



**CAPTURE DE MOUVEMENT PAR FUSION DE POSES MULTIVUES  
POUR UNE RÉADAPTATION EN ENVIRONNEMENT VIRTUEL**

**Mémoire**

**Michaël Bonenfant**

**Maîtrise en génie électrique**  
Maître ès sciences (M. Sc.)

Québec, Canada

© Michaël Bonenfant, 2017



# Résumé

Plusieurs recherches ont démontré les avantages de la réadaptation physique assistée par la robotique ainsi que la réadaptation en environnement virtuel. Les exercices de réadaptation sont fréquemment effectués sur un système coûteux de capture de mouvement basé sur le suivi de marqueurs fixés sur le patient. L'utilisation de ce type de système nécessite du temps et l'expertise d'un clinicien. Le projet que nous proposons a pour but de supporter de tels exercices de réadaptation en plus de pallier aux problèmes des systèmes de capture de mouvement conventionnels très coûteux. Le système développé comprend un mécanisme à câbles intégré en environnement virtuel procurant un retour haptique et visuel au patient. L'estimation de la pose humaine sans marqueur est effectuée en temps réel pour permettre l'analyse du mouvement, l'affichage de l'avatar du patient ainsi que son interaction avec l'environnement virtuel. Le système de vision fait usage de capteurs 3D peu coûteux afin d'offrir un dispositif abordable pour la plupart des cliniques de réadaptation. Un tel système permet une reconfigurabilité importante afin d'optimiser les exercices de réadaptation aux besoins du patient. Ce mémoire couvre les étapes du développement du système, notamment l'approche de fusion de pose selon des points de vue multiples basée sur le filtrage particulaire. L'analyse des résultats présentés est prometteuse à différents niveaux sachant que le système est aussi très flexible et adaptable à différents contextes autres que la réadaptation, comme le milieu du divertissement.

# Abstract

Several works have shown the benefits of robotic assisted physical rehabilitation as well as rehabilitation in a virtual environment. Rehabilitation exercises are often performed using an expensive motion capture system based on markers attached to the patient. The use of this type of system requires time and expertise from a clinician. The project we are introducing aims to support such rehabilitation exercises in addition to avoid some of the problems of an expensive conventional motion capture systems. The developed system includes a cable mechanism integrated into a virtual environment providing haptic and visual feedback to the patient. The estimation of the human pose without marker is carried out in real-time to allow the analysis of the movement, the display of the patient avatar as well as its interaction with the virtual environment. The vision system exploit inexpensive 3D sensors to provide an affordable device for most rehabilitation clinics. That system allows enough reconfigurability to optimize the rehabilitation exercises to the needs of the patient. This thesis covers the stages of the system development, including the fusion approach based on multiple views based on particle filtering. The analysis of the results presented is promising, knowing that the system is also very flexible and adaptable to different contexts other than rehabilitation, such as the entertainment industry.

# Table des matières

<b>Résumé</b>	<b>iii</b>
<b>Abstract</b>	<b>iv</b>
<b>Table des matières</b>	<b>v</b>
<b>Liste des figures</b>	<b>viii</b>
<b>Liste des tableaux</b>	<b>xii</b>
<b>Remerciements</b>	<b>xiii</b>
<b>Introduction</b>	<b>1</b>
<b>Revue de littérature</b>	<b>4</b>
1.1 Capteurs 3D .....	4
1.1.1 Caméra temps de vol .....	5
1.1.2 Kinects .....	6
Kinect v1 .....	6
Kinect v2 .....	7
1.1.3 Vicon.....	8
1.2 Calibration .....	9
1.2.1 Calibration intrinsèque.....	9
1.2.2 Calibration extrinsèque .....	10
1.2.3 Alignement d'information 3D.....	11
Iterative Closest Point .....	13
1.3 Estimation de pose humaine .....	14
1.3.1 Estimation de pose avec marqueurs .....	15
1.3.2 Estimation de pose sans marqueurs .....	16
Algorithme de Microsoft .....	16
1.4 Fusion de données .....	18
1.4.1 Filtre particulaire.....	18
<b>Méthodologie</b>	<b>20</b>

2.1	Description du système proposé .....	21
2.2	Calibration et alignement.....	23
2.2.1	Calibration intrinsèque des Kinects.....	25
2.2.2	Calibration extrinsèque des Kinects.....	26
	Alignement brut.....	27
	Alignement fin.....	31
2.2.3	Calibration du mécanisme à câbles .....	33
2.2.4	Calibration du casque VR.....	37
2.3	Fusion des données de pose du sujet .....	39
2.3.1	Traitement des données à fusionner .....	40
	Acquisition et transfert .....	40
	Alignement.....	41
	Correspondance.....	41
	Mesure de qualité des données.....	42
2.3.2	Contraintes reliées au squelette.....	44
2.3.3	Synchronisation et interpolation des poses.....	46
	Fusion des poses en temps réel.....	47
	Fusion des poses en post-traitement .....	47
2.3.4	Filtrage et intégration - Détails d'implantation .....	48
2.4	Environnement virtuel.....	50
2.4.1	Casque VR .....	51
2.4.2	Scène .....	52
2.4.3	Effecteur du mécanisme à câbles.....	53
2.4.4	Avatar du sujet .....	54
	<b>Résultats expérimentaux</b> .....	<b>57</b>
3.1	Spécifications du montage .....	57
3.2	Matériel de validation .....	59
3.2.1	Calibration du système de validation.....	60
3.3	Acquisition des données de suivi du squelette.....	62
3.3.1	Détails des exercices effectués .....	63
	Pose fixe.....	63
	Marche .....	63

Marche statique .....	63
Flexion .....	63
Manutention.....	63
Manutention Plus.....	64
Rotation droite.....	64
Rotation gauche .....	64
Squat.....	64
Squat Plus.....	64
3.4 Méthode de validation des données.....	64
3.4.1 Alignement global .....	66
3.4.2 Alignement temporel.....	67
3.4.3 Alignement des modèles de squelette .....	68
3.5 Analyse des résultats expérimentaux .....	74
Robustesse de la méthode de fusion.....	75
Corrélation des résultats de fusion et de validation .....	78
Fiabilité des mesures de position et d'angle.....	81
Erreur angulaire et erreur de position.....	82
<b>Conclusion</b>	<b>89</b>
Suivi et travaux futurs.....	89
<b>Annexe A Spécifications des caméras Bonita Vicon</b>	<b>91</b>
<b>Annexe B Nexus: logiciel de capture de données Vicon</b>	<b>93</b>
<b>Annexe C Modèle du Squelette de l'algorithme d'estimation de pose de Microsoft</b>	<b>95</b>
<b>Annexe D Données supplémentaires</b>	<b>98</b>
<b>Références</b>	<b>102</b>

# Liste des figures

Figure 1 – Utilisation du montage avec environnement virtuel simplifié. (a) Utilisateur immergé en réalité virtuelle manipulant le mécanisme à câbles (b) Représentation simplifiée de la scène (table, boîte et avatar) selon le point de vue d'une caméra virtuelle (c) Vue selon la perspective de l'utilisateur portant le casque de réalité virtuelle. ....	3
Figure 1.1 – Classification des méthodes d'extraction d'information de forme en 3 dimensions. Informations provenant de <a href="#">[1]</a> Fig. 1.1. ....	5
Figure 1.2 – Caméra Vicon observant des marqueurs sphériques rétro réfléchissants fixés sur un objet. En rouge: une des caméras Vicon. En bleu: l'objet suivi. En jaune: un des marqueurs.....	9
Figure 1.3 – Projection de perspective d'un point selon les transformations entre les repères image, caméra et monde.....	12
Figure 1.4 – Classification des technologies de capture de mouvement humain. Informations provenant de <a href="#">[18]</a> Fig. 2. ....	14
Figure 1.5 – Reconstruction des membres inférieurs d'un sujet suivi par marqueurs avec le logiciel Nexus de Vicon. ....	16
Figure 1.6 – Identification des joints sur une pose humaine à partir d'une image de profondeur selon la méthode <a href="#">[20]</a> . (a) Information de l'image de profondeur (b) Déduction de la distribution des segments du corps par pixel. (c) Proposition de la position des joints déduits. Image provenant de <a href="#">[20]</a> Figure 5. ....	17
Figure 2.1 – Schéma du système de capture de mouvement en réalité virtuelle avec retour haptique développé. ....	22
Figure 2.2 – Transformation entre le repère global et les repères de deux caméras K1 et K2.....	24
Figure 2.3 – Cible trouée vue par un capteur Kinect. (a) Cible vue par la caméra RGB (b) Cible vue par le capteur de profondeur et segmentée en distance. (c) Points détectés sur la cible. ....	26
Figure 2.4 – Identification d'une sphère selon les repères de 4 caméras non-alignées, chaque couleur représente l'information provenant d'un point de vue différent. (a) Nuages de points (b) Sphères identifiées dans les nuages de points par la méthode	

RANSAC.....	29
Figure 2.5 – Alignement d'information provenant de quatre caméras, chaque couleur représente un point de vue différent. (a) Nuages de points alignés (b) Nuages de points alignés et sphères identifiées (c) Sphères identifiées seules.....	30
Figure 2.6 – Nuage de points d'une personne vue selon quatre points de vue alignés. Chaque couleur représente un point de vue différent.....	31
Figure 2.7 – Ambiguïté lors du raffinement de l'alignement par ICP. Chaque couleur représente un point de vue différent.(a) Nuage de points aligné représentant une sphère avant l'application de l'algorithme ICP (b) Nuage de points incorrectement corrigé par l'algorithme ICP. ....	32
Figure 2.8 – Marqueurs fixés sur l'effecteur du mécanisme à câbles permettant l'alignement avec le système de vision, identifiés par M1, M2 et M3.....	34
Figure 2.9 – Marqueurs sphériques rétro réfléchissants fixés sur un objet vu par un capteur Kinect. (a) Marqueurs vus par la caméra RGB (b) Marqueurs vus par la caméra IR.....	35
Figure 2.10 – Identification du point de croisement entre deux lignes dans l'espace $\mathbb{R}^3$ . Le point $c$ représentant le centre du marqueur sphérique $\mathbf{M}$ est le point milieu du segment délimité par $b_0$ et $a_0$ .....	37
Figure 2.11 – Vue de dessus de l'alignement de l'orientation horizontale du casque de réalité virtuelle.....	38
Figure 2.12 – Visualisation du bénéfice de la redondance de l'information selon différents points de vue.....	40
Figure 2.13 – Courbe parabolique du poids selon un critère de qualité.....	43
Figure 2.14 – Arborescence représentant l'ordre de la fusion de l'information basée sur l'estimation de la position selon les différents points de vue et la différence entre la longueur de l'os connue et la distance mesurée au joint précédent.....	45
Figure 2.15 – Représentation de la sphère créée par la mesure du joint précédent $j$ et de la longueur de l'os connu $l$ . La position du joint suivant devrait se retrouver près de la surface de cette sphère.....	46
Figure 2.16 – Calcul de probabilité selon la fonction de vraisemblance. (a) Distance $d_j$ d'une particule $x^i$ à la surface de la sphère créée par la longueur de l'os $l$ selon le centre donné par la position du joint précédent $j$ . (b) Distance pondérée $d_{jk}$ d'une particule $x^i$	

aux estimations des capteurs $y_k$ . L'indice $k$ représente chacun des capteurs.....	50
Figure 2.17 – Reconstruction numérique fidèle à la réalité d'une salle de laboratoire de réadaptation. ....	53
Figure 2.18 – Représentation simplifiée (sans scène) de l'interaction du sujet en environnement virtuel avec l'effecteur (caisse de lait). L'image est présentée selon le point de vue du sujet portant le casque de réalité virtuelle. ....	54
Figure 2.19 – Affichage de l'avatar du sujet effectuant une tâche de manutention. (a) Affichage du squelette de l'avatar sur le moniteur du PC de contrôle et d'analyse de l'expérience. (b) Affichage de l'avatar dans le casque de réalité virtuelle porté par le sujet effectuant une tâche. ....	55
Figure 3.1 – Vue de dessus du schéma du montage expérimental utilisé lors de la prise des données.....	58
Figure 3.2 – Vue de côté du schéma du montage expérimental pour un capteur Kinect utilisé lors de la prise des données. ....	59
Figure 3.3 – Capteurs utilisés pour l'acquisition de données. (a) Caméra Vicon Bonita [58]. (b) Capteur Kinect V2 de Microsoft [59]. ....	60
Figure 3.4 – Objet utilisé pour la calibration entre le système de vision développé et le système de suivi de marqueurs Vicon. ....	61
Figure 3.5 – Effet de l'interpolation et du ré-échantillonnage sur une courbe de mesures dans le temps. Position d'un joint avant (a) et après (b) ré-échantillonnage. Notez que cette opération est effectuée avant l'alignement spatial, l'alignement temporel et la correction du biais de modèle. Mesures de position sur l'axe X global du joint de la hanche droite pour l'exercice manutention. ....	65
Figure 3.6 – Effet de l'alignement global sur une courbe de mesures dans le temps. Position d'un joint avant (a) et après (b) alignement global. Notez que cette opération est effectuée avant l'alignement temporel et la correction du biais de modèle. Mesures de position du joint de la hanche droite sur l'axe global X pour l'exercice manutention. .	66
Figure 3.7 – Effet de l'alignement temporel sur une courbe de mesures dans le temps. Position d'un joint avant (a) et après (b) alignement temporel. Notez qu'au début de l'exercice, le sujet devait lever et redescendre trois fois la jambe. Cette procédure n'est pas nécessaire à l'alignement mais permet de bien visualiser le résultat. Mesures de position du joint du genou droit sur l'axe global Y pour l'exercice pose fixe. ....	68
Figure 3.8 – Visualisation des deux modèles utilisés. En bleu, le modèle du squelette	

après fusion des données par notre système de vision. En orange, le modèle (bas du corps) de validation par le système Vicon. .... 69

Figure 3.9 – Transformations entre les repères global  $\mathbf{M}$  et local  $\mathbf{R}\mathbf{t}_j$ . (a) Transformation  $\mathbf{T}_j$  permettant de calculer le joint fusionné  $\mathbf{f}_j$  dans le repère du joint de validation  $\mathbf{t}_j$ . (b) Agrandissement du segment et représentation du biais par le vecteur  $\mathbf{d}_j$  selon la position  $\mathbf{t}_j$  du joint de validation et l'orientation  $\mathbf{R}_s$  du segment de validation associé. 71

Figure 3.10 – Effet de l'alignement local d'un segment sur une courbe de mesures dans le temps. Position d'un joint avant (a) et après (b) alignement local. Le sujet devait lever et redescendre trois fois la jambe en début d'exercice. Mesures de position du joint de la cheville droite sur l'axe global Z pour l'exercice pose fixe. .... 73

Figure 3.11 – Mesure de l'angle des jambes gauche (a) et droite (b). Les courbes orange et bleu sont respectivement les mesures observées par le système de validation et le système de vision proposé. Les autres courbes sont les mesures d'angles observés par les capteurs Kinects disposés autour du sujet. Mesures d'angle des jambes pour l'exercice marche statique. .... 76

Figure 3.12 – Mesure de la position d'un joint dans le temps. Les courbes orange et bleu sont respectivement les mesures observées par le système de validation et le système de vision proposé. Les autres courbes sont les mesures observées par les capteurs Kinects disposés autour du sujet. Mesures de position du joint de la cheville gauche sur l'axe global X pour l'exercice manutention. .... 77

Figure 3.13 – Variation d'angle d'un segment dans le temps. Mesures d'angle de la jambe droite pour l'exercice marche statique (a) et du dos par rapport au plan du sol pour l'exercice squat (b). .... 80

Figure 3.14 – Suivi d'un joint en position dans le temps au cours d'un exercice. Mesures de position du joint du genou gauche sur l'axe global X, Y, et Z (dans le même ordre (a), (b) et (c)) pour l'exercice marche statique. .... 84

Figure 3.15 – Suivi de l'angle des jambes dans le temps. Les deux graphes représentent respectivement la jambe gauche (a) et la jambe droite (b) observées selon les deux systèmes. Mesures d'angles des jambes pour l'exercice marche statique. .... 86

Figure 3.16 – Suivi de l'angle du dos dans le temps au cours d'un exercice. Les graphes représentent l'angle du dos observé individuellement par les deux systèmes selon la même échelle que la [Figure 3.15](#) (a) et selon une échelle étirée (b). Mesures d'angles du dos par rapport au sol pour l'exercice marche statique. .... 87

# Liste des tableaux

Table 1.1 – Spécifications techniques de la Kinect v1. Informations provenant de [3], [4], [5].	7
Table 1.2 – Spécifications techniques de la Kinect v2. Informations provenant de [2].	8
Table 1.3 – Performance de systèmes d'estimation de pose. Informations provenant de [18] Table 1.	15
Table 2.1 – Spécifications techniques du casque de réalité virtuelle Oculus Rift DK2. Informations provenant de [41].	51
Table 3.1 – Abréviation des noms de joints et segments.	75
Table 3.2 – Erreurs angulaires moyennes sur l'exercice <i>Marche</i> (degrés)	78
Table 3.3 – Erreurs angulaires moyennes sur l'exercice <i>Marche statique</i> (degrés)	78
Table 3.4 – Erreur angulaire sur l'ensemble des exercices (degrés)	78
Table 3.5 – Corrélation des angles mesurés ( $r$ )	79
Table 3.6 – Fiabilité en position pour l'exercice <i>Marche statique</i> selon un seuil de 10 cm	81
Table 3.7 – Fiabilité en position pour l'exercice <i>Manutention</i> selon un seuil de 10 cm	81
Table 3.8 – Fiabilité en position pour l'exercice <i>Squat</i> selon un seuil de 10 cm	82
Table 3.9 – Fiabilité en position pour l'exercice <i>Marche</i> selon un seuil de 10 cm	82
Table 3.10 – Fiabilité en mesure d'angle du dos selon différents seuils	82
Table 3.11 – Distance euclidienne entre les joints observés par les deux systèmes (cm)	83
Table 3.12 – Erreur angulaire jambes (degrés)	85
Table 3.13 – Erreur angulaire dos (degrés)	86

# Remerciements

Je tiens à sincèrement remercier mon directeur de maîtrise, Denis Laurendeau qui m'a toujours soutenu et dirigé dans mes recherches tout en me laissant la liberté d'explorer et d'expérimenter à ma guise. En plus d'être une source inestimable de connaissances et de bons conseils, il représente un mentor pour moi dans le domaine de la vision artificielle et de la recherche scientifique.

Je veux aussi remercier le personnel de recherche du Laboratoire de Vision et Système Numérique pour toute l'expertise, la collaboration, les bons conseils et les discussions intéressantes s'éternisant parfois. Je remercie chaleureusement mes collègues et amis du LVSN, du CIRRIS, d'UMANX et de l'Université Laval qui m'ont permis de me surpasser, mais surtout d'apprécier les innombrables heures passées à travailler, discuter et résoudre des problèmes en leur compagnie. J'espère que certains se reconnaîtront personnellement dans ces remerciements.

Enfin, j'aimerais remercier ma famille, mes proches et mes amis, mais surtout les personnes se retrouvant dans ces trois catégories à la fois. Une mention spéciale à mon frère Francis Bonenfant pour son appui illimité. Vous m'avez permis de passer au travers de plusieurs épreuves et de moments difficiles.

# Introduction

Le développement des capteurs 3D abordables tel que la Kinect de Microsoft a encouragé et facilité la recherche et le développement en vision artificielle tout en apportant de nouveaux défis. Leur accessibilité et leur versatilité ont permis l'élaboration de plusieurs systèmes de vision 3D qui, sans cette nouvelle technologie, n'aurait pas été envisageable avec un budget limité. À l'instar de sa popularité dans le monde du divertissement, ce capteur est très intéressant dans le domaine de la recherche en vision artificielle car, bien qu'il ne rivalise pas avec les caméras et autres systèmes coûteux de plus haute qualité, il offre un très bon rendement à condition d'exploiter les algorithmes de traitement des données 3D appropriés.

Ces capteurs de profondeur permettant l'enregistrement d'information physique sans interférence avec la scène observée sont très intéressants lors de l'analyse biomécanique du mouvement humain. La capture de mouvement en trois dimensions est souvent associée aux systèmes de suivi de marqueurs et autres capteurs tel que les accéléromètres. Bien que ce type de capture de mouvement basé sur des systèmes de suivi de marqueurs soit très répandu et fiable, ces systèmes présentent différents défauts. Un de ces problèmes est l'exigence du placement des marqueurs sur le sujet, ce qui entraîne une augmentation du temps de préparation, un risque d'interférence lors de certains mouvements ainsi qu'un comportement possiblement différent dans les mouvements effectués.

Différents algorithmes de capture de mouvements basés sur l'information 3D sans l'utilisation de marqueurs ont été proposés au cours des dernières années. Ces solutions présentent généralement des avantages et des inconvénients selon différents facteurs tels que la scène observée, le matériel utilisé et l'objectif visé par l'application. Microsoft propose un algorithme d'estimation de pose humaine exploitant l'information de distance d'une seule caméra et fonctionnant en temps réel. Le résultat, bien que loin d'être parfait, est exploitable dans certaines circonstances pour l'analyse biomécanique du mouvement humain.

L'objectif du projet présenté dans ce mémoire est d'utiliser plusieurs de ces capteurs bon marché et de les combiner pour augmenter la fiabilité et la précision de la capture du mouvement de patients en réadaptation évoluant dans un espace défini et restreint. Ces mesures seront utilisées dans l'analyse biomécanique de sujets présentant des maux de dos.

L'approche choisie dans nos recherches est d'estimer la pose d'un sujet à l'aide d'un algorithme de capture de mouvement sous différents points de vue observés chacun par une caméra Kinect et de fusionner l'information de chaque point de vue plutôt que d'estimer une seule pose selon l'information 3D de tout les points de vue combinés. En se servant de la redondance de l'information, cette stratégie permet d'éviter plusieurs mauvaises estimations de position de joints du modèle du sujet dont on fait le suivi. La procédure offre une robustesse à l'occlusion et un espace observé modulable selon les besoins. Il est aussi avantageux de pré-traiter individuellement l'information provenant de chaque capteur Kinect et de ne transférer que ce qui est nécessaire à la fusion. Le défi avec l'approche présentée plus loin est d'élaborer une méthode de fusion d'information 3D en temps réel qui permette l'amélioration de la qualité de la capture de mouvement, c'est-à-dire d'atteindre une bonne précision de mesure de la pose pour notre application de réadaptation tout en étant robuste aux occlusions et aux déplacements entre les différents points de vue des caméras couvrant l'espace de travail.

Ce projet est issu d'une collaboration multidisciplinaire entre les domaines de la vision artificielle, de la robotique et de la biomécanique. En effet, l'analyse du mouvement est effectuée dans un contexte de réadaptation en environnement virtuel comportant certaines interactions physiques, effectuées sur un robot à câbles, permettant un retour haptique et visuel chez le patient pendant une expérience. Voir la [Figure 1](#) pour un aperçu du montage et la [Figure 2.1](#) qui présente plus en détail les différentes composantes du système.

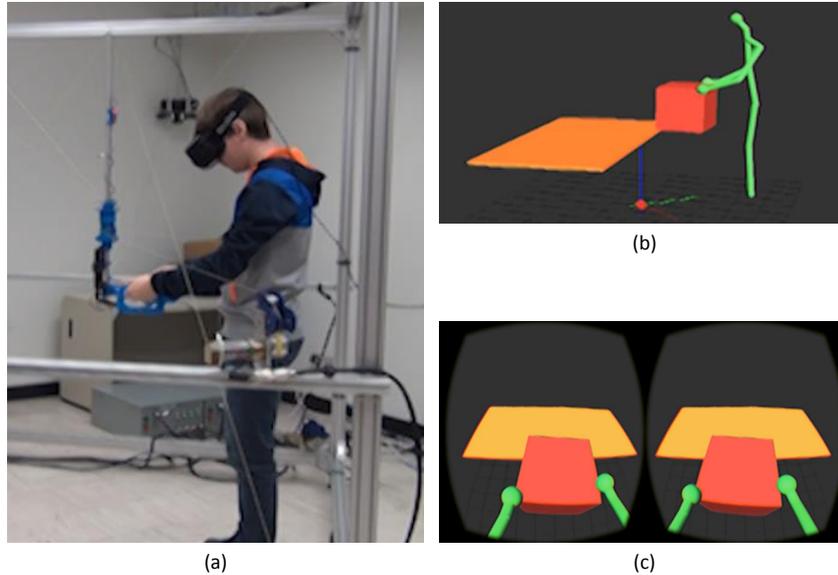


Figure 1 – Utilisation du montage avec environnement virtuel simplifié. (a) Utilisateur immergé en réalité virtuelle manipulant le mécanisme à câbles (b) Représentation simplifiée de la scène (table, boîte et avatar) selon le point de vue d'une caméra virtuelle (c) Vue selon la perspective de l'utilisateur portant le casque de réalité virtuelle.

Le mécanisme à câbles, conçu par un autre membre de l'équipe, sert à simuler un objet, une boîte que le sujet soulève et dépose sur une étagère de l'espace de travail virtuel. Grâce à un dispositif de visualisation en réalité virtuelle (Oculus Rift 2), le sujet perçoit une scène réelle. Le système de suivi du mouvement développé dans ce travail fournit des informations en temps réel sur la pose du sujet. Ces informations sont analysées par l'équipe de spécialistes en biomécanique pour évaluer le sujet souffrant de maux de bas de dos.

Le mémoire est présenté comme suit. Le Chapitre 1 couvre la revue de littérature sur les systèmes et méthodes reliés à la solution proposée. Dans le Chapitre 2 sera présentée la méthodologie utilisée afin de résoudre le problème de capture de mouvement multivues en temps réel pour un sujet évoluant dans l'espace de travail du mécanisme à câbles. Le Chapitre 3 présente et discute les résultats expérimentaux et évalue la performance du système de capture de mouvement puis la conclusion et le suivi du projet seront présentés à la suite. Les références et annexes compléteront le document.

# Chapitre 1

## Revue de littérature

Ce chapitre décrit les outils, le matériel, les méthodes et les algorithmes reliés au projet. Dans l'ordre seront présentés les capteurs 3D, la calibration, l'estimation de pose humaine puis la fusion de données.

### 1.1 Capteurs 3D

Plusieurs techniques sont utilisées pour faire l'acquisition de l'information 3D, que ce soit pour la mesure et la reconstruction d'objet, l'interprétation d'une image 2D ou l'analyse de la démarche d'une personne en réadaptation. Le choix de la méthode de capture de données brutes doit être adapté au traitement spécifique, l'inverse est aussi vrai, afin d'obtenir le résultat optimal. Il est possible de diviser ces méthodes en deux grandes classes, passives et actives. Ces deux classes sont elles-mêmes divisées une deuxième fois en deux autres classes, point de vue unique et points de vue multiples ([Figure 1.1](#)).

