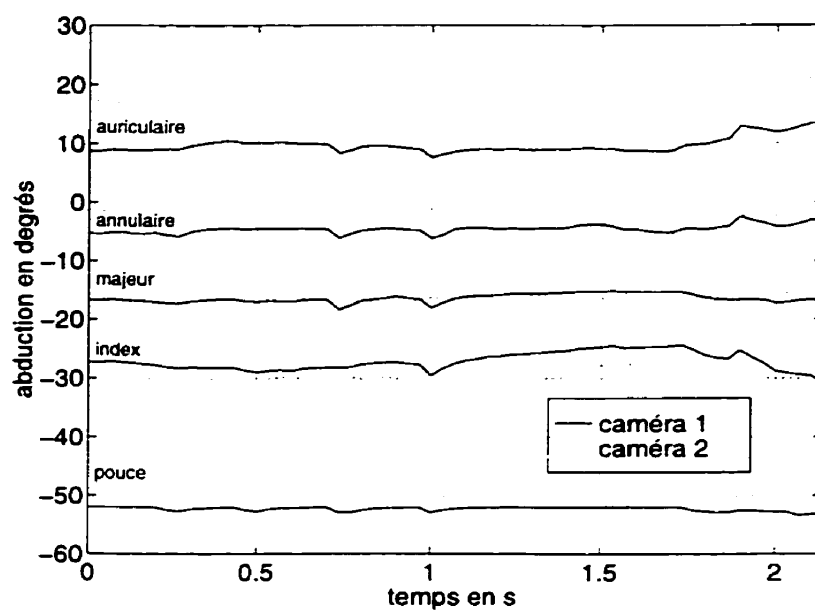


Figure 4.20 : Angle d'abduction de chaque doigt

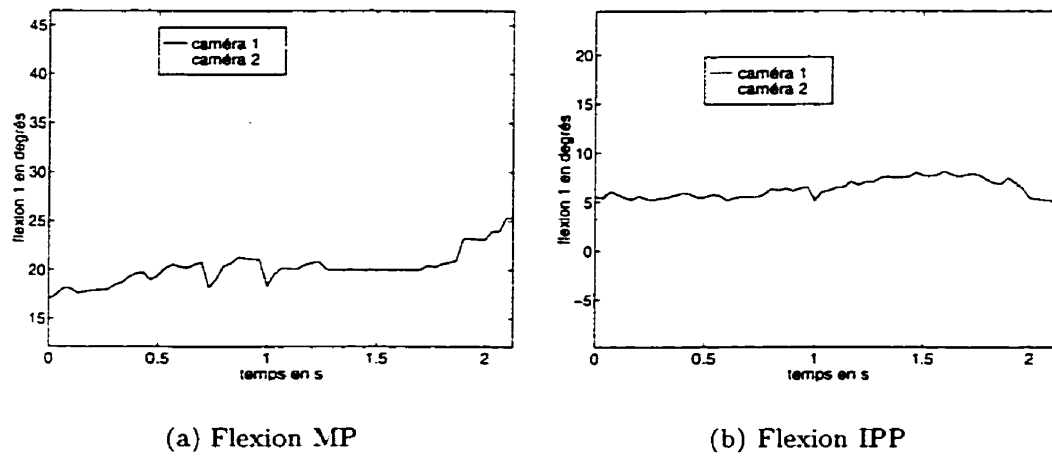


Figure 4.21 : Reconstruction du pouce pour les deux séquences

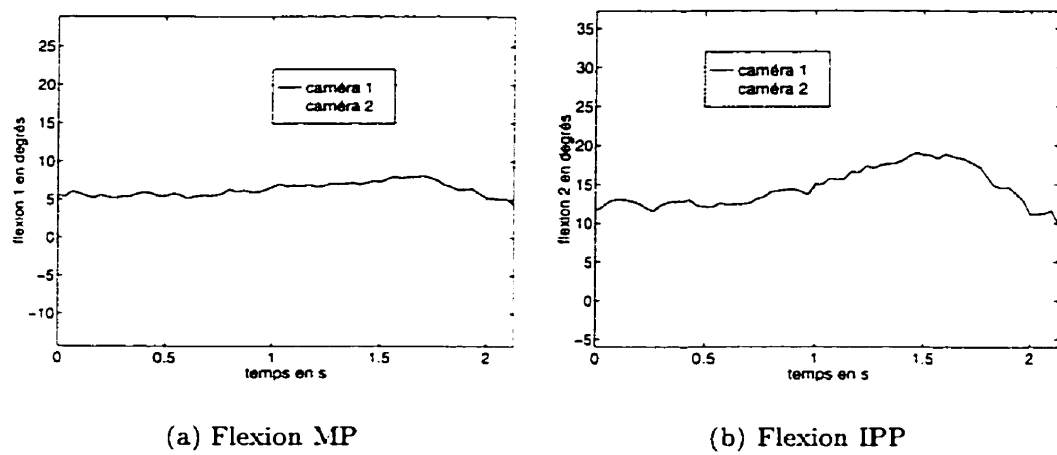


Figure 4.22 : Reconstruction de l'index pour les deux séquences

Les erreurs entre les deux reconstructions sont analysées aux tableaux 4.12 à 4.16.

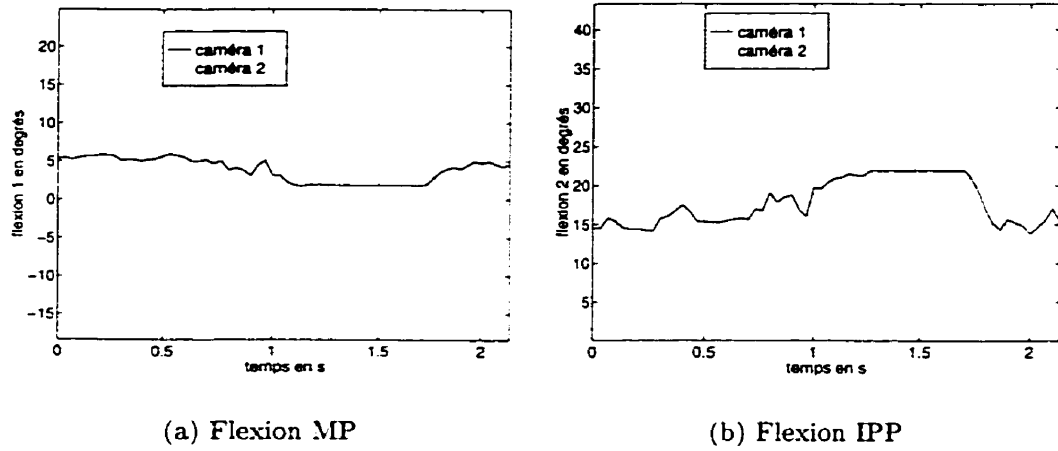


Figure 4.23 : Reconstruction du majeur pour les deux séquences

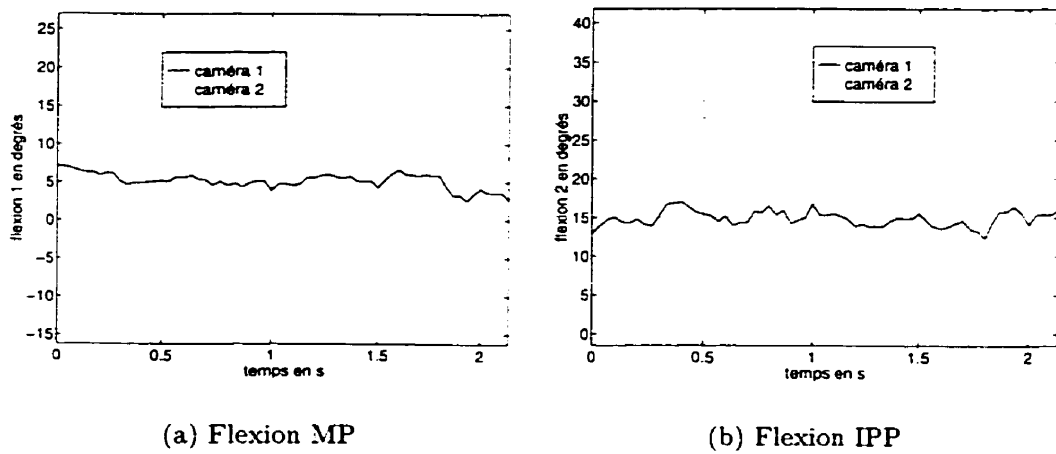


Figure 4.24 : Reconstruction de l'annulaire pour les deux séquences

L'erreur en abduction est limitée autour de 2 degrés sauf pour le cas de l'auriculaire qui présente une différence plus marquée autour de 7 degrés. L'erreur en flexion pour l'angle  $q_1$  est très faible pour tous les doigts hormis le pouce avec 11.5 degrés. Cependant, l'erreur en angle  $q_1$  du pouce n'affecte pas les propriétés de saisie de la

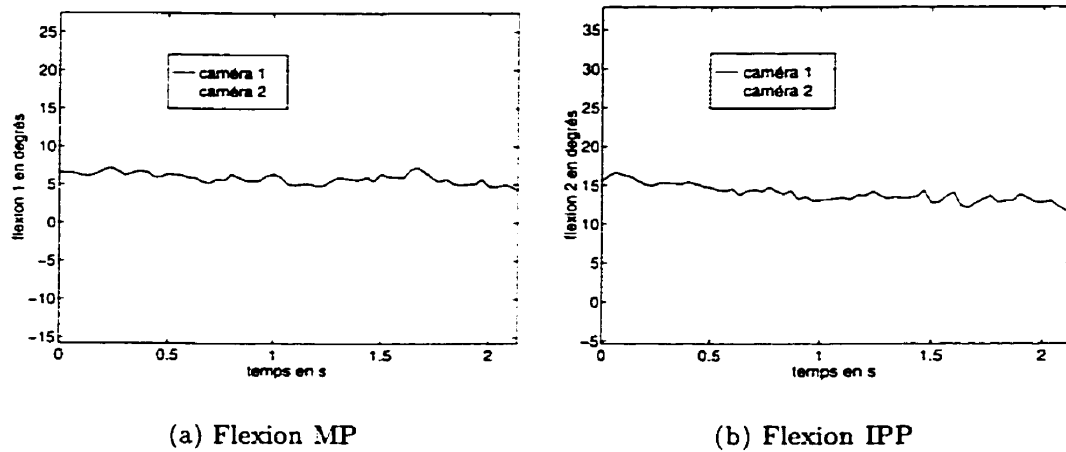


Figure 4.25 : Reconstruction de l'auriculaire pour les deux séquences

Tableau 4.12 : Pouce

grandeur	$q_0$	$q_1$	$q_2$
différence moyenne en degrés	2.2	11.5	1.1
biais en degrés	2.2	11.5	0.1
différence maximale en degrés	4.2	18.0	3.0
écart type en degrés	0.8	4.9	0.7

posture de la main mais décale le pouce par rapport au plan de la paume. C'est la faible erreur en angle  $q_2$  de 1 degré environ qui permet de dire que la seconde flexion du pouce est conservée, celle-là même qui précise la nature de la pince (sphérique ou planaire par exemple). En ce qui concerne le cas des autres doigts, l'erreur en flexion pour chacun des deux angles est inférieure à celle observée sur la séquence de synthèse. Cela ne veut pas dire que la reconstruction est plus précise. Il se peut que les valeurs obtenues de chacune des caméras soient plus éloignées des valeurs exactes. Néanmoins, on peut dire que les deux postures reconstruites à partir de

Tableau 4.13 : Index

grandeur	$q_0$	$q_1$	$q_2$
différence moyenne en degrés	2.9	3.1	2.2
biais en degrés	2.8	1.1	0.1
différence maximale en degrés	8.5	7.8	7.3
écart type en degrés	1.9	1.9	1.8

Tableau 4.14 : Majeur

grandeur	$q_0$	$q_1$	$q_2$
différence moyenne en degrés	0.6	2.4	8.3
biais en degrés	0.5	1.6	7.8
différence maximale en degrés	2.1	4.4	14.1
écart type en degrés	0.5	1.2	3.6

chacune des caméras sont très voisines, ce qui nous permet d'avancer que la méthode de reconstruction des variables d'articulation de la main est valable et ne dépend pas du point de vue de la caméra.

Tableau 4.15 : Annulaire

grandeur	$q_0$	$q_1$	$q_2$
différence moyenne en degrés	2.7	2.2	7.1
biais en degrés	2.6	0.1	7.1
différence maximale en degrés	4.4	5.8	13.2
écart type en degrés	1.3	1.5	4.1

Tableau 4.16 : Auriculaire

grandeur	$q_0$	$q_1$	$q_2$
différence moyenne en degrés	6.9	0.9	3.3
biais en degrés	6.9	0.6	3.3
différence maximale en degrés	9.2	2.6	7.4
écart type en degrés	2.1	0.6	1.4

## Chapitre 5

# Traitement des occlusions

La reconstruction de la posture de la main repose entièrement sur la détection d'indices de couleur dont est revêtu le gant. Il peut arriver qu'un ou plusieurs de ces indices soit masqué à cause de la posture particulière de la paume ou des doigts à un instant donné. En cas d'absence d'information sensorielle, comme cela peut arriver, il importe de trouver une solution qui permette de reconstruire malgré tout la posture totale de la main. Nous avons donc eu recours à des méthodes de raisonnement heuristique afin de retrouver l'information géométrique en l'absence de données sensorielles. Afin de bénéficier de suffisamment d'information pour assurer le positionnement correct de la paume, celle-ci est supposée visible dans sa totalité. Nous traiterons donc uniquement le cas des occlusions des extrémités des doigts.



## 5.1 Conditions d'existence d'une occlusion

Lors de la détection des indices de couleur sur le gant, il se peut qu'une ou plusieurs taches ne soient pas détectées, soit parce qu'elles sont trop petites, soit parce qu'elles sont cachées. L'observation des mouvements de la main nous permet d'affirmer que l'extrémité d'un doigt est susceptible d'être masquée par le doigt lui-même, par un autre doigt ou bien encore par la paume. Dans la majeure partie des cas de non-détection, il apparaît que l'extrémité du doigt occulté est cachée par le doigt lui-même ou par un autre doigt avant d'être masquée par la paume. En conséquence, nous étudierons les cas d'occlusion de l'extrémité d'un doigt par un autre doigt.

Tout d'abord, nous définissons un ordre de visibilité des doigts. Il s'agit d'un tableau qui classe les doigts par ordre croissant de profondeur, c'est à dire en commençant par le doigt le plus proche de la caméra et en terminant par le doigt le plus éloigné. Un doigt peut alors être caché par n'importe quel doigt le précédant dans ce tableau. En se basant sur des observations du mouvement de la main, deux situations distinctes à l'origine d'occultations sont apparues, pendant le mouvement d'approche de la main d'une part, et au cours de la phase de préhension de l'objet d'autre part.

### 5.1.1 Occultations pendant la phase d'approche

Dans ce premier cas, la main se dirige en direction de l'objet avec une vitesse croissante. Les doigts s'écartent légèrement pour se préparer à la saisie, (Jeannerod, 1984), par contre ils bougent relativement peu en flexion/extension. Deux classes d'occlusion

peuvent alors être relevées.

- Rotation de la main autour de l'axe du poignet

Il se peut que la main effectue alors une rotation autour du poignet selon l'axe le long de la paume. Les doigts les plus éloignés de la caméra sont alors occultés par les doigts plus proches, voir figure 5.1.



(a) Main entièrement visible

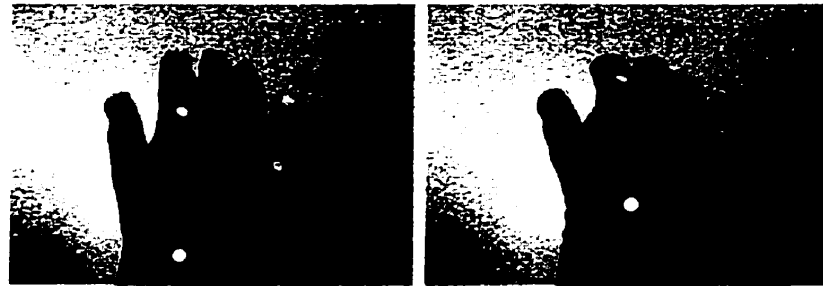
(b) Main entièrement visible

(c) Annulaire et auriculaire occultés

Figure 5.1 : La main tourne légèrement autour du poignet

- Rotation de la main autour de l'axe perpendiculaire à celui du poignet

Dans ce cas, c'est la partie dorsale de la paume qui tourne et finit par se cacher de la caméra. C'est comme si la main vue de haut pivotait en direction du sol, entraînant ainsi une occultation. Les doigts peuvent alors être occultés par eux-mêmes ou encore par la paume, voir figure 5.2.