Contrairement aux techniques passives qui ne se basent que sur l'éclairage ambiant, les techniques actives profitent d'une illumination contrôlée, temporellement ou spatialement, de manière à simplifier le processus de capture de l'information. Les systèmes multivues sont constitués de plusieurs capteurs et/ou illuminateurs qui permettent de mettre à profit la triangulation de l'information observée en assumant que la pose des composantes du système est connue [\[1\]](#).

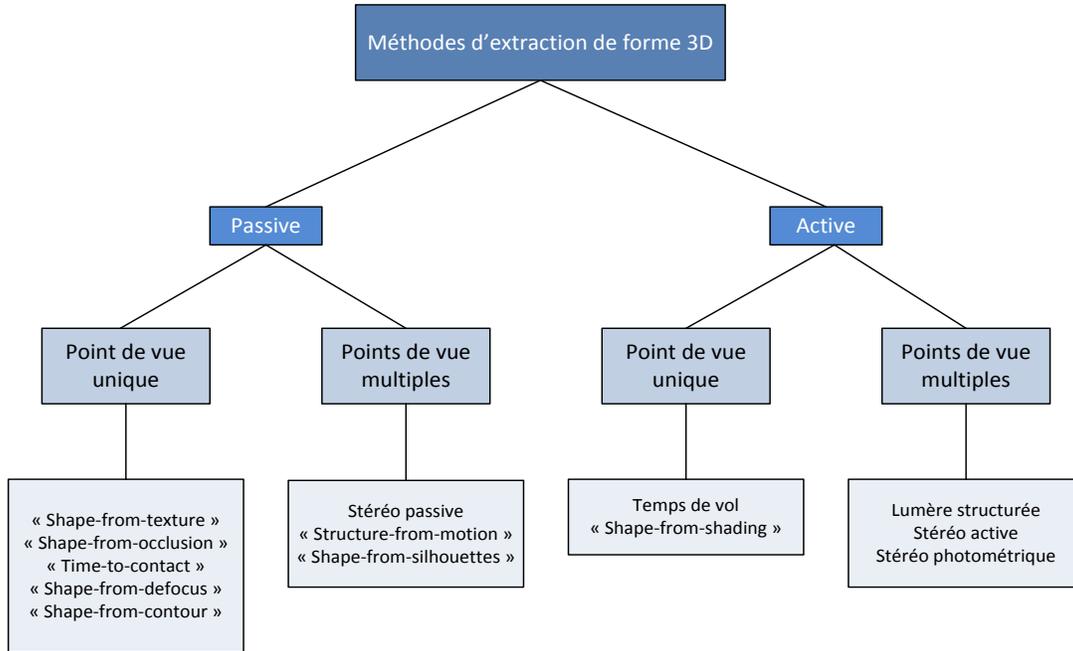


Figure 1.1 – Classification des méthodes d'extraction d'information de forme en 3 dimensions. Informations provenant de [1] Fig. 1.1.

### 1.1.1 Caméra temps de vol

La caméra de type temps de vol, mieux connue sous le nom de "Time of Flight" (ToF), est classée parmi les capteurs actifs à simple point de vue, car, bien que ce capteur comprenne une caméra et un illuminateur, l'objectif n'est pas de tirer avantage de la position de son illuminateur comme dans le cas de la triangulation avec laser [1], [6]. Cette méthode est basée sur la mesure du temps de déplacement de la lumière entre la scène observée et le capteur. Comme la vitesse de la lumière est une constante connue, la distance est directement proportionnelle au temps que prend la lumière pour quitter le projecteur et retourner vers la caméra.

Deux types de technologies existent dans ce domaine, le premier type de capteur mesure directement le temps entre l'envoi et le retour d'une lumière pulsée, généralement une lumière infrarouge. Le deuxième type étant habituellement moins coûteux, mais aussi moins précis sur de grandes distances, mesure le déphasage de l'onde sinusoïdale transmise vers la scène [1]. Ce type de caméra peut fournir une image de distance en temps réel et est une solution compacte, car elle ne nécessite pas de points de vue multiples ni même une séparation ("Baseline") entre

les composantes, comme pour la stéréoscopie.

### 1.1.2 Kinects

La Kinect de Microsoft est un dispositif rassemblant différents capteurs dont plusieurs microphones, une caméra couleur et une caméra de profondeur. Deux versions de ces caméras ont été produites, la première exploite une méthode de mesure de distance basée sur la technologie de lumière structurée et la seconde bénéficie d'un capteur ToF. Bien que la Kinect soit dédiée en premier lieu au marché du divertissement, il est possible d'interfacer ses différents capteurs (caméra RGB, caméra de distance, microphones) directement et donc de l'utiliser pour des projets de recherche et d'ingénierie en adaptant et en améliorant ses capacités. Son arrivée a provoqué un impact important dans le domaine de la recherche en vision artificielle, médicale et robotique [3], [7].

#### Kinect v1

La première génération de Kinects incorpore un projecteur de lumière structurée, une caméra infrarouge, une caméra couleur ainsi qu'une série de quatre microphones. Afin de permettre le plein contrôle de ces capteurs, Microsoft et autres groupes (OpenNI / OpenKinect) offrent un "Software Development Kit" (SDK) permettant de prendre le contrôle de ses capteurs et offrant différents algorithmes tels que l'estimation de pose d'une personne ou la détection de visages et d'expressions faciales [7], [8]. La [Table 1.1](#) ci-dessous résume les caractéristiques de la Kinect v1.

Table 1.1 – Spécifications techniques de la Kinect v1. Informations provenant de [3], [4], [5].

	Communication	USB
Caméra Distance	Type	Lumière structuré
	Résolution	640 x 480
	Fréquence	30 Hz
	Champ de vision horizontal	57 degrés
	Champ de vision vertical	43 degrés
	Portée effective	0.8 - 3.5m
Caméra Couleur	Type	RGB
	Résolution	1280 x 960
	Fréquence	30 Hz
	Champ de vision horizontal	62 degrés
	Champ de vision vertical	48.6 degrés

## Kinect v2

La deuxième génération de Kinects comprend pratiquement tous les avantages de la version précédente, mais utilise la technologie de capteur ToF décrite précédemment. Ce changement de méthode d'acquisition de l'image de distance distingue grandement la qualité de l'information en comparaison à la version précédente. En effet, sa résolution d'image est supérieure et sa précision de mesure en distance est doublée à courte portée et multipliée par dix à longue portée (6m+). La Kinect v2 est aussi plus robuste à la lumière ambiante (artificielle ou naturelle) et offre des performances améliorées en reconstruction 3D ainsi qu'en estimation de pose [8]. La [Table 1.2](#) présente les caractéristiques de la Kinect v2.

Table 1.2 – Spécifications techniques de la Kinect v2. Informations provenant de [2].

	Dimensions	249 x 66 x 67 mm
	Communication	USB 3.0
Caméra Distance	Type	Temps de vol
	Résolution	512 x 424
	Fréquence	30 Hz
	Champ de vision horizontal	70.6 degrés
	Champ de vision vertical	60.0 degrés
	Portée effective	0.5 – 4.5m
Caméra Couleur	Type	RGB
	Résolution	1920 x 1080
	Fréquence	30 Hz
	Champ de vision horizontal	84.1 degrés
	Champ de vision vertical	53.8 degrés

### 1.1.3 Vicon

Le système de capture de mouvement de Vicon permet un suivi de marqueurs passif en 3D à haute fréquence et à haute résolution. Vicon offre différentes caméras et logiciels, mais le principe de suivi reste généralement très similaire, plusieurs caméras infrarouges détectent et suivent les marqueurs sphériques placés dans la scène, sur des objets et/ou personnes, puis l'information spatiale de chaque marqueur est obtenue par la triangulation de ces observations. Cette technologie de capture de mouvement est utilisée dans les domaines du divertissement, de l'ingénierie et des sciences biomédicales pour des applications comme l'analyse biométrique de la posture, la réalité virtuelle et augmentée, la mesure de pose d'un robot ou d'une pièce mécanique, pour n'en nommer que quelques-une [9]. Voir la [Figure 1.2](#) pour un aperçu du montage du système Vicon.

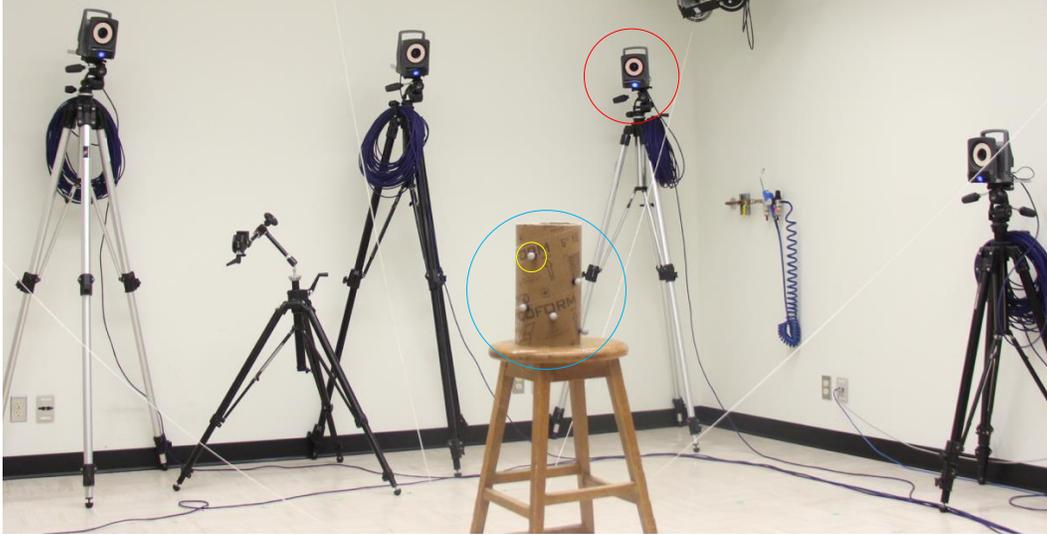


Figure 1.2 – Caméra Vicon observant des marqueurs sphériques rétro-réfléchissants fixés sur un objet. En rouge: une des caméras Vicon. En bleu: l'objet suivi. En jaune: un des marqueurs.

## 1.2 Calibration

Dans tout système de vision artificielle, la calibration des caméras est une étape nécessaire. Il est important de s'assurer de la qualité de la calibration, car l'ensemble des mesures dépend de celle-ci. Sans entrer dans les détails des différentes méthodes et étapes de calibration, [10], [11], [12], [13] la solution recherchée est une matrice de projection  $\mathbf{P}$  permettant de passer de l'espace monde  $\mathbf{R}^3$  vers l'espace image  $\mathbf{R}^2$ . Elle est composée de deux éléments principaux, la matrice de calibration intrinsèque  $\mathbf{K}$  et la matrice de calibration extrinsèque  $[\mathbf{R}|\mathbf{t}]$ .

$$\mathbf{P} = \mathbf{K}[\mathbf{R}|\mathbf{t}] \quad (1.1)$$

### 1.2.1 Calibration intrinsèque

Cette première étape permet de retrouver les paramètres réels de la caméra ainsi que de corriger la distorsion de la lentille afin d'obtenir une bonne correspondance entre la scène observée et l'image acquise. Les paramètres contenus dans cette matrice de calibration  $\mathbf{K}$  de taille  $3 \times 3$  sont la distance focale  $f_x, f_y$ , le centre optique  $x_0, y_0$  et la déviation oblique des axes  $s$ .

$$\mathbf{K} = \begin{bmatrix} f_x & s & x_0 \\ 0 & f_y & y_0 \\ 0 & 0 & 1 \end{bmatrix} \quad (1.2)$$

Bien que le modèle théorique de caméra sténopé ne présente logiquement qu'une seule distance focale, ici  $f_x$  et  $f_y$  incluent un facteur d'agrandissement modélisant de possibles défauts du capteur. Cette matrice est une approximation de la projection réelle entre le monde 3D et l'image 2D. Par contre, il est possible d'obtenir une meilleure précision en corrigeant certaines erreurs causées par la distorsion des lentilles de la caméra. L'une des méthodes les plus courantes pour corriger la distorsion radiale et tangentielle est détaillée dans [\[10\]](#).

### 1.2.2 Calibration extrinsèque

Lorsque les paramètres intrinsèques de la caméra connus, il est simple de passer de l'espace 3D (monde) à l'espace 2D (image) en appliquant une multiplication matricielle sur l'information à projeter en se basant sur le modèle sténopé. Par contre, il est parfois nécessaire de connaître la position d'un point par rapport à un autre référentiel que celui de la caméra. La matrice  $[\mathbf{R} \ \mathbf{t}]$  représente la transformation rigide effectuée pour passer du système de coordonnées de la caméra vers un nouveau système de coordonnées. Ce dernier peut être une autre caméra, un objet dans la scène ou bien une référence considérée comme pose initiale absolue. Les paramètres contenus dans cette matrice sont la rotation, représentée par la matrice de rotation 3x3  $\mathbf{R}$  et la translation, représentée par le vecteur de translation  $\mathbf{t}$ .

$$[\mathbf{R}|\mathbf{t}] = \begin{bmatrix} r_{11} & r_{12} & r_{13} & tx \\ r_{21} & r_{22} & r_{23} & ty \\ r_{31} & r_{32} & r_{33} & tz \end{bmatrix} \quad (1.3)$$

Pour adapter les calculs à la géométrie projective, il est possible d'utiliser les coordonnées homogènes, cette transformation permet d'appliquer la rotation et la translation en une seule multiplication matricielle. La matrice 3x4  $[\mathbf{R} \ | \ \mathbf{t}]$  est maintenant représentée par la matrice 4x4  $\mathbf{T}$ . Cette matrice de transformation rigide sera aussi appelée matrice de calibration extrinsèque.

$$\mathbf{T} = \begin{bmatrix} r_{11} & r_{12} & r_{13} & t_x \\ r_{21} & r_{22} & r_{23} & t_y \\ r_{31} & r_{32} & r_{33} & t_z \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad (1.4)$$

Pour simplifier le calcul de la matrice de projection, il est aussi possible de modifier la matrice de calibration intrinsèque en ajoutant une colonne vide afin d'obtenir une matrice 3x4 qu'il est possible d'utiliser dans le calcul en coordonnées homogènes. Connaissant la matrice de calibration intrinsèque  $\mathbf{K}$  et extrinsèque  $\mathbf{T}$ , un point 3D  $\mathbf{p}_m$  représenté dans un référentiel quelconque peut être projeté sur le plan image 2D afin d'obtenir sa position dans l'image  $\mathbf{p}_i$ .

$$\mathbf{p}_i = \begin{bmatrix} u \\ v \\ w \end{bmatrix} = \mathbf{KT}\mathbf{p}_m = \begin{bmatrix} f_x & s & x_0 & 0 \\ 0 & f_y & y_0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} r_{11} & r_{12} & r_{13} & t_x \\ r_{21} & r_{22} & r_{23} & t_y \\ r_{31} & r_{32} & r_{33} & t_z \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x_m \\ y_m \\ z_m \\ 1 \end{bmatrix} \quad (1.5)$$

### 1.2.3 Alignement d'information 3D

Les étapes présentées dans les paragraphes précédents permettent de passer de différents systèmes de coordonnées, que ce soit d'un espace à trois dimensions vers un autre espace à trois dimensions ou d'une projection vers un espace de dimension inférieure. Sur la [Figure 1.3](#), la matrice de calibration intrinsèque permet de passer du repère image  $\mathbf{I}$  vers le repère caméra  $\mathbf{C}$  puis la matrice de calibration extrinsèque (Transformation  $\mathbf{T}$ ) permet de passer du repère  $\mathbf{C}$  vers le repère monde  $\mathbf{M}$ . L'ensemble de ces transformations permet la projection d'un point  $\mathbf{p}_i$  de l'image vers le point  $\mathbf{p}_m$  dans le repère monde.

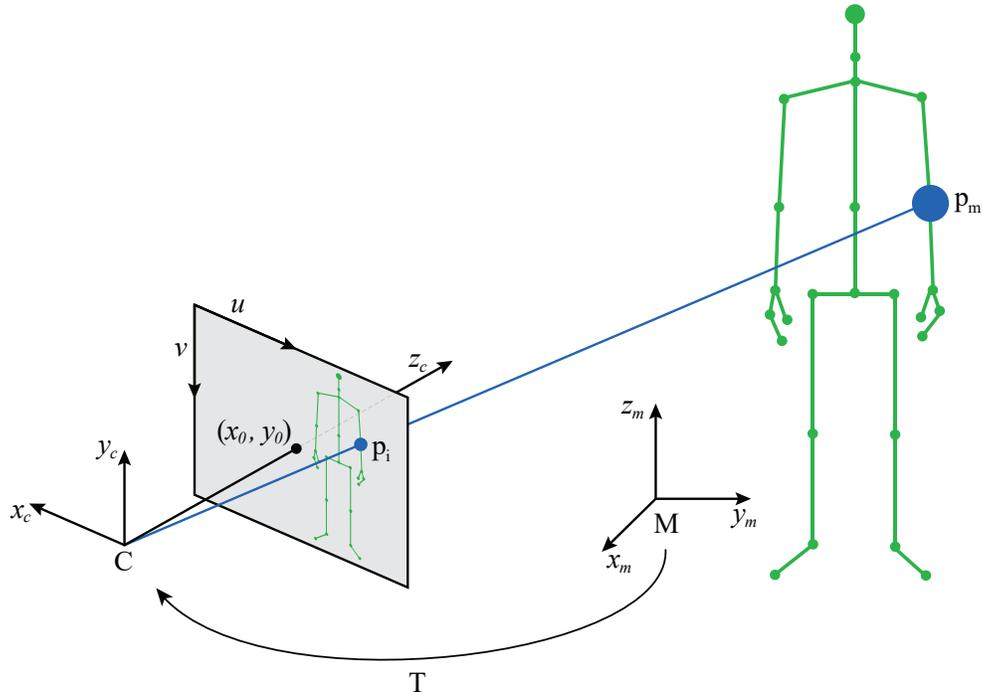


Figure 1.3 – Projection de perspective d'un point selon les transformations entre les repères image, caméra et monde.

Dans tout les cas, la matrice de calibration extrinsèque représentant la transformation entre deux poses est obtenue par l'alignement de données observées et de manière inverse permet d'aligner l'information d'un système de coordonnées vers un autre. Ce type d'alignement est très utilisé en vision artificielle afin de reconstruire un objet ou une scène à partir d'information obtenue de différents points de vue.

Pour trouver la transformation entre deux systèmes, il est nécessaire de connaître la correspondance entre au moins trois points, car il n'existe qu'une seule transformation rigide expliquant ce changement de repère. Par contre, cela n'est possible que pour un système théorique parfait. Un système de vision ne peut pas trouver la position d'un point de manière exacte et on observe une erreur causée par le bruit de mesure et la résolution limitée du capteur. Beaucoup d'autres facteurs peuvent aussi causer des erreurs, la distorsion de lentille discutée précédemment par exemple.

Plusieurs méthodes d'alignement d'information 3D ont été présentées [14] et nécessitent généralement un alignement initial qui sera raffiné par une procédure d'optimisation. L'estimation de cette transformation peut être obtenue facilement par l'utilisation de points

de référence connus, une cible de calibration par exemple. En assumant que nous avons deux ensembles de points correspondants  $\mathbf{A}=\{a_1, a_2, \dots, a_n\}$  et  $\mathbf{B}=\{b_1, b_2, \dots, b_n\}$ , chacun exprimé dans un système de coordonnées différent, il est possible d'estimer la transformation rigide optimale  $[\mathbf{R} \mid \mathbf{t}]$  entre ces ensembles. Une approche généralement utilisée est la minimisation de l'erreur par la méthode des moindres carrés [15], [16].

$$(\mathbf{R}, \mathbf{t}) = \operatorname{argmin} \sum_{i=1}^n \|(\mathbf{R}a_i + \mathbf{t}) - b_i\|^2 \quad (1.6)$$

Cette solution offre une estimation de matrice servant de condition initiale à plusieurs algorithmes d'alignement fin.

### Iterative Closest Point

Un de ces algorithmes classés comme alignement fin [14] qui est très utilisé en reconstruction 3D et en robotique est "Iterative Closest Point" (ICP) [17]. C'est une approche simple de type "Force brute" qui cherche itérativement un minimum local de l'erreur entre deux ensembles de points. Un alignement initial de bonne précision est donc un élément essentiel à la qualité du résultat final et à la convergence de l'algorithme. Voir l'Algorithme 1 pour plus d'information sur son implémentation.

---

#### Algorithme 1 Iterative Closest Points

---

Point set  $A$  and  $B$

**do**

**for each** points  $a_i$  in  $A$  **do**

    Find the closest point  $b_i \in B$

**end for each**

  Compute the rigid transformation  $T$  that minimizes  $\sum_{i=1}^N \|b_i - Ta_i\|^2$

  Apply the transformation  $A = T \cdot A$

**while** Convergence condition not satisfied

---

### 1.3 Estimation de pose humaine

L'estimation de la pose humaine est le fait de mesurer le plus fidèlement possible la position des joints et la configuration des articulations du squelette d'un sujet analysé. La mesure de la pose humaine pour la réadaptation est un sujet qui intéresse les chercheurs depuis longtemps [18], [19] et qui a progressé rapidement dans les dernières années avec les avancées en vision artificielle grâce aux algorithmes d'estimation de pose sans marqueurs et aux capteurs 3D [20]. C'est un outil important qui permet de travailler la mobilité de patients atteints de certaines maladies chroniques ou en rémission d'un accident vasculaire cérébral [18], par exemple. Plusieurs approches sont utilisées pour faire le suivi de joints avec et sans marqueurs. La Figure 1.4 présente un classement des technologies utilisées pour l'estimation de pose humaine.

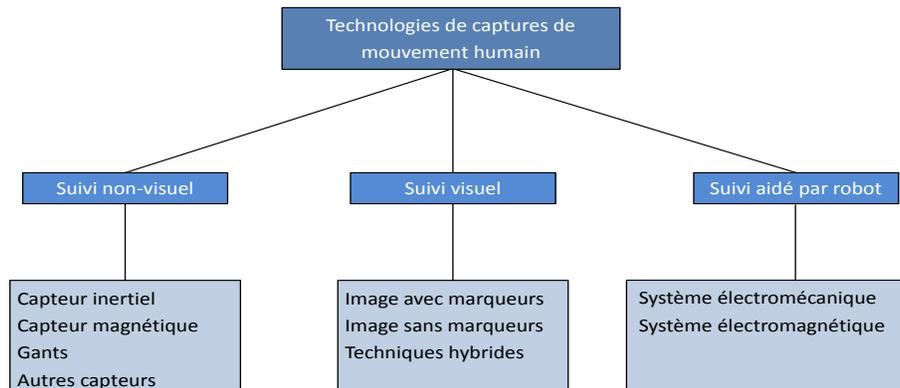


Figure 1.4 – Classification des technologies de capture de mouvement humain. Informations provenant de [18] Fig. 2.

Aucune de ces méthodes n'est parfaite, mais chacune présente des avantages et des inconvénients qu'il est possible d'adapter aux besoins de l'application visée. Dans le cas de notre projet, la minimisation des coûts est importante, mais son effet négatif sur la précision de la mesure doit être limité. Pour un résumé des caractéristiques de base de ces systèmes, voir la Table 1.3.

Table 1.3 – Performance de systèmes d'estimation de pose. Informations provenant de [18]

Table 1.

Système	Précision	Compacte	Calculs	Coût	Désavantages
Inertiel	Haute	Oui	Efficace	Bas	Dérive
Magnétique	Moyenne	Oui	Efficace	Bas	Matériaux ferromagnétiques
Ultrason	Moyenne	Non	Efficace	Bas	Occlusion
Gant	Haute	Oui	Efficace	Moyen	Posture partielle
Marqueurs	Haute	Non	Inefficace	Moyen	Occlusion
Sans marqueurs	Haute	Oui	Inefficace	Bas	Occlusion
Combinatoire	Haute	Non	Inefficace	Haut	Multidisciplinaire
Robot	Haute	Non	Inefficace	Haut	Mouvement limité

### 1.3.1 Estimation de pose avec marqueurs

Les systèmes d'estimation de pose humaine basés sur le suivi de marqueurs permettent de minimiser l'incertitude grâce au placement de marqueurs selon une disposition connue et idéale. Ces marqueurs peuvent être passifs, actifs ou hybrides, c'est-à-dire un marqueur qui ne génère pas de lumière, mais va réfléchir une lumière émise par un illuminateur [18]. Le Vicon est classé parmi ces systèmes hybrides car ses marqueurs sphériques reflètent la lumière infrarouge émise par un illuminateur placé sur les caméras. Une fois la position des marqueurs obtenue, il est simple de reconstruire la configuration des articulations du sujet observé (Figure 1.5). Le placement des marqueurs peut être une tâche complexe afin d'obtenir une pose biométriquement correcte et l'occlusion de certains marqueurs est pratiquement inévitable durant la prise de mesures. Par contre, certains algorithmes ont été développés pour améliorer la précision, la fiabilité et éliminer les erreurs de mouvement des marqueurs [21].

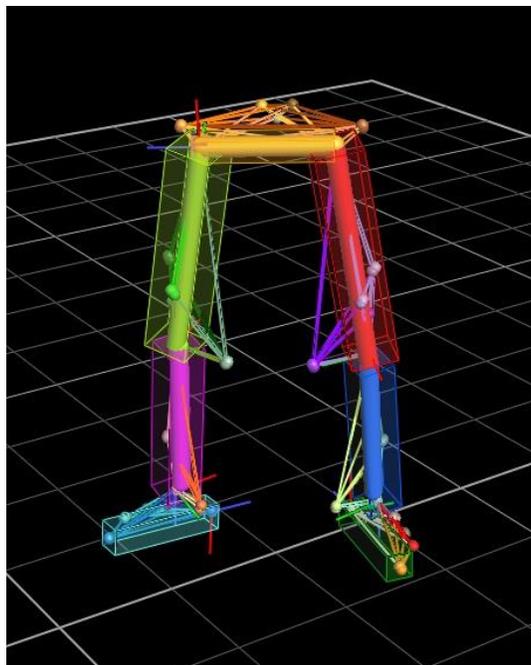


Figure 1.5 – Reconstruction des membres inférieurs d'un sujet suivi par marqueurs avec le logiciel Nexus de Vicon.

### 1.3.2 Estimation de pose sans marqueurs

La complexité de cette méthode est causée par la nécessité d'identifier la position des joints de la personne observée sur une image. Ces algorithmes peuvent être utilisés en temps réel ou en post-traitement. Les techniques dépendent du niveau de précision désiré, par exemple si le but est d'estimer seulement le centre de masse, la position de la tête et des mains ou encore la position et l'orientation de tous les segments principaux. L'algorithme dépend aussi généralement du capteur utilisé pour l'acquisition d'informations. Certaines approches tirent avantage d'un suivi sur plusieurs images dans le temps, d'autres se servent de vues multiples ou bien d'images 3D [18], [20].