(a) Main entièrement visible

(b) Index, majeur et annulaire occultés

Figure 5.2 : Cas où le doigt s'auto-occulte

### 5.1.2 Occultations pendant la phase de saisie

Dans ce cas, la main est plutôt immobile dans sa phase de préhension. Les doigts se referment progressivement autour de l'objet dans un mouvement de flexion. On peut distinguer deux sortes d'occlusions selon le moment dans la phase de préhension.

- Début de la fermeture des doigts

Un doigt est alors généralement occulté par un des doigts précédents dans le tableau de visibilité (figure 5.3).

- Fin de la saisie

Il se peut qu'un doigt se masque lui-même s'il vient se coller autour de l'objet (voir figure 5.4). Dans ce cas, il devient difficile de prévoir la suite du mouvement des doigts, à moins de connaître précisément la position et la taille de l'objet à saisir.

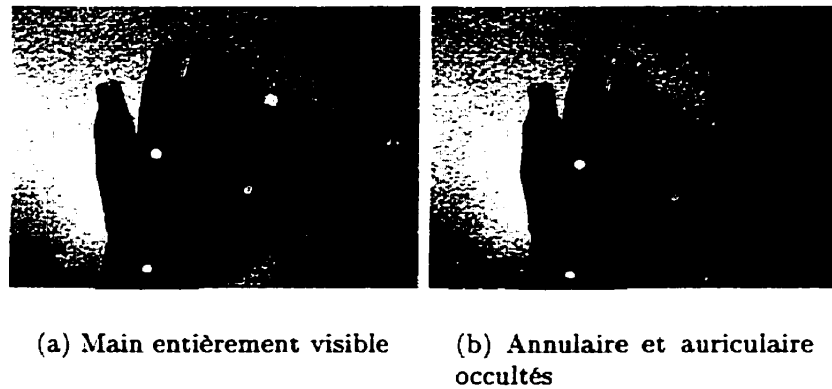


Figure 5.3 : Le poing se ferme

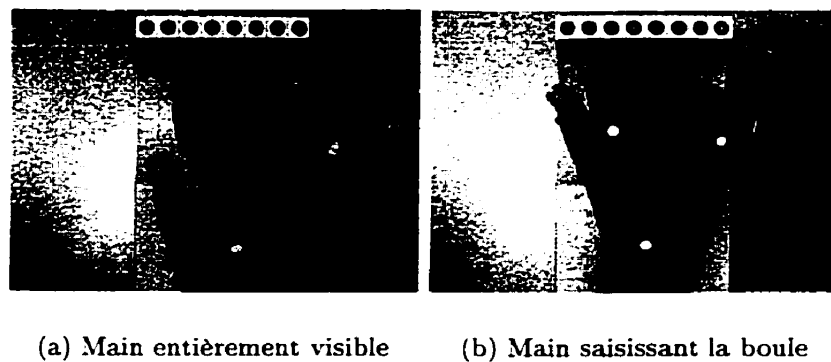


Figure 5.4 : Cas où la main enveloppe l'objet

## 5.2 Expressions mathématiques des conditions d'occlusion

On suppose que l'extrémité d'un doigt peut être assimilée à une sphère. Elle peut être cachée soit par l'extrémité d'un autre doigt, modélisée également par une sphère, soit par une des différentes phalanges (phalange, phalangette, phalangine) du doigt, modélisées chacune par un cylindre. Dans ce cas, l'extrémité d'un doigt peut être cachée

soit par le doigt lui-même, soit par un doigt extérieur. La figure 5.5 illustre le cas d'une sphère cachée aussi bien par une autre sphère que par un cylindre (symbolisant ici une phalange) selon le point de vue.



Figure 5.5 : Modélisation du doigt et de son extrémité

### 5.2.1 Occlusion d'une sphère par une sphère

L'extrémité du doigt est assimilée à la sphère  $S_2$  de centre  $C_2$  et de rayon  $r_2$ . Supposons qu'elle soit cachée par l'extrémité d'un autre doigt, représentée par la sphère  $S_1$  de centre  $O_1$  et de rayon  $r_1$  (figure 5.6).

Soit  $O$  le centre optique de la caméra. La sphère  $S_2$  est cachée par la sphère  $S_1$  si et seulement si la sphère  $S_2$  est à l'intérieur du cône de centre  $O$  enveloppant la sphère  $S_1$ . Soit  $O_2$  la projection de  $C_2$  sur la droite  $(OO_1)$  et  $d_2$  la distance de  $O_2$  au cône :

$$d_2 = r_1 \frac{OO_2}{OO_1}.$$

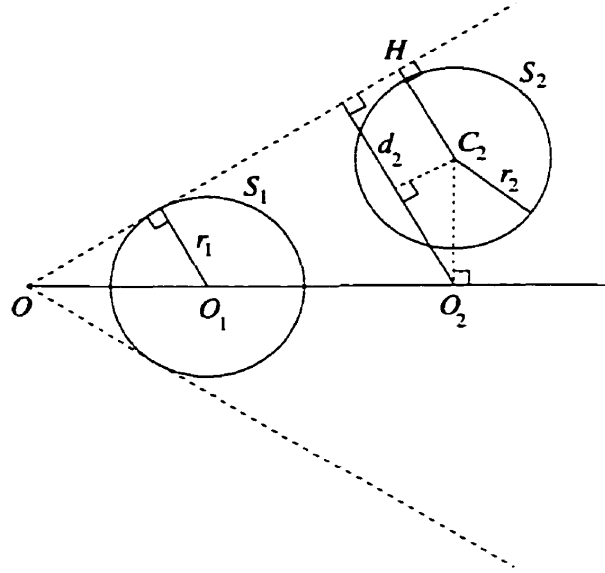


Figure 5.6 : Occlusion d'une sphère par une sphère

La sphère  $S_2$  est cachée par la sphère  $S_1$  si et seulement si la distance de  $C_2$  au cône,  $C_2H$ , est supérieure au rayon  $r_2$  de la sphère  $S_2$ . Or, la distance  $C_2H$  est égale à  $d_2$  moins la projection de  $O_2C_2$  sur  $d_2$ . La condition nécessaire et suffisante de l'occlusion de la sphère  $S_2$  par la sphère  $S_1$  s'écrit donc :

$$d_2 - \frac{O_2C_2}{OO_1} \sqrt{OO_1^2 - r_1^2} \geq r_2.$$

### 5.2.2 Occlusion d'une sphère par un cylindre

Étudions tout d'abord le cas d'une sphère  $S_2$  de centre  $C_2$  et de rayon  $r_2$  cachée par une des phalanges d'un autre doigt. La phalange est assimilée à un cylindre défini par ses extrémités  $b_1$  et  $b_2$  et son rayon  $r_1$ . La sphère  $S_2$  est masquée par le cylindre si et seulement si elle est à l'intérieur du dièdre formé par les 2 plans tangents au

cylindre en  $O$ .

Soit  $(\pi)$  le plan passant par  $O$ , perpendiculaire à la droite  $(b_1 b_2)$ . Ce plan coupe le cylindre en un cercle de centre  $P$  et de rayon  $r_1$  qui permet de déterminer les normales  $n_1$  et  $n_2$  aux plans tangents  $\tau_1$  et  $\tau_2$  à ce cercle passant par  $O$ . Soient  $Q_1$  et  $Q_2$  les projections respectives de  $C_2$  sur les plans  $\tau_1$  et  $\tau_2$ . Le centre optique de la caméra est suffisamment loin des doigts pour que l'on puisse considérer que l'angle formé par les deux plans  $\tau_1$  et  $\tau_2$  est inférieur à  $\pi/2$ . Ce qui amène à l'assertion suivante : le centre  $C_2$  de la sphère  $S_2$  est à l'intérieur du dièdre si et seulement si l'angle  $(\vec{C_2 Q_1}, \vec{C_2 Q_2}) > \pi/2$ . La sphère  $S_2$  est à l'intérieur du dièdre formé par  $\tau_1$  et  $\tau_2$  si et seulement si  $C_2$  est à l'intérieur du dièdre et les distances de  $C_2$  à chacun de ces plans sont supérieures à  $r_2$  (figure 5.7).

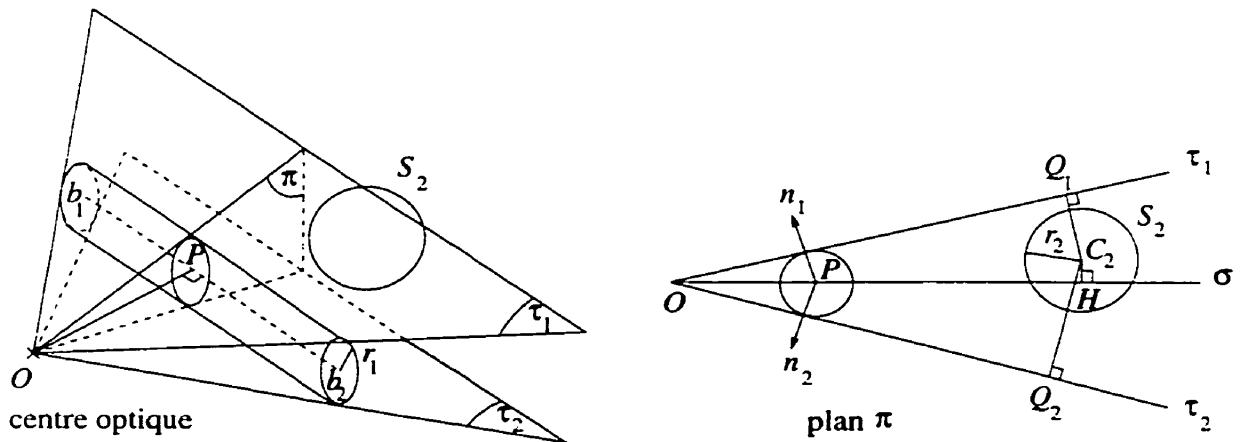


Figure 5.7 : Première condition d'occlusion d'une sphère par un cylindre

Cette condition C1 s'écrit :

$$\begin{cases} \vec{C_2Q_1} \cdot \vec{C_2Q_2} < 0 \\ C_2Q_1 \geq r_2 \\ C_2Q_2 \geq r_2 \end{cases}$$

Il reste à s'assurer que la sphère ne se projette pas à côté du cylindre, c'est-à-dire en dehors des extrémités  $b_1$  et  $b_2$ . Étant donné qu'en réalité une phalange est reliée soit à une autre phalange soit à l'extrémité du doigt, on modélise les extrémités du cylindre par deux sphères pour assurer la continuité en cas d'occlusion par la jonction de deux cylindres. Le cylindre est alors fermé par les deux sphères de centre  $b_1$  et  $b_2$  et de rayon  $r_1$ . Soit  $(\sigma)$  le plan défini par les points  $O$ ,  $b_1$  et  $b_2$  et  $H$  la projection de  $C_2$  sur  $(\sigma)$ , figure 5.7 et figure 5.8. Si la première condition est vérifiée et si  $H$

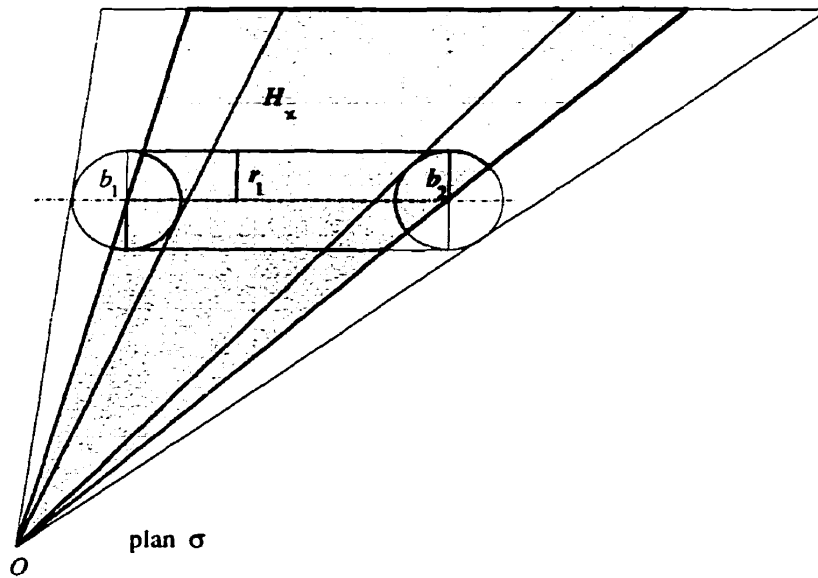


Figure 5.8 : Seconde condition d'occlusion d'une sphère par un cylindre

appartient à la portion de plan grisée définie par les deux droites  $(Ob_1)$  et  $(Ob_2)$ , alors



la sphère  $S_2$  est masquée par le cylindre complété des sphères centrées en  $b_1$  et  $b_2$ . Il se peut également que  $H$  soit en dehors de la portion grisée mais que la sphère  $S_2$  soit masquée par l'une des deux sphères centrées en  $b_1$  et  $b_2$ . On vérifie alors l'occlusion de  $S_2$  par les sphères  $(b_1, r_1)$  et  $(b_2, r_1)$  selon la méthode vue précédemment.

Cette condition C2 s'écrit :

$$\left\{ \begin{array}{l} \mathcal{A}(O, H, b_1) * \mathcal{A}(O, H, b_2) < 0 \text{ ou} \\ S_2 \text{ masquée par la sphère } (b_1, r_1) \text{ ou} \\ S_2 \text{ masquée par la sphère } (b_2, r_1) \end{array} \right.$$

où  $\mathcal{A}(A, B, C)$  désigne l'aire signée du triangle  $(A, B, C)$ . La sphère  $S_2$  est masquée par le cylindre si les conditions C1 et C2 sont vérifiées.

Étudions enfin le cas où l'extrémité du doigt est cachée par les phalanges du même doigt. Soit  $O$  le centre optique,  $P$  l'extrémité du doigt et  $Q$  le point du doigt à la limite de visibilité, comme illustré à la figure 5.9. La ligne de vue  $OQ$  fait alors un angle négligeable par rapport au plan de mouvement du doigt. Afin de vérifier cette hypothèse, on définit l'angle  $\theta$  formé par la ligne de vue ( $OQ$ ) et le plan de mouvement du doigt défini par son vecteur normal  $\vec{n}$ .

$$\theta = \arcsin \frac{|\vec{OQ} \cdot \vec{n}|}{\|\vec{OQ}\| \cdot \|\vec{n}\|}$$

On calcule alors l'angle minimal de visibilité  $\phi$  défini par

$$\phi = \arctan \frac{r_1 - r_3}{l_0},$$

où  $l_0$  est la somme des projections des longueurs des trois phalanges dans le plan de mouvement du doigt. voir figure 3.7.

$$l_0^2 = [d_1 + d_2 \cos(q_2) + d_3 \cos(q_2 + q_3)]^2 + [d_2 \sin(q_2) + d_3 \sin(q_2 + q_3)]^2$$

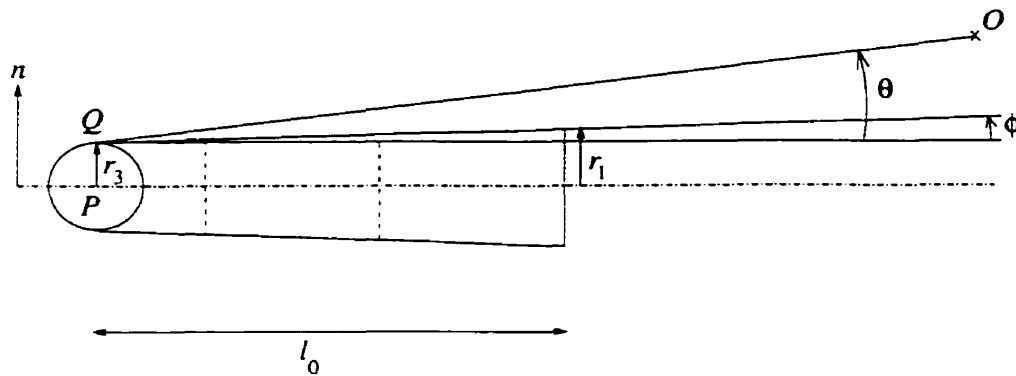


Figure 5.9 : Condition d'auto-occlusion d'un doigt

La condition d'occlusion s'écrit alors  $\theta < \phi$ .

### 5.3 Reconstruction en cas d'occlusions

Afin de reconstruire adéquatement la posture de la main dans l'espace malgré l'absence de certains indices de couleurs, une heuristique a été développée afin de prendre en compte les occlusions. Elle consiste à mettre en évidence la classe d'occlusions con-

cernée, mouvement d'approche de la main ou bien fermeture du poing.

Soit  $V_A$ , la vitesse de la main et  $V_{q_1}$  et  $V_{q_2}$  les vitesses moyennes de l'angle  $q_1$  et  $q_2$  respectivement pour tous les doigts à un instant donné. Si  $V_A$  dépasse un certain seuil  $S_{V_A}$  déterminé par expérimentation, on suppose que l'on est dans la phase d'approche de la main, c'est-à-dire dans la phase à vitesse rapide. Dans ce cas, les angles de flexion du doigt occulté sont considérés comme constants. On fait donc la prédiction que les angles  $q_1$  et  $q_2$  du doigt occulté sont les mêmes que pour l'image précédente. Par contre, si  $V_A$  est inférieure à  $S_{V_A}$ , on fait l'hypothèse que la main se trouve dans sa phase de préhension d'objet. Sa vitesse est alors relativement faible et les doigts se referment progressivement autour de l'objet à saisir. Dans ce cas, on fait la prédiction que les angles de flexion augmentent. On effectue une prédiction linéaire basée sur la vitesse des angles  $V_{q_1}$  et  $V_{q_2}$  à l'instant précédent. Une fois la nouvelle valeur des angles de flexion assignée, on vérifie que la nouvelle configuration des angles conduit effectivement à une situation d'occlusion du doigt considéré. Les équations de cinématique directe donnent alors la position prédite de l'extrémité du doigt occulté. On teste tout d'abord la possibilité d'auto-occlusion par le doigt lui-même. Si elle n'est pas validée, on teste la possibilité d'occlusion par chacun des doigts du tableau de visibilité qui précèdent le doigt occulté jusqu'à ce que la validation soit confirmée. On commence par le test d'occlusion par la sphère, puis par chacun des cylindres du doigt en cours, élément du tableau de visibilité.

Il peut arriver malgré tout que la validation échoue. Dans ce cas, si l'on avait fait

l'hypothèse de la situation d'approche, on teste les conditions de la phase de saisie. Si l'on se trouvait déjà dans ce dernier cas, on incrémente itérativement les angles de flexion au cas où leur vitesse augmente plus vite que prévu. On fait l'essai un certain nombre de fois et l'on effectue les étapes de validation pour chaque nouvelle configuration des angles de flexion. Si après un nombre fixé d'essais, la validation a échoué, on conserve la configuration de l'image précédente en espérant qu'une prochaine image visible viendra corriger la configuration erronée. L'algorithme appliqué est décrit en figure 5.10.

## **5.4 Résultats**

### **5.4.1 Validation de l'algorithme du traitement des occlusions**

Plusieurs séquences d'images de synthèse ont été créées à l'aide de la librairie graphique OpenGL afin de valider la reconstruction de la posture de la main en cas d'occlusions d'un ou de plusieurs doigts. Les diverses séquences illustrent chacun des types d'occlusions à savoir la fermeture du poing, la rotation de la paume en mouvement et le cas d'auto-occultation. Huit séquences en tout ont été analysées, chacune comprenant vingt-cinq images et présentant des occlusions significatives. Les moyennes des valeurs absolues des erreurs pour chaque variable d'articulation obtenues sur chacune des séquences sont repertoriées dans les tableaux 5.1 à 5.5. Les grandeurs prises en compte sont les suivantes : moyenne des valeurs absolues des erreurs, biais (valeur

```

si  $V_A(t) > S_{V_A}$ 
  phase d'approche => flexion constante
   $q_1(t) = q_1(t-1)$ 
   $q_2(t) = q_2(t-1)$ 
si auto-occlusion
  validation=VRAI
sinon pour chaque doigt précédent dans le tableau de visibilité
  si occlusion par une sphère
  validation=VRAI
  sinon pour chaque phalange
  si occlusion par un cylindre
  validation=VRAI
  sinon
  échec_occlusion=VRAI
sinon ou si échec_occlusion
  phase de saisie => flexion augmente
  tant que validation fausse
   $q_1(t) = q_1(t-1) + V_{q_1}(t-1) \cdot \Delta t$ 
   $q_2(t) = q_2(t-1) + V_{q_2}(t-1) \cdot \Delta t$ 
  si auto-occlusion
  validation=VRAI
  sinon pour chaque doigt précédent dans le tableau de visibilité
  si occlusion par une sphère
  validation=VRAI
  sinon pour chaque phalange
  si occlusion par un cylindre
  validation=VRAI

```

Figure 5.10 : Algorithme de décision en cas d'occlusions

absolue de la différence entre la moyenne de la variable initiale et la moyenne de la variable reconstruite), valeur absolue de l'erreur maximale et écart type de la valeur absolue de l'erreur.

L'erreur sur l'angle d'abduction est proche de l'erreur obtenue sans occlusions (chapitre 4)

Tableau 5.1 : Pouce

grandeur	$q_0$	$q_1$	$q_2$
erreur moyenne en degrés	4.8	8.1	12.7
biais en degrés	4.3	6.9	12.7
erreur maximale en degrés	7.6	22.3	21.9
écart type en degrés	2.0	6.5	4.9

Tableau 5.2 : Index

grandeur	$q_0$	$q_1$	$q_2$
erreur moyenne en degrés	2.1	5.5	5.2
biais en degrés	1.9	4.6	2.3
erreur maximale en degrés	2.8	10.0	10.0
écart type en degrés	0.4	2.3	2.4

sauf pour les deux derniers doigts. En effet, sur les séquences étudiées, l'annulaire et l'auriculaire sont les doigts le plus souvent occultés. Étant donné que l'algorithme d'estimation de la pose d'un doigt occulté ne prend pas en compte l'angle d'abduction et conserve la dernière valeur calculée, un décalage en abduction peut se produire si le doigt est occulté sur un certain nombre d'images consécutives. Il en résulte une erreur moyenne légèrement plus élevée autour de 6 degrés. Si l'on observe les résultats concernant les angles de flexion, il apparaît là-encore que ce sont l'annulaire et

Tableau 5.3 : Majeur

grandeur	$q_0$	$q_1$	$q_2$
erreur moyenne en degrés	3.1	5.6	4.1
biais en degrés	3.1	5.2	2.2
erreur maximale en degrés	3.6	10.9	9.6
écart type en degrés	0.2	2.3	2.6

Tableau 5.4 : Annulaire

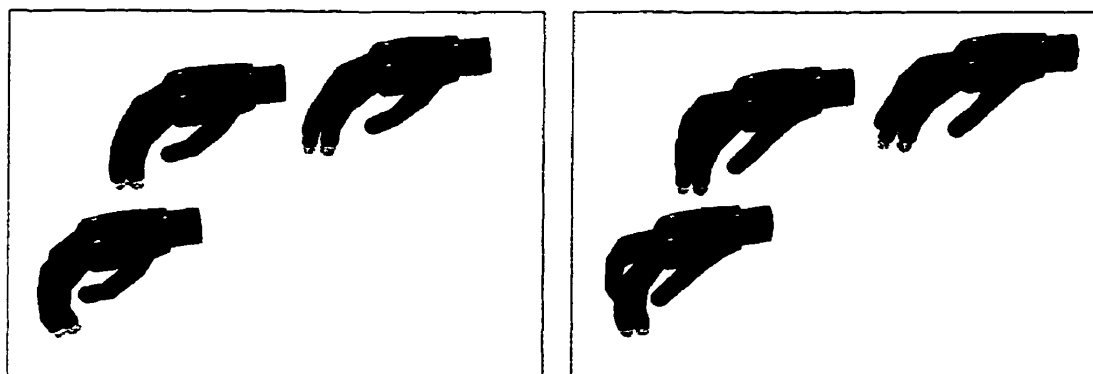
grandeur	$q_0$	$q_1$	$q_2$
erreur moyenne en degrés	5.3	10.5	7.1
biais en degrés	5.3	10.0	5.1
erreur maximale en degrés	6.2	20.7	16.9
écart type en degrés	0.7	4.3	4.4

Tableau 5.5 : Auriculaire

grandeur	$q_0$	$q_1$	$q_2$
erreur moyenne en degrés	7.2	16.1	10.8
biais en degrés	7.2	16.1	9.6
erreur maximale en degrés	10.0	29.2	18.1
écart type en degrés	1.5	6.1	3.9

l'auriculaire qui présentent les erreurs les plus élevées autour de 13 degrés pour  $q_1$  et 9 degrés pour  $q_2$ . Au total, c'est une erreur supplémentaire de 1 degré pour l'annulaire et de 4 degrés pour l'auriculaire qui vient s'ajouter pour les angles  $q_1$  par rapport au cas sans occlusions. Ce résultat est tout à fait acceptable si l'on garde en tête que l'information sensorielle nécessaire fait défaut au processus de reconstruction de la posture de la main dans le cas traité. Trois portions de séquences reconstruites sont illustrées sur les figures 5.11 à 5.13, mettant en lumière les trois catégories d'occlusions citées plus haut.

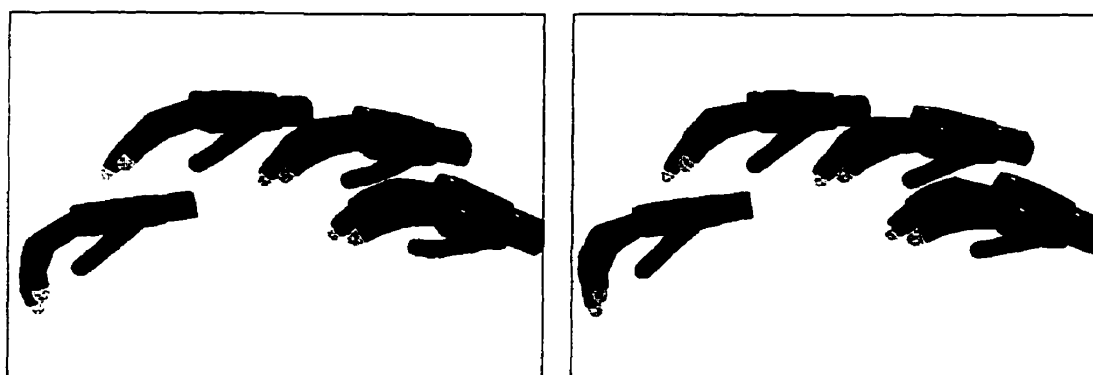
Des anomalies sont clairement visibles, comme la seconde flexion trop importante de l'annulaire et de l'auriculaire pour la séquence de la figure 5.11. En outre, il est une faiblesse de l'algorithme à souligner : un doigt est considéré soit visible soit occulté, mais pas partiellement occulté. Ceci implique qu'un doigt peut disparaître subitement



(a) Images originales

(b) Images reconstruites

Figure 5.11 : Cas d'occlusions lors de la fermeture du poing



(a) Images originales

(b) Images reconstruites

Figure 5.12 : Cas d'occlusions lors de la rotation du poignet

derrière un autre quand le seuil de pixels de la tache déclenchant la détection n'est plus atteint. Le cas de l'auto-occlusion est difficile à traiter comme le montre la figure 5.13 étant donné que l'occlusion peut être géométriquement validée pour une infinité de valeurs possibles en flexion. Aucun élément ne permet alors de savoir si la



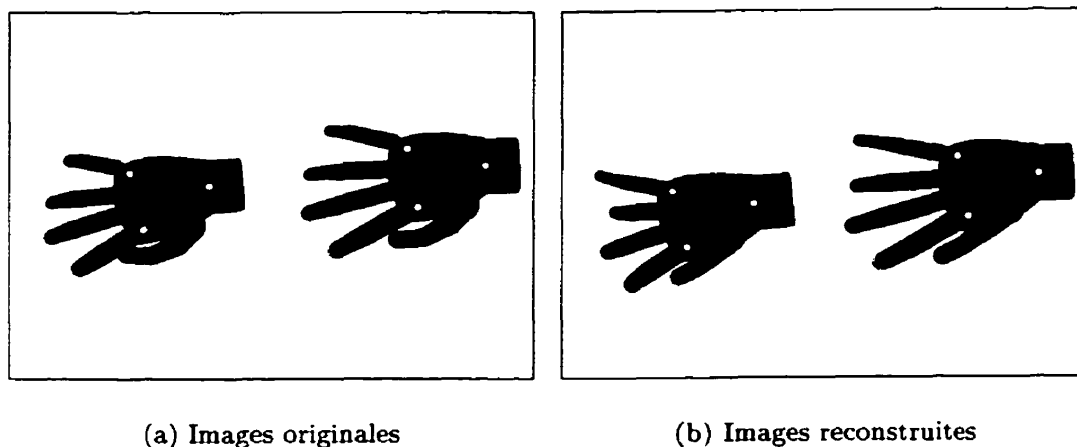


Figure 5.13 : Cas d'auto-occlusions

main se ferme davantage ou non dans un tel cas. Enfin, l'occlusion due à la rotation du poignet sur la figure 5.12 est envisagée de façon correcte puisque l'annulaire et l'auriculaire conservent une posture presque identique à celle de la séquence initiale. En définitive, ce test met en évidence la consistance du processus de reconstruction même en présence d'occlusions. Si le recouvrement de la posture de la main n'est pas parfait, l'algorithme de traitement des occlusions permet tout de même d'envisager de traiter n'importe quelle séquence de la main en mouvement sans contrainte excessive sur l'orientation des doigts par rapport à la caméra.

## Chapitre 6

### Étude de gestes

Ce chapitre présente l'interface graphique de détection visuelle et de reconstruction de la posture de la main développée dans le cadre de l'étude de certains gestes de préhension.

#### **6.1 Interface Humain-Machine de détection de la posture de la main en temps réel**

Afin de disposer d'un banc d'expérimentation d'utilisation facile et immédiate, l'interface de reconnaissance visuelle de la posture de la main a été optimisée afin de fonctionner en temps réel. La main en mouvement est filmée par une caméra couleur Watec202-B. Les indices de couleur sont détectés puis la posture de la main est calculée par la méthode de reconstruction décrite au chapitre 4. Le modèle graphique OpenGL de la

main reconstruite est alors affiché au fur et à mesure que les images de la main prises par la caméra sont analysées. Les données de toutes les variables du modèle d'état de la main sont sauvegardées dans un fichier en vue de l'analyse de gestes.

Les étapes de calibration de la caméra et des longueurs des phalanges de la main sont effectuées à part. La distance focale de la caméra ainsi que son angle d'ouverture sont calibrés à l'aide de l'algorithme de Tsai-Wilson (Tsai, 1987; Wilson, 1994). L'étape de calibration de la main se fait à partir d'une image de la main prise à plat. La position et l'orientation de la paume sont calibrées à l'aide de la pose du triangle disposé sur le dessus du gant, ce qui donne le plan de la main. La détection des extrémités de couleur des doigts donne la longueur réelle totale de chaque doigt. Pour chaque doigt, la longueur de chaque phalange est ensuite déterminée par conservation des proportions entre les trois phalanges mesurées initialement sur le doigt de la main réelle. Il est ainsi possible de calibrer les longueurs des phalanges pour différents usagers afin de généraliser l'usage de l'interface.

### **6.1.1 Plateforme utilisée**

L'interface de suivi visuel et de reconstruction de la posture de la main a été implantée sur un Power Macintosh 7500 équipé d'un processeur dual PowerPC 604e 180Mhz et d'une carte d'acquisition vidéo intégrée. Sur le processeur principal est traitée l'acquisition asynchrone des images de la caméra au taux de 30 images par seconde. L'affichage du modèle graphique OpenGL de la main est également pris en charge

par ce processeur. Le second processeur est, lui, responsable de l'analyse des images et du processus de reconstruction des variables d'articulation de la main. Dès qu'une image est disponible, il l'analyse et met à jour les nouvelles valeurs des angles du modèle de la main. Le processus principal est en attente permanente de nouvelles valeurs pour rafraîchir l'affichage du modèle graphique de la main. Ainsi, l'affichage de la main graphique ne ralentit par les étapes de détection d'indices de couleur et de reconstruction des angles de la main et présente l'avantage de s'exécuter en temps réel. L'utilisateur peut donc vérifier au cours de son mouvement que la main est effectivement bien reconstruite et peut valider le fichier de résultats en sortie du programme.

### **6.1.2 Détails de l'algorithme de recouvrement de la posture de la main**

Il est désirable que le programme de reconstruction de la posture de la main fonctionne de façon robuste, que la main soit effectivement visible ou non, afin de permettre un suivi continu de la main. Pour ce faire, nous allons définir à l'aide d'une machine à états finis les différents états dans lesquels peut se trouver la main ainsi que les transitions qui assurent le passage d'un état à un autre. Au départ, la main est totalement invisible. La première reconstruction impose de détecter à la fois la paume et tous les doigts afin d'initialiser le processus de reconstruction. Une fois que la main est totalement visible, elle peut passer dans quatre états différents :

- paume visible et doigts tous visibles

- paume visible et certains doigts masqués
- paume masquée et doigts tous visibles
- paume masquée et certains doigts masqués

On ajoute à l'état de visibilité de la paume et des doigts la nature de l'algorithme appliqué pour la reconstruction des angles des doigts. Comme nous le verrons par la suite, selon le degré de confiance que l'on a dans la précision de la posture de la paume reconstruite, on applique l'algorithme de reconstruction des angles des doigts soit sur l'intervalle total des angles d'abduction, soit sur un intervalle réduit. Tous ces éléments conduisent au diagramme d'état illustré à la figure 6.1. On définit les variables suivantes :

PV=paume visible

DTV=doigts tous visibles

DIT=calcul des angles des doigts sur l'intervalle d'abduction total

DIR=calcul des angles des doigts sur un intervalle d'abduction réduit

PD=paume disparue.