#### Algorithme de Microsoft

La méthode développée par Microsoft et publiée en 2013 [20] offre une précision supérieure à l'état de l'art ainsi qu'une performance en temps réel avec une bonne corrélation sur des données réelles et synthétiques. L'algorithme, implémenté sur le SDK ("Software Development

Kit") de Microsoft, est l'un des plus utilisés et des plus documentés ces dernières années [22], [23], [24], [25]. La méthode utilise une approche reposant sur la reconnaissance d'objet afin de créer une représentation intermédiaire des segments du squelette d'une personne (Figure 1.6). Cela simplifie le problème de l'estimation de pose en un problème plus simple de classification par pixel. L'ensemble de données d'entraînement est large et varié, ce qui aide à l'invariance de l'approche à la pose du sujet, à la forme du corps et aux vêtements portés.

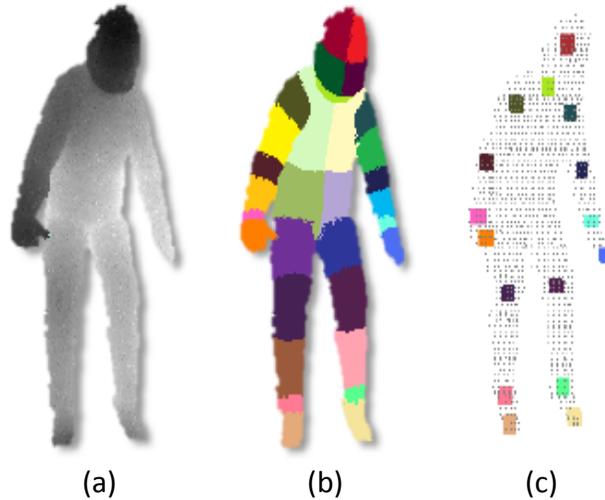


Figure 1.6 – Identification des joints sur une pose humaine à partir d'une image de profondeur selon la méthode [20]. (a) Information de l'image de profondeur (b) Déduction de la distribution des segments du corps par pixel. (c) Proposition de la position des joints déduits. Image provenant de [20] Figure 5.

Bien que la méthode fonctionne bien pour retrouver la position des joints d'une personne avec une précision acceptable dans les bonnes conditions, cela n'est pas le cas dans toutes les poses et tous environnements. En effet, comme la méthode est basée sur l'estimation de pose sur une image de profondeur, il est nécessaire que le sujet soit bien orienté par rapport à la caméra et à une distance acceptable selon la qualité des données du capteur. Un des problèmes observés [22] est l'auto-occlusion des joints dans plusieurs poses ainsi que l'occlusion des joints par des objets dans la scène. La variabilité de la précision est très dépendante de la pose du sujet et du point de vue. Nous verrons qu'il est possible de corriger certains de ces problèmes et d'augmenter la précision grâce à l'ajout de plusieurs points de vue et à un modèle anthropométrique simple du sujet.

## 1.4 Fusion de données

L'intégration de données multiples pouvant avoir une représentation différente, une fréquence d'acquisition différente et provenir de plusieurs capteurs plus ou moins précis exploite un processus de fusion de données. L'objectif est d'utiliser des données hétérogènes (en fréquence, précision ou représentation) et d'en extraire un résultat fiable. Augmenter le nombre de données crée une redondance qui permet d'éviter les occlusions et de réduire l'ambiguïté. Différentes méthodes de fusion de données [26] sont utilisées pour un large éventail d'applications dans des domaines variés comme la robotique, la vision artificielle, la finance ou les réseaux de capteurs.

### 1.4.1 Filtre particulaire

Les techniques de fusion de données basées sur la méthode de Monte-Carlo sont parmi les plus populaires et rendent de très bons résultats dans l'approximation de probabilités [30]. Le filtre particulaire [27], [45] est une implémentation récursive de la méthode de Monte-Carlo séquentielle (SMC) [28] permettant de résoudre le problème du filtrage de données. Ce filtre est souvent utilisé comme alternative au filtre de Kalman lorsque le bruit de mesure n'est pas Gaussien et le système non-linéaire.

Le principe de fonctionnement est basé sur un ensemble de particules échantillonnées qui représentent la fonction de densité de probabilité "Probability Density Function" (PDF), chacune de ces particules contient un ensemble des valeurs représentant l'état du système estimé ainsi qu'un poids. La particule possédant le poids le plus grand après une itération de l'algorithme représente la plus grande probabilité de l'état du système. Dans notre application de suivi d'une personne, le système est représenté par un joint (articulation) du squelette et son état est la position de celui-ci ainsi que la longueur du segment associé (os). Le poids de chaque particule est modifié lorsque l'on mesure l'état du système. Il est nécessaire de constamment ré-échantillonner de nouvelles particules et d'éliminer celles ayant un poids trop faible afin d'éviter la dégénérescence, une situation où seulement quelques particules ont un poids significatif [29]. Voir l'Algorithme 2 pour plus d'information sur l'implémentation du filtre particulaire de type "Bootstrap" [48], [49].

---

**Algorithm 2** Bootstrap particle filter

---

Initialization

**for**  $i = 1$ :Number of particles **do**

    Sample particle  $x_0^i$  from the initial distribution  $p(x_0)$  based of prior knowledge or set to zero.

**end for**

Filtering

**for**  $t = 1$ :Filtering iterations **do**

**for**  $i = 1$ :Number of particles **do**

        Predict state  $x_t^i$  from the transition function  $p(x_t|x_{t-1}^i)$

        Measure weights  $w_t^i$  from the likelihood function  $p(y_t|x_t^i)$

**end for**

    Normalise current weights  $w_t$

    Resample particles  $\{x_t^i, w_t\}$  according to the weights  $w_t$

**end for**

---

## Chapitre 2

# Méthodologie

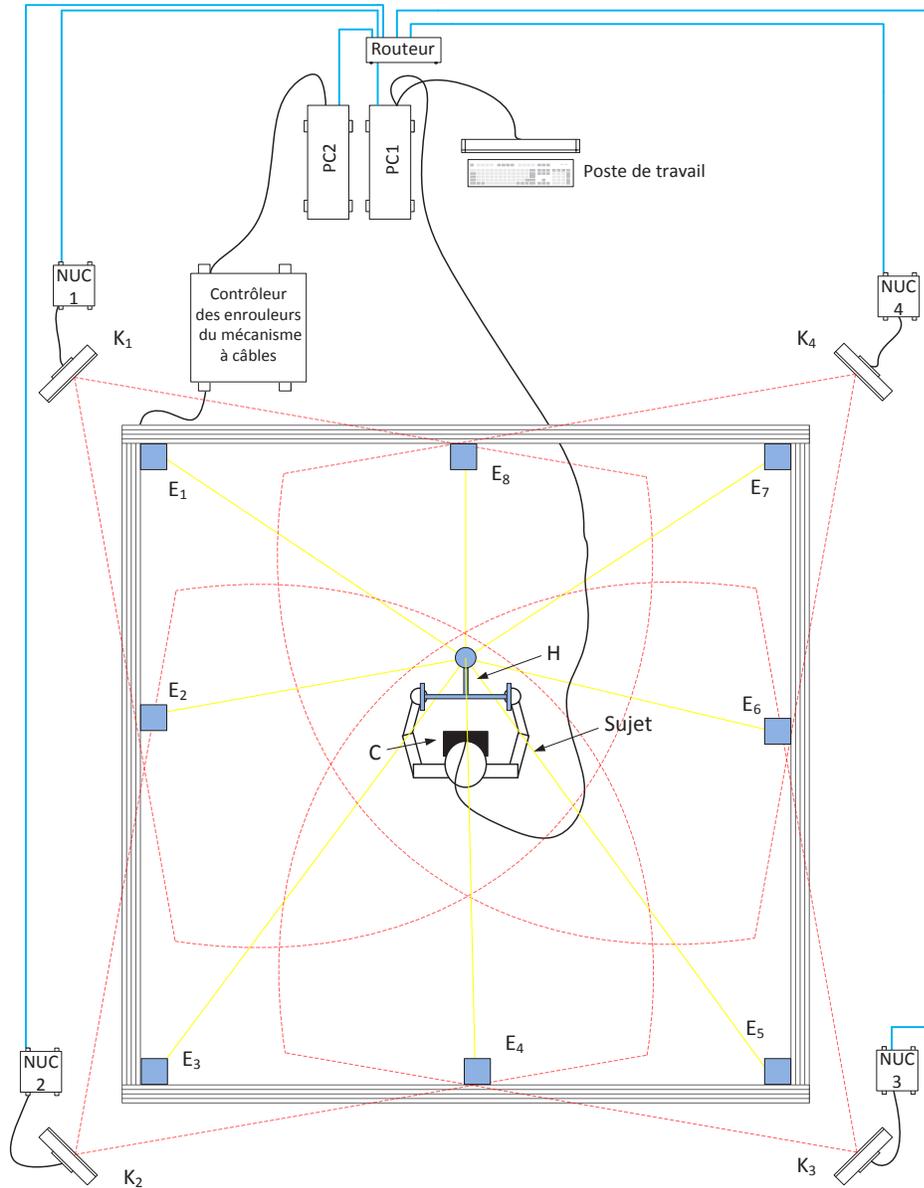
L'objectif du développement de ce projet de mesure de pose en environnement virtuel est de permettre une réadaptation plus fonctionnelle, adaptée et moins coûteuse.

L'utilisation d'un robot dans un contexte de réadaptation semble augmenter l'efficacité des techniques conventionnelles [31]. Cette approche permet d'adapter de manière optimale la tâche au patient en offrant un environnement et un retour haptique configurable. Avec l'utilisation du robot parallèle à câbles, la conception d'un environnement virtuel dans lequel le patient est invité à évoluer devient un point essentiel permettant l'immersion dans l'espace de travail choisi. Un autre point est essentiel pour compléter le système, soit le suivi en temps réel de la pose du sujet. Cette mesure de la pose est utilisée à différentes étapes du processus. Premièrement, elle permet de connaître la position de la tête du sujet dans l'espace de travail afin que l'image projetée dans le casque de réalité virtuelle via lequel l'environnement est affiché corresponde au point de vue de la personne. Il est aussi nécessaire d'afficher la pose des membres de la personne en réalité virtuelle afin qu'elle puisse naturellement interagir avec l'environnement. Finalement, l'information de pose des membres et de position des articulations pourra être analysée biométriquement afin de caractériser la performance motrice du patient. À partir de maintenant, nous utiliserons les termes "Squelette" pour décrire l'information associée à la pose des segments du sujet et "Avatar" pour décrire le squelette affiché dans l'environnement virtuel.

Comme le système doit être abordable pour le domaine de la réadaptation, le choix d'un robot parallèle à entraînement par câbles, de capteurs de profondeur de type Kinects ainsi que du casque de réalité virtuelle Oculus est approprié. Bien que ces systèmes soient peu coûteux, ils présentent certains défis dont les solutions sont détaillées dans ce chapitre.

## 2.1 Description du système proposé

Le système développé comprend plusieurs composants d'acquisition, de traitement et d'affichage d'information ([Figure 2.1](#)). L'utilisation simultanée de plusieurs capteurs Kinects et d'un casque de réalité virtuelle Oculus RIFT 2 ajoute certaines complications. Chaque capteur Kinect doit être connecté à un Hub USB 3.0 indépendant et il est aussi recommandé d'utiliser un Hub USB 3.0 pour connecter le casque VR afin d'optimiser la performance d'affichage. Afin de ne pas limiter le système selon le nombre de caméras utilisées, une approche modulaire a été choisie. Cette approche permet d'éviter certaines limitations logicielles et matérielles. Par exemple, le système d'exploitation Windows qui n'accepte qu'une seule Kinect par PC ou le problème de la quantité de données à traiter qui augmente avec le nombre de caméras. Dans le cas du présent système, 4 capteurs Kinects sont utilisés. Une solution simple et élégante a été retenue, celle d'utiliser un ordinateur compact "Intel Next Unit of Computing" (NUC) associé à chaque capteur Kinect.



$K_i$  : Capteurs Kinects  
 $E_i$  : Enrouleurs  
 C : Casque Oculus RIFT2  
 H : Effecteur du mécanisme

Figure 2.1 – Schéma du système de capture de mouvement en réalité virtuelle avec retour haptique développé.

Ces ordinateurs NUC serviront à l'acquisition, au prétraitement et au transfert des données des capteurs Kinects. Cette solution augmente légèrement les coûts associés au projet, mais permet de modifier facilement le nombre de caméras utilisées afin de couvrir l'espace de travail. Cette compartimentation en sous-systèmes (NUC et capteur) est aussi

mise à profit par la redistribution des calculs. L'image de profondeur captée par une Kinect est transmise de chaque caméra par protocole USB vers chaque NUC, puis traitée selon les besoins (extraction du squelette, filtrage du nuage de points ou alignement) puis les données sont ensuite envoyées à l'ordinateur principal par l'intermédiaire d'un logiciel serveur se servant du protocole Ethernet.

Ces données sont reçues par un ordinateur central (PC) qui servira à la fusion des informations provenant de chaque unité NUC, à la gestion de l'environnement virtuel, à l'affichage et à l'interface utilisateur. Le casque VR est directement branché par USB sur ce PC central et la communication avec le robot à câbles se fait par Ethernet via un commutateur réseau "Network Switch".

## 2.2 Calibration et alignement

Afin d'intégrer les données provenant des différents sous-systèmes, il est nécessaire qu'elles soient toutes représentées dans le même repère de coordonnées. Ce repère peut être n'importe lequel des systèmes de coordonnées de caméras, mais il est plus pratique d'en définir un nouveau qui sera appelé repère global. Chacun des systèmes doit donc être calibré sur ce dernier, c'est-à-dire que l'on doit calculer la matrice de transformation rigide (Section [1.2.2](#)) qui explique la rotation et la translation entre le repère local à chaque caméra Kinect et du casque Oculus vers le repère global.

Il n'est pas nécessaire que tous les composants du système (robot, caméra, casque) soient calibrés directement à partir de ce système de coordonnées global. En effet, il faut déterminer la transformation entre au moins un des composants et le repère global puis il est possible de calibrer les autres composants entre eux. Comme le décrit l'équation 2.1, en calculant la série de transformations qui mène au repère, nous pouvons obtenir la matrice de calibration extrinsèque (Section [1.2](#)) vers ce repère. Voir la [Figure 2.2](#).

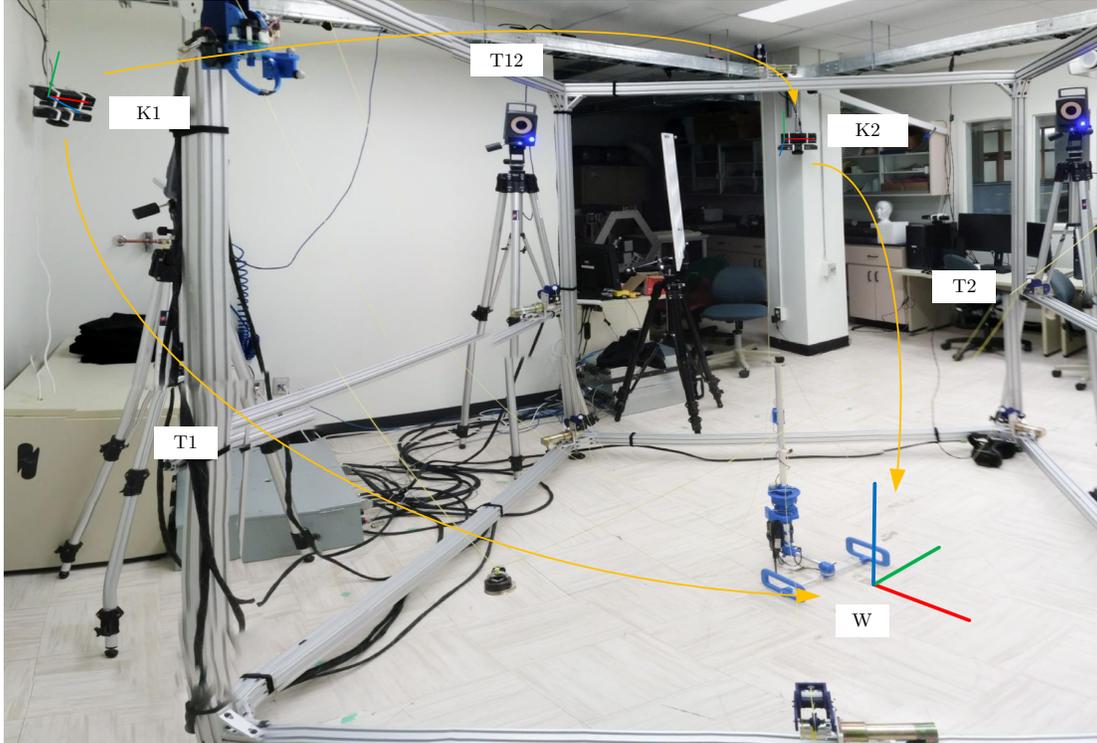


Figure 2.2 – Transformation entre le repère global et les repères de deux caméras K1 et K2.

Si la caméra K2 est calibrée par rapport au repère global  $\mathbf{W}$  selon la matrice de transformation rigide  $\mathbf{T2}$  et que la caméra K1 est calibrée par rapport à la caméra K2 selon la matrice de transformation rigide  $\mathbf{T12}$ , multiplier les matrices  $\mathbf{T12}$  et  $\mathbf{T2}$  permet de retrouver la matrice  $\mathbf{T1}$ . Inverser une matrice de calibration changera la direction de la transformation, c'est donc la pose d'un repère vue selon l'autre repère.

$$\mathbf{T1} = \mathbf{T12} \cdot \mathbf{T2} \quad (2.1)$$

Par la même logique  $\mathbf{T12}$  est directement obtenu si l'on connaît  $\mathbf{T1}$  et  $\mathbf{T2}$ .

$$\mathbf{T12} = \mathbf{T1} \cdot \mathbf{T2}^{-1} \quad (2.2)$$

Pour notre système, une première étape d'alignement initial est faite pour calibrer toutes les caméras dans un repère global  $\mathbf{W}$  selon les transformations  $\mathbf{T}_i$ . Ensuite, pour l'alignement fin les autres caméras sont calibrées sur une caméra utilisée comme référence (Équation (2.1) et (2.2)), celle-ci est reportée sur le repère global  $\mathbf{W}$ . Cette approche permet de l'imiter

l'accumulation d'erreur lors de l'alignement de l'information entre une série de caméras.

D'une manière semblable, le robot et le casque sont calibrés par rapport aux caméras Kinects. Comme le robot à câbles et le casque VR Oculus sont mobiles dans l'espace, ils doivent être calibrés à chaque utilisation. Les enrouleurs du robot sont fixes ([Figure 2.1](#)), mais un raffinement est nécessaire pour mesurer la pose de l'effecteur qui est manipulé par le sujet dans les expériences.

### 2.2.1 Calibration intrinsèque des Kinects

Avant de calibrer les caméras entre elles dans le montage, chacune doit être calibrée individuellement afin de permettre la projection de l'information 2D dans le monde 3D qui servira à la fusion des squelettes après alignement. Ce processus est appelé calibration intrinsèque. Il est possible d'estimer les paramètres intrinsèques de la caméra ToF du capteur Kinect comme détaillé dans la section [1.2.1](#) en se servant de l'image de profondeur. À noter que la caméra infrarouge est aussi (capteurs et lentilles) celle utilisée pour l'acquisition de l'information de profondeur, il serait donc possible d'estimer les mêmes paramètres à partir de l'image IR ou de l'image de profondeur. Par contre, l'image de profondeur permet de faciliter la segmentation ou la détection, car elle fournit justement la distance au capteur.

Comme la caméra ToF ne capture pas la texture (couleur), il n'est pas possible d'utiliser une cible de calibrage avec un patron de points visibles (cercles ou damier) comme on procède habituellement pour calibrer les caméras 2D classiques. Nous utilisons plutôt un plan avec un patron de cercles troués à distance régulière et connue de manière précise ([Figure 2.3\(a\)](#) et [\(b\)](#)). L'information de distance observée à travers les trous étant différente de l'information observée sur le plan, il est possible d'identifier le centre des cercles dans l'image de profondeur. Plusieurs images de la cible sont prises pour couvrir complètement le champ de vision de la caméra de manière à permettre l'estimation des paramètres de distorsion radiale de la lentille. Les points identifiés sont ensuite utilisés pour retrouver les paramètres intrinsèques (spécifiés dans la section [1.2.1](#)) de la caméra grâce à la méthode de calibration de Zhang [\[11\]](#).

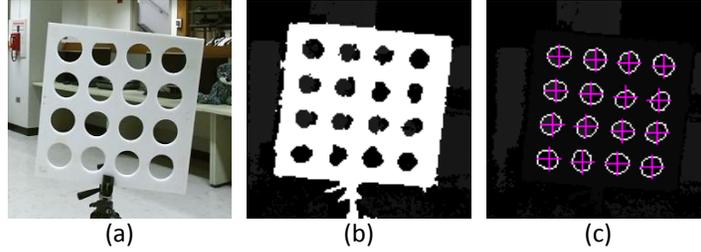


Figure 2.3 – Cible trouée vue par un capteur Kinect. (a) Cible vue par la caméra RGB (b) Cible vue par le capteur de profondeur et segmentée en distance. (c) Points détectés sur la cible.

L'identification de la position du centre des cercles dans l'image de profondeur (Figure 2.3(c)) est obtenue en segmentant la distance afin de ne conserver que l'information rapprochée de la cible, donc éliminer l'arrière-plan (Figure 2.3(b)). On cherche ensuite l'ellipse expliquant le cercle vu de perspective pour estimer son centre sur l'image selon la méthode des moindres carrés proposée par Fitzgibbon [46]. Avec la liste des points de référence (cible physique) et la liste des points observés (cible image) on estime les paramètres intrinsèque de la caméra et les paramètres de distorsion de la lentille avec les calculs basés sur les méthodes de Zhang [11] et Bouguet [32].

Afin de permettre la calibration extrinsèque, il est nécessaire d'obtenir l'information des points de l'image 2D dans l'espace 3D. Pour obtenir un point 3D  $\mathbf{pm}_{(x,y,z)}$  à partir d'un point sélectionné dans l'image il est premièrement nécessaire d'appliquer la correction de la distorsion de lentille selon les paramètres fournis par la calibration intrinsèque (Section 1.2.1) puis de multiplier le vecteur constitué du pixel  $\mathbf{pi}_{(u,v,d)}$  avec l'inverse de la matrice de calibration  $\mathbf{K}$ , la valeur de  $d$  étant la distance donnée par le capteur ToF. Les variables  $u$  et  $v$  indiquent la position en pixels dans l'image.

$$\mathbf{p}_m = \mathbf{K}^{-1} \begin{bmatrix} u \\ v \\ d \end{bmatrix} \quad (2.3)$$

### 2.2.2 Calibration extrinsèque des Kinects

Cette procédure consiste à estimer la transformation rigide (matrice de calibration

extrinsèque) qui explique la pose (position et orientation) de chaque caméra par rapport au repère global. La calibration extrinsèque se fait en déterminant la transformation alignant de l'information vue par les caméras dans leurs repères respectifs. Dans notre système l'opération se fait en deux étapes, un alignement initial afin d'obtenir une bonne estimation de la matrice de calibration, puis une étape de raffinement pour optimiser cette transformation. Cette calibration en deux temps est une solution au problème de convergence des algorithmes de calibration fins [14], [17], par un préalignement. L'alignement doit être fait sur l'information observée dans la scène donc après avoir été projetée selon la matrice de calibration intrinsèque et la valeur de profondeur mesurée par le capteur ToF. La projection de tous les pixels est représentée par un nuage de points ("Point Cloud") en trois dimensions.

### **Alignement brut**

L'alignement brut, ou alignement initial, sert principalement de base à l'algorithme d'alignement fin présenté par la suite. L'approche retenue est d'aligner le centre de trois sphères vues par toutes des caméras Kinects. Chaque caméra observe une section de la surface de chaque sphère qui n'est pas nécessairement la même section que celles vues par les autres caméras selon leur point de vue respectif. Par contre, le centre de ces sphères devrait être le même pour toutes les caméras si la surface des sphères est bien identifiée dans chaque vue.

Pour estimer la position du centre des sphères à partir de points d'une section de leur surface nous utilisons la méthode "RANdom Sample Consensus" (RANSAC) [33]. Cet algorithme itératif (Algorithme 3) estime les paramètres d'un système bruité contenant des valeurs aberrantes en se basant sur un modèle mathématique. Dans notre cas le modèle est l'équation cartésienne de la sphère  $(x-x_0)^2+(y-y_0)^2+(z-z_0)^2 = r^2$ .

---

**Algorithme 3** RANSAC

---

Point set  $A$

Distance to model threshold  $\epsilon$

Minimum inliers threshold  $\tau$

**do**

Randomly select the minimum number of points  $a_k \in A$  required to compute the model  $M$ .

Estimate the equation of the model  $M$  from random points  $a_k$ .

Identify the inliers  $a_i \in A$  that fit the equation of the model  $M$  under a tolerance  $\epsilon$ .

**if** Number of inliers  $a_i$  exceeds the threshold  $\tau$  **then**

Re-estimate the equation of the model  $M$  from inliers  $a_i$

Set *convergence* = *true*

**end if**

**while** *convergence* = *false*

---

Malgré la robustesse généralement associée à l'approche RANSAC, il est d'abord nécessaire de filtrer grossièrement le nuage de point de manière à éliminer le plus possible de points aberrants (i.e., qui ne représentent pas la surface de la sphère) avant d'appliquer l'optimisation. Ce filtrage se fait simplement en ajustant une boîte limitant la recherche de points autour de la sphère.

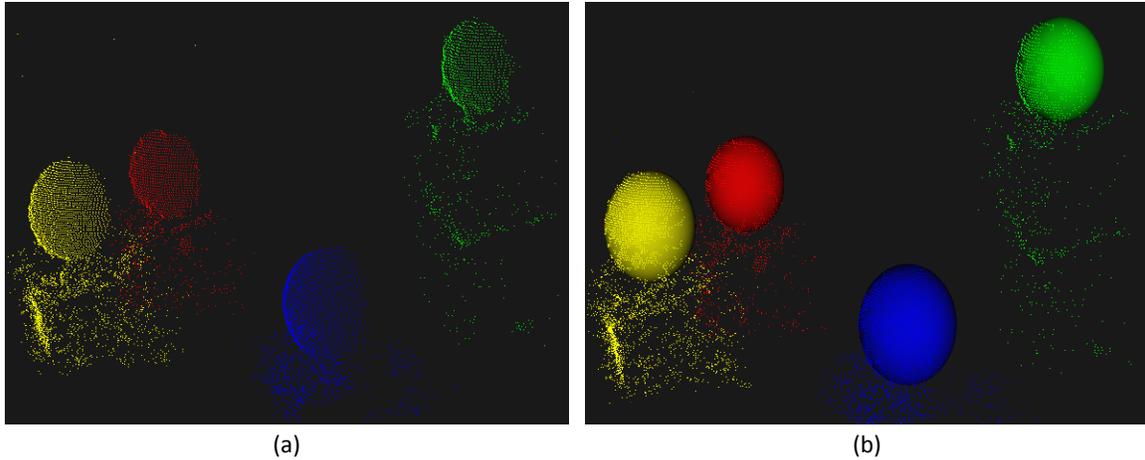


Figure 2.4 – Identification d'une sphère selon les repères de 4 caméras non-alignées, chaque couleur représente l'information provenant d'un point de vue différent. (a) Nuages de points (b) Sphères identifiées dans les nuages de points par la méthode RANSAC.