À partir de l'état **nouveau**, la main doit passer dans l'état [PV,DTV] pour être affichée, faisant alors appel à la procédure de calcul des angles des doigts sur l'intervalle d'abduction total DIT. Si la paume est entièrement détectée sur un certain nombre d'images consécutives et si tous les doigts sont visibles, le calcul des angles se fait sur un intervalle d'abduction réduit DIR. Si un ou plusieurs doigts sont masqués, la main

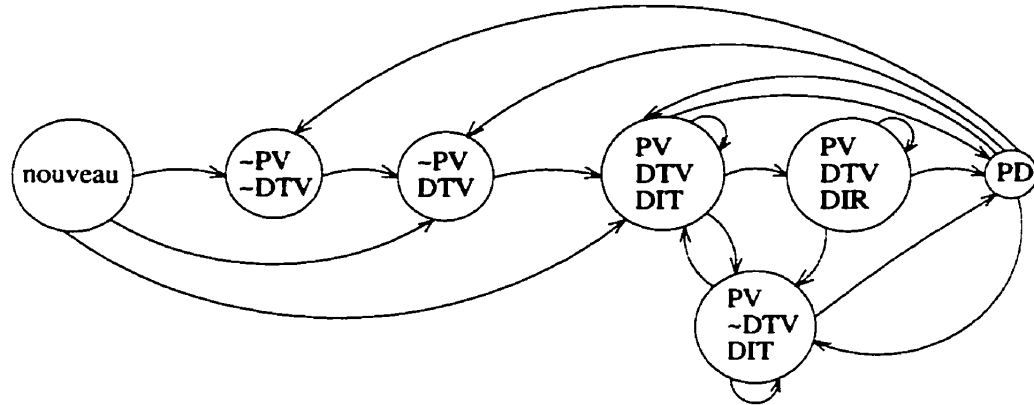


Figure 6.1 : Diagramme des différents états dans lesquels peut se trouver la main passe dans l'état  $[PV, DTV, DIT]$  et requiert l'algorithme de traitement des occlusions. La main doit repasser dans l'état  $[DTV, DIT]$  avant de pouvoir bénéficier de l'algorithme DIR. De chacun de ces états, la paume peut disparaître et la main passe dans l'état PD, auquel cas elle n'est plus affichée.

On se propose de définir une variable **NB\_PAUMES** qui détermine le nombre minimal d'images consécutives sur lesquelles la paume est entièrement visible nécessaire afin d'assurer un certain niveau de confiance en la détection de la paume. De cette variable dépend ensuite le choix de l'algorithme de reconstruction appliqué à la configuration des doigts. Si la paume a été détectée dans sa totalité sur un nombre d'images supérieur à **NB\_PAUMES** et si tous les doigts sont visibles, on calcule l'angle d'abduction de chacun des doigts sur un intervalle réduit centré autour de la valeur précédente de l'angle d'abduction, ceci afin d'accélérer les calculs d'optimisation de l'angle d'abduction. On nomme cette procédure **reconstruction\_doigts\_réduite**. Si ce n'est pas le cas, on reconstruit chaque doigt[i] visible sur son intervalle total

autorisé en abduction à l'aide de la procédure `reconstruction_doigt[i]_totale`. Ensuite, chaque `doigt[i]` occulté est reconstruit au moyen de la procédure de traitement des occlusions `reconstruction_occlusion[i]`. La main est alors affichée. Si ensuite, il arrive qu'un des indices de la paume soit masqué, on affiche la posture précédente de la main tant que la paume n'est pas de nouveau visible en totalité. L'algorithme regroupant ces différents cas est illustré sur l'organigramme de la figure 6.2.

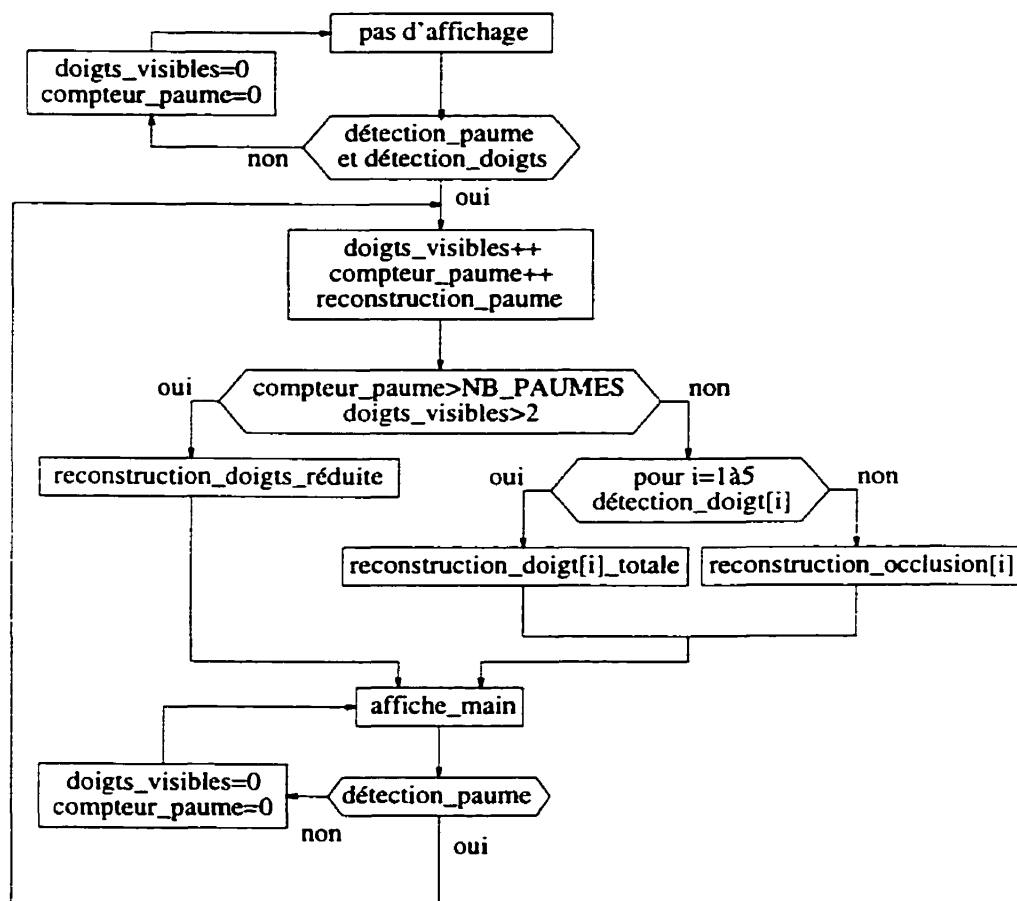


Figure 6.2 : Organigramme de l'interface de reconstruction de la posture de la main

### 6.1.3 Améliorations apportées à l'étape de détection des taches

Lors de la prise d'une séquence de la main en mouvement à l'aide de la caméra, il apparaît un flot de mouvement dans l'image dû à la trop faible fréquence de fermeture de l'obturateur de la caméra. Il en résulte un bavement au niveau des taches à détecter (figure 6.3). Étant donné l'importance de la bonne estimation de la posture de la

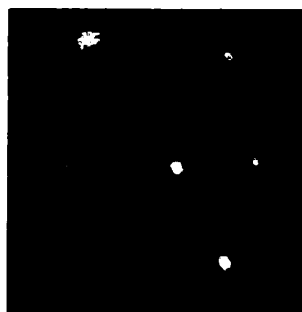


Figure 6.3 : Étalement dans l'image des taches à détecter

paume, il est crucial de positionner le plus précisément possible les centroïdes des taches blanches du triangle de la paume. Afin de résoudre ce problème, on estime dans la tache la position du vrai centroïde approximé au niveau du centre du cercle contenu dans la tache à l'extrémité de la direction de mouvement de la main. Soit  $A$  la position courante du point du triangle au niveau du poignet et  $A_{\text{prec}}$  celle sur l'image précédente. Le vecteur  $\overrightarrow{A_{\text{prec}}A}$  constitue une bonne approximation de la direction de l'axe principal de l'ellipse sur l'image suivante. Sur cette dernière, on détecte le centroïde de la tache complète ainsi que le point de l'ellipse à l'extrême droite en direction du vecteur  $\overrightarrow{A_{\text{prec}}A}$ . Le vrai centroïde du cercle est calculé en retirant au point à l'extrême droite la hauteur maximale de l'ellipse le long de l'axe principal de



l'ellipse (figure 6.4).

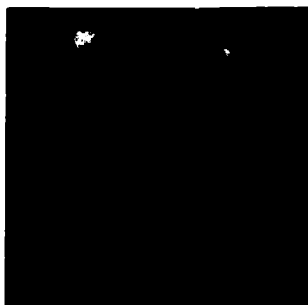


Figure 6.4 : Estimation du vrai centroïde de la tache à l'intersection des axes

## 6.2 Étude de plusieurs gestes de préhension

À l'aide de l'interface précédente, il est envisageable d'enregistrer plusieurs séquences de gestes afin d'étudier les trajectoires des angles de la main pour chacune. On cherche à étudier les caractéristiques propres à chacun des gestes dans l'éventail suivant :

- saisie d'une tasse par le haut par une prise sphérique (figure 6.5(a))
- saisie d'une tasse par le côté par une prise cylindrique (figure 6.5(b))
- saisie d'une trousse posée à plat par une prise cylindrique (figure 6.5(c))

Une série de séquences est analysée pour chaque geste et les angles de toutes les variables d'articulations de la main consignés dans un fichier.

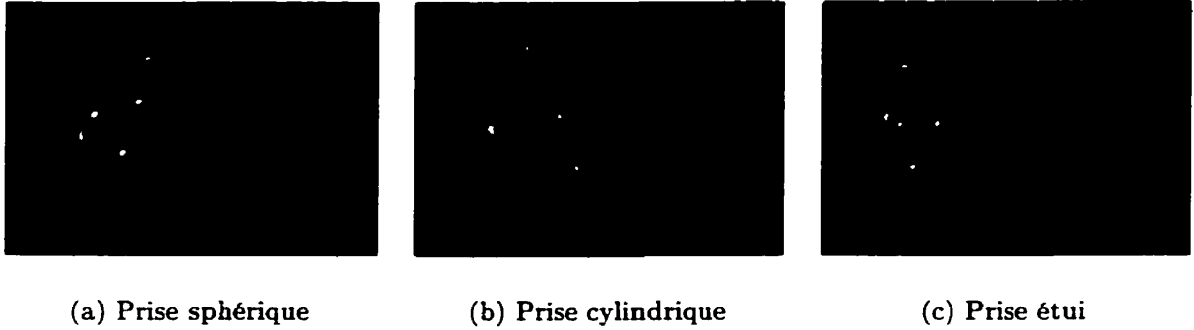


Figure 6.5 : Différentes prises

### 6.2.1 Mise à l'échelle temporelle

Pour une prise donnée, toutes les séquences s'y référant ne s'étendent pas sur le même intervalle de temps. Le geste peut non seulement être effectué avec une vitesse variable mais les instants de début ou de fin sont assignés de façon plus ou moins précise. Il importe donc de ramener toutes les séquences d'un même type de prise à un intervalle de temps identique. Des caractéristiques angulaires pourront alors en être extraites. On utilise la technique dite de "Time Warping" (dilatation ou contraction temporelle) afin de ramener toutes les séquences d'une même prise à la longueur moyenne des séquences. Si  $N$  est la moyenne des longueurs de séquences, on transforme chaque séquence de longueur  $M$  en une séquence de longueur  $N$ , figure 6.6. La valeur de l'angle  $q_N$  à l'indice  $i$  de la séquence de longueur  $N$  s'écrit en conséquence :

$$q_N[i] = q_M \left[ \left\lfloor i \frac{M-1}{N-1} \right\rfloor \right] + \left\{ i \frac{M-1}{N-1} - \left\lfloor i \frac{M-1}{N-1} \right\rfloor \right\} \left\{ q_M \left[ \left\lfloor i \frac{M-1}{N-1} \right\rfloor + 1 \right] - q_M \left[ \left\lfloor i \frac{M-1}{N-1} \right\rfloor \right] \right\}$$

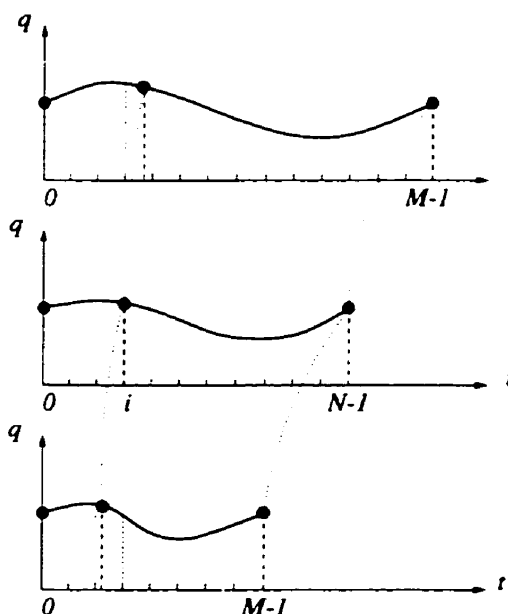


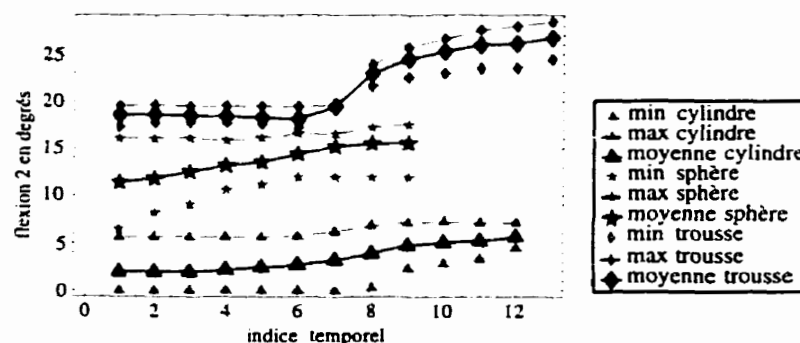
Figure 6.6 : Mise à l'échelle temporelle de toutes les séquences

avec interpolation linéaire de l'angle entre l'indice entier le plus proche et l'indice réel.

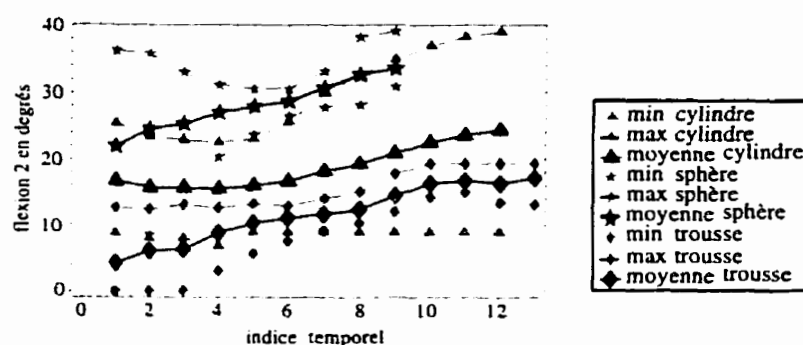
### 6.2.2 Reconnaissance de gestes

Une fois toutes les séquences d'une prise spécifiée ramenées sur le même intervalle, on étudie les trajectoires des moyennes de chaque angle ainsi que les valeurs maximales et minimales rencontrées. Ceci permet d'obtenir une bande à l'intérieur de laquelle sont circonscrites les valeurs d'un angle donné. L'étude de chacune de ces bandes pour tous les types de prise permet d'étudier les angles les plus facilement distinguables. Cette étude nous a conduit à proposer un critère de reconnaissance des trois types de gestes définis plus tôt. On choisit le second angle de flexion du pouce  $q_{p_2}$  et de l'index  $q_{i_2}$  respectivement. La figure 6.7 illustre les plages de variation des seconds angles

de flexion du pouce et de l'index respectivement pour les trois catégories de gestes sélectionnées. Le trait plein représente la moyenne de l'angle sur toutes les séquences ramenées à un unique intervalle temporel, tandis que les pointillés correspondant indiquent les valeurs minimales et maximales rencontrées. Pour chaque type de prise,



(a) Pouce



(b) Index

Figure 6.7 : Moyennes des angles pour les trois types de prise

on dénote par  $M_p$  et  $M_i$  les trajectoires moyennes des angles  $q_{p2}$  et  $q_{i2}$  (représentées en trait plein sur la figure 6.7). Pour analyser la séquence d'un geste indéterminé, on propose la mesure de ressemblance suivante : d'une part, la distance de l'angle

moyen  $q_{p_2}$  de cette séquence à la moyenne  $M_p$  et d'autre part, la distance de l'angle moyen  $q_{i_2}$  de cette séquence à la moyenne  $M_i$ . Plus précisément, pour une séquence indéterminée  $g[k]$ , on effectue la mise à l'échelle sur la longueur  $N$  propre à chaque prise de référence puis on calcule les mesures  $d_p$  et  $d_i$  comme suit :

$$d_p = \frac{1}{N} \sum_{k=0}^{N-1} [(g_p[k] - M_p[k])^2]$$

$$d_i = \frac{1}{N} \sum_{k=0}^{N-1} [(g_i[k] - M_i[k])^2] .$$

On calcule les distances  $d_p$  et  $d_i$  pour chaque type de prise. Si la prise correspondant à la distance  $d_p$  minimale et la prise correspondant à la distance  $d_i$  minimale sont identiques, on suppose que la prise a été reconnue et on affecte cette prise au geste considéré. Sinon, on calcule la somme  $d$  des distances  $d_p$  et  $d_i$  pour chaque type de prise. Si la distance  $d$  minimale présente un écart d'au moins 10% par rapport aux distances  $d$  des deux autres types de prise, on affecte la prise correspondante au geste inconnu, sinon on ne peut se prononcer de façon certaine.