Le point central de la sphère est donc obtenu avec la méthode RANSAC dans le repère de chacune des caméras, voir la [Figure 2.4](#). Pour estimer la matrice de transformation rigide entre le repère respectif de chaque caméra et le repère global, il est nécessaire que les ensembles de points à apparier pour l'estimation contiennent au minimum trois points. Il est possible d'utiliser plus de points pour répartir (moyenner) l'erreur d'alignement et améliorer la robustesse, mais comme dans notre cas la calibration sera raffinée, il est acceptable d'en utiliser seulement trois. Précisons que plutôt que de prendre trois sphères différentes qui seraient difficiles à différencier dans les images, il est préférable d'observer la même sphère à trois positions différentes. C'est ce que nous avons adopté comme approche.

Cette technique nous permet de définir un système de coordonnées global sans effort supplémentaire. Nous utilisons donc trois positions de sphère et les définissons de la manière suivante:

- Point 1: Position de l'origine.
- Point 2: Direction de l'axe X.
- Point 3: Direction de l'axe Y.

Chaque caméra peut observer la sphère aux trois positions qu'elle occupe. Il n'est donc pas nécessaire de passer par un repère intermédiaire et il est possible d'estimer directement

la transformation entre les caméras et le repère global défini comme expliqué ci-dessus. Voir la [Figure 2.2](#) pour visualiser ces transformations sur le système développé.

Trouver la transformation rigide optimale entre le repère d'une des caméras et le repère global revient à résoudre l'équation suivante:

$$\mathbf{B} = \mathbf{R} \cdot \mathbf{A} + \mathbf{t} \quad (2.4)$$

$\mathbf{A}=\{a_1, a_2, \dots, a_n\}$  et  $\mathbf{B}=\{b_1, b_2, \dots, b_n\}$  sont deux ensembles de points 3D correspondants observés dans l'image et fixés sur l'objet de calibrage (ici la sphère à ses trois positions),  $\mathbf{R}$  est la matrice de rotation et  $\mathbf{t}$  le vecteur de translation.

Comme détaillé dans [\[17\]](#), il est possible de résoudre cette équation par la décomposition en valeurs singulières ("Singular Value Decomposition") (SVD)[\[34\]](#). Il suffit de calculer la matrice de covariance des deux ensembles de points et d'appliquer la méthode SVD pour retrouver les composants de la matrice de rotation  $\mathbf{R}$  et du vecteur de translation  $\mathbf{t}$ . On reconstruit la matrice de calibration extrinsèque (homogène 4x4) en combinant ces résultats comme il est détaillé dans l'équation 1.3 et 1.4. Chacune des matrices obtenues permet d'aligner l'ensemble de points correspondant au point de vue d'une caméra. La [Figure 2.5](#) présente le résultat de cet alignement sur quatre points de vue.

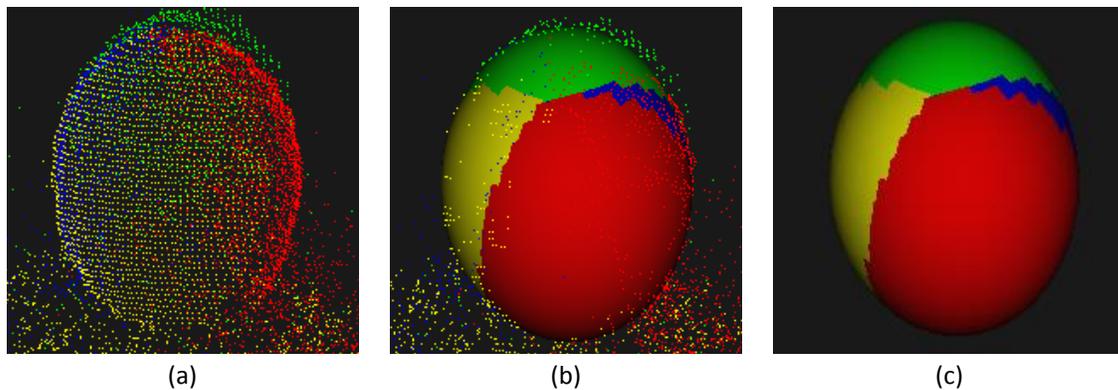


Figure 2.5 – Alignement d'information provenant de quatre caméras, chaque couleur représente un point de vue différent. (a) Nuages de points alignés (b) Nuages de points alignés et sphères identifiées (c) Sphères identifiées seules.

## Alignement fin

Malgré que la pose des caméras soit bien estimée grâce à la méthode d'alignement précédente, l'impact d'un raffinement peut-être considérable sur le résultat final ([Figure 2.6](#)).

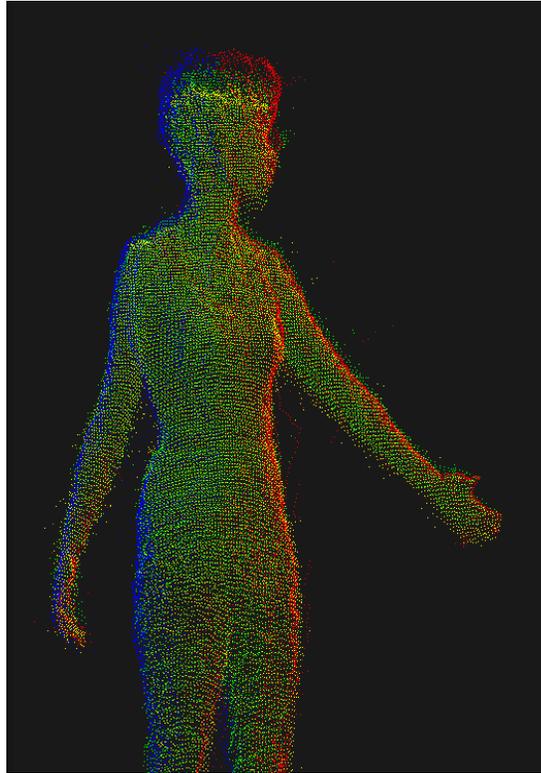


Figure 2.6 – Nuage de points d'une personne vue selon quatre points de vue alignés. Chaque couleur représente un point de vue différent.

L'algorithme ICP décrit dans la revue de littérature (Algorithme 1) est la solution retenue pour optimiser l'alignement des nuages de points, ce qui permettra d'améliorer la précision de la matrice de calibration extrinsèque obtenue par l'approche décrite à la section précédente.

Si la configuration spatiale des caméras le permet (recouvrement des champs de vision pour toutes les caméras), il est idéal d'effectuer le calibrage extrinsèque de toutes les caméras par rapport à une autre caméra considérée comme la référence. De cette manière, l'erreur de calibration ne sera pas accumulée dans la série de transformations. Pour aider à la convergence de l'algorithme vers une bonne solution, nous filtrons grossièrement les points superflus et ne conservons que la zone centrale observable de tous les points de vue pour ensuite calibrer selon la procédure décrite ci-dessous.

Après avoir ajouté un objet facilement identifiable dans la scène (cible, plan, boîte, etc.), nous accumulons plusieurs images (nuages de points) sur chaque caméra tout en déplaçant l'objet à différentes poses entre chaque enregistrement de données. Trois ou quatre poses suffisent généralement à donner un bon résultat. Il faut s'assurer que l'objet a une géométrie qui ne cause pas d'ambiguïté dans l'alignement par ICP, par exemple un plan d'une épaisseur non négligeable vu des deux côtés sera aligné comme si les deux faces faisaient partie de la même surface d'épaisseur nulle. Un autre exemple simple d'ambiguïté est la sphère qui présente une surface pratiquement identique, peu importe le point de vue. Si cette propriété est exploitée pour le calibrage grossier décrit à la section précédente, elle s'avère nuisible à la phase de raffinement. L'algorithme aura tendance à aligner ces surfaces comme si le point de vue était le même (voir la [Figure 2.7](#) pour une illustration de cette situation). Pour contrer cela, l'objet observé doit présenter une forme non symétrique ou orientée judicieusement entre chaque prise d'image de manière à ne laisser qu'une seule solution optimale.

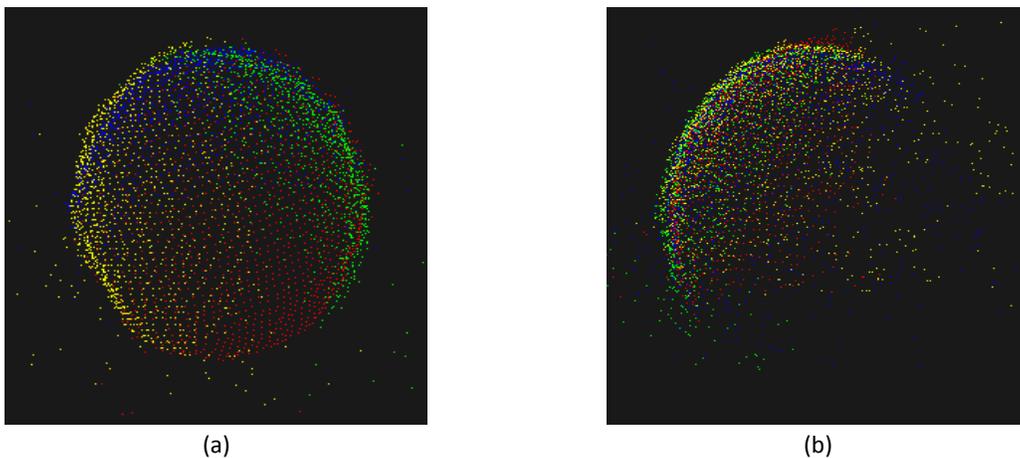


Figure 2.7 – Ambiguïté lors du raffinement de l'alignement par ICP. Chaque couleur représente un point de vue différent. (a) Nuage de points aligné représentant une sphère avant l'application de l'algorithme ICP (b) Nuage de points incorrectement corrigé par l'algorithme ICP.

Une fois les images acquises, nous appliquons la méthode ICP pour raffiner la calibration brute obtenue par la méthode des trois sphères. Cela aura pour effet d'améliorer l'alignement de l'information donnée par les caméras, c'est à dire, les squelettes obtenus par l'algorithme d'estimation de pose. Cette étape est cruciale pour le processus de fusion qui nécessite des positions de joints rapprochés dans l'espace global.

### 2.2.3 Calibration du mécanisme à câbles

La pose de l'effecteur du mécanisme à câbles qui est utilisé pour créer un retour haptique au sujet en cours de réadaptation est connue précisément par rapport à sa position au démarrage du système. Cette position est connue dans le repère du mécanisme, mais pas dans celui du système de vision. Une calibration est donc nécessaire afin de connaître la pose de l'effecteur dans le repère du système de vision. De plus, l'origine du système de coordonnées de l'effecteur du mécanisme est différente à chaque utilisation (démarrage du système mécanique), car elle dépend justement de la position initiale du robot. Il est donc nécessaire de faire la calibration avec le système de vision à chaque fois que le système est remis en marche.

L'approche utilisée pour la calibration du mécanisme est très similaire à celle utilisée pour la calibration initiale des caméras. Elle consiste à résoudre par SVD un système matriciel pour obtenir la transformation rigide entre deux ensembles de points, la position de ces points devant être connue précisément dans les deux référentiels. Il n'est pas possible de se servir de la technique des centres des sphères exploitée pour le calibrage extrinsèque du système de vision, car les fixer sur l'effecteur n'est pas réaliste. La méthode d'alignement de nuages de points par ICP n'est pas non plus idéale à cause de la taille et de la forme réduite de l'effecteur qui est conçu pour ne pas entraver la scène.

Une solution est d'exploiter des points visibles par les caméras et connus par le système mécanique. Pour cela, nous avons fixé trois marqueurs sphériques rétro réfléchissants sur le mécanisme, voir la [Figure 2.8](#). La position physique de ces marqueurs est déterminée lors de la conception mécanique de l'effecteur. Leur position dans le repère du robot est donc connue en tout temps.

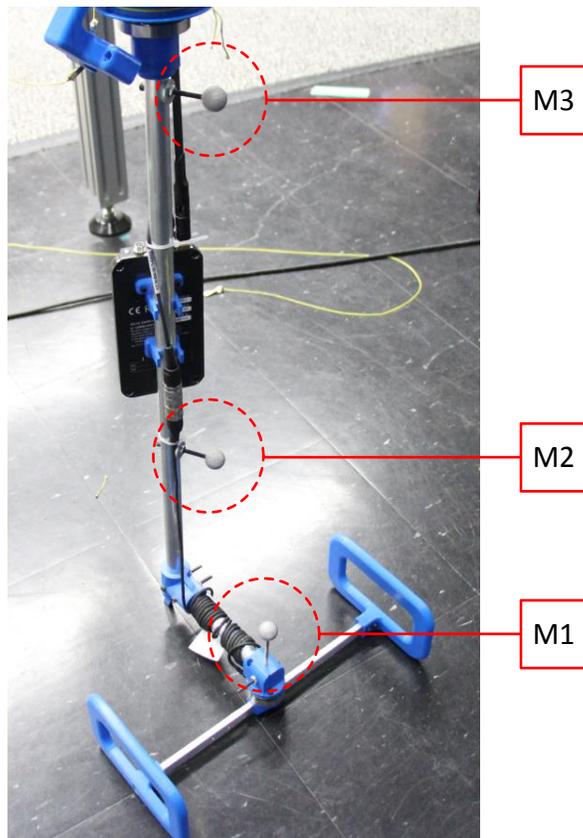


Figure 2.8 – Marqueurs fixés sur l'effecteur du mécanisme à câbles permettant l'alignement avec le système de vision, identifiés par M1, M2 et M3.

Pour retrouver la position de ces marqueurs dans l'origine globale du système (représentant la scène), nous devons pouvoir les observer par un composant calibré sur ce repère (global). La Kinect peut observer les marqueurs avec la caméra couleur ou avec la caméra infrarouge (IR). On remarque sur la [Figure 2.9](#) qu'il est possible de tirer avantage de l'émetteur IR du ToF de la Kinect pour identifier les marqueurs spécialement conçus pour réfléchir la lumière infrarouge. Comme mentionné précédemment, le capteur IR est utilisé par le ToF pour générer l'image de profondeur, nous pouvons donc utiliser les mêmes paramètres intrinsèques et extrinsèques obtenus lors de la calibration du capteur de profondeur (ToF).

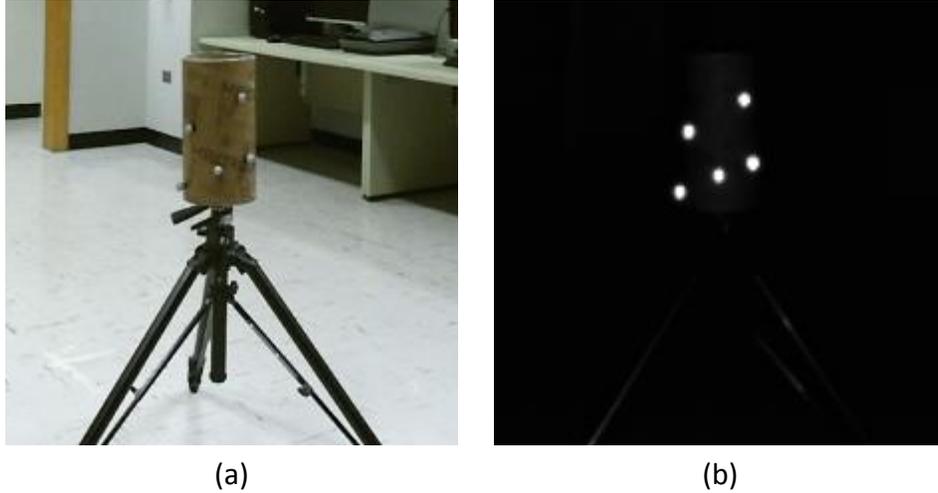


Figure 2.9 – Marqueurs sphériques rétro réfléchissants fixés sur un objet vu par un capteur Kinect. (a) Marqueurs vus par la caméra RGB (b) Marqueurs vus par la caméra IR.

Afin d'éliminer les sources de fausses détections de marqueurs par les caméras IR, une opération de masquage est appliquée. Les marqueurs sont d'abord cachés ou retirés de la scène. Ensuite, nous enregistrons plusieurs images IR puis elles sont seuillées et additionnées en une image binaire représentant les zones d'émission de lumière infrarouge. Une opération morphologique d'ouverture (érosion et dilatation) est aussi appliquée à l'image pour bien masquer les zones bruitées. Ce masque doit être refait si la scène change ou si les caméras sont déplacées, mais sinon il sera conservé sur plusieurs utilisations.

Une fois le masque obtenu, les marqueurs sur le robot sont réexposés et rendu visibles aux caméras. Le masque est ensuite appliqué sur les images provenant des caméras pour faciliter la détection des marqueurs. Un algorithme de détection de contour [47] est utilisé pour identifier la position des marqueurs dans l'image. L'algorithme cherche les pixels connectés puis en trouve le centre de masse. Si le rayon est assez grand pour correspondre à un marqueur selon la taille et la distance connue des caméras, le point est retenu.

L'identification individuelle des trois marqueurs est nécessaire, car nous avons besoin de la correspondance des points vus par les caméras avec la position des marqueurs dans le repère du robot. Ce problème est résolu d'une manière simple et efficace, les marqueurs sont positionnés sur l'effecteur selon des hauteurs physiques différentes qui nous permettent de les différencier selon leur position dans l'image. Pour que cette méthode d'identification fonctionne, il faut

s'assurer que l'orientation de chaque caméra et la position des trois marqueurs ne laissent aucune place à l'ambiguïté. Il est possible de placer l'effecteur dans une position idéale lors de l'alignement si cela facilite l'opération d'identification.

Après avoir détecté sans ambiguïté la position des trois marqueurs dans au moins deux images, il est nécessaire de trianguler leurs positions dans la scène en projetant une ligne passant par le point image vers l'espace global. Théoriquement, le point où deux lignes projetées provenant de deux points de vue (caméras) se rencontrent correspond à la position physique du centre du marqueur. En pratique à cause du bruit d'acquisition et du caractère discret des images, les lignes dans l'espace 3D ne se rencontrent généralement pas en un point, mais il est possible de trouver l'endroit où ces lignes se croisent au plus proche en cherchant la plus petite distance entre ces deux projections.

Comme détaillé dans [35], le segment le plus court entre les lignes sera perpendiculaire à ces deux lignes. On définit  $a_1$  et  $a_2$  comme deux points sur la ligne **A** ainsi que  $b_1$  et  $b_2$  comme deux points sur la ligne **B**. Ces points limites doivent nécessairement englober l'intersection cherchée (plus court segment entre les lignes) pour obtenir un résultat valable. Les points  $a_0$  et  $b_0$  sont les extrémités de ce segment (Figure 2.10). Si le produit scalaire de ces vecteurs est égal à zéro (lignes perpendiculaires), il est possible de calculer la valeur des points du segment d'intersection selon les équations suivantes:

$$(\mathbf{a}_0 - \mathbf{b}_0) \cdot (\mathbf{a}_2 - \mathbf{a}_1) = 0 \quad (2.5)$$

$$(\mathbf{a}_0 - \mathbf{b}_0) \cdot (\mathbf{b}_2 - \mathbf{b}_1) = 0 \quad (2.6)$$

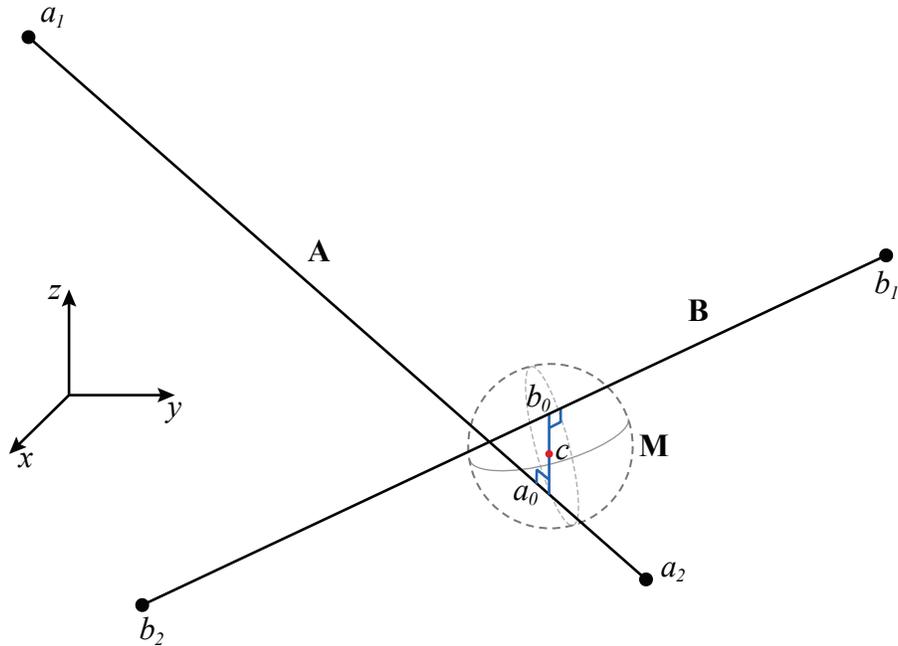


Figure 2.10 – Identification du point de croisement entre deux lignes dans l'espace  $\mathbb{R}^3$ . Le point  $c$  représentant le centre du marqueur sphérique  $\mathbf{M}$  est le point milieu du segment délimité par  $b_0$  et  $a_0$ .

On conserve ces points et on refait l'opération de triangulation pour chaque combinaison de deux lignes (une ligne par caméra pour un marqueur). Si la longueur du segment d'intersection est plus élevée qu'un seuil fixé selon le bruit de mesure, le point d'intersection triangulé  $c$  représentant le centre du marqueur est rejetés. La position moyenne de tous les points  $c$  retenus représente l'estimation de la position d'un marqueur. On refait l'opération pour obtenir la position de chacun des trois marqueurs puis, comme pour l'alignement brut, on cherche la transformation optimale entre les deux ensembles de points (repère robot et repère global) selon les équation [\(2.5\)](#) et [\(2.6\)](#), la résolution détaillée est décrite dans [\[50\]](#).

## 2.2.4 Calibration du casque VR

Le dernier composant à calibrer est le casque de réalité virtuelle. Le fonctionnement du système de réalité virtuelle est expliqué de manière plus détaillée dans la section 2.4, mais pour ce qui est de la calibration nous devons pour l'instant savoir que le suivi de la position du casque sera assuré par les caméras Kinect et que l'orientation sera basée sur un capteur

inertiel (IMU) à l'intérieur du casque.

La calibration entre l'orientation du casque et de celle la scène permet d'aligner le point de vue (caméra virtuelle) de l'image projetée dans le casque avec la scène virtuelle et ses composantes réelles (mécanisme à câbles et squelette du sujet). Une procédure simple et fonctionnelle est proposée pour effectuer ce calibrage. Le casque est d'abord mis en marche afin d'avoir accès à l'information d'orientation provenant du gyroscope (rotation dans le plan horizontal) et de l'accéléromètre (inclinaison dans le plan horizontal). Puis on positionne manuellement le casque le long d'un axe du repère global  $x_W$ . On enregistre la valeur d'orientation  $\theta_0$  donnée par le casque selon l'axe de son repère local  $x_O$  puis on la retire des prochaines mesures afin de réinitialiser son orientation sur  $x_W$ .

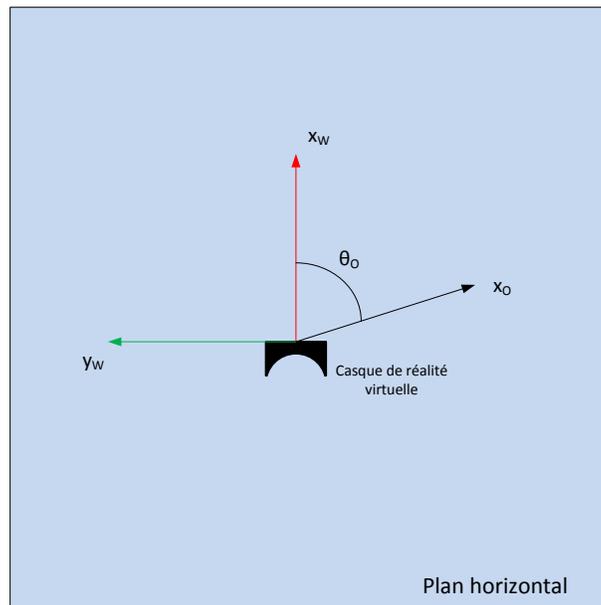


Figure 2.11 – Vue de dessus de l'alignement de l'orientation horizontale du casque de réalité virtuelle.

Cette procédure doit être répétée à chaque fois que le casque est mis en marche, car un gyroscope perd son point de référence initial lorsque le casque est mis hors tension. Bien que l'inclinaison dans le plan horizontal ("Tilt") mesurée par l'accéléromètre est une référence fiable parce-qu'elle est basée sur la direction de l'accélération gravitationnelle, la rotation autour de l'axe vertical ne peut pas être assurée par la direction du nord magnétique donnée par le magnétomètre à cause des multiples interférences dans le champ magnétique mesuré par l'IMU du casque.

## 2.3 Fusion des données de pose du sujet

La fusion de données est le processus de combinaison de données provenant de différentes sources. Ces données peuvent être plus ou moins fiables et provenir de différents types de capteurs. Cette approche de fusion est souvent nécessaire lorsqu'un système n'est que partiellement observable ou que la fiabilité de l'information n'est pas garantie.

L'objectif spécifique à notre système est de réduire l'incertitude des mesures de position des joints du squelette du sujet obtenus par une seule Kinect en combinant les informations provenant de plusieurs Kinects. Voir le schéma du montage de la [Figure 2.1](#) montrant l'arrangement spatial des caméras Kinects dont nous voulons fusionner les informations.

L'algorithme d'extraction du squelette développé par Microsoft [\[20\]](#) fournit un très bon résultat en temps réel, mais n'est pas suffisant pour une application comme la nôtre. Le fait que la méthode de [\[20\]](#) soit basée sur des images de profondeur recueillies d'un seul point de vue est trop limitatif pour notre application, notamment à cause des occlusions. En utilisant ce même algorithme selon plusieurs points de vue, on augmente la couverture spatiale et réduisons l'impact des occlusions par la redondance de l'information (voir la [Figure 2.12](#)). Nous prenons donc avantage de cela pour augmenter la couverture (espace observé), la fiabilité (identification) et la précision (mesure de position) du système de suivi de mouvement. Avec ces améliorations, le choix d'utiliser l'algorithme d'estimation de pose du squelette proposé par Microsoft devient une solution efficace et adaptée à nos besoins pour l'application de réadaptation.