### 6.2.3 Résultats obtenus

Une série de nouvelles séquences des trois prises ont été analysées afin de servir de test pour la mesure de ressemblance proposée. Une tasse a servi d'objet pour les prises sphériques et cylindriques et un étui à lunettes a été saisi pour la prise cylindrique à plat. Les résultats obtenus avec les distances  $d_p$  et  $d_i$  sont présentés au tableau 6.1.

Comme ce tableau le démontre, les critères  $d_p$  et  $d_i$  pris séparément permettent de

Tableau 6.1 : Verdict des distances  $d_p$  et  $d_i$  obtenues pour chaque geste de test

séquence	$d_p$ minimale	$d_i$ minimale	verdict
cylindre 1	cylindre	cylindre	reconnu
cylindre 2	cylindre	étui	dilemme
cylindre 3	cylindre	étui	dilemme
cylindre 4	cylindre	cylindre	reconnu
cylindre 5	cylindre	sphère	dilemme
sphère 1	sphère	sphère	reconnu
sphère 2	sphère	sphère	reconnu
sphère 3	cylindre	sphère	dilemme
sphère 4	sphère	sphère	reconnu
sphère 5	sphère	sphère	reconnu
étui 1	sphère	étui	dilemme
étui 2	sphère	étui	dilemme
étui 3	sphère	cylindre	dilemme
étui 4	sphère	étui	dilemme
étui 5	sphère	étui	dilemme

reconnaître la prise dans 40% des cas uniquement. De plus, ni l'un ni l'autre ne suffit à lui seul pour prendre une décision. On calcule donc la somme  $d$  des distances  $d_p$  et  $d_i$  pour chaque type de prise et on sélectionne la prise pour laquelle la distance  $d$  est minimale. Les résultats obtenus sont retranscrits au tableau 6.2.

Les prises cylindriques et sphériques ont toutes été reconnues avec succès, par contre on dénombre une indétermination en ce qui concerne la prise de l'étui non distinguable d'une prise cylindrique (écart inférieur à 10%). Le critère proposé, fort simple au demeurant, donne donc de très bons résultats. Les prises sphériques et cylindriques sont facilement distinguables, comme le prouvent les scores importants obtenus par les mauvaises prises. Ce n'est pas le cas pour la prise de l'étui dont la distance est proche de celle du cylindre. En effet, l'étude des angles obtenus révèle que les angles

Tableau 6.2 : Distance de ressemblance obtenue pour chaque geste de test

séquence	<i>d</i> sphère	<i>d</i> cylindre	<i>d</i> étui	reconnu
cylindre 1	210.8	71.2	444.1	oui
cylindre 2	455.3	130.7	412.5	oui
cylindre 3	690.1	220.8	437.7	oui
cylindre 4	202.4	55.1	393.3	oui
cylindre 5	151.1	83.4	483.8	oui
sphère 1	32.4	249.2	390.5	oui
sphère 2	34.0	216.1	382.4	oui
sphère 3	330.2	651.3	1314.3	oui
sphère 4	89.3	298.6	351.1	oui
sphère 5	56.6	345.7	595.7	oui
étui 1	206.8	189.2	89.5	oui
étui 2	322.1	160.6	104.9	oui
étui 3	189.8	136.1	138.3	incertitude
étui 4	314.3	161.6	116.5	oui
étui 5	312.0	224.3	76.2	oui

de flexion de l'index sont plus proches de la moyenne du cylindre, ce qui s'explique par le fait que ces deux prises peuvent être relativement semblables au niveau de la posture de l'index. Le pouce, lui, a l'avantage d'être dans une position distincte pour chaque type de prise, c'est pourquoi son second angle de flexion joue un rôle déterminant dans la reconnaissance des gestes.

#### 6.2.4 Conclusion

L'étude de reconnaissance des gestes a été menée sur trois types de gestes distincts. Le critère de ressemblance a été choisi en tenant compte des différences les plus marquées issues de l'observation des trajectoires des gestes. Il va sans dire que ce critère est arbitraire et qu'une étude plus fouillée faisant intervenir les vitesses angulaires ou bien

les corrélations inter-doigts conduirait à des résultats plus significatifs. Cependant, cette brève étude a le mérite de démontrer que différentes formes de prises sont facilement distinguables et reconnaissables par l'étude des variables d'articulations de la main. On peut donc exploiter le système de suivi de la posture de la main pour reconnaître les gestes effectués par l'opérateur et ainsi mieux comprendre le mouvement humain. L'application qui suit concerne le contrôle d'un manipulateur robotique dans une tâche de saisie d'objet à partir de la trajectoire du mouvement de la main réalisé par un opérateur. L'étude des gestes permet de tirer parti de l'information sur la prise effectuée par l'humain afin de guider le robot de façon plus appropriée au moment de la saisie de l'objet.



## Chapitre 7

# Application au contrôle d'une pince robotique

Comme démontré au chapitre 6, il est possible de distinguer différentes prises particulières effectuées par un opérateur muni du gant. Dans le domaine du contrôle robotique, il peut être intéressant d'exploiter la connaissance tirée du mouvement humain afin de guider un bras manipulateur dans une tâche de saisie d'objet. On rejoint alors le domaine de l'apprentissage par l'exemple. Ce chapitre propose justement une application du suivi de la posture de la main au contrôle d'une pince robotique en charge de saisir un objet. Le contrôle du robot se ferait une fois le geste terminé par l'opérateur. Une telle procédure serait pertinente, par exemple, dans le cas de délais importants entre l'envoi de la commande au robot et son exécution. On peut songer à des applications de contrôle de véhicules ou de manipulateurs par téléopération

comme il existe dans des environnements hostiles (espace, mines). En l'absence de pince robotique disponible au laboratoire, les expérimentations ont été développées en simulation. La première partie de ce chapitre expose les techniques de contrôle des manipulateurs robotiques basées sur un asservissement visuel à l'aide d'une ou plusieurs caméras.

## **7.1 Revue bibliographique sur l'asservissement visuel**

Le contrôle des manipulateurs robotiques par asservissement visuel s'est largement développé au cours des deux dernières décennies. La difficulté inhérente au contrôle traditionnel des manipulateurs dans l'espace des coordonnées généralisées ou l'espace opérationnel réside dans l'absence de boucle de rétroaction sur la valeur réelle des variables d'articulation ou sur la position du segment terminal du manipulateur dans l'espace cartésien. Un tel contrôle suppose également que le but à atteindre ou la trajectoire à suivre sont connus de façon précise dans l'espace d'état choisi pour le contrôle. Cette procédure se heurte assez vite à des difficultés dans le cas d'environnements non contraints, de manipulateurs non calibrés ou de tâches trop générales. L'asservissement dans l'image consiste, lui, à fermer la boucle de rétroaction sur la position et l'orientation dans l'image de l'embout terminal plutôt que sur les variables d'articulation du manipulateur, voir figure 7.1.

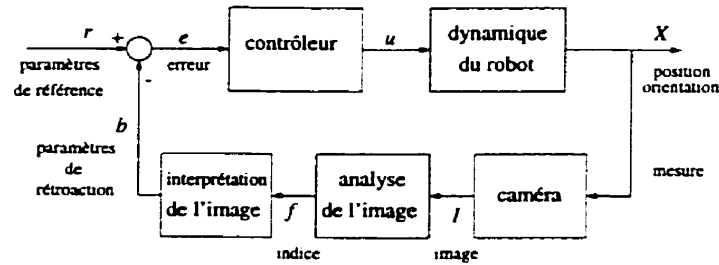


Figure 7.1 : Structure du contrôleur dans le cas d'asservissement visuel

Une revue soulignant les principales caractéristiques du contrôle par asservissement visuel a été dressée par Hutchinson, Hager et Corke (1996). Reprenant les travaux (Weiss, Sanderson et Neuman, 1987; Sanderson et Weiss, 1983), ils dressent une synthèse des différentes structures de contrôleurs basés sur l'image. Ils distinguent tout d'abord deux modes d'asservissement dépendant de la représentation des paramètres utilisés dans la boucle de rétroaction:

- **Contrôle basé sur la position (“position-based control”)**

L'asservissement est alors basé sur la position et les indices dans l'image sont interprétés en termes de position et orientation de la cible par rapport à la caméra. Ceci implique la connaissance du modèle géométrique de la cible et la calibration de la caméra par rapport au robot.

- **Contrôle basé sur l'image (“image-based servoing”)**

L'asservissement est cette fois basé dans l'image, auquel cas les paramètres de rétroaction sont directement les traits saillants dans l'image.

D'autre part, les systèmes asservis dans l'image sont caractérisés par leur mode de

contrôle des joints:

- en boucle ouverte

C'est le cas du **“suivi visuel”**. Il n'y a pas de contrôle des joints. L'envoi de la commande est faite à vide et la vision est utilisée uniquement pour stabiliser le mécanisme.

- en boucle fermée

C'est ce que l'on nomme **“look-and-move”**. Les angles des joints sont dans ce cas contrôlés de façon dynamique, le plus souvent à l'aide de l'interface de commande du robot qui permet directement le contrôle en vitesse ou en positions incrémentales.

### 7.1.1 Contrôle basé sur l'image

Différentes implantations ont été envisagées selon que les caractéristiques de la caméra et du manipulateur sont connues ou non. Le plus généralement, la cinématique du robot est supposée connue. La caméra peut être fixée dans l'espace de travail (Skaar, Brockman et Hanson, 1987). Les paramètres de la caméra sont alors inconnus mais l'allure de la trajectoire réelle de la cible, en l'occurrence une balle, est connue (trajectoire parabolique). Il s'agit d'intercepter la balle dont la trajectoire est estimée en temps réel en même temps que celle de la pince du manipulateur. L'intersection des 2 trajectoires est calculée puis la commande en angles correspondante envoyée au bras manipulateur.

Dans (Feddema et Mitchell, 1989), la caméra est fixée sur l'embout terminal du manipulateur (configuration "eye-in-hand") mais les auteurs ne connaissent pas la trajectoire réelle des indices dans l'image et la planifient donc à chaque nouvelle acquisition d'image. La tâche consiste à suivre un joint en détectant son trou central. Leur système est basé sur un module visuel de suivi de trajectoire dans l'image pour contrôler le segment terminal du manipulateur sur lequel est fixée la caméra. Le taux d'acquisition d'images peut varier au cours du temps. Le planificateur est prêt à prendre en compte tout changement de but dans l'image, ceci à tout instant, et à assurer la continuité en vitesse de la trajectoire des indices. Ce qui conduit à la mise en place d'un contrôleur dynamique basé sur l'image de type 'look-and-move' et dont l'équation est la suivante:

$$\dot{\theta} = \mathcal{J} \dot{f}$$

où  $f$  est le vecteur des indices dans l'image,  $\theta$  le vecteur des variables d'articulation et  $\mathcal{J}$  la matrice jacobienne englobant le jacobien cinématique, la transformation du repère de la caméra dans le repère de base du robot et la transformation perspective de la caméra.

Suivant le même schéma, d'autres auteurs introduisent par dessus le modèle dynamique du robot dans le module du contrôleur, (Espiau, Chaumette et Rives, 1992), afin de calculer les accélérations des angles de commande des joints.

Le choix des traits saillants dans l'image a également son importance dans la capacité de contrôler le manipulateur efficacement. Espiau et al. (1992) font une revue des

différentes primitives qui existent (points, lignes ou cercles).

L'avantage de cette approche est un délai réduit par l'absence d'interprétation des indices dans l'image. D'autre part la précision en positionnement du système est moins sensible aux erreurs de calibration de la caméra. Par contre, ce type de contrôle est sujet aux instabilités et singularités dans la matrice jacobienne liant les angles des joints aux points dans l'image. En définitive, il semble plus approprié aux tâches où le modèle géométrique fait défaut, comme la téléopération par exemple.

### **7.1.2 Contrôle basé sur la position**

Des travaux récents remettent en question la pertinence de l'information en 2D uniquement et préfèrent considérer la géométrie en 3D à partir des indices. Cette étape est généralement coûteuse en temps de calcul et se heurte aux difficultés de reconstruction de la pose d'un objet à partir de l'information dans l'image. Cependant elle bénéficie des améliorations de la vitesse des processeurs spécialisés, des progrès en stéréo et des senseurs 3D actifs.

Les travaux de Tsai et Lenz (1989) proposent une méthode de calibration de la caméra fixée sur le segment terminal du robot par rapport au référentiel du robot. Cette technique permet de bouger la caméra en fonction de la scène à observer. Andersson (1989) a mis au point le contrôle automatique d'un robot joueur de ping-pong afin d'intercepter la balle, prouvant ainsi la faisabilité du contrôle en trajectoire basée sur la position. Horaud, Dornaika et Espiau (1998) élargissent leurs travaux à l'usage

de 2 caméras stéréo indépendantes du robot et exploitent la géométrie 3D projective dans une tâche d'alignement de la pince avec un objet.

Wunsch et Hirzinger (1997) exploitent l'information dans l'espace afin de calculer la configuration spatiale de l'embout terminal par rapport à l'objet et de prévoir les occlusions. Le suivi est dans ce cas appuyé par un filtre de Kalman pour assurer la prédiction des grandeurs de position et orientation.

La connaissance de la position et orientation de la cible a l'avantage de ramener le contrôle de la pince dans l'espace cartésien, ce qui a été longuement étudié en robotique. Cependant, cette technique présente l'inconvénient de reposer sur le calcul fastidieux de la pose de l'objet en 3D à partir de l'image, d'une part, et sur une calibration précise de la caméra par rapport au robot d'autre part.

## 7.2 Modèle du manipulateur

En ce qui concerne notre étude, le choix du manipulateur s'est porté sur un robot industriel courant : le robot CRS A-460, dont un exemplaire se trouve au laboratoire. Il s'agit d'un robot manipulateur à six degrés de liberté, robuste et relativement facile à contrôler (figure 7.2).

Il est constitué de quatre membrures et d'un segment terminal articulés par 6 joints. Le premier angle,  $\theta_1$ , constitue la rotation autour de la base. Les articulations  $\theta_2$  et  $\theta_3$  réalisent les flexions des seconde et troisième membrures, le quatrième joint correspond à la rotation,  $\theta_4$ , de la troisième membrure. Enfin,  $\theta_5$  permet la flexion

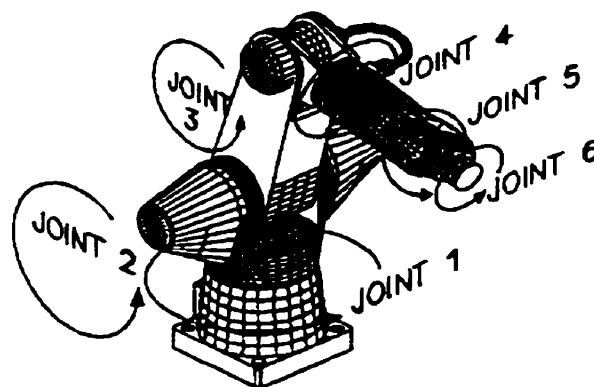


Figure 7.2 : Définition des articulations du robot CRS A-460 (tiré de la documentation technique du robot CRS A-460)

de la dernière membrure et  $\theta_6$  la rotation du segment terminal au bout duquel se fixe l'outil.

### 7.2.1 Cinématique directe

Un modèle graphique de ce robot a été implanté sous la librairie OpenGL et sa cinématique a été développée afin de l'animer. Les repères associés à chaque segment sont assignés de façon systématique selon le formalisme de Denavit-Hartenberg (figure 7.3).

Tableau 7.1 : Paramètres de Denavit-Hartenberg

# segment	rotation $\theta_i$	distance $d_i$	longueur $a_i$	courbure $\alpha_i$
1	$\theta_1$	$d_1$	0	$-\pi/2$
2	$\theta_2 - \pi/2$	0	$a_2$	0
3	$\theta_3$	0	0	$-\pi/2$
4	$\theta_4$	$d_4$	0	$\pi/2$
5	$\theta_5$	0	0	$-\pi/2$
6	$\theta_6$	$d_6$	0	0



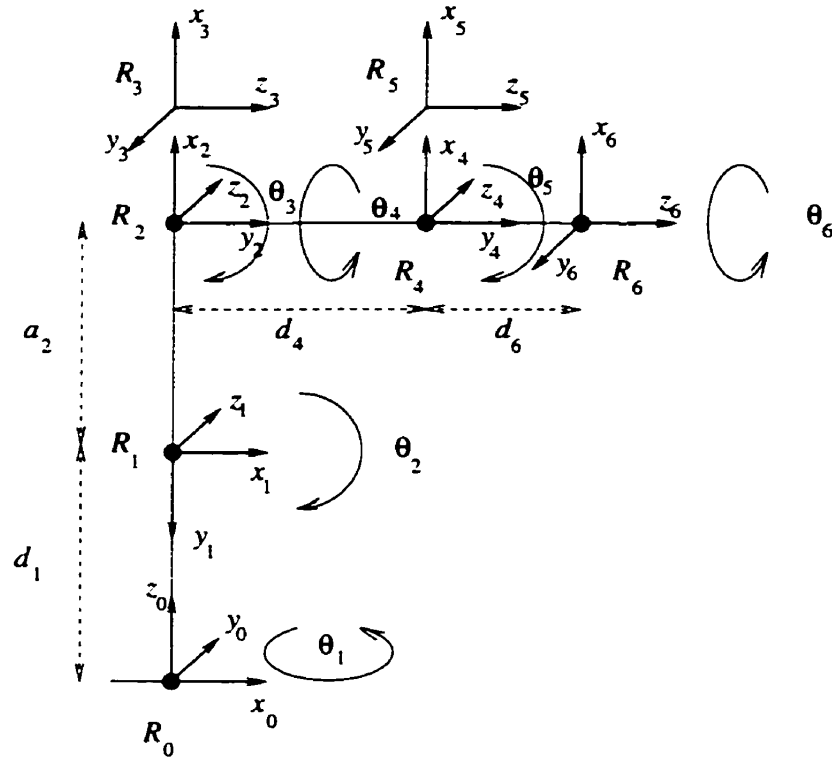


Figure 7.3 : Modèle cinématique du robot

Soit  $M[i-1, i]$  dénotant la matrice de transformation homogène décrivant la position et l'orientation du référentiel  $R_i$  par rapport au référentiel  $R_{i-1}$ ,  $M[i-1, i]$  s'exprime à l'aide des paramètres de Denavit-Hartenberg de la façon suivante :

$$M[i-1, i] = R[z, \theta_i] T[z, d_i] T[x, a_i] R[x, \alpha_i].$$

$M[0, 6]$  permet ainsi d'obtenir la position et l'orientation du référentiel  $R_6$  lié à la pince dans le référentiel de base  $R_0$ . La position est donnée par les trois coordonnées  $x$ ,  $y$  et  $z$  dans le repère  $R_0$  et l'orientation par les trois angles d'Euler : lacet (rotation par rapport à  $z_0$ ), tangage (rotation par rapport à  $y_0$ ) et roulis (rotation par rapport

à  $x_0$ ).

### 7.2.2 Cinématique inverse

Supposons spécifiées la position et l'orientation du référentiel  $R_6$  par rapport au référentiel de base  $R_0$ . Le problème de cinématique inverse consiste à déterminer les angles  $\theta_1$  à  $\theta_6$  qui réalisent la configuration désirée. Cependant, étant donné que le robot comporte six degrés de liberté, il est impossible de déterminer une solution analytique au problème de cinématique inverse. On procède par la méthode communément appelée "resolved motion rate control" qui exploite la matrice jacobienne obtenue dans le cas de déplacements incrémentaux (Schilling, 1990).

Soit  $q \in \mathbb{R}^6$  le vecteur des variables d'articulation  $\theta_i$  et  $x \in \mathbb{R}^6$  la configuration du repère de l'outil par rapport au repère de base du robot. Les équations de cinématique directe peuvent se formuler de la façon suivante :

$$x = w(q).$$

En différenciant cette équation, on obtient :

$$\dot{x} = V(q)\dot{q},$$

où  $V(q)$  correspond à la matrice jacobienne de l'outil :

$$V_{ij}(q) = \frac{\partial w_i(q)}{\partial q_j}.$$

Ainsi, pour un déplacement incrémental  $dx$ , l'incrément  $dq$  à ajouter au vecteur des variables d'articulation s'écrit :

$$dq = V^{-1}(q)dx.$$

### 7.3 Modèle de la pince

Tout comme le manipulateur, la pince a été créée sous la librairie OpenGL. Il s'agit d'une réplique de la pince BarrettHand de la compagnie Barrett Technology Inc. Elle est constituée de trois doigts, dont le pouce qui reste fixé à sa base. Les deux autres peuvent tourner autour du plan de la paume. De plus chaque doigt possède deux degrés de liberté en flexion (figure 7.4).

Deux prises distinctes ont été programmées, à savoir la prise sphérique et la prise cylindrique. Dans le cas de la prise sphérique, les doigts sont écartés à angles équidistants de 120 degrés et se referment progressivement. En ce qui concerne la prise cylindrique, les deux doigts mobiles effectuent une rotation de 180 degrés pour se coller aux côtés du pouce et tous se referment légèrement. L'intérêt de ces deux prises réside dans le fait que l'on peut aisément les distinguer lorsqu'effectuées au

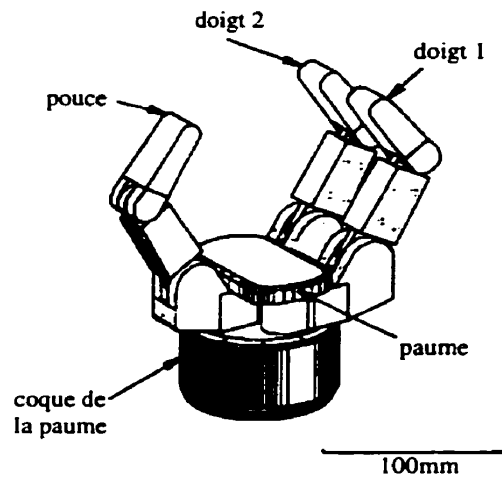


Figure 7.4 : Modèle de la pince Barrett (tiré de la documentation technique de Barrett Technology Inc.)

préalable par un opérateur humain. Trois postures distinctes sont illustrées à la figure 7.5.

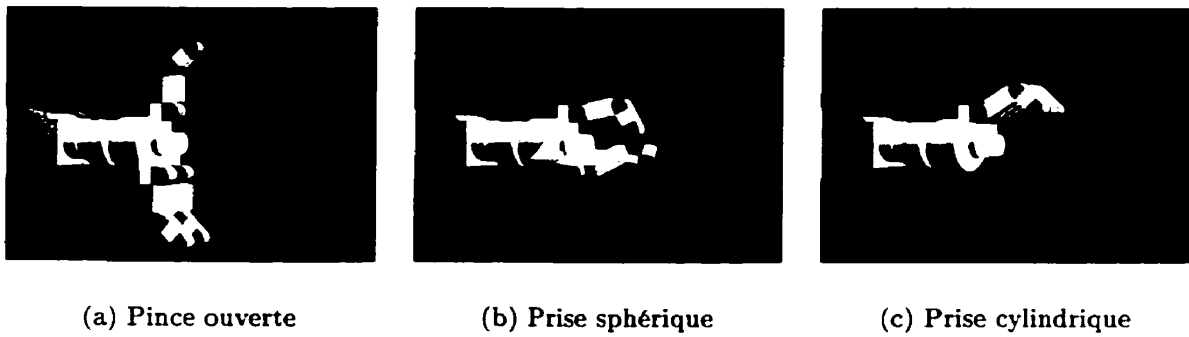


Figure 7.5 : Différentes configurations de la pince graphique

## 7.4 Contrôle de la pince par suivi de la main

L'idée est la suivante. Soit un opérateur effectuant une saisie d'objet à l'aide du gant filmé par une caméra. Supposons connues la position et l'orientation d'un repère attaché à l'objet par rapport au repère de la caméra. Lors du mouvement de la main, les positions et orientations successives de la main sont calculées par rapport au repère de la caméra. Supposons maintenant que l'on veuille guider une pince robotique de sorte qu'elle effectue la même tâche que l'opérateur en passant par les mêmes configurations intermédiaires que la main a suivies. On s'attarde ici sur la position du poignet et l'orientation de la paume afin de guider la pince et non sur les configurations individuelles de chaque doigt. Il suffit de considérer une caméra placée avec la même configuration que celle de la main par rapport à l'objet que le robot doit saisir. Les placements successifs de la pince par rapport à la caméra sont identiques à ceux qu'a suivis la main. Leur connaissance permet ainsi de contrôler le robot.

### 7.4.1 Principes

La première étape consiste à calibrer la caméra par rapport à une grille de calibration liée à l'objet à saisir. Soit  $C_m$  le repère lié à la caméra filmant la main,  $R_g$  le repère lié à la grille de calibration et  $R_t$  le repère lié à l'objet. La calibration de la caméra permet de déterminer la matrice de transformation  $M[C_m, R_g]$  grâce à des techniques classiques de calibration (Tsai, 1987; Wilson, 1994) à partir de l'image de la grille 7.6.

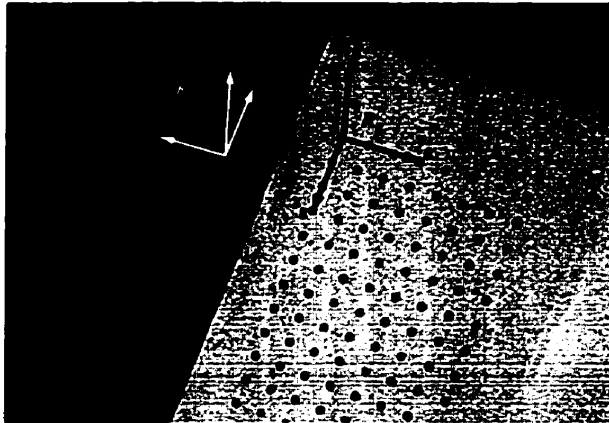


Figure 7.6 : Repères de la grille de calibration et de l'objet

On suppose connue la matrice de transformation  $M[R_g, R_t]$  donnant la position et orientation de  $R_t$  par rapport à  $R_g$ . En pratique, le repère lié à l'objet peut être choisi par rapport à la base du robot, comme illustré à la figure 7.7. On peut donc

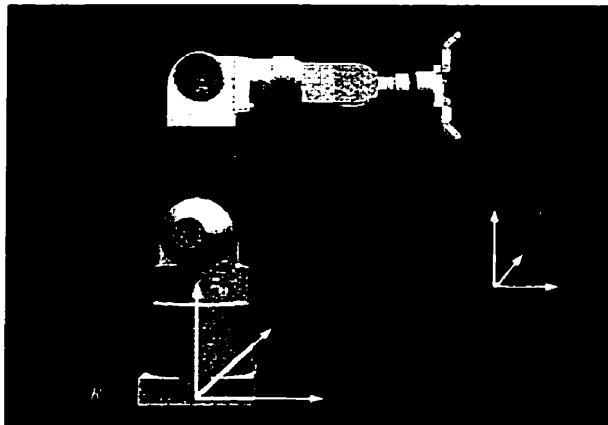


Figure 7.7 : Repère de l'objet par rapport à la base du robot

placer la caméra virtuelle  $C_0$  liée au robot qui permettra en théorie de visualiser la pince du même point de vue que la main par rapport à  $C_m$ . Il est à noter que  $M[C_m, C_0]$  est arbitraire.

La transformation qui permet de positionner la caméra liée au robot  $C_0$  par rapport au repère de base du robot  $R_0$  est donnée par

$$M[R_0, C_0] = M[R_0, R_t] M[R_t, R_g] M[R_g, C_m] M[C_m, C_0].$$

Vient maintenant l'étape du contrôle proprement dit. Soit  $A_i$  le repère attaché à la main correspondant à la position et à l'orientation de la paume dans le référentiel de la caméra  $C_m$ .  $A_i$  est déterminé à chaque instant par reconstruction de la posture de la main et donne l'expression de  $M[C_m, A_i]$ . Il faut rajouter une transformation au repère de la main car la pince ne correspond pas tout à fait à la physiologie de la main. Il s'agit d'ajuster le mouvement de la pince par rapport à celui de la main selon le type de saisie, sphérique ou cylindrique. Cette dernière transformation donne la relation  $M[A_i, R_6]$ . En définitive le contrôle du robot se fait par assignation des positions et orientations successives de l'outil par rapport à la base du robot données par la matrice de transformation  $M[R_0, R_6]$  :

$$M[R_0, R_6] = M[R_0, C_0] M[C_0, C_m] M[C_m, A_i] M[A_i, R_6].$$

C'est-à-dire

$$M[R_0, R_6] = M[R_0, R_t] M[R_t, R_g] M[R_g, C_m] M[C_m, A_i] M[A_i, R_6],$$

où l'on voit bien que le repère de la caméra  $C_0$  n'intervient pas et constitue donc bien un intermédiaire virtuel de visualisation.

### 7.4.2 Illustration

Deux séquences de la main réelle ont été prises afin d'illustrer les deux types de saisie. Les postures successives de la main ont ensuite été assignées au robot simulé. Les figures 7.8 et 7.9 illustrent les résultats obtenus sur une prise sphérique et sur une prise cylindrique respectivement ainsi que les configurations successives du manipulateur robotique muni de la pince.



### 7.4.3 Précision de la prise

La matrice de transformation  $M[A_i, R_6]$  qui assigne à la posture de la main une position et orientation de la pince est fixée selon le type de prise, sphérique ou cylindrique. Afin de valider cette transformation ainsi que la précision obtenue par la pince, des tests effectués sur plusieurs séquences des deux types de gestes ont été menés. De plus, plusieurs critères ont été retenus afin d'analyser la précision de la configuration de la pince au moment de la fermeture.

#### 7.4.3.1 Prise sphérique

En ce qui concerne la prise sphérique, le point qui est manipulé est le centre du poignet de la pince. Sa position en hauteur doit être surélevée par rapport à la tasse pour éviter tout heurt et sa projection sur le plan de repos de la tasse doit être le plus proche possible du centre de la tasse de sorte que les doigts de la pince se referment bien autour de la tasse. De plus l'orientation de l'axe du poignet de la pince doit être similaire à l'axe vertical de la tasse. On étudie donc la position du centre du poignet de la pince par rapport à la tasse ainsi que l'orientation de l'axe du poignet par rapport à l'axe  $z$  du repère de la tasse. La hauteur de la tasse vaut 0.092m. Idéalement, la position de la pince devrait se situer en  $[0, 0, 0.18]$ m et l'angle entre les axes devrait être nul. Quatre séquences calibrées ont été analysées et les résultats obtenus sont illustrés au tableau 7.2.

L'erreur est de l'ordre de quelques cm en ce qui concerne la position de la pince par

Tableau 7.2 : Écart de la position et de l'orientation de la pince par rapport à la configuration idéale lors de la fermeture sphérique

séquence	$x$ en m	$y$ en m	$z$ en m	angle en degrés
1	-0.028	-0.034	0.014	16
2	-0.012	0.007	-0.003	5.5
3	-0.008	0.000	-0.003	8
4	0.003	0.021	0.001	12.5

rapport au centre de la tasse. Le placement en hauteur est par contre impeccable. L'erreur en angles est de 10.5 degrés en moyenne, ce qui est moins précis. Seule la première prise présente un axe pivoté de façon marquée par rapport à l'axe de la tasse mais la prise est malgré tout valide, puisque c'est elle qui est illustrée à la figure 7.10.

Dans tous les cas, la pince est en mesure de saisir l'objet. Afin de mieux visualiser les positions finales du centre du poignet de la pince, les trajectoires effectuées par la pince sont retracées à la figure 7.11.