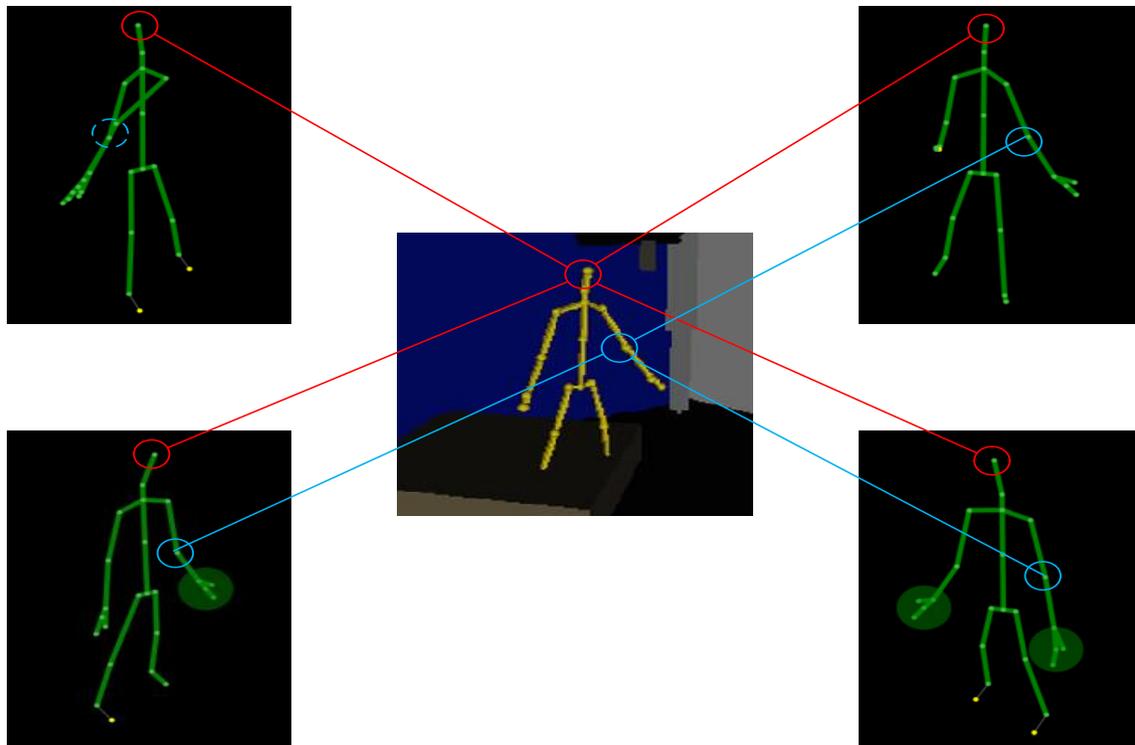


Figure 2.12 – Visualisation du bénéfice de la redondance de l'information selon différents points de vue.

L'approche de fusion de données retenue est l'approximation par filtrage particulaire pour sa flexibilité dans le filtrage de systèmes non-linéaires et son adéquation à la complexité du mouvement humain. La théorie du filtre particulaire est présentée brièvement en section [1.4.1](#) puis son implémentation est présentée en section [2.3.4](#).

### 2.3.1 Traitement des données à fusionner

#### Acquisition et transfert

La première étape de la méthode de fusion de donnée est l'extraction des données brutes, c'est-à-dire l'estimation de la position des joints du squelette selon chacun des points de vue. Les sous-systèmes NUC reçoivent l'image de profondeur du capteur Kinect et calculent la position des 25 joints représentant la pose de la personne observée en 2D grâce à l'algorithme d'estimation de pose utilisé. Ces sous-systèmes NUC sont tous synchronisés sur la même

horloge et un temps associé à chacune des mesures est enregistré. La transmission de l'information entre chaque NUC et le PC principal se fait ensuite par communication Ethernet. Ce PC principal est utilisé pour les opérations de fusion, d'enregistrement et d'affichage.

## Alignement

Comme mentionné, l'information de position des joints est représentée dans le repère de l'image de profondeur des capteurs Kinects sur chacun des NUC. Après avoir transféré l'information sur le PC principal, on projette les points (2D) dans la scène (3D) vue par le repère local de chacune des caméras selon la matrice de calibration intrinsèque calculée précédemment (voir l'équation (2.3) et la Figure 1.3). Ensuite on aligne ces points sur le repère global commun aux différents points de vue selon la matrice de calibration extrinsèque (voir les équations (1.3), (1.4) et la Figure 2.2). Il est possible d'effectuer ce processus en parallèle sur la liste de tous les joints comme le décrivent les équations suivantes:

$$\mathbf{P}_{ki} = \begin{bmatrix} x_1 & x_2 & \dots & x_n \\ y_1 & y_2 & \dots & y_n \\ z_1 & z_2 & \dots & z_n \\ 1 & 1 & \dots & 1 \end{bmatrix}_{4 \times N} \quad (2.7)$$

$$\mathbf{P}_{mi} = \mathbf{T}_{ki} \mathbf{P}_{ki} \quad (2.8)$$

$\mathbf{P}_{ki}$  représente l'ensemble des positions des joints observés en 3D dans le repère local de chacun des (indice  $i$ ) capteurs Kinects.  $\mathbf{T}_{ki}$  est la matrice de calibration pour chacun des repères locaux et le repère global.  $\mathbf{P}_{mi}$  représente ces mêmes points mais représentés dans le repère global.

## Correspondance

Un point essentiel à la fusion est d'établir la correspondance correcte entre les joints identifiés pour chaque caméra. Bien que l'algorithme d'estimation de pose fournisse l'identification du joint, cette identification n'est pas toujours valide et l'on doit éliminer les fausses correspondances autrement, la fusion fournit des résultats peu robustes. Par exemple, l'algorithme d'extraction du squelette ne fait pas de différence entre une personne vue de face et une

personne vue de dos, c'est-à-dire que les membres de gauche et les membres de droite seront inversés (effet miroir) si la personne est vue de dos. Il serait possible d'inverser simplement l'information de certaines caméras en ayant une connaissance *a priori* de l'orientation du sujet, mais cela ne serait pas fonctionnel pour un suivi robuste aux déplacements dans la scène.

L'approche développée est d'ajouter la notion de squelette de référence. Les nouvelles positions de joints obtenus sont constamment comparées à ce squelette de référence. Si l'information inversée est plus proche de la référence, le squelette comparé est inversé. Cette méthode ne demande pas beaucoup de temps de calcul à cause du nombre limité de joints et est assez flexible pour inclure d'autres changements que l'image miroir, par exemple la présence de plusieurs squelettes. Pour l'instant, le système ne supporte pas le mode multi-utilisateurs, mais est robuste aux multiples personnes dans l'espace de travail par comparaison des squelettes avec celui de référence. La référence peut être choisie selon le besoin, dans notre application la référence utilisée est la pose précédente du squelette fusionné.

Cette référence est fonctionnelle dans tous les cas pour le suivi du squelette, car la fréquence d'acquisition est assez élevée pour éviter l'ambiguïté lors de la comparaison (inversion subite). Par contre, afin d'initialiser une orientation voulue ou bien pour certaines applications comportant un montage où toutes les caméras font face à l'utilisateur, il est possible d'utiliser le squelette d'une des caméras comme référence ou bien même un squelette vu par un autre système, dans le cas de prise de mesure de validation par exemple.

## Mesure de qualité des données

Afin d'améliorer le résultat du processus de fusion, il est possible d'ajouter de l'information en entrée, c'est-à-dire de qualifier l'information à fusionner. Cet indice de qualité représente le niveau de certitude de l'information reçue par chacun des sous-systèmes (NUC et Kinect). Une mesure de qualité est calculée en fonction de plusieurs critères et sert à attribuer une confiance (un poids) à chacune des mesures (position d'un joint) utilisée comme entrée par l'algorithme de fusion de données par filtrage particulière. Trois mesures de qualité sont utilisées, nous appelons ces valeurs: qualité de l'image, qualité du squelette et qualité du joint. Les poids associés à chacune sont bornés entre 0 et 1 puis ajustés en fonction de la qualité observée comme présenté sur la [Figure 2.13](#). Selon nos observations, une courbe parabolique approxime bien la confiance des mesures pour les différents critères de qualités utilisés. Par contre, une étude plus complète des critères de qualité serait nécessaire afin d'identifier une

fonction plus représentative de la confiance selon chacun d'eux.

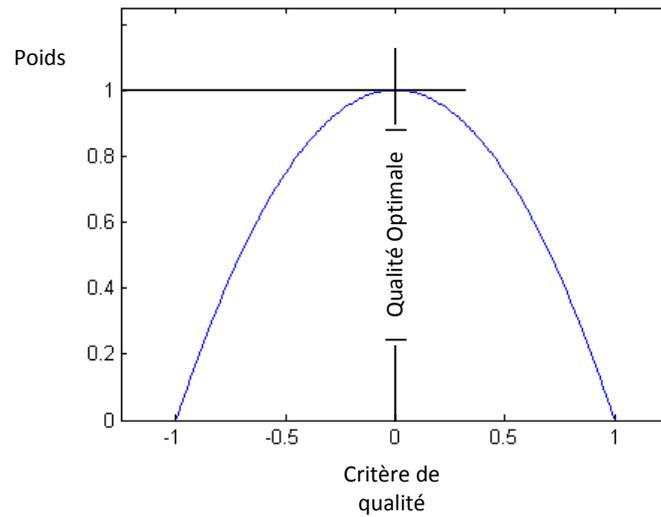


Figure 2.13 – Courbe parabolique du poids selon un critère de qualité.

La qualité de l'image diminue avec la distance du squelette à la caméra, la qualité est maximale lorsque le squelette se trouve dans la zone définie comme optimale de la mesure de profondeur pour la Kinect. La qualité du squelette est ajustée en fonction du nombre de joints bien identifiés et diminue donc en fonction du nombre d'occlusions. La qualité du joint est ajustée en fonction de sa distance au joint correspondant sur le squelette de référence ainsi que son état de mesure.

Nous appellerons ces états, *position observée* lorsque le joint est mesuré et directement dans le champ de vue de la caméra, *position déduite* lorsque le joint est caché du champ de vue de la caméra (occlusion), mais estimé selon l'ensemble du squelette puis *position inconnue* lorsqu'il n'y a aucune information de position associée à un joint.

La mesure de distance au joint de référence peut représenter une vitesse de déplacement anormale ou un mauvais alignement. Une position déduite réduira encore la qualité de la mesure et une position inconnue la réduira au minimum.

Un dernier perfectionnement permet d'éviter l'intégration de données instables variant entre la position déduite et la position observée, nous pouvons appeler ce phénomène "jitter". Lorsqu'un joint varie de manière répétitive entre l'état inconnu, déduit ou observé, il y a de

grandes chances que le joint soit erroné. Pour éviter ce cas, la valeur associée à la qualité doit pouvoir diminuer rapidement, mais augmenter plus graduellement. Pour appliquer ce principe, on vérifie la valeur de qualité précédente, si la valeur présente est plus grande que la valeur précédente, il faut moyenniser ces données afin d'incrémenter lentement la qualité, sinon utiliser directement la nouvelle valeur qui est nécessairement plus basse que la précédente.

### 2.3.2 Contraintes reliées au squelette

L'algorithme d'estimation de la pose utilisé et proposé par Microsoft est entraîné avec une base de données très large et variée, ce qui permet de rendre l'application que nous développons flexible à différents types de morphologie pour les sujets humains qui l'utilisent. Il est aussi conçu pour fonctionner sur une image de distance et ne prend pas en compte l'information temporelle. Cette approche peut causer des défauts dans le suivi de joints, défauts qui peuvent être corrigés en ajoutant certaines contraintes sur la longueur fixe des segments entre les joints. Le filtrage particulière combiné au système d'acquisition multivues (4 points de vues dans notre cas) permet de réduire le bruit dans la détection de la position des joints du squelette et rend le suivi robuste aux occlusions. Comme observé dans [44], la longueur des segments (os du squelette) entre les joints mesurés varie beaucoup. Ajouter une contrainte au filtre particulière le forçant à respecter la longueur de ces segments permet de réduire leur variation, ce qui améliore les résultats.

Une étape de calibration du squelette a été implantée et consiste à mesurer la longueur des segments (os) du sujet avant la prise des données. On demande donc à la personne de prendre une position idéale pour la mesure des segments (jambes et dos légèrement fléchis) puis on vérifie la longueur des os observés après la fusion pendant quelques secondes, 2 à 3 secondes semble suffisant pour réduire le bruit de mesure. Ces données sont moyennées puis enregistrées dans un fichier qui sera utilisé pour raffiner les mesures lors du processus de fusion suivant.

Pour appliquer la contrainte de longueur des segments lors de la fusion, on fixe un joint comme la référence zéro qui sera traité en premier sans prendre en compte l'information de longueur des os lors de chaque itération de filtrage des joints. Les autres joints seront combinés par la suite selon leur position dans l'arborescence définie pour le squelette ([Figure 2.14](#)).

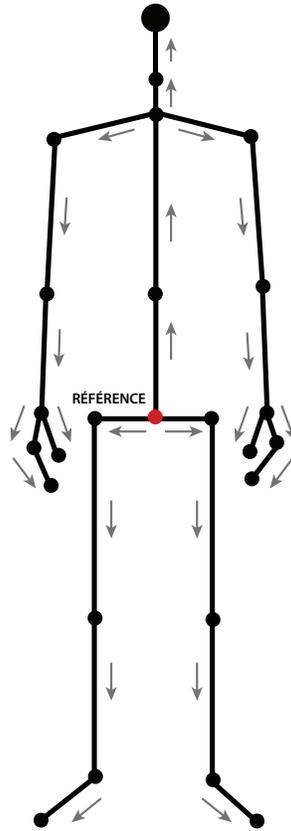


Figure 2.14 – Arborescence représentant l'ordre de la fusion de l'information basée sur l'estimation de la position selon les différents points de vue et la différence entre la longueur de l'os connue et la distance mesurée au joint précédent.

En suivant l'arborescence, la position de chaque joint distal (plus loin du centre du corps) est calculée par le filtre particulaire et peut être utilisée pour contraindre la position du joint suivant. En pratique, lors de l'étape d'observation de l'état du système (observation de nouvelles mesures par les capteurs) et de l'ajustement de poids de chaque particule, on incrémente le poids associé aux particules ayant une valeur de position qui est près de la surface de la sphère formée par la longueur de l'os (rayon) et la position du joint précédent (centre). Sur la [Figure 2.15](#), la longueur de l'os est définie par  $l$  et le joint précédent par  $j$ . Le nouveau joint devrait en théorie se trouver près de la surface de cette sphère. Modifier la valeur du poids associé à cette contrainte et le choix de la distance à la surface tolérée permet de rendre la contrainte plus ou moins sévère.

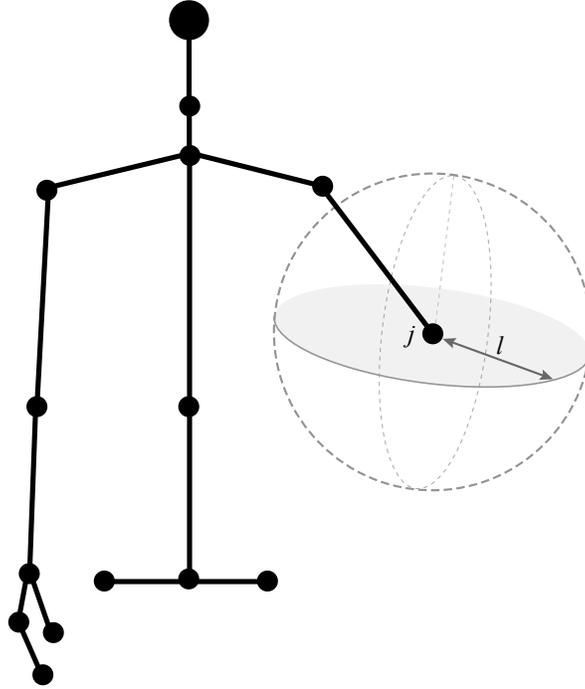


Figure 2.15 – Représentation de la sphère créée par la mesure du joint précédent  $j$  et de la longueur de l'os connu  $l$ . La position du joint suivant devrait se retrouver près de la surface de cette sphère.

### 2.3.3 Synchronisation et interpolation des poses

Afin que le filtre particulaire puisse intégrer les mesures, l'information temporelle est nécessaire. Comme expliqué précédemment, chaque mesure acquise voit une étiquette temporelle ("timestamp") lui être attribuée. L'horloge des sous-systèmes NUC est synchronisée avec celle du PC principal via un protocole réseau (NTP) qui assure la synchronisation temporelle nécessaire à la fusion. Bien que les ordinateurs soient synchronisés sur un temps de référence, les images provenant des caméras à 30Hz n'arrivent pas toutes au même moment. Des approches différentes doivent être utilisées pour combiner ces données en temps réel ou en post-traitement. Le mode post-traitement est défini par l'enregistrement des données brutes lors de l'exercice et la fusion de façon hors ligne. Cet enregistrement peut se faire en plus du traitement temps réel.

## **Fusion des poses en temps réel**

Un facteur très important en réalité virtuelle est de minimiser le délai entre une action (mouvement du robot, de l'avatar ou du point de vue) et l'affichage sur le casque VR afin que l'utilisateur perçoive les rétroactions en temps réel. Le compromis que nous avons choisi de faire pour satisfaire cette contrainte est de sacrifier une certaine précision dans la mesure pour minimiser le délai. Pour cela, nous effectuons une itération de filtrage particulière à une fréquence supérieure ou égale à 30Hz en sélectionnant les données de position des joints les plus récentes. Avec cette méthode, l'interpolation de la position des joints n'est pas permise et l'extrapolation pourrait fausser les calculs à cause du mouvement non-linéaire humain et du bruit dans l'information de profondeur. La perte de précision en position par le traitement en temps réel provient donc de l'impossibilité d'aligner parfaitement les données de positions des joints dans le temps sans créer un délai.

## **Fusion des poses en post-traitement**

Lors d'une séance de réadaptation, les données de position des joints provenant des NUC sont alignées spatialement puis enregistrées dans un fichier avec un temps associé. L'enregistrement permet d'améliorer la fiabilité de ces mesures en les fusionnant hors ligne ce qui permet une meilleure analyse biomécanique et offre au thérapeute une meilleure évaluation des problèmes du sujet. En interpolant les données, on s'assure de fusionner les positions de joints alignées sur un même temps et on évite ainsi une erreur causée par une mesure décalée temporellement. L'interpolation de type "Cubic Spline" est utilisée pour sa simplicité et son efficacité dans l'interpolation du mouvement. Une fois que l'interpolation est faite, on ré-échantillonne l'information sur la courbe spline cubique afin d'obtenir des données d'entrée du filtre alignées dans le temps et à intervalle fixe. Cette interpolation est effectuée sur le vecteur de position ainsi que sur les coefficients de qualité associés.

Comme expliqué dans la section [2.3.2](#), pour stabiliser la pose selon la contrainte de longueur des os, il est nécessaire de fusionner les données de position des joints de manière séquentielle. Contrairement au traitement temps réel qui est effectué en parallèle. Lors de l'observation des mesures et de la mise à jour des poids associés à chaque particule, la longueur du segment est prise en compte. C'est-à-dire que l'on vérifie si la distance entre la position associée à chaque particule du filtre et la position du joint précédent correspond à la distance enregistrée lors de la calibration du squelette (enregistrement des longueurs d'os). Les joints du squelette sont donc mis à jour selon les mesures des sous-systèmes en respect

avec la longueur des os.

Un autre avantage du traitement hors ligne est qu'il est acceptable de sacrifier la vitesse de calcul pour une meilleure précision. Par conséquent, le nombre de particules du filtre est augmenté jusqu'à ce que le gain de précision soit négligeable par rapport à la précision du capteur et de l'algorithme.

### **2.3.4 Filtrage et intégration - Détails d'implantation**

Le filtre particulaire utilisé dans ce travail est le type générique présenté en revue de littérature [1.4.1](#). L'initialisation de l'état des particules du filtre est simple, car on suppose qu'aucun des joints n'est observé au début de la prise de mesure (sujet en dehors de la zone observée). L'état initial est donc un vecteur position zéro avec un poids associé nul. Il est ensuite nécessaire d'initialiser certains paramètres du filtre, la variance du bruit d'échantillonnage ainsi que le nombre de particules. Ces deux paramètres sont dépendants l'un de l'autre.

La variance représente la taille de la zone couverte à chaque itération, autrement dit, la variabilité de l'état des particules lors d'une transition (itération de filtrage). Dans notre cas, c'est la distance entre la position précédente et la position à estimer. La règle empirique des trois sigmas permet d'estimer la variance nécessaire selon la taille de la fenêtre à observer. Par exemple, 95% des données sont généralement contenues dans une plage de plus ou moins 2 écarts types autour de la moyenne d'une distribution normale. En connaissant le temps entre les itérations et en fixant une vitesse de déplacement maximale pour chaque joint, on peut estimer la valeur d'écart-type et donc obtenir le paramètre de variance pour le filtre associé à chaque joint. Si nécessaire, il est possible de modifier dynamiquement la variance en fonction des observations précédentes et de la fixer sur une tendance calculée selon la vitesse de déplacement courante. Cela à l'effet d'augmenter la précision lorsque l'état est stable et de conserver une observation valide en changement rapide, mais en sacrifiant la précision, car la variance sera dépendante de la vitesse du joint.

Le nombre de particules représente la taille de l'échantillonnage utilisé pour approximer la position. Généralement, plus le nombre de particules utilisées est élevé, plus la précision de la prédiction sera élevée. Ce paramètre est dépendant de la variance, car un nombre restreint

de particules échantillonnées sur une grande plage (variation élevée) offre une probabilité réduite de représenter le système réel. La complexité d'un filtre particulaire est linéaire en fonction du nombre de particules ( $O(N)$ ). C'est donc la capacité de calcul du système qui limite le nombre de particules ainsi que la saturation de la précision lorsque le nombre de particules optimal est atteint.

Après l'initialisation du filtre, les étapes à effectuer pour chaque itération de filtrage sont celles mentionnée dans l'Algorithme 2 présenté en section [1.4.1](#). Ces étapes sont la transition, la pondération, la normalisation et le ré-échantillonnage.

Comme mentionné précédemment, il existe plusieurs variantes de filtrage basées sur les méthodes de Monte-Carlo et le filtre choisi ici est assez général. Il serait possible d'utiliser d'autres variantes en appliquant la même procédure. L'étape qui concerne spécifiquement notre système est la pondération, soit lorsque le poids de chaque particule est mis à jour en fonction de l'observation des mesures par la fonction de vraisemblance:

$$w_t^i = p(y_t|x_t^i) \quad (2.9)$$

$w$  représente le poids de chaque particule (probabilité de représenter l'état du système),  $y$  est la mesure observée (information des capteurs) et  $x$  est l'état de la particule (prédiction de la position du joint).

La fonction de vraisemblance  $p(y_t|x_t)$  est l'inverse de la fonction de densité de probabilité postérieure  $p(x_t/y_t)$ . Cette fonction donne une information sur la probabilité qu'a l'état courant d'être près de l'état réel en fonction de l'observation. Cette observation est la distance pondérée aux mesures de capteurs alignées ainsi que la différence de longueur d'os.

$$p(y_t|x_t^i) = \alpha \sum_{k=1}^K dy_k + \beta |d_j - l| \quad (2.10)$$

Tel que montré à la [Figure 2.16](#),  $d_y$  est la distance entre une observation  $y$  de capteur et une position donnée par une particule  $x$ . La distance au joint précédent dans la hiérarchie est représenté par  $d_j$  et la longueur calibrée du segment par  $l$ . Les constantes  $\alpha$  et  $\beta$  permettent d'ajuster l'importance du respect de l'une ou l'autre des mesures.

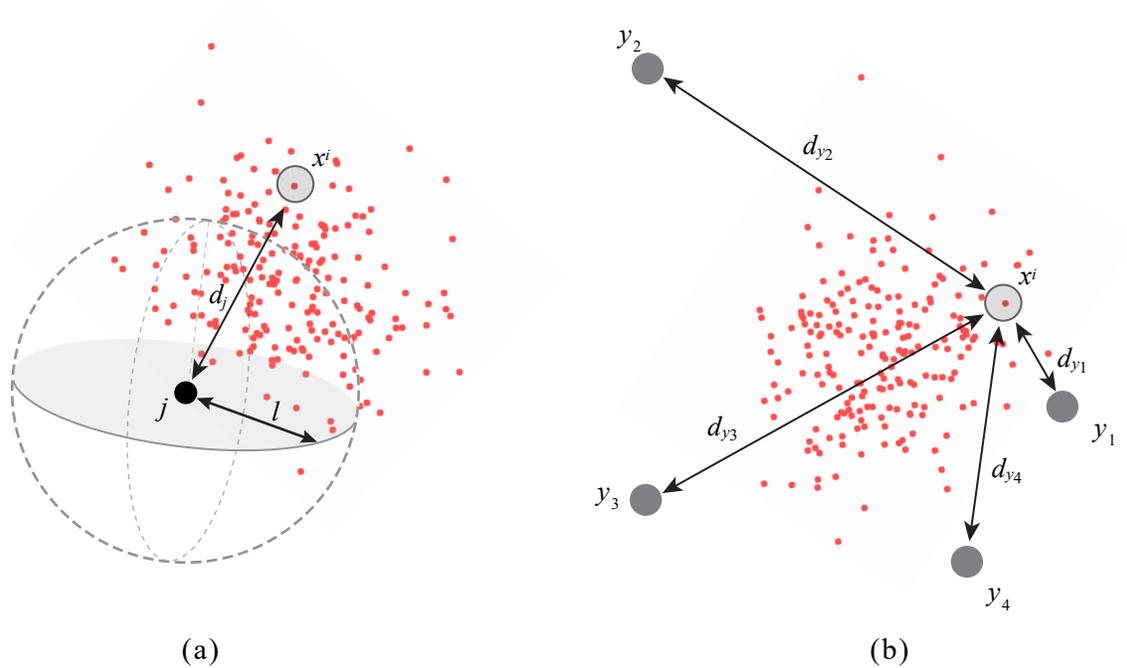


Figure 2.16 – Calcul de probabilité selon la fonction de vraisemblance. (a) Distance  $d_j$  d'une particule  $x^i$  à la surface de la sphère créée par la longueur de l'os  $l$  selon le centre donné par la position du joint précédent  $j$ . (b) Distance pondérée  $d_{y_k}$  d'une particule  $x^i$  aux estimations des capteurs  $y_k$ . L'indice  $k$  représente chacun des capteurs.