Les figures 7.11(a,b) illustrent deux vues en 3D de la trajectoire du point de contrôle de la pince pour les différentes séquences. La figure 7.11(c) retrace la projection de la trajectoire sur le plan  $(x, y)$  afin de situer le point d'arrivée par rapport au centre de la tasse. Enfin, la figure 7.11(d) présente l'évolution de l'angle formé par l'axe de la pince par rapport à l'axe  $z$  de la tasse. Elle permet d'analyser la précision de l'orientation de la pince au moment de sa fermeture et savoir si les trois doigts vont faire contact en même temps.

### 7.4.3.2 Prise cylindrique

Pour le cas de la prise cylindrique, il importe que le cylindre formé par la fermeture des doigts présente un axe parallèle à l'axe de la tasse et que le centre du poignet soit suffisamment proche d'un bord de la tasse pour la saisir. On s'attarde donc sur la projection du centre du poignet par rapport au cercle formant la base de la tasse, à la position du poignet par rapport à la mi-hauteur de la tasse ainsi que l'orientation de l'axe du cylindre par rapport à l'axe  $z$  du repère de la tasse. La hauteur de la pince vaut 0.092m et son rayon vaut 0.037m. La hauteur de la pince devrait se situer autour de 0.045m, la projection du centre du poignet devrait être localisée à un rayon de 0.14m et l'angle entre l'axe du cylindre formé par la pince et l'axe  $z$  de la tasse devrait être nul. Les résultats obtenus sur cinq séquences sont illustrés au tableau 7.3.

Dans le cas présent, le centre du poignet se projette sur l'axe  $y$  à environ 0.14m de

Tableau 7.3 : Écart de la position et de l'orientation de la pince par rapport à la configuration idéale lors de la fermeture cylindrique

séquence	$\sqrt{x^2 + y^2}$ en m	$z$ en m	angle en degrés
1	-0.009	-0.013	12
2	-0.001	-0.005	9
3	0.004	-0.038	5.5
4	0.004	-0.028	9
5	0.000	-0.017	3.5

la tasse, ce qui assure une saisie efficace de la tasse. De plus, l'axe de formation du cylindre est adéquat pour une saisie stable avec une erreur angulaire moyenne de 8 degrés. Il arrive cependant que la pince saisisse la tasse un peu trop par le bas (cas de

l'écart en  $z$  de  $-0.038\text{m}$ ). La figure 7.12 présente le résultat obtenu sur la première séquence.

La figure 7.13 présente les courbes des trajectoires de l'extrémité de la pince pour toutes les séquences.

Sur les figures 7.13(a,b), on retrouve deux vues en 3D de la trajectoire du centre de la pince pour chaque séquence. La projection sur le plan  $(x, y)$  est présentée à la figure 7.13(c) afin de localiser le point d'arrivée à l'extérieur du cercle formé par la base de la tasse (à environ  $0.14\text{m}$  du centre). Finalement, la figure 7.13(d) retrace l'écart de l'angle entre l'axe du cylindre formé par la pince et l'axe  $z$  de la tasse. Si l'axe du cylindre est dévié, la prise ne sera pas stable. En ce qui concerne les résultats obtenus, l'axe du cylindre est bien positionné pour assurer le contact de chacun des doigts avec la tasse. Cependant, la position de la pince se trouve décalée vers le bas, ce qui ne nuit pas ici à la stabilité de la prise mais qui peut entraîner une collision avec le support de la tasse. Dans ce cas, il importe d'ajouter un module de détection d'obstacles pour éviter ce genre d'inconvénient. La prise serait alors ajustée de façon adéquate.

#### 7.4.4 Conclusion

Il apparaît donc que la correspondance choisie entre la posture de la main et celle de la pince pour un type de prise donné est cohérente sur plusieurs séquences et ne dépend pas de l'analyse d'une séquence en particulier. Il est donc possible de contrôler la

pince en fonction des données de position et orientation de la main si la catégorie de prise a été reconnue. Bien sûr, l'étape critique reste la fin de la saisie, étant donné que les données peuvent faire défaut en ce qui concerne la posture de la main en présence d'occlusions. En pratique, il serait préférable d'ajouter un contrôle basé sur l'image afin de procéder à une saisie plus fine et de contrôler plus précisément la fin de la saisie.



(a) main origi-  
nale

(b)

(c)



(d) main  
graphique



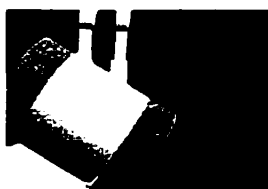
(e)



(f)



(g) pince vue du  
même point de  
vue



(h)



(i)



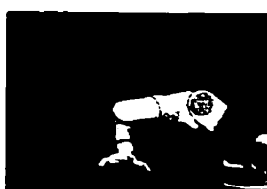
(j)



(k) pince vue  
d'un second  
point de vue



(l)



(m)



(n)

Figure 7.8 : Mouvement de la main et de la pince pour une prise sphérique

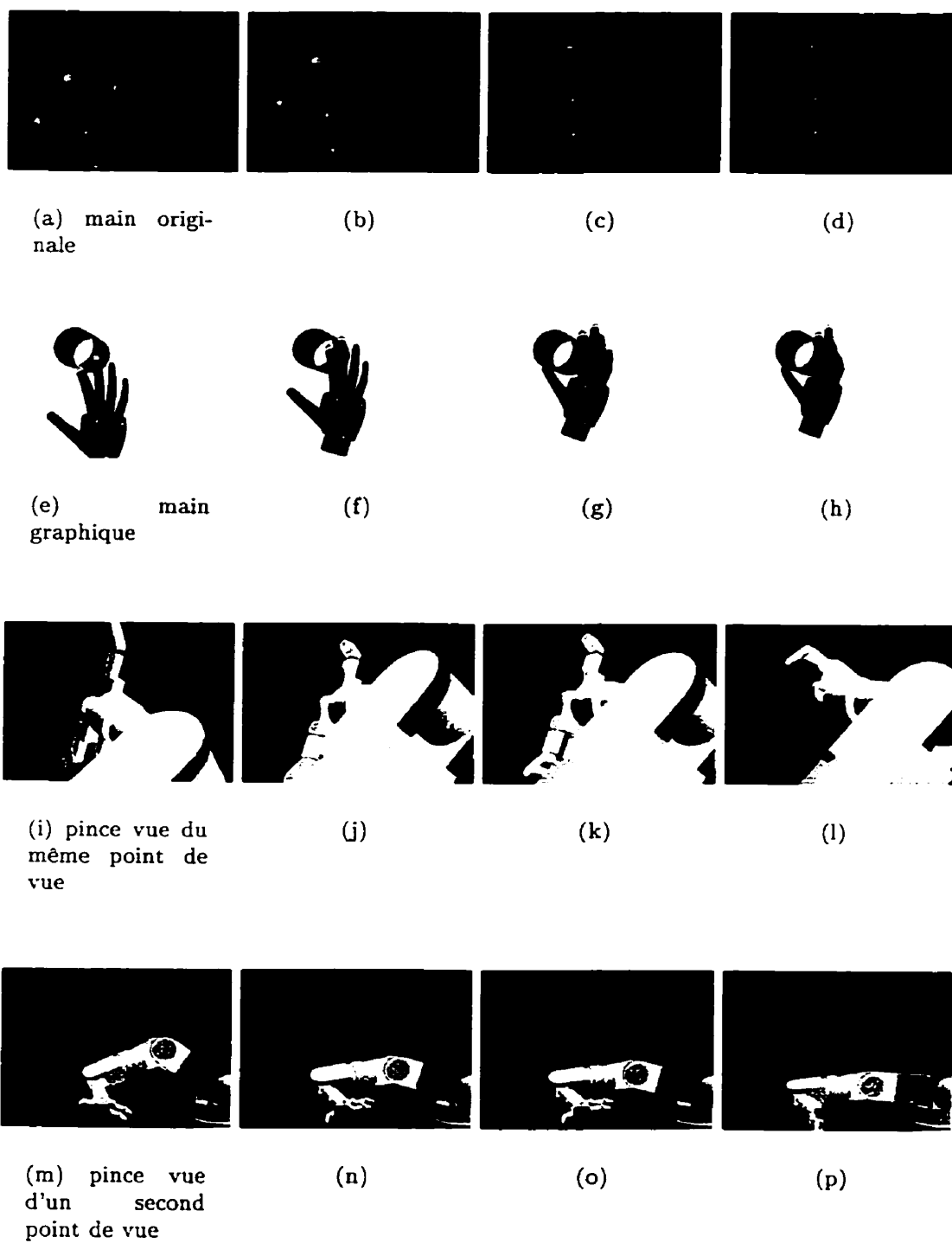


Figure 7.9 : Mouvement de la main et de la pince pour une prise cylindrique



Figure 7.10 : Exemple de configuration de la prise sphérique



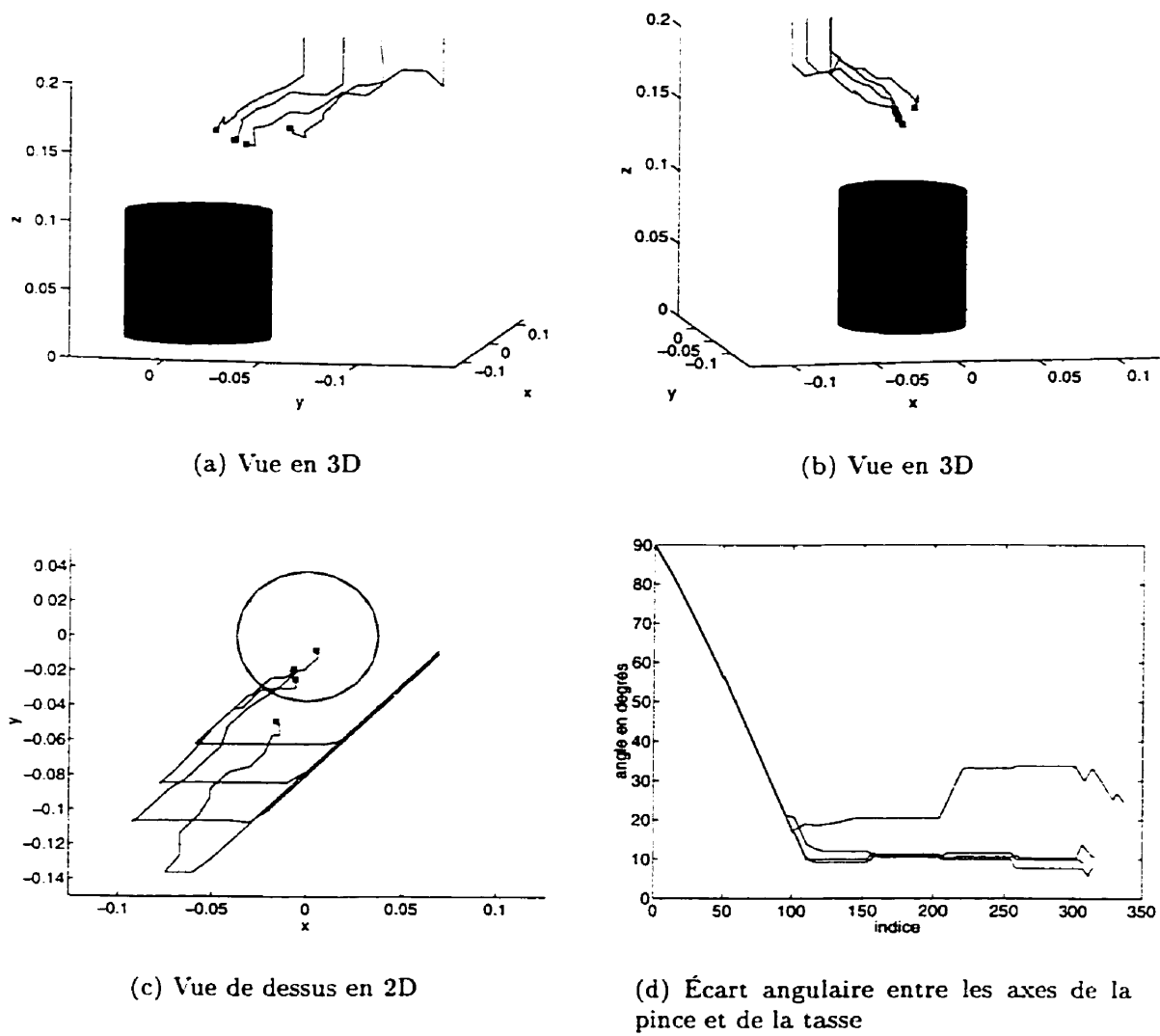
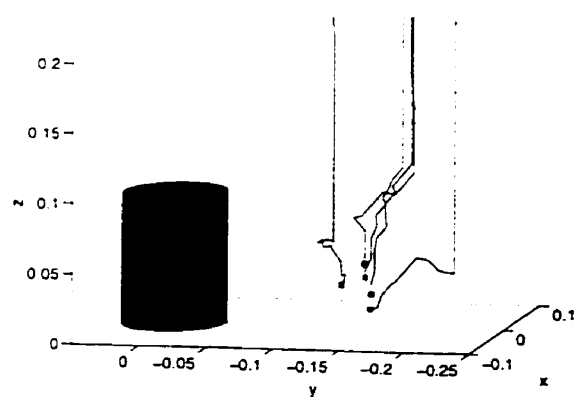


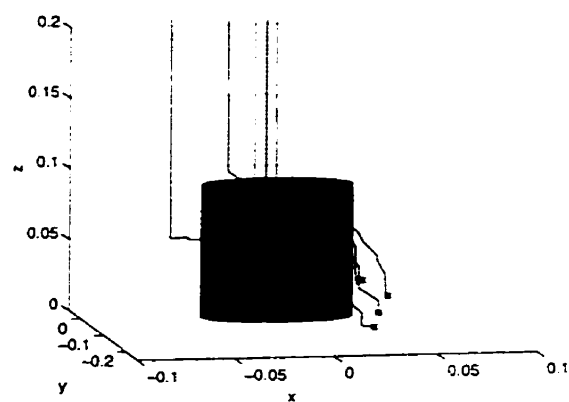
Figure 7.11 : Trajet de la pince pour différentes séquences d'une prise sphérique



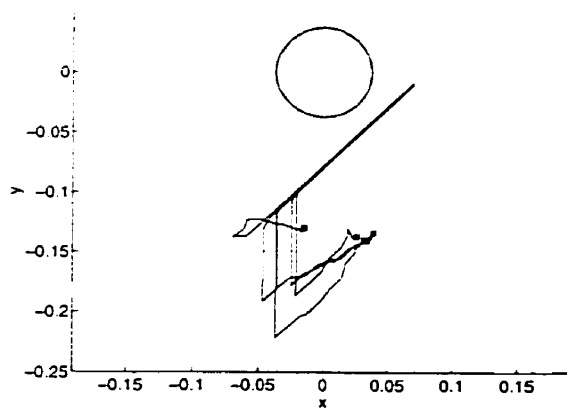
Figure 7.12 : Exemple de configuration de la prise cylindrique



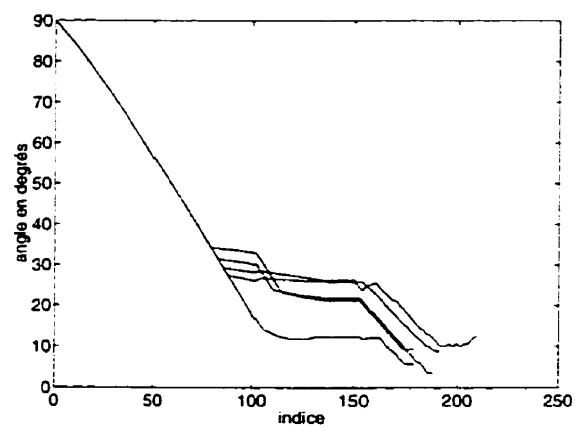
(a) Vue en 3D



(b) Vue en 3D



(c) Vue de dessus en 2D



(d) Écart angulaire entre l'axe du cylindre de la pince et celui de la tasse

Figure 7.13 : Trajet de la pince pour différentes séquences d'une prise cylindrique

## Chapitre 8

### Conclusion

Le domaine de l'Interface Humain-Machine est un univers en ébullition ouvrant les portes à des applications nombreuses et variées afin de rehausser le degré d'interaction entre l'utilisateur et la machine. Entre la reconnaissance vocale ou bien la détection de visée du regard, l'interprétation des gestes suscite un intérêt marqué pour son utilisation intuitive et naturelle. La main peut en effet servir à assigner des instructions à l'ordinateur, à contrôler le mouvement d'un objet à l'écran ou bien encore à manipuler un objet réel et offrir l'éventail de ses habiletés. Parmi les techniques de détection des variables de configuration de la main, on peut citer les gants munis de capteurs, les suiveurs magnétiques ou bien les techniques de détection visuelle. La difficulté majeure de la reconnaissance de la posture de la main réside dans le fait que la main est constituée d'un ensemble de chaînes cinématiques articulées, caractérisées par un nombre important de degrés de liberté et sujettes aux occlusions. Les systèmes

existant présentent souvent l'inconvénient de contraindre les mouvements du sujet, de manquer de précision et de ne pas répondre en temps réel.

Dans le cadre de ce projet, nous nous sommes intéressés à la détection visuelle de la posture de la main en mouvement à l'aide d'une unique caméra couleur filmant des indices d'intérêt disposés sur la main. Nous avons tout d'abord développé un modèle original de la cinématique de la main sous forme d'un squelette à vingt-six degrés de liberté pour lequel les équations de cinématique directe et inverse ont été explicitées. Cette première contribution fait l'objet du chapitre 3. L'avantage de ce modèle est d'être à la fois relativement simple afin d'extraire les équations de cinématique inverse indispensables à l'étape de reconstruction des variables d'articulation de la main tout en conservant les degrés de mobilité naturelle de la main, en particulier le mouvement de la pince effectuée par le pouce.

Le chapitre 4 présente la méthode envisagée de détection de la posture de la main à partir d'indices de couleur disposés sur le gant dont celle-ci est revêtue. Cette méthode est basée sur le modèle cinématique de la main et exploite l'information minimale nécessaire à la reconstruction des variables d'articulation de la main. L'étape de reconstruction fait appel à un processus d'optimisation de l'angle d'abduction et aux équations de cinématique inverse. Le choix de la méthode vise l'efficacité en terme de coût de calcul, ce qui permet d'aboutir à la mise en place d'une interface de suivi de la posture de la main fonctionnant en temps réel. Afin de disposer d'une interface robuste, fonctionnant peu importe la posture de la main, un algorithme de

détection des occlusions a été développé au chapitre 5. Il consiste à prédire la posture des doigts masqués et à valider la nouvelle position des doigts en vérifiant l'occlusion par une méthode de visibilité géométrique. La pertinence des résultats obtenus permet d'envisager l'utilisation de ce système afin d'exploiter plus en détail l'information contenue dans la trajectoire de la main.

La dernière partie du mémoire concerne les applications du système de suivi de la main. La première étude concernant l'analyse de certains gestes de préhension est décrite au chapitre 6. Plusieurs types de saisie d'objets ont été analysés à partir de séquences de prise de l'objet par l'opérateur. Des critères de caractérisation de ces gestes, basés sur les trajectoires de quelques angles seulement, ont été mis en lumière, permettant de distinguer des propriétés particulières propres à chaque type de préhension. Cette étape confirme la possibilité de reconnaître des gestes de saisie spécifiques à partir du modèle d'état de la main.

Les propriétés de dextérité de la main humaine ont été mises à profit au chapitre 7 dans l'application de contrôle d'une pince robotique basée sur le modèle guide de l'humain. La calibration de la caméra liée à l'espace de travail de l'opérateur par rapport à l'objet saisi par l'opérateur permet de connaître les configurations de la main par rapport à l'objet. Un robot situé dans un espace éloigné, en charge de saisir un objet similaire dont la position lui est connue, peut ainsi bénéficier de l'information de trajectoire de la main transformée dans le repère de l'objet qu'il doit saisir. Des simulations ont démontré la faisabilité de cette technique originale d'"enseignement

par l'exemple".

En conclusion, ce travail propose trois principales contributions novatrices:

- Sur le plan théorique, nous avons modélisé la cinématique de la main à l'aide d'un modèle squelette à vingt-six degrés de liberté, manipulable en cinématique directe et inverse.
- Sur le plan pratique, nous avons implanté une interface de détection visuelle et de reconstruction de la posture de la main fonctionnant de façon robuste et en temps réel.
- Sur le plan expérimental, nous avons appliqué l'interface au contrôle d'une pince robotique guidée par le mouvement humain exploitant l'étude préliminaire des caractéristiques de plusieurs gestes de préhension.

Pour finir, le système développé dans ce mémoire présente l'avantage d'être multidisciplinaire. En effet, le gant revêtu des indices de couleur constitue un moyen de contrôle à multi degrés de liberté. Il peut se transformer tout à tour en un périphérique d'entrée pour des jeux vidéo permettant de diriger des objets en position et en orientation, en un outil de modélisation graphique 3D pour de la sculpture ou encore en outil de contrôle à plusieurs degrés de liberté redéfinissables à partir des vingt-six variables offertes par la main. Pour ce faire, il serait souhaitable d'apporter certaines améliorations quant à la qualité du modèle reconstruit, en particulier estimer plus précisément les angles d'abduction et traiter de façon exhaustive toutes les

occlusions possibles.



## Bibliographie

- AHMAD, T., TAYLOR, C., LANITIS, A. et COOTES, T. (1997), Tracking and recognising hand gestures. using statistical shape models, *Image and Vision Computing* **15**, 345–352.
- ANDERSSON, R. (1989), Dynamic sensing in a ping-pong playing robot. *ITRA* **5**(6). 728–739.
- BOBICK, A. et WILSON, A. (1995), A state-based technique for the summarization and recognition of gesture, dans *Proceedings of the IEEE International Conference on Computer Vision*, Cambridge, MA, pp. 1093–1099.
- BURDEA, G. (1996), *Force and Touch Feedback for Virtual Reality*, Wiley-Interscience, New York.
- CARD, S., ENGLISH, W. et BURR, B. (1978), Evaluation of a mouse rate-controlled isometric joystick, step keys and text keys for text selection on a CRT. *Ergonomics* **21**, 601–613.

- CIPOLLA, R., OKAMOTO, Y. et KUNO, Y. (1993), Robust structure from motion using motion parallax, dans *Proceedings of the IEEE International Conference on Computer Vision*, Berlin, Germany, pp. 374–382.
- DARRELL, T. et PENTLAND, A. (1993), Space-time gestures, dans *IEEE conference on Computer Vision and Pattern Recognition*, New York, pp. 335–340.
- DAVIS, J. et BOBICK, A. (1997), The representation and recognition of human movement using temporal templates, dans *IEEE conference on Computer Vision and Pattern Recognition*, Puerto Rico, pp. 928–934.
- DAVIS, J. et SHAH, M. (1994), Recognizing hand gestures, dans *Proceedings of the European Conference on Computer Vision*, Stockholm, Sweden, pp. 331–340.
- DEMENTHON, D. et DAVIS, L. (1990), New exact and approximate solutions of the three-point perspective problem, dans *Proceedings of the IEEE International Conference on Robotics and Automation*, Vol. 1, Cincinnati, OH, pp. 40–45.
- ESPIAU, B., CHAUMETTE, F. et RIVES, P. (1992), A new approach to visual servoing in robotics, *IEEE Transactions on Robotics and Automation* **8**(3), 313–326.
- FEDDEMA, J. et MITCHELL, O. (1989), Vision-guided servoing with feature-based trajectory generation, *IEEE Transactions on Robotics and Automation* **5**(5), 691–700.

- FITTS, P. (1954), The information capacity of the human motor system in controlling the amplitude of movement, *Journal of Experimental Psychology* **47**, 381-391.
- FOLEY, J., DAM, A. V., FEINER, S. et HUGHES, J. (1996), *Computer Graphics : Principles and Pratique*, Addison-Wesley Systems Programming Series.
- GAVRILA, D. et DAVIS, L. (1995), Towards 3-D model-based tracking and recognition of human movement: a multi-view approach, dans *International Workshop on Face and Gesture Recognition*, Zurich.
- GIBSON, J. (1977), The theory of affordances, dans R. Shaw et J. Bransford, éditeurs. *Perceiving, Acting and Knowing: Toward an Ecological Psychology*, John Wiley and Sons, NewYork.
- GONCALVES, L., BERNARDO, E. D., URSELLA, E. et PERONA, P. (1995). Monocular tracking of the human arm in 3D, dans *Proceedings of the IEEE International Conference on Computer Vision*, Cambridge, MA, pp. 764-770.
- HAND, C. (1994), A survey of 3-D input devices, rapport technique CS TR94/2, Department of Computer Science, De Montfort University, Leicester. UK.
- HORAUD, R., DORNAIKA, F. et ESPIAU, B. (1998), Visually guided object grasping, *IEEE Transactions on Robotics and Automation* **14**(4), 525-532.
- HUTCHINSON, S., HAGER, G. et CORKE, P. (1996), A tutorial on visual servo control, *IEEE Transactions on Robotics and Automation* **12**(5), 651-670.

- JEANNEROD, M. (1984), The timing of natural prehension movements, *Journal of Motor Behavior* **16**(3), 235–254.
- JEANNEROD, M. (1994), Object oriented action, dans K. Bennett et U. Castiello, éditeurs, *Insights into the Reach to Grasp Movement*, Advances in Psychology 105, North Holland.
- JO, K., KUNO, Y. et SHIRAI, Y. (1998), Context-based recognition of manipulative hand gestures for human computer interaction, dans *Proceedings of the 3rd Asian Conference on Computer Vision*, Vol. 2, Hong Kong, China, pp. 368–375.
- KAKADIARIS, I. et METAXAS, D. (1995), 3D human body model acquisition from multiple views, dans *Proceedings of the Fifth IEEE International Conference on Computer Vision*, Boston, MA, pp. 618–623.
- KENDON, A. (1986), Current issues in the study of gesture, dans P. P. J. Nespoulos et A. Lecours, éditeurs, *The biological foundations of gestures: Motor and Semiotic Aspects*, Lawrence Erlbaum Associates, Hillsdale, NJ, pp. 23–47.
- KJELDSEN, R. et KENDER, J. (1997), Interaction with on-screen objects using visual gesture recognition, dans *IEEE conference on Computer Vision and Pattern Recognition*, Puerto Rico, pp. 788–793.
- KOFFKA, K. (1935), *Principles of Gestalt psychology*, Harcourt, Brace, and World, New York.

- KUCH, J. et HUANG, T. (1995), Vision based hand modeling and tracking for virtual teleconferencing and telecollaboration, dans *Proceedings of the IEEE International Conference on Computer Vision*, Cambridge, MA, pp. 666–671.
- KUNESCH, E., BINKOFSKI, F. et FREUND, H.-J. (1989), Invariant temporal characteristics of manipulative hand movements, *Experimental Brain Research* **78**, 539–546.
- KUNO, Y., SAKAMOTO, M., SAKATA, K. et SHIRAI, Y. (1994), Vision-based human interface with user-centered frame, dans *Proceedings of the IEEE International Conference on Intelligent Robots and Systems*, Munich, Germany, pp. 2023–2029.
- LATHUILLIÈRE, F. et HERVÉ, J.-Y. (1999a), Visual tracking of hand posture in a robot control application, dans *Proceedings of the 12<sup>th</sup> conference on Vision Interface*, Trois-Rivières, Canada, pp. 188–195.
- LATHUILLIÈRE, F. et HERVÉ, J.-Y. (1999b), Visual tracking of hand posture with occlusion handling, rapport technique GRPR-RT 9902, GRPR, École Polytechnique de Montréal.
- LATHUILLIÈRE, F. et HERVÉ, J.-Y. (2000a), Human computer interface based on visual hand posture tracking, soumis à EUSIPCO 2000.
- LATHUILLIÈRE, F. et HERVÉ, J.-Y. (2000b), Visual tracking of hand posture with occlusion handling, soumis à ICPR 2000.

- LEE, J. et KUNII, T. (1995), Model-Based Analysis of Hand Posture, *IEEE Computer Graphics and Applications* **15**(5), 77–86.
- LIEN, C. et HUANG, C. (1998), The model-based dynamic hand posture identification using genetic algorithm, dans *Proceedings of the 3rd Asian Conference on Computer Vision*, Vol. 1, Hong Kong, China, pp. 706–713.
- MILLAR, R. et CRAWFORD, G. (1996), A mathematical model for hand-shape analysis, dans *Progress in Gestural Interaction, Proceedings of Gesture Workshop*.
- NECHYBA, M. et XU, Y. (1996), On the fidelity of human skill models, dans *Proceedings of the IEEE International Conference on Robotics and Automation*, Vol. 3, Minneapolis, MN, pp. 2688–2693.
- NESPOULOUS, J. et LECOURE, A. (1986), Gestures: Nature and function, dans P. P. J. Nespoulos et A. Lecours, éditeurs, *The Biological Foundations of Gestures: Motor and Semiotic Aspects*, Lawrence Erlbaum Associates, Hillsdale, NJ, pp. 49–62.
- PAVLOVIC, V., SHARMA, R. et HUANG, T. (1997), Visual interpretation of hand gestures for Human-Computer Interaction: A review, *IEEE Transactions on Pattern Analysis and Machine Intelligence* **19**(7), 677–694.
- PLAMONDON, R. (1995), A kinematic theory of rapid human movements, *Biological Cybernetics* **72**, 295–320.

- PLAMONDON, R. et ALIM, A. (1997), Speed/accuracy trade-offs in target-directed movements, *Behavioral and Brain Sciences* **20**(2), 279–349.
- QUEK, F. (1994), Toward a vision-based hand gesture interface, dans *Proceedings of the Virtual Reality System and Technology Conference*, Singapore, pp. 17–29.
- RABINER, L. (1989), A tutorial on hidden Markov models and selected applications in speech recognition. *Proceedings of the IEEE* **77**(2), 257–286.
- REHG, J. et KANADE, T. (1993), DigitEyes: Vision-based human hand tracking, rapport technique CMU-CS-93-220, School of Computer Science, Carnegie Mellon University, Pittsburgh, PA.
- REHG, J. et KANADE, T. (1995), Model-based tracking of self-occluding articulated objects, dans *Proceedings of the IEEE International Conference on Computer Vision*, Cambridge, MA, pp. 612–617.
- RIGOLL, G. et KOSMALA, A. (1997), New improved feature extraction methods for real-time high performance image sequence recognition, dans *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing*, Vol. 4, Munich, Germany, pp. 2901–2904.
- RIJPKEMA, H. et GIRARD, M. (1991), Computer animation of knowledge-based human grasping, *Computer Graphics, Proc. SIGGRAPH* **25**(4), 339–347.

- SANDERSON, A. et WEISS, L. (1983), Adaptive visual servo control of robots, dans A. Pugh, éditeur, *Robot Vision*, IFS, Bedford, UK, pp. 107–116.
- SCHILLING, R. (1990), *Fundamentals of Robotics : Analysis and Control*. Prentice-Hall.
- SHIMADA, N., SHIRAI, Y., KUNO, Y. et MIURA, J. (1998), 3-D pose estimation and model refinement of an articulated object from a monocular image sequence, dans *Proceedings of the 3rd Asian Conference on Computer Vision*, Vol. 1, Hong Kong, China, pp. 672–679.
- SKAAR, S., BROCKMAN, W. et HANSON, R. (1987), Camera-space manipulation, *The International Journal of Robotics Research* **6**(4), 20–32.
- TSAI, R. (1987), A versatile camera calibration technique for high-accuracy 3D machine vision metrology using off-the-shelf TV cameras and lenses, *IEEE Journal of robotics and Automation* **3**(4), 323–344.
- TSAI, R. et LENZ, R. (1989), A new technique for fully autonomous and efficient 3D robotics hand/eye calibration, *IEEE Transactions on Robotics and Automation* **5**(3), 345–358.
- TUBIANA, R. (1981), *The Hand*. Vol. 1, Sanders, Philadelphia, PA.



- WEISS, L., SANDERSON, A. et NEUMAN, C. (1987), Dynamic sensor-based control of robots with visual feedback, *IEEE Journal of Robotics and Automation* 3(5), 404–417.
- WELMAN, C. (1993), Inverse kinematics and geometric constraints for articulated figure manipulation, Master's thesis, Simon Fraser University.
- WILSON, R. (1994), Modeling and Calibration of Automated Zoom Lenses. Thèse de doctorat, Carnegie Mellon University, Pittsburg, PA.
- WUNSCH, P. et HIRZINGER, G. (1997), Real-time visual tracking of 3-D objects with dynamic handling of occlusion, dans *Proceedings of the IEEE International Conference on Robotics and Automation*, Albuquerque, NM, pp. 2868–2873.
- YAMAMOTO, M. et KOSHIKAWA, K. (1991), Human motion analysis based on a robot arm model, dans *IEEE conference on Computer Vision and Pattern Recognition*, Hawaii, USA, pp. 664–665.
- YAMATO, J., OHYA, J. et ISHII, K. (1992), Recognizing human action in time-sequential images using hidden markov model, dans *Proceedings of the IEEE International Conference on Robotics and Automation*, Nice, France, pp. 379–385.
- YANG, J., XU, Y. et CHEN, C. (1994), Gesture interface: Modeling and learning, dans *Proceedings of the IEEE International Conference on Robotics and Automation*, Vol. 2, San Diego, California, pp. 1747–1752.

YANG, J., XU, Y. et CHEN, C. (1997), Human action learning via hidden markov model, dans *IEEE International Conference on Systems, Man, and Cybernetics*, Vol. 27-1, Hyatt Orlando, Orlando, Florida, pp. 34-44.