Après l'étape de pondération, il est possible de déterminer un état (position) probable pour chaque itération. Ce nouvel état est celui associé à la particule comportant le plus grand poids de l'échantillonnage, c'est cette valeur qui sera utilisée comme mesure de position d'un joint. L'opération de filtrage est effectuée sur chaque joint dans le temps pour finalement obtenir l'estimation de la pose du sujet pour toute de l'expérience.

## 2.4 Environnement virtuel

L'utilisation de la réalité virtuelle est un point clé de notre projet, car elle permet de travailler plusieurs aspects de la motricité d'un patient en réadaptation de manière sécuritaire et adaptable aux besoins du traitement afin d'optimiser chaque séance. Des améliorations significatives ont été observées lors de la réadaptation basée sur la réalité virtuelle en comparaison aux approches conventionnelles sur des patients ayant subi un accident vasculaire cérébral (AVC) [36]. Il est aussi démontré que la réadaptation motrice est améliorée avec

l'interaction physique ("Haptic feedback") en environnement virtuel et permet d'augmenter l'interactivité et la motivation du sujet [37], [38], [39], [40].

### 2.4.1 Casque VR

Le casque de réalité virtuelle permet de projeter une image de l'environnement désiré directement devant les yeux du sujet. Ce type de visualisation permet une immersion complète en temps réel dans l'environnement adapté selon la tâche de réadaptation. Le dispositif utilisé est l'Oculus Rift (Development Kit 2). Il a été retenu pour sa performance et son coût peu élevé. Voir la [Table 2.1](#) pour les spécifications techniques de l'Oculus.

Table 2.1 – Spécifications techniques du casque de réalité virtuelle Oculus Rift DK2. Informations provenant de [\[41\]](#).

Prix	~350\$ USD
Type d'affichage	OLED
Résolution (par œil)	960 x 1080
Pixels par pouce	441
Fréquence d'affichage	75 Hz
Champ de vision horizontal	90°
Communication	USB et HDMI
Capteurs	Gyroscope Accéléromètre Magnétomètre
Poids	440g

L'utilisation de l'accéléromètre de l'Oculus pour le calcul de la position de la tête du sujet n'est pas idéale à cause de la dérive de la mesure avec le temps qui cause une accumulation d'erreur. Le casque est conçu pour être suivi en position par une caméra externe basée sur une série de marqueurs (DEL IR placé sous sa surface). Par contre, comme le champ de vision limité de cette caméra ne permet pas de voir le casque en tout temps dans l'espace de travail délimité par le robot à câbles du montage de réadaptation, il est nécessaire de faire le suivi du casque avec notre propre système. Le processus de fusion présenté dans la section 2.3 permet d'estimer la position du centre de la tête. C'est ce point qui est utilisé pour le positionnement du casque en temps réel. Il est nécessaire de calculer la translation entre le casque et le centroïde de la tête afin de reporter la caméra virtuelle à l'endroit réel selon le

squelette observé. Cette donnée est mesurée manuellement sur la personne, mais il est possible d'utiliser des paramètres génériques sans trop influencer l'expérience.

La pose de la tête est complétée en ajoutant les mesures du gyroscope et de l'accéléromètre qui fournissent une estimation de l'orientation. Le gyroscope permet d'obtenir directement l'orientation dans le plan horizontal (rotation autour de l'axe vertical) en exploitant le principe de conservation du moment angulaire. L'accéléromètre permet d'obtenir l'inclinaison dans le plan horizontal (rotation autour des deux autres axes) en se basant sur la direction de l'accélération gravitationnelle. Voir la [Figure 2.11](#) afin de mieux visualiser les axes et le plan mentionnés. Le positionnement (translation) assuré par le système de caméras est limité à une fréquence de 30Hz, l'orientation est pour sa part assurée par le capteur inertiel qui fonctionne à une fréquence supérieure à l'affichage (75Hz). L'utilisateur ne ressent généralement pas de délais entre le mouvement et l'affichage, car la translation de la caméra influence beaucoup moins le champ de vision que l'orientation. En effectuant un suivi global dans l'espace de travail, on évite de l'obligation de déplacement virtuellement le point de vue du sujet dans l'espace de travail. Cela élimine une grande source de malaise pour l'utilisateur en réalité virtuelle ("Cybersickness") causé par un conflit d'information sensorielle entre le flux optique et le système vestibulaire qui contribue à la sensation de mouvement [\[42\]](#). En effet le sujet se déplace réellement dans l'espace et son point de vue correspond au point de vue de la caméra virtuelle projetant l'information de la scène dans le casque.

### **2.4.2 Scène**

Afin de comparer le comportement du sujet en réadaptation avec et sans l'utilisation de réalité virtuelle, nous avons synthétisé une scène virtuelle basée sur le laboratoire où sera effectuée l'expérience de traitement ([Figure 2.17](#)).



Figure 2.17 – Reconstruction numérique fidèle à la réalité d'une salle de laboratoire de réadaptation.

Une fois l'évaluation du système en environnement contrôlé terminé, il est possible de modifier l'environnement selon les besoins en affichant un milieu de travail typique ou un espace public par exemple. Nous avons utilisé les logiciels libres Krita pour les textures et Blender pour la modélisation en 3D. L'affichage de la scène dans le casque et en vue externe est basé sur la librairie OpenSceneGraph.

### 2.4.3 Effecteur du mécanisme à câbles

Pour permettre un retour visuel correspondant au retour haptique généré par le mécanisme à câbles, la position de l'effecteur réel et celle de l'effecteur virtuel observé par le sujet grâce au casque Oculus doivent correspondre. Nous connaissons la pose du robot selon notre système de coordonnées en passant par la matrice de calibration calculée. On affiche un objet choisi selon l'interaction physique désirée pour la tâche pour laquelle le sujet est évalué. Dans le cas présenté, deux poignées sont fixées au robot et représentent une boîte affichée dans l'espace virtuel ([Figure 2.18](#)).

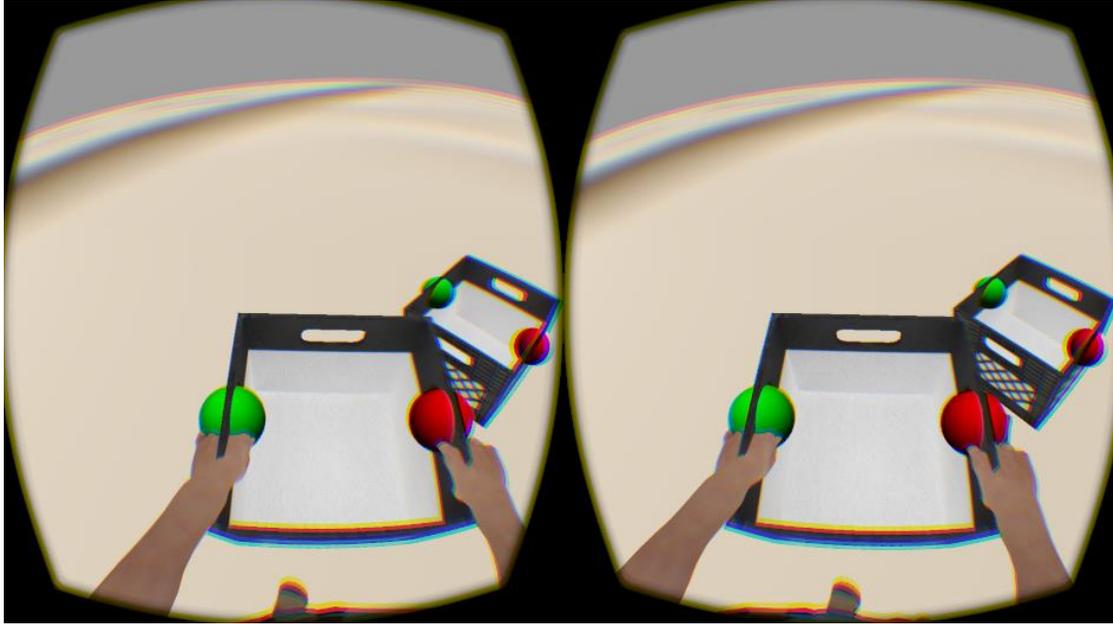
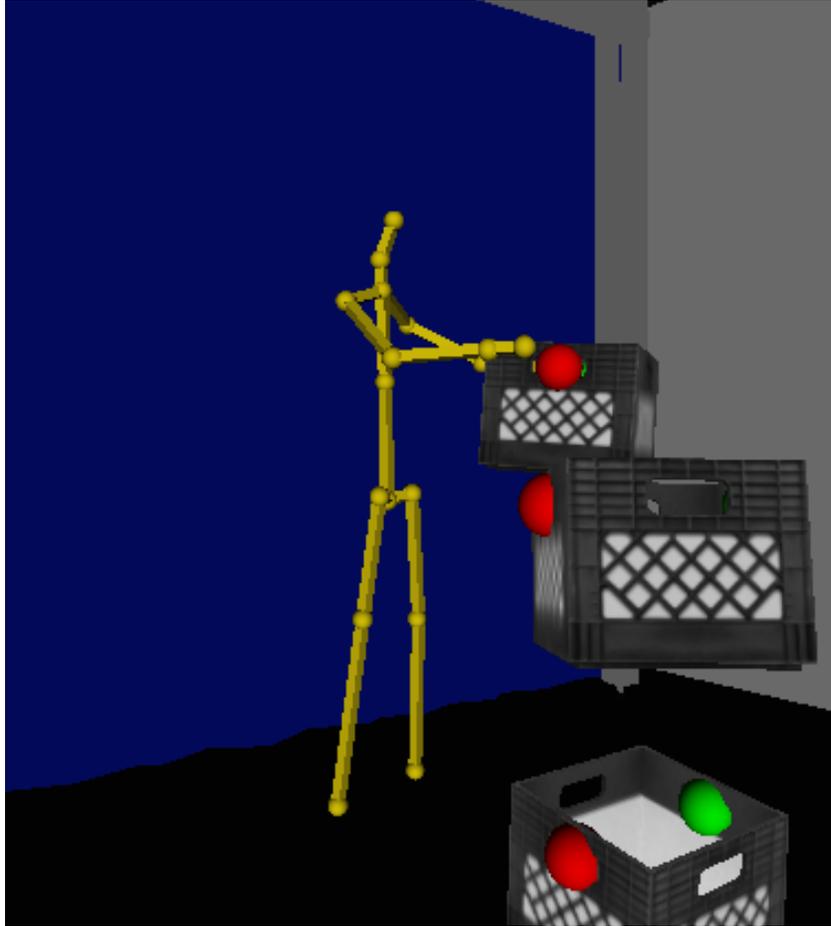


Figure 2.18 – Représentation simplifiée (sans scène) de l'interaction du sujet en environnement virtuel avec l'effecteur (caisse de lait). L'image est présentée selon le point de vue du sujet portant le casque de réalité virtuelle.

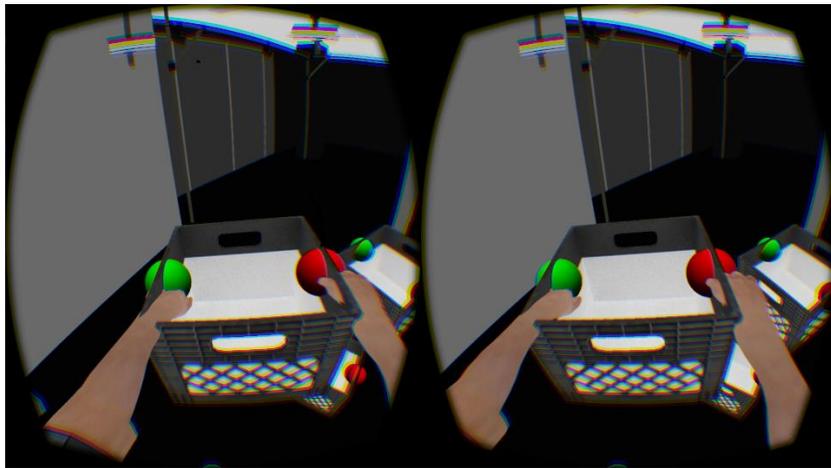
Lorsque l'utilisateur interagit avec l'effecteur du robot, ce dernier produit une résistance qui est modulée en fonction des paramètres voulus par l'expérimentateur (poids, friction, etc.). Sa position est mesurée par le système mécanique puis transmise au système d'affichage en réalité virtuelle en temps réel.

#### 2.4.4 Avatar du sujet

L'immersion dans l'environnement virtuel est essentielle pour que le sujet puisse effectuer ses tâches de manière naturelle comme il le ferait dans un environnement de travail ou de réadaptation réel. L'affichage de l'avatar (corps virtuel) en réalité virtuelle est important pour la proprioception, c'est-à-dire le modèle mental que le sujet se fait de ses membres en relation avec l'environnement ([Figure 2.19](#)). Ces membres virtuels doivent correspondre à leur équivalent réel pour permettre la concordance entre la perception visuelle et le système sensoriel [\[43\]](#).



(a)



(b)

Figure 2.19 – Affichage de l'avatar du sujet effectuant une tâche de manutention. (a) Affichage du squelette de l'avatar sur le moniteur du PC de contrôle et d'analyse de l'expérience. (b) Affichage de l'avatar dans le casque de réalité virtuelle porté par le sujet effectuant une tâche.

Il est possible de faire fonctionner le processus de fusion des poses en temps réel (fréquence de 30Hz limitée par l'acquisition d'images des Kinects) en adaptant le filtre particulaire pour que tous les joints soient calculés en parallèle. L'estimation de la position de la tête de l'avatar est utilisée pour déplacer le champ de vision du sujet. La position du champ de vision est légèrement filtrée (moyenne mobile) pour éviter un tremblement dans l'affichage. Par contre, la pose de la tête estimée par le processus de fusion (Section [2.3](#)) est déjà assez lisse et ne doit pas être fortement filtrée, ce qui induirait un délai désagréable entre le mouvement de la tête et le changement de point de vue. L'affichage de la position des autres joints et segments offre une perception visuelle au sujet permettant une interaction naturelle avec l'effecteur. La modélisation de l'avatar texturé a été effectuée en utilisant le logiciel Blender avec le module d'exportation vers le format OSG utilisable par la librairie OpenSceneGraph.

# Chapitre 3

## Résultats expérimentaux

Cette section présente et analyse les résultats d'estimation de poses fixes et dynamiques obtenues avec le système développé, voir la section [3.3.1](#) détaillant les exercices effectués lors de la prise de données pour l'analyse des résultats expérimentaux. Les données analysées sont capturées selon un montage de caméras détaillé à la section suivante. Les données sont traitées et analysées en post-traitement en utilisant des paramètres identiques pour chaque expérience.

Nous avons choisi le système de capture Vicon (section [1.1.3](#)) basé sur l'identification et la mesure de position de marqueurs sphériques rétro réfléchissants à l'infrarouge disposés sur le sujet comme mesure de vérité terrain. La méthode de calibration des systèmes est présentée dans la section [3.2.1](#). Le traitement et la méthode d'analyse des données sont faits de manière à démontrer le niveau de précision biométrique de l'algorithme d'estimation de pose choisi ainsi que la précision et la fiabilité de l'estimation de la position des joints et de l'angle des segments calculés par la méthode de fusion de données proposée.

### 3.1 Spécifications du montage

Afin d'évaluer la qualité de l'estimation de pose du système développé, nous avons retiré les facteurs pouvant influencer le résultat de la capture d'information tels que le mécanisme à câble et le casque de réalité virtuelle. Les caméras sont disposées de manière uniforme afin d'obtenir des données comparables entre les expériences, c'est-à-dire que la disposition des points de vue est pensée de manière générale et n'est pas optimisée en fonction de la tâche effectuée. Voir la [Figure 3.1](#) présentant un schéma du montage de prise de données expérimentales.

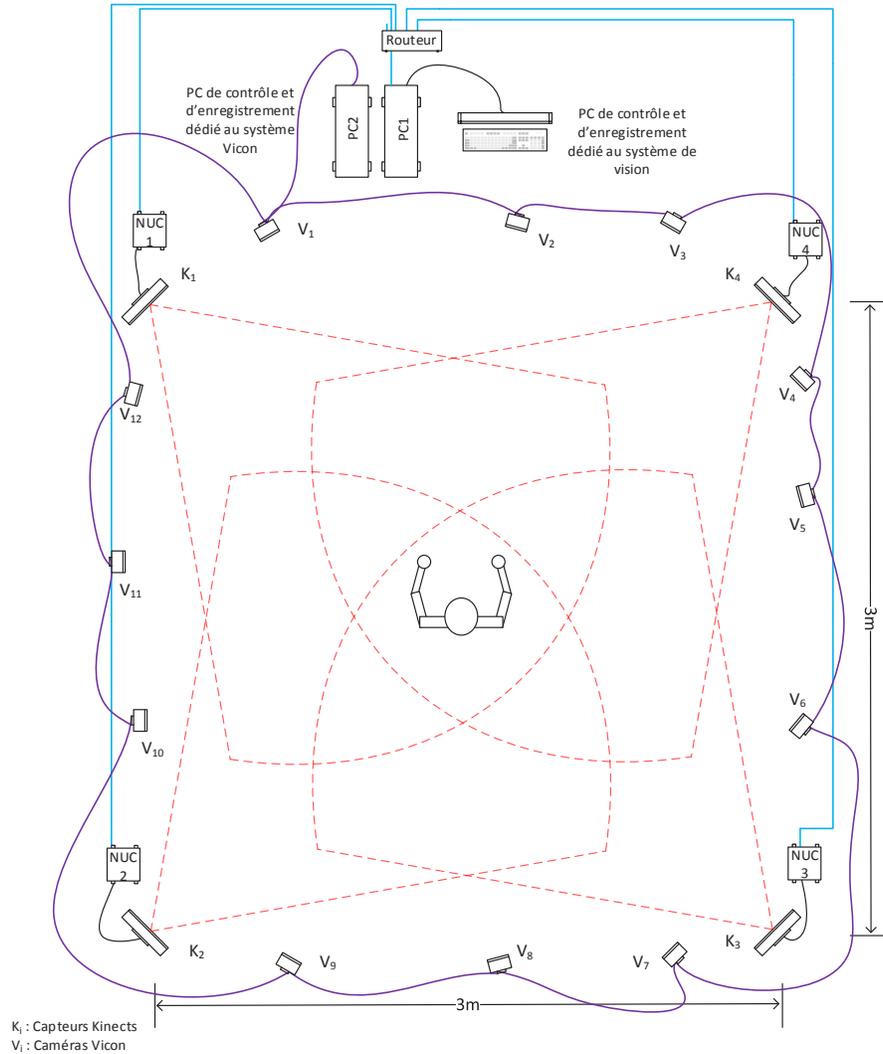


Figure 3.1 – Vue de dessus du schéma du montage expérimental utilisé lors de la prise des données.

Quatre capteurs Kinect sont disposés uniformément en périphérie de la zone de travail observée. Cette disposition forme un carré comportant des côtés de 3m et des points de vue horizontaux dirigés vers le centre de l'espace de travail. L'orientation des caméras est choisie de manière à obtenir un champ de vision vertical supérieur qui est parallèle au sol, offrant aussi une observation constante en hauteur et une limite d'observation au plancher qui est fixée par l'amplitude du champ de vision vertical (voir [Figure 3.2](#)). La hauteur des caméras fixées à 185mm est choisie en fonction de la grandeur du sujet observé. La disposition présente restreint l'observation au sol (membres inférieurs) en fonction de la hauteur et de la distance des caméras et donc de la grandeur du sujet.

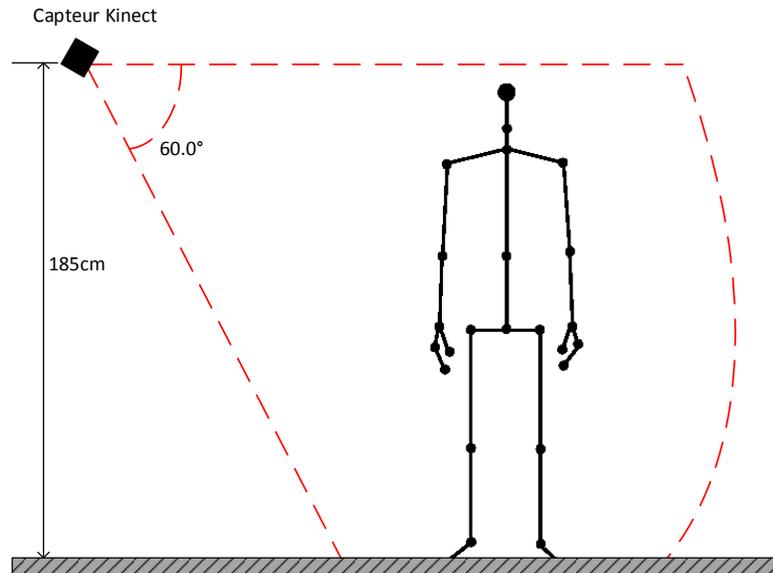


Figure 3.2 – Vue de côté du schéma du montage expérimental pour un capteur Kinect utilisé lors de la prise des données.

La méthode de fusion de l'information choisie (voir section [2.3](#) sur le filtrage particulière) est robuste aux occlusions et il n'est pas nécessaire que tous les joints du sujet soient observés selon tous les points de vue tout au long de l'expérience de suivi du sujet. La disposition des capteurs est aussi flexible et peut être optimisée en fonction de la tâche choisie (déplacements dans l'espace de travail, rotations du sujet, poses fixes, etc.). L'utilisateur peut donc sortir de la zone observée par un sous-ensemble des caméras sans empêcher l'estimation de la position des joints.

## 3.2 Matériel de validation

Le système choisi pour fournir les données de vérité terrain pour l'évaluation de l'estimation de pose du système proposé est le système de capture de mouvement Vicon. Sa précision et sa fiabilité en font une solution bien adaptée à l'analyse de nos résultats. Il permet de définir un modèle de squelette fidèle au sujet et de le suivre dans le temps.

Plusieurs études dans le domaine de la biomécanique mentionnent le Vicon comme choix de système de génération de données de vérité terrain, notamment concernant l'analyse de

l'estimation de la pose par des capteurs Kinect [51], [52], [53], [54], [55], [56], [57]. Le choix d'utiliser le Vicon comme modèle de validation et l'algorithme développé par Microsoft [20] fournissant le modèle initial à notre méthode de fusion permet de vérifier l'apport de notre système de suivi par rapport à ces nombreuses études en évitant beaucoup de facteurs d'erreur (Modèles de squelettes et système de mesure). Nous devons tout de même noter que la plupart de ces études sont basées sur l'utilisation de la première génération de capteurs Kinect. Depuis, l'algorithme de suivi et la qualité de l'information de distance ont été améliorés avec la version V2 utilisée dans notre projet. Voir la section 1.1.2 pour plus de détails sur la différence entre les deux générations de capteurs.

Un ensemble de caméras Vicon de type Bonita (voir Annexe A pour les spécifications des Bonita) et quatre capteurs de type Kinect V2 (Figure 3.3) ont été utilisés pour capturer le mouvement du sujet dans l'espace de travail.



Figure 3.3 – Capteurs utilisés pour l'acquisition de données. (a) Caméra Vicon Bonita [58]. (b) Capteur Kinect V2 de Microsoft [59].

### 3.2.1 Calibration du système de validation

Le système de capture de mouvement Vicon comprend un ensemble de caméras qui doivent être calibrées ensemble afin de permettre le suivi de marqueurs rétro réfléchissants dans l'infrarouge. Pour effectuer cette calibration, nous utilisons la méthode simple décrite par Vicon dans le guide d'utilisation et avec le matériel de calibration fourni.

Une fois que la calibration du système Vicon est satisfaisante, nous calibrons ce système avec notre système de vision. Pour effectuer cette calibration, nous devons associer de l'information vue par les deux systèmes. Le système Vicon est conçu pour l'identification et la mesure en position de marqueurs sphériques. Il est possible d'observer ces marqueurs avec notre système de vision mais comme la résolution des caméras du capteur Kinect n'est pas très grande ([Table 1.2](#)) une autre approche est utilisée.

Des marqueurs Vicon sont fixés sur un objet sphérique (ballon) vu par notre système de vision ([Figure 3.4](#)) et par le système Vicon. Tel que détaillé dans le chapitre précédent (section [2.2](#)), nous identifions le centre de la sphère observée dans le nuage de points aligné de chacun des quatre capteurs de profondeur. Nous effectuons la même opération d'identification du centre de la sphère sur les marqueurs observés par le Vicon. Comme la précision de ce système est grande, seulement quelques points (marqueurs) sont nécessaires afin de retrouver la sphère formée par ceux-ci. Le rayon des deux sphères observées est différent, mais le centre est théoriquement le même.



Figure 3.4 – Objet utilisé pour la calibration entre le système de vision développé et le système de suivi de marqueurs Vicon.

Lorsque le centre de la sphère est obtenu selon le référentiel de notre système de vision et du système Vicon, il est nécessaire de répéter l'opération au moins deux autres fois afin d'obtenir deux ensembles de trois points permettant d'optimiser une transformation contraignant les 6 degrés de liberté d'un corps rigide en 3 dimensions. Ceci dit, l'utilisation

d'un plus grand nombre de points (déplacement et mesure du centre de la sphère) répartis uniformément dans l'espace de travail est recommandée pour assurer une calibration fiable.

Pour obtenir la matrice de calibration exprimant la transformation rigide entre les deux systèmes de coordonnées nous minimisons l'erreur entre ces positions grâce à la méthode d'optimisation présentée dans le chapitre 2 section [1.2.3](#) selon l'équation [\(1.6\)](#).

### 3.3 Acquisition des données de suivi du squelette

Afin d'obtenir une mesure d'erreur valide, il est nécessaire de définir les modèles de squelettes à comparer (modèle fusionné par notre méthode et modèle utilisé pour la validation), c'est-à-dire la position relative au squelette (contrairement à la position absolue dans l'espace) de chacun des joints observés. Le modèle défini comme vérité terrain mesuré par le Vicon doit représenter le plus fidèlement possible la position biométrique des joints réels des membres de la personne. Les marqueurs Vicon sont fixés à des endroits déterminés par des experts en biomécanique, puis la position des joints réels est déduite selon la position et l'orientation de ces marqueurs. Cela nous permet de comparer notre système à un modèle fidèle à la pose humaine.

Il faut noter que le biais entre la position réelle des joints et la position mesurée des joints n'est pas nécessairement une conséquence de l'erreur de suivi causée par le système développé, mais dépend de la définition du modèle estimé par l'algorithme d'estimation de pose choisi et varie pour chaque modèle. Une méthode permettant de séparer cet offset de l'erreur de suivi de pose sera détaillée dans la section [3.4](#).

La position des joints de chacun des squelettes est enregistrée à la fréquence limitée par les capteurs Kinects de 30 Hz. Ces données sont associées à un temps ("timestamp") qui est synchronisé entre les quatre sous-systèmes NUC. L'étape de synchronisation est décrite à la section [2.3.3](#). Le logiciel Nexus du système de mesure Vicon est utilisé pour enregistrer les données à une fréquence de 100 Hz, mais n'est pas synchronisé avec le système de vision. Les données enregistrées par le système Vicon et celles fusionnées provenant du système développé sont alignées dans le temps selon la méthode présentée à la section suivante.

### **3.3.1 Détails des exercices effectués**

Pour la capture de ces résultats expérimentaux, seulement le bas du corps est mesuré par le système de validation. Le positionnement des marqueurs Vicon et le traitement des données de validation sont effectués grâce à l'expertise du Centre interdisciplinaire de recherche en réadaptation et intégration sociale (CIRIS). Au début de l'enregistrement de chaque exercice, le sujet soulève et dépose le pied 3 fois afin de faciliter la vérification de l'alignement temporel.

#### **Pose fixe**

Le sujet est debout, centré dans l'espace observé, les jambes tendues, le dos droit et les bras le long du corps. Aucun mouvement n'est effectué. Cet exercice sert pour la calibration des modèles utilisés lors de la validation ainsi que la calibration de la longueur des segments utilisés lors de la fusion en post-traitement.

#### **Marche**

Le sujet marche lentement en cercle dans l'espace observé autour du point central.

#### **Marche statique**

Le sujet marche sur place tout en conservant une position fixe au centre de l'espace observé. Il effectue donc des flexions simultanées de la hanche et du genou avec alternance entre les membres droit et gauche avec un mouvement de balancement des bras. Cet exercice pouvant se rapprocher de la marche sur tapis roulant est effectué afin d'évaluer le potentiel du système en estimation de pose sur tapis roulant.

#### **Flexion**

Le sujet effectue une simulation de prise d'un objet à environ 75 cm du sol, avec une flexion du dos. Une fois la position atteinte, elle est maintenue pendant 20 secondes.

#### **Manutention**

Le sujet effectue une simulation du mouvement de prise d'un objet à environ 75 cm du sol, avec flexion du dos et une légère flexion des genoux. Il effectue ensuite une simulation de dépôt de l'objet à environ 1m30 du sol face à lui. Le dos reste droit avec une flexion des épaules à 45

## **Manutention Plus**

Le sujet effectue une simulation du mouvement de prise d'un objet au sol, avec une grande flexion du dos et une légère flexion des genoux puis simule le dépôt de l'objet à environ 1m30 du sol face à lui. Le dos reste droit avec une flexion des épaules à 45°.

## **Rotation droite**

Le sujet effectue une simulation du mouvement de prise d'un objet à environ 75cm du sol à sa droite avec une flexion et une torsion du dos. Il effectue ensuite une simulation de dépôt de l'objet à environ 1m30 du sol face à lui.

## **Rotation gauche**

Le sujet effectue une simulation du mouvement de prise d'un objet à environ 75cm du sol à sa gauche avec une flexion et une torsion du dos. Il effectue ensuite une simulation de dépôt de l'objet à environ 1m30 du sol face à lui.

## **Squat**

Le sujet effectue une flexion des jambes, légèrement écartées vers l'extérieur, conservant le dos droit. Une fois la pose atteinte, elle est conservée lors de l'exercice.

## **Squat Plus**

Le sujet effectue une flexion des jambes d'amplitude significativement plus grande (plus de 150%) que pour l'exercice Squat. Il maintient les jambes légèrement écartées vers l'extérieur et le dos droit. Une fois la pose atteinte, elle est conservée lors de l'exercice.

## **3.4 Méthode de validation des données**

Afin de démontrer la performance du système de vision proposé, nous avons défini une méthode de validation basée sur la précision et la fidélité des résultats obtenus après traitement. Ce traitement comprend l'alignement global de la position des joints mesurés par les deux systèmes en utilisant la matrice de calibration obtenue selon la méthode présentée à la section [3.2.1](#), le ré-échantillonnage des données sur une fréquence identique nécessaire à la correspondance temporelle, la synchronisation temporelle ainsi que la calibration du modèle de squelette généré selon le modèle du squelette de validation.

## Interpolation et ré-échantillonnage

Afin de comparer les deux ensembles de données, nous devons identifier une correspondance pour chaque position des joints dans le temps. La fréquence d'acquisition du système Vicon a été fixée à 100Hz concernant les données présentées. La fréquence de fusion (nombre d'itérations de filtrage par seconde) n'est pas nécessairement faite à la même fréquence (nous avons utilisé 30Hz pour cette prise de données). Pour augmenter le nombre de données résultant de la fusion à une fréquence égale à celle de l'acquisition des données de validation obtenue par le système Vicon, nous avons choisi la méthode d'interpolation par spline cubique puis de ré-échantillonner celle-ci à intervalles de 10 millisecondes correspondant à la fréquence de 100Hz du Vicon.

Cette opération est appliquée sur chaque joint fusionné ainsi que sur chaque joint vu individuellement par les capteurs Kinects, ces données permettront de quantifier la robustesse à l'occlusion et l'effet de la fusion sur la précision. Un exemple de résultat après interpolation et ré-échantillonnage est présenté à la [Figure 3.5](#). L'effet observé est un étirement de la courbe de données fusionnées pour l'adapter à la fréquence d'échantillonnage du système de validation (Vicon).

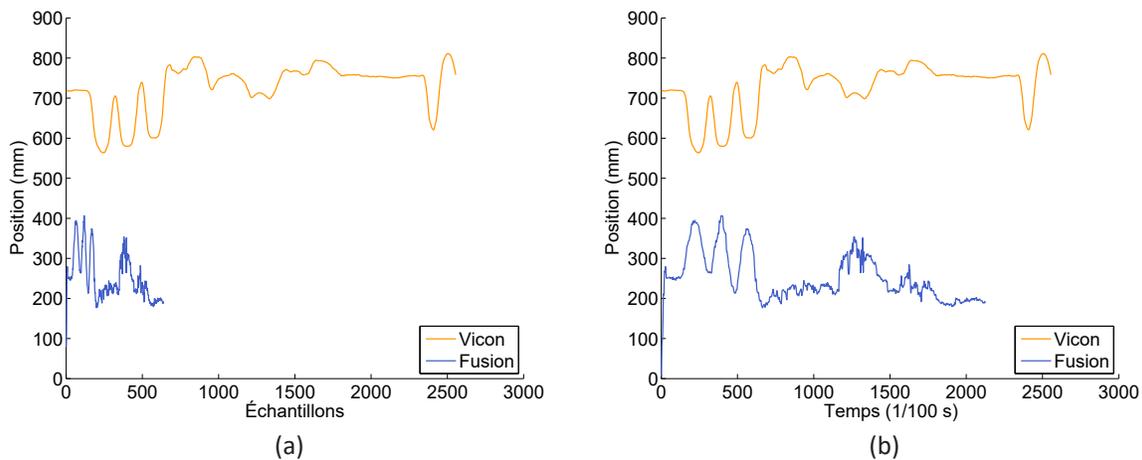


Figure 3.5 – Effet de l'interpolation et du ré-échantillonnage sur une courbe de mesures dans le temps. Position d'un joint avant (a) et après (b) ré-échantillonnage. Notez que cette opération est effectuée avant l'alignement spatial, l'alignement temporel et la correction du biais de modèle. Mesures de position sur l'axe X global du joint de la hanche droite pour l'exercice manutention.

### 3.4.1 Alignement global

L'alignement global est essentiel à la comparaison des systèmes, car il permet de représenter les données dans un même repère, celui-ci étant le référentiel du système Vicon fixé lors de sa calibration. Appliquer cette transformation rigide est simple, il suffit de multiplier les positions des joints vus par le système de vision par la matrice de calibration selon l'équation suivante:

$$\mathbf{F}_M = \begin{bmatrix} f_{m_{x1}} & f_{m_{x2}} & \dots & f_{m_{xn}} \\ f_{m_{y1}} & f_{m_{y2}} & \dots & f_{m_{yn}} \\ f_{m_{z1}} & f_{m_{z2}} & \dots & f_{m_{zn}} \\ 1 & 1 & \dots & 1 \end{bmatrix}_{4 \times N} = \mathbf{C}_{4 \times 4} \mathbf{F}_{K_{4 \times N}} \quad (3.1)$$

Ces opérations sont effectuées en coordonnées homogènes afin de faciliter les opérations mathématiques. La matrice  $\mathbf{C}$  de taille  $4 \times 4$  est celle obtenue par la calibration détaillée en section 3.2.1. Les vecteurs  $f_m$  contenus dans la matrice  $\mathbf{F}_K$  sont les positions d'un joint dans le temps vu par le système de vision développé. La matrice  $\mathbf{F}_M$  résultante contient les positions du même joint dans le temps, mais aligné au repère du système de validation Vicon. Encore une fois, cet alignement est effectué sur chaque joint pour les données initiales (avant la fusion) et fusionnées. Un exemple de cet alignement entre les données fusionnées et le système de validation Vicon est présenté à la [Figure 3.6](#).

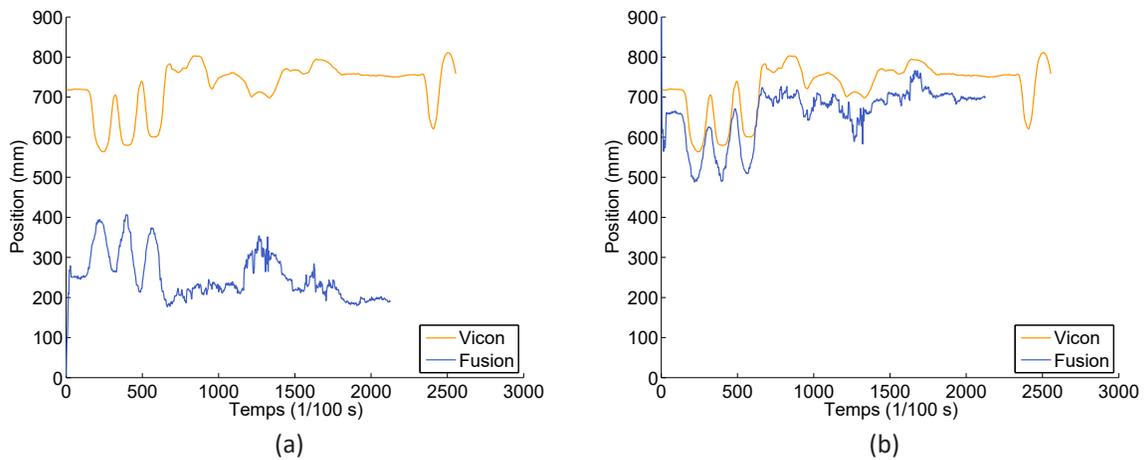


Figure 3.6 – Effet de l'alignement global sur une courbe de mesures dans le temps. Position d'un joint avant (a) et après (b) alignement global. Notez que cette opération est effectuée avant l'alignement temporel et la correction du biais de modèle. Mesures de position du joint de la hanche droite sur l'axe global X pour l'exercice manutention.

On conserve toujours un biais temporel ainsi qu'un biais dans l'espace causé par les différences dans les modèles de squelettes comparés. La méthode de correction de ces deux biais sera présentée plus loin.

### **3.4.2 Alignement temporel**

Bien que nous ayons synchronisé les sous-systèmes NUC entre eux (voir la section [2.3.3](#) pour plus de détails) pour permettre la fusion, la différence temporelle de notre système par rapport au système de validation Vicon n'est pas connue. Comme nous avons effectué un ré-échantillonnage des données obtenues par notre système, ce désalignement est expliqué par une simple translation dans la dimension du temps.

Cette translation étant limitée entre le temps initial et final d'enregistrement, nous avons donc choisi d'aligner temporellement les deux courbes selon une approche simple dite de "force brute". On calcule la distance euclidienne moyenne entre les points des deux courbes (fusion et validation) pour chacune des positions possibles dans le temps puis on identifie la distance minimale. La translation associée à cette distance minimale correspond au délai dans le temps entre les deux ensembles de données. On l'utilise donc pour décaler l'une des deux courbes et donc l'aligner temporellement comme il est possible d'observer sur la [Figure 3.7](#).

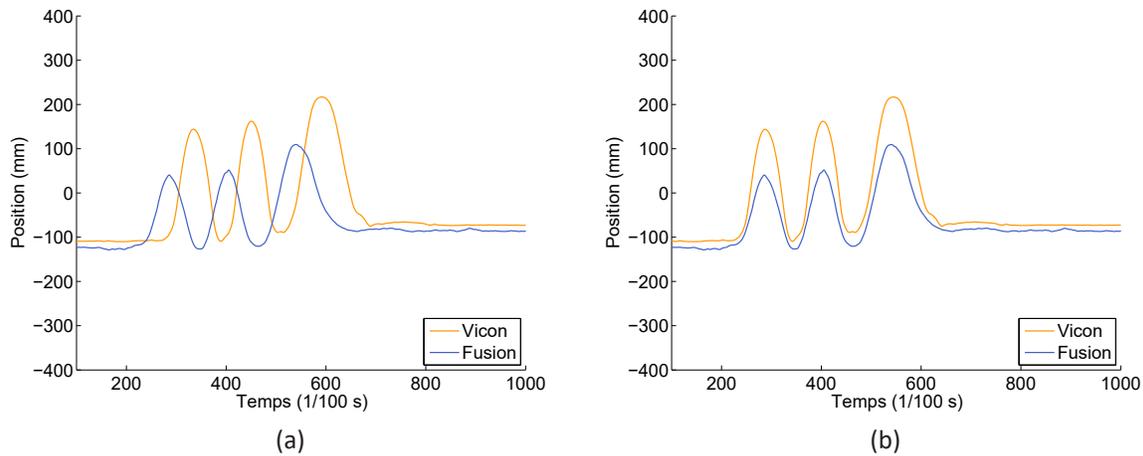


Figure 3.7 – Effet de l'alignement temporel sur une courbe de mesures dans le temps. Position d'un joint avant (a) et après (b) alignement temporel. Notez qu'au début de l'exercice, le sujet devait lever et redescendre trois fois la jambe. Cette procédure n'est pas nécessaire à l'alignement mais permet de bien visualiser le résultat. Mesures de position du joint du genou droit sur l'axe global Y pour l'exercice pose fixe.

Lorsque la correspondance temporelle des courbes est connue selon la translation identifiée, les sections qui ne se recouvrent pas sont éliminées. Nous coupons aussi les 2 premières et dernières secondes de chaque prise de données, cela permet d'éliminer les mouvements indésirables au début et à la fin de l'exercice en plus de permettre une bonne stabilité dans la convergence du filtre particulaire qui est initialisé en position zéro tel que mentionné à la section [2.3.4](#).

### 3.4.3 Alignement des modèles de squelette

Un alignement des modèles de squelette permet de séparer l'erreur de suivi des joints et le biais de position du joint défini par le modèle utilisé (Annexe C) par rapport au modèle de validation (voir [Figure 3.8](#)). Pour notre système, nous avons utilisé le squelette de l'algorithme d'estimation de pose présenté par Microsoft [\[20\]](#) pour différentes raisons énumérées plus tôt (sections [2.3.2](#) et [3.2](#)). Par contre, la méthode de fusion de données n'est pas contrainte par cet algorithme spécifique et l'utilisation d'un autre algorithme d'estimation de pose changerait le modèle reconstruit après la fusion.

Il n'est pas valide d'évaluer la qualité du suivi sans faire la différence entre l'erreur causée par le biais du modèle et l'erreur de suivi de la position d'un joint dans le temps. Le but de l'étape de validation est d'évaluer la qualité du suivi, mais il sera aussi possible d'évaluer les qualités de l'estimation biométrique du modèle reconstruit à partir de l'algorithme utilisé. Nous avons choisi un modèle de validation se rapprochant le plus possible du modèle biométrique. Par le fait même, il nous sera possible de qualifier le modèle utilisé tout en corrigeant localement la position des joints estimés par notre système de vision selon la position du modèle de validation grâce à une étape de calibration de modèle. La [Figure 3.8](#) démontre bien que cette étape est cruciale à l'évaluation du système. Sans cette correction, on induirait un biais local constant (qu'il ne serait pas possible de corriger globalement) dans notre évaluation de la qualité du suivi. Ce biais peut être vu comme une erreur systématique causée par la différence dans la définition de la position des joints sur un modèle de squelette.

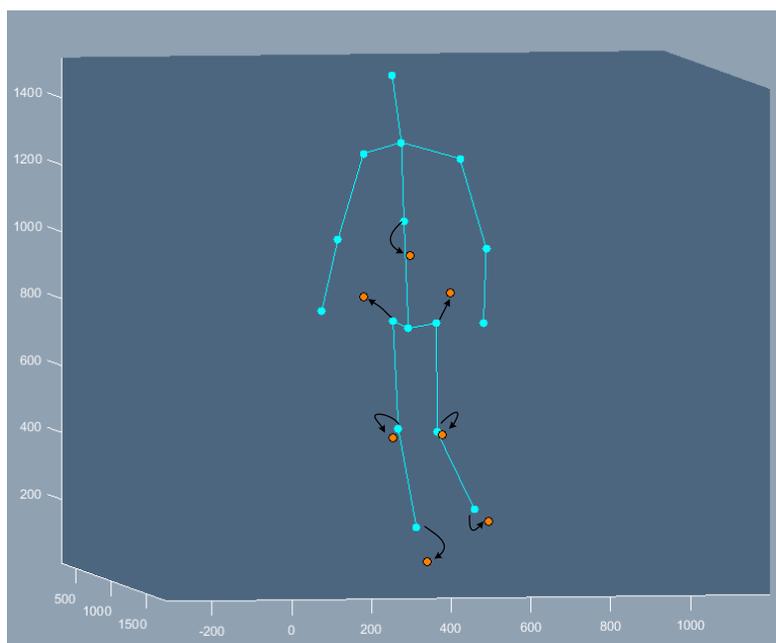


Figure 3.8 – Visualisation des deux modèles utilisés. En bleu, le modèle du squelette après fusion des données par notre système de vision. En orange, le modèle (bas du corps) de validation par le système Vicon.

La méthode suivante permet de retrouver les différences locales sur chaque segment puis de modifier le modèle utilisé pour notre système (Annexe C) afin de le projeter sur le modèle biométrique obtenu après traitement des données de validation. Il est nécessaire de mentionner que cette correction de modèle est seulement effectuée à des fins de validation et ne peut être appliquée sans un modèle de validation, car l'orientation précise des segments

doit être connue, ce qui n'est pas le cas pour notre système. L'information obtenue selon cette méthode pourra tout de même être utilisée pour l'analyse et l'amélioration du modèle initial utilisé (Algorithme d'estimation de pose [\[20\]](#)). Cette méthode pourra aussi servir de comparatif entre les différents algorithmes d'estimation de pose humaine.

Afin de corriger localement la position d'un joint sur un segment, il est nécessaire de calculer sa position sur le segment par rapport au joint réel. Ce biais (translation selon la position réelle du joint) n'est pas constant dans le référentiel global car l'erreur systématique est fonction de l'orientation du segment. Nous devons donc utiliser un référentiel local sur le joint pour évaluer cette valeur. Les marqueurs Vicon fixés sur les segments observés nous permettent d'obtenir l'information nécessaire au calcul du biais puis d'appliquer la correction sur les données traitées selon notre méthode. La [Figure 3.9](#) permet de visualiser les transformations nécessaires à cet alignement local détaillé par la suite.

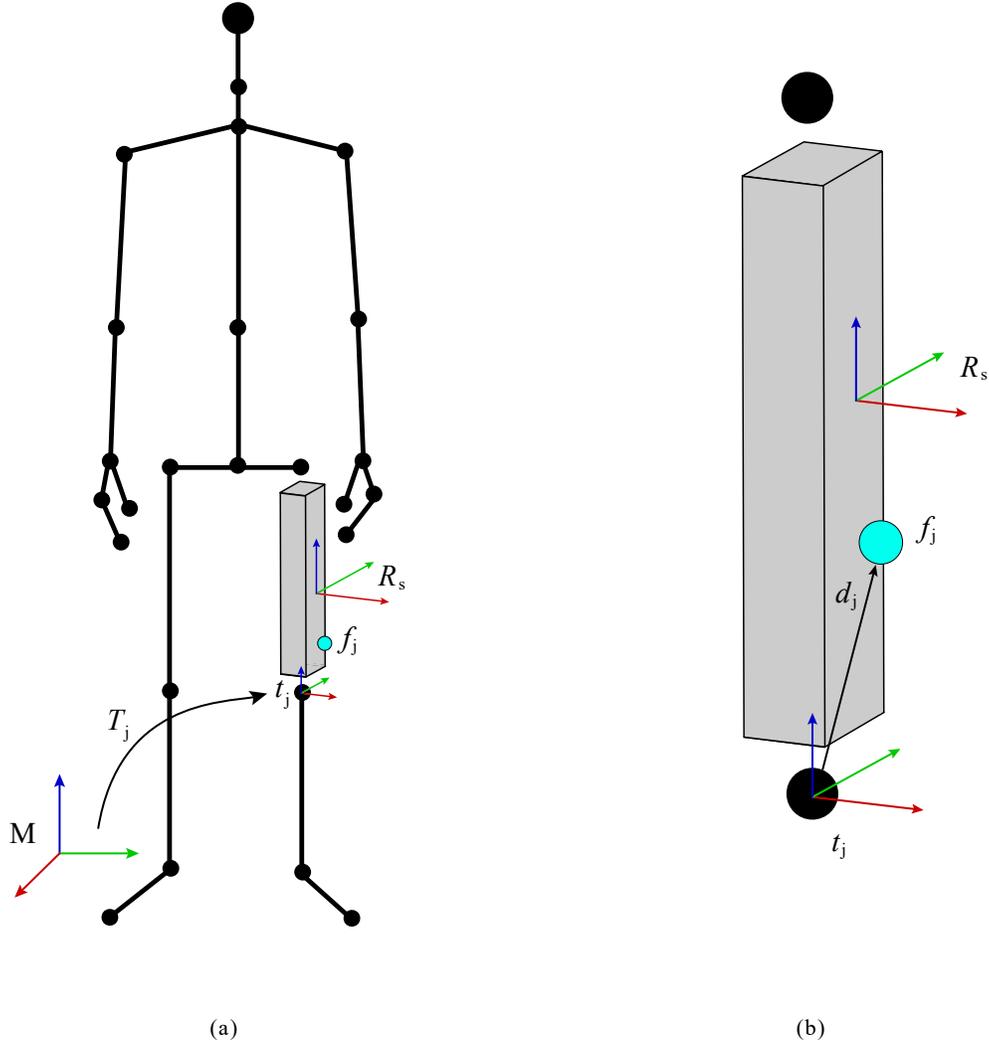


Figure 3.9 – Transformations entre les repères global  $\mathbf{M}$  et local  $\mathbf{Rt}_j$ . (a) Transformation  $\mathbf{T}_j$  permettant de calculer le joint fusionné  $\mathbf{f}_j$  dans le repère du joint de validation  $\mathbf{t}_j$ . (b) Agrandissement du segment et représentation du biais par le vecteur  $\mathbf{d}_j$  selon la position  $\mathbf{t}_j$  du joint de validation et l'orientation  $\mathbf{R}_s$  du segment de validation associé.

La première étape de cette correction de modèle peut être considérée comme une calibration locale des joints ou, formulé différemment, une calibration des modèles. Nous devons calculer la différence de position entre notre modèle et le modèle de validation pour chaque joint en utilisant un référentiel local. Ce référentiel est celui du segment associé au joint mesuré, il est obtenu selon les trois marqueurs rétro réfléchissants (triade) fixés sur le sujet et suivis par le système Vicon. Comme la triade est fixée de manière rigide sur le segment, il est possible de déduire l'orientation du segment associé en fonction de la position des joints adjacents et de certaines hypothèses biométriques. Pour obtenir la matrice de pose locale  $\mathbf{T}$

(Figure 3.9) de chaque joint  $\mathbf{j}$  pour chaque mesure  $\mathbf{n}$ , nous associons la position du joint  $\mathbf{t}_j$  à l'orientation du segment  $\mathbf{R}_s$  correspondant comme le démontre l'équation suivante:

$$\mathbf{T}_{\mathbf{j},\mathbf{n}} = [\mathbf{R}_{s,\mathbf{n}} | \mathbf{t}_{j,\mathbf{n}}] = \begin{bmatrix} r_{11s,\mathbf{n}} & r_{12s,\mathbf{n}} & r_{13s,\mathbf{n}} & t_{x_{j,\mathbf{n}}} \\ r_{21s,\mathbf{n}} & r_{22s,\mathbf{n}} & r_{23s,\mathbf{n}} & t_{y_{j,\mathbf{n}}} \\ r_{31s,\mathbf{n}} & r_{32s,\mathbf{n}} & r_{33s,\mathbf{n}} & t_{z_{j,\mathbf{n}}} \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad (3.2)$$

Ensuite, nous devons retrouver la position du joint fusionné vu par notre système de vision par rapport au référentiel local. Comme le système de validation Vicon est calibré avec le nôtre, nous avons déjà les données de position de tous les joints selon un référentiel global. L'équation suivante nous permet de retrouver la position de chaque joint fusionné  $\mathbf{F}_L$  vue par le référentiel local  $\mathbf{V}_j$  à partir du même joint  $\mathbf{F}_M$  vu par le référentiel global  $\mathbf{M}$ .

$$\mathbf{F}_{Lj,\mathbf{n}} = \mathbf{T}_{j,\mathbf{n}}^{-1} \mathbf{F}_{Mj,\mathbf{n}} = \begin{bmatrix} r_{11s,\mathbf{n}} & r_{12s,\mathbf{n}} & r_{13s,\mathbf{n}} & t_{x_{j,\mathbf{n}}} \\ r_{21s,\mathbf{n}} & r_{22s,\mathbf{n}} & r_{23s,\mathbf{n}} & t_{y_{j,\mathbf{n}}} \\ r_{31s,\mathbf{n}} & r_{32s,\mathbf{n}} & r_{33s,\mathbf{n}} & t_{z_{j,\mathbf{n}}} \\ 0 & 0 & 0 & 1 \end{bmatrix}^{-1} \begin{bmatrix} fm_{x_{j,\mathbf{n}}} \\ fm_{y_{j,\mathbf{n}}} \\ fm_{z_{j,\mathbf{n}}} \\ 1 \end{bmatrix} \quad (3.3)$$

Cette position locale  $\mathbf{F}_L$  représente le joint du modèle fusionné par le système développé selon le repère fixé sur le joint du modèle de validation  $\mathbf{V}_j$ . Nous voulons retirer ce biais pour ne laisser que l'erreur de suivi du système développé. Pour cela, nous enregistrons le vecteur de translation représentant le biais lors d'un exercice de calibration où le sujet prend une pose fixe pendant quelques secondes au centre de l'espace observé puis nous utilisons ce biais constant pour projeter le modèle de squelette fusionné sur le modèle de validation. Le biais  $\mathbf{d}_j$  (Figure 3.9) est obtenu simplement en calculant la valeur moyenne des vecteurs de translation observés dans le temps lors de l'exercice de calibration. La fréquence d'acquisition des mesures  $\mathbf{n}$  est de 100Hz et la moyenne est individuellement calculé pour chaque joint  $\mathbf{j}$ .

$$\mathbf{d}_j = \sum_{\mathbf{n}=1}^{\mathbf{N}} \frac{\mathbf{F}_{Lj,\mathbf{n}}}{\mathbf{N}} \quad (3.4)$$

La deuxième étape de la correction du modèle est d'appliquer la transformation qui éliminera le biais mesuré lors de la première étape que nous avons nommée calibration locale des modèles. Nous calculons donc encore une fois la position du joint  $\mathbf{F}_L$  vue localement pour

chaque segment et pour chaque mesure.

$$\mathbf{F}_{L_{j,n}} = \mathbf{T}_{j,n}^{-1} \mathbf{F}_{M_{j,n}} \quad (3.5)$$

La correction de la position des joints du modèle fusionné par le biais calculé est ensuite effectuée en retirant cette dernière valeur.

$$\mathbf{F}_{LC_{j,n}} = \mathbf{F}_{L_{j,n}} - \mathbf{d}_j \quad (3.6)$$

L'opération suivante permet de finalement retrouver la position corrigée dans le repère global. Le repère local de chaque segment nous servant seulement à effectuer la correction mais l'erreur sera ensuite calculé dans l'espace global.

$$\mathbf{F}_{MC_{j,n}} = \mathbf{T}_{j,n} \mathbf{F}_{LC_{j,n}} \quad (3.7)$$

La [Figure 3.10](#) démontre l'effet de cet alignement local sur une courbe de position dans le temps selon une seule des trois dimensions vue selon le référentiel global.

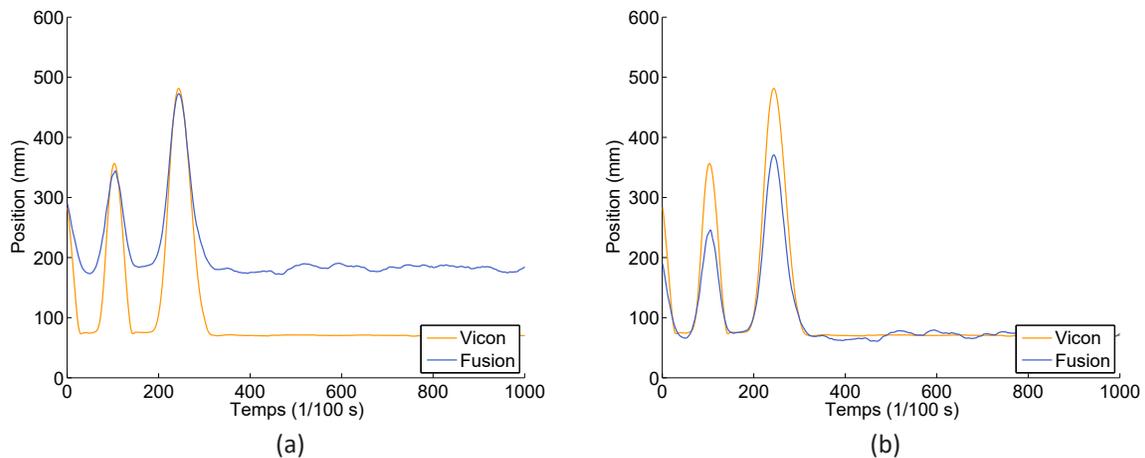


Figure 3.10 – Effet de l'alignement local d'un segment sur une courbe de mesures dans le temps. Position d'un joint avant (a) et après (b) alignement local. Le sujet devait lever et redescendre trois fois la jambe en début d'exercice. Mesures de position du joint de la cheville droite sur l'axe global Z pour l'exercice pose fixe.

Le filtrage particulière à tendance à rendre des résultats statistiquement probables mais bruités, nous effectuons donc un simple filtrage à moyenne mobile sur les données fusionnées.

### 3.5 Analyse des résultats expérimentaux

Les résultats présentés dans cette section servent de base d'analyse de la performance de notre système. Nous allons utiliser ces données pour vérifier dans quel cas le système performe bien et l'adapter aux exercices de réadaptation en réalité virtuelle avec retour haptique. Nous avons vérifié la qualité de mesure de la pose fournie en comparant les résultats avec ceux acquis par le système de validation. L'apport de la méthode de fusion est démontré en comparant les résultats avec les données initiales fournies au filtre particulière par les quatre capteurs Kinect utilisés.

La robustesse de la méthode de fusion aux données bruitées et erronées des capteurs est d'abord présentée par la robustesse de la mesure de position des joints et des angles des segments. Ensuite, nous démontrons la forte corrélation du suivi de la pose dans le temps en comparaison avec le système de validation. Puis le niveau de fidélité des résultats est présenté pour chacun des joints selon le pourcentage de mesures sous un seuil d'erreur en position et en angle. Finalement, nous présentons l'erreur angulaire de la pose des segments (jambes et dos) ainsi que l'erreur absolue dans le repère de coordonnées global pour la mesure de position de chacun des joints et chacun des exercices.

L'angle utilisé pour définir une jambe est celui contenu entre les trois joints de la cheville, du genou et de la hanche. L'angle du dos est défini par le vecteur contenu entre le bas et le centre de la colonne vertébrale par rapport au plan du sol. L'information est enregistrée lors des exercices présentés dans la section [3.3.1](#) et traités de manière hors ligne selon la démarche présentée à la section [3.4](#). Dans le but de faciliter l'analyse de la qualité des résultats pour différents types de mouvement, les exercices effectués sont courts et spécifiques, c'est-à-dire que le sujet effectue une tâche bien définie de manière répétitive (simulation de prise d'objet, marche, flexions, etc.). Pour ne pas biaiser les résultats en fonction de l'exercice, la position des 4 capteurs Kinects n'est pas optimisée, mais plutôt répartie de manière uniforme autour du centre de l'espace de travail. Les abréviations de la [Table 3.1](#) seront utilisées dans ce chapitre.

Table 3.1 – Abréviation des noms de joints et segments

CG	Cheville Gauche
GG	Genou Gauche
HG	Hanche Gauche
CD	Cheville Droite
GD	Genou Droit
HD	Hanche Droite
D	Dos
JG	Jambe Gauche
JD	Jambe Droite

### Robustesse de la méthode de fusion

L'approche de fusion de données de position de joint par filtrage particulaire utilisée est très robuste aux mauvaises données d'entrée du filtre. Ces données sont les observations des joints estimés selon l'algorithme d'estimation de pose pour chacun des quatre capteurs. Cette information initiale est parfois erronée, c'est pour pallier à ce problème que le filtrage particulaire est utilisé et permet de fournir une estimation de position réaliste. La [Figure 3.11](#) présente un exemple de la robustesse de l'approche proposée dans des conditions difficiles d'estimation de pose. On observe des informations en partie très faussées provenant des quatre capteurs Kinects, mais la sortie du filtre est stable et fournit une information réaliste et utilisable.

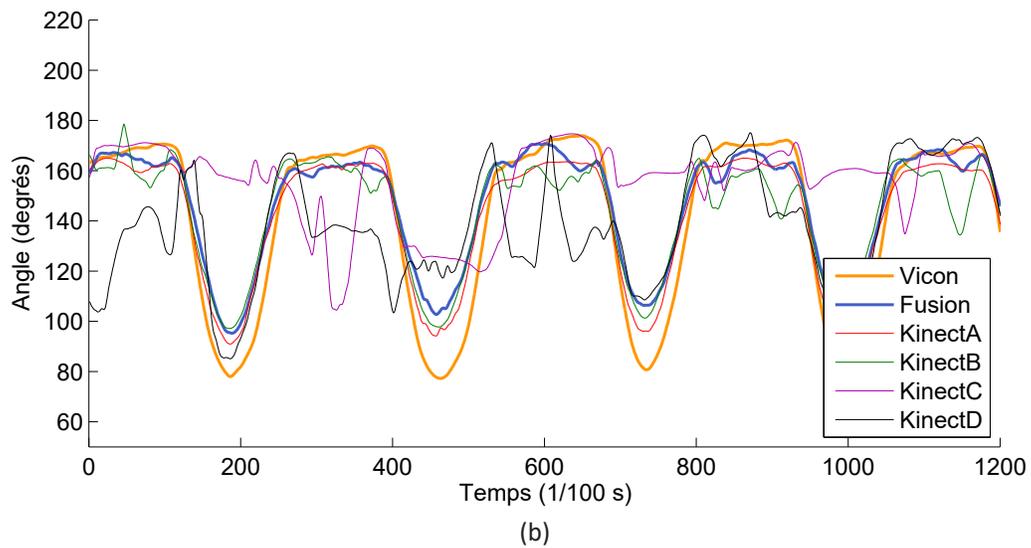
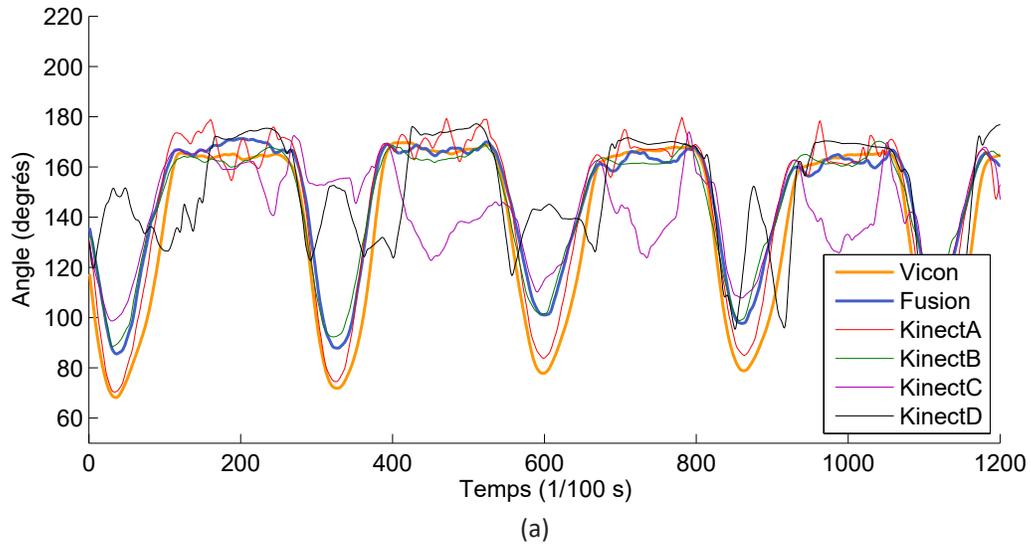


Figure 3.11 – Mesure de l'angle des jambes gauche (a) et droite (b). Les courbes orange et bleu sont respectivement les mesures observées par le système de validation et le système de vision proposé. Les autres courbes sont les mesures d'angles observés par les capteurs Kinects disposés autour du sujet. Mesures d'angle des jambes pour l'exercice marche statique.

Les valeurs erronées provenant des sous-systèmes sont en partie causées par la configuration des capteurs (autour du sujet) qui permet de généraliser le montage de manière à ne pas avantager un exercice plutôt qu'un autre, par exemple la marche statique plutôt que dynamique. Cette configuration nous permet aussi de prendre en compte l'apport en stabilité de la méthode de fusion.

La [Figure 3.12](#) présente elle aussi la robustesse de la méthode de fusion proposée, mais sur la mesure de position d'un joint plutôt que de la mesure de l'angle. Le joint observé était en position stable lors de la section de l'exercice présenté. Malgré cela, chacun des capteurs a fourni des informations erronées qui ont été filtrées comme désiré.

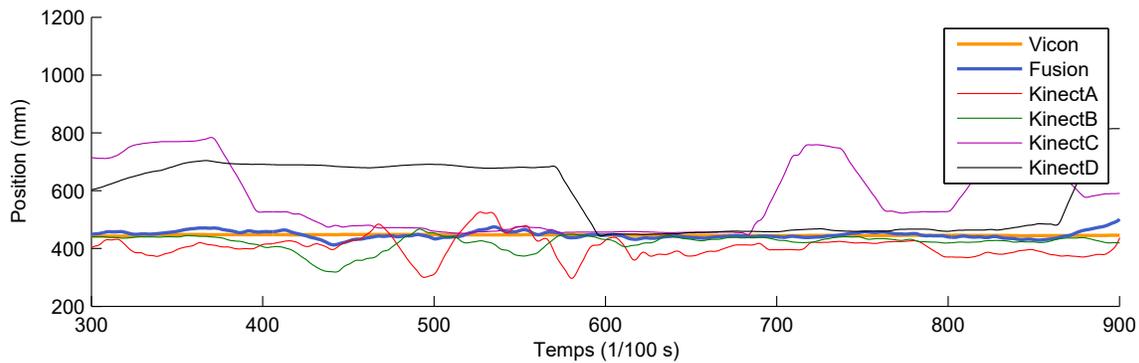


Figure 3.12 – Mesure de la position d'un joint dans le temps. Les courbes orange et bleu sont respectivement les mesures observées par le système de validation et le système de vision proposé. Les autres courbes sont les mesures observées par les capteurs Kinects disposés autour du sujet. Mesures de position du joint de la cheville gauche sur l'axe global X pour l'exercice manutention.

Les tables suivantes ([Table 3.2](#), [Table 3.3](#) et [Table 3.4](#)) présentent les erreurs angulaires moyennes des segments observés pour les exercices de *Marche*, de *Marche statique* et pour l'ensemble des exercices. À des fins de comparaison entre les données initiales provenant individuellement des sous-systèmes, les résultats moyens des quatre Kinects (KA, KB, KC, KD) sont présentés au côté des résultats de l'approche proposée (Fusion). Voir la [Table 3.12](#) et la [Table 3.13](#) pour les résultats de l'approche proposée par exercice.

Table 3.2 – Erreurs angulaires moyennes sur l'exercice *Marche* (degrés)

	JD	JG	D
KA	14,21	14,68	1,58
KB	12,72	12,11	2,54
KC	18,67	12,53	1,82
KD	12,27	13,89	3,31
<b>Fusion</b>	<b>9,56</b>	<b>10,93</b>	<b>1,95</b>

Table 3.3 – Erreurs angulaires moyennes sur l'exercice *Marche statique* (degrés)

	JD	JG	D
KA	8,54	8,60	2,05
KB	11,16	13,16	5,47
KC	24,47	24,46	2,10
KD	20,08	21,78	3,11
<b>Fusion</b>	<b>11,00</b>	<b>10,18</b>	<b>2,02</b>

Table 3.4 – Erreur angulaire sur l'ensemble des exercices (degrés)

	Moyenne			Écart-type		
	JD	JG	D	JD	JG	D
Kinects	20,93	17,24	6,27	11,65	8,96	3,78
<b>Fusion</b>	<b>10,68</b>	<b>10,51</b>	<b>5,91</b>	<b>4,22</b>	<b>3,58</b>	<b>3,42</b>

Bien que le résultat obtenu par filtrage particulière soit dépendant de la qualité des données en entrée, la méthode présente une robustesse aux données erronées qui nous permet d'obtenir des résultats supérieurs aux sous-systèmes individuels et indépendants de l'orientation du sujet (voir les résultats des exercices *Marche* ([Table 3.2](#)) et *Marche statique* ([Table 3.3](#))).

## Corrélation des résultats de fusion et de validation

Nous avons calculé le coefficient de corrélation de Pearson ( $r$ ) qui repose sur le calcul de la covariance entre deux ensembles de données correspondants. Dans le cas présenté, ce coefficient est calculé pour les mesures d'angle des segments dans le temps entre les données de validation et les données de l'approche proposée.

Un coefficient de 0 ne présente aucune corrélation et un coefficient absolu de 1 présente une corrélation linéaire parfaite. Un coefficient plus petit que 0 est interprété comme une corrélation négative et un coefficient plus grand que 0 est interprété comme une corrélation

positive. Il est possible d'interpréter les valeurs absolues de ce coefficient en se basant sur le guide suivant proposé par Evans JD [61].

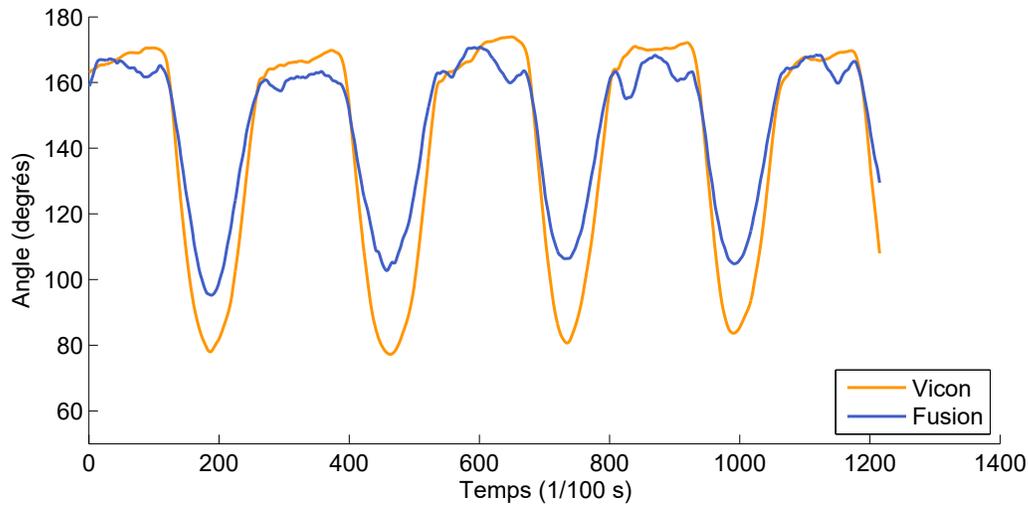
- $0.00 \leq r \leq 0.19$  : Corrélation très faible
- $0.20 \leq r \leq 0.39$  : Corrélation faible
- $0.40 \leq r \leq 0.59$  : Corrélation modérée
- $0.60 \leq r \leq 0.79$  : Corrélation forte
- $0.80 \leq r \leq 1.00$  : Corrélation très forte

La [Table 3.5](#) présente les résultats de corrélation entre les données fusionnées par notre méthode et les données de validation pour les angles des segments observés sur différents exercices. La corrélation sur les exercices dont l'angle des segments ne varie pratiquement pas dans le temps (jambes fixes ou dos fixe), n'est pas calculée car elle représente le bruit de mesure plutôt que la qualité de suivi du mouvement. On constate que la corrélation est très forte dans presque toutes les situations.

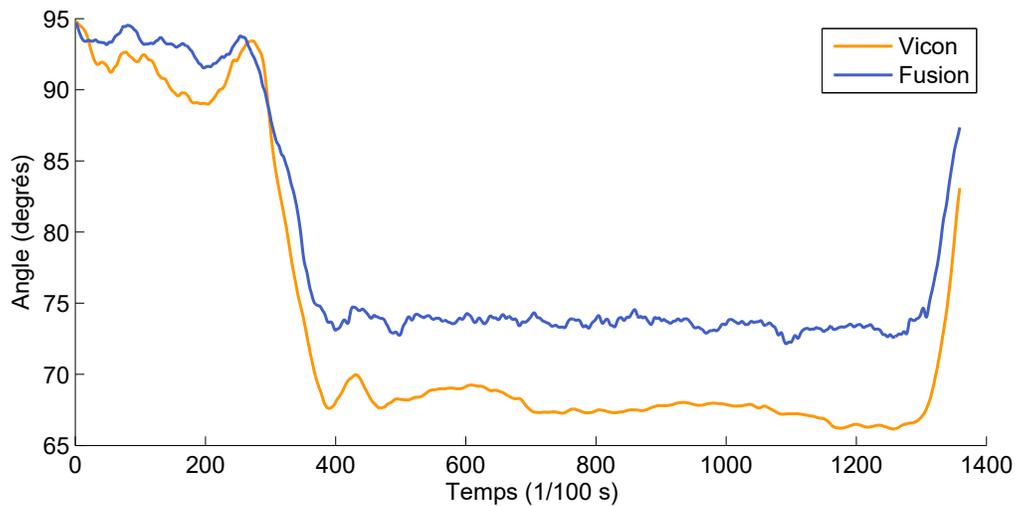
Table 3.5 – Corrélation des angles mesurés ( $r$ )

	JD	JG	D
Manutention	0,92	0,96	0,97
Manutention +	0,87	0,85	0,96
Marche	0,78	0,59	0,82
Marche Statique	0,98	0,98	0,88
Squat	0,93	0,90	0,99
Squat +	0,98	0,90	0,99
Moyenne	0,91	0,86	0,94

La [Figure 3.13](#) présente deux exemples de mesure d'angle dans le temps pour deux exercices différents. On peut observer une corrélation très forte entre les deux courbes malgré une amplitude biaisée par la définition des modèles de squelettes.



(a)



(b)

Figure 3.13 – Variation d'angle d'un segment dans le temps. Mesures d'angle de la jambe droite pour l'exercice marche statique (a) et du dos par rapport au plan du sol pour l'exercice squat (b).

La longueur des pas, le rythme et la direction du mouvement (changement de position et d'angle dans le temps) calculés selon la méthode proposée représentent très bien les données de validation. Le biais en mesure absolue observé varie selon la pose du sujet et est causé en partie par la définition de la position des joints de la méthode d'extraction de squelettes qui sera différente en fonction de l'algorithme utilisé. La corrélation plus faible (modérée à forte) observée sur la mesure de l'angle des jambes pour l'exercice *Marche* doit être étudiée plus en détail lors de prochaines prises de données, mais la vitesse de déplacement des joints plus élevée est une explication possible.

## Fiabilité des mesures de position et d'angle

Afin de représenter la fiabilité des données nous avons calculé le pourcentage de mesure présentant une erreur d'estimation de position de joint et d'angle de segment sous certains seuils. Ce résultat permet d'identifier la stabilité du suivi dans l'espace de travail. La position des joints estimés individuellement par les sous-systèmes selon l'algorithme d'estimation de pose utilisé [20], est souvent faussée par des occlusions ou des poses du sujets plus complexes à identifier.

Notre approche utilise la redondance de l'information ainsi que différents points de vue pour estimer la position la plus probable des joints selon les données fournies. Pour bien identifier l'apport de la méthode de fusion proposée, les résultats provenant de chacun des sous-systèmes (KA, KB, KC, KD) sont présentés en plus des résultats de l'approche proposée (Fusion). La [Table 3.6](#), la [Table 3.7](#), la [Table 3.8](#) et la [Table 3.9](#) présentent le pourcentage de mesure de position de joint ayant une erreur sous un seuil de 10 cm pour les données initiales et les données fusionnés selon différents exercices. Le seuil de 10 cm est fréquemment utilisé comme référence afin de représenter la validité de la détection des joints d'un algorithme d'estimation de pose [62], [63], [64]. La [Table 3.10](#) présente le pourcentage de mesure d'angle du dos ayant une erreur sous différents seuils pour plusieurs exercices.

Table 3.6 – Fiabilité en position pour l'exercice *Marche statique* selon un seuil de 10 cm

	CG	GG	HG	CD	GD	HD	D	Moyenne
KA	30,45%	97,86%	61,81%	67,41%	100,00%	100,00%	100,00%	79,65%
KB	66,26%	91,28%	100,00%	37,45%	100,00%	100,00%	98,93%	84,84%
KC	29,14%	79,09%	100,00%	45,02%	48,40%	100,00%	100,00%	71,66%
KD	69,38%	45,51%	95,97%	60,08%	81,23%	88,72%	100,00%	77,27%
<b>Fusion</b>	<b>93,17%</b>	<b>100,00%</b>	<b>100,00%</b>	<b>90,70%</b>	<b>100,00%</b>	<b>100,00%</b>	<b>100,00%</b>	<b>97,70%</b>

Table 3.7 – Fiabilité en position pour l'exercice *Manutention* selon un seuil de 10 cm

	CG	GG	HG	CD	GD	HD	D	Moyenne
KA	6,16%	28,26%	19,02%	0,59%	55,28%	66,03%	45,25%	31,51%
KB	33,57%	92,98%	84,59%	56,00%	83,80%	71,21%	55,87%	68,29%
KC	47,54%	85,44%	60,26%	23,34%	38,16%	77,77%	65,31%	56,83%
KD	23,80%	15,67%	88,79%	71,74%	79,15%	80,59%	56,92%	59,52%
<b>Fusion</b>	<b>84,85%</b>	<b>99,74%</b>	<b>84,79%</b>	<b>81,05%</b>	<b>91,93%</b>	<b>81,18%</b>	<b>61,70%</b>	<b>83,61%</b>

Table 3.8 – Fiabilité en position pour l'exercice *Squat* selon un seuil de 10 cm

	CG	GG	HG	CD	GD	HD	D	Moyenne
KA	0,74%	66,59%	0,00%	0,74%	68,87%	29,36%	18,69%	26,43%
KB	34,07%	94,19%	99,56%	42,31%	96,69%	71,08%	77,48%	73,63%
KC	97,57%	93,30%	12,07%	0,00%	4,64%	100,00%	27,37%	47,85%
KD	0,00%	0,00%	98,31%	89,85%	90,88%	21,41%	26,20%	46,66%
<b>Fusion</b>	<b>100,00%</b>	<b>100,00%</b>	<b>99,34%</b>	<b>91,61%</b>	<b>93,89%</b>	<b>100,00%</b>	<b>55,78%</b>	<b>91,52%</b>

Table 3.9 – Fiabilité en position pour l'exercice *Marche* selon un seuil de 10 cm

	CG	GG	HG	CD	GD	HD	D	Moyenne
KA	8,05%	45,23%	4,03%	6,04%	45,77%	4,70%	0,00%	16,26%
KB	34,09%	62,55%	63,36%	15,97%	65,77%	76,24%	97,99%	59,42%
KC	44,70%	48,86%	58,66%	48,46%	65,50%	100,00%	94,90%	65,87%
KD	66,44%	65,10%	82,68%	70,60%	66,31%	79,19%	93,15%	74,78%
<b>Fusion</b>	<b>69,66%</b>	<b>80,94%</b>	<b>100,00%</b>	<b>82,01%</b>	<b>90,60%</b>	<b>100,00%</b>	<b>100,00%</b>	<b>89,03%</b>

La [Table 3.10](#) permet de valider la fiabilité de la mesure d'angle du dos sous différents seuils ainsi que d'analyser l'influence des différents exercices sur la qualité du résultat.

Table 3.10 – Fiabilité en mesure d'angle du dos selon différents seuils

Seuils (degrés)	5.0°	7,5°	10.0°	15.0°
Marche Statique	98,93%	100,00%	100,00%	100,00%
Manutention	53,57%	72,33%	87,15%	95,54%
Squat	36,57%	99,04%	100,00%	100,00%
Marche	98,79%	100,00%	100,00%	100,00%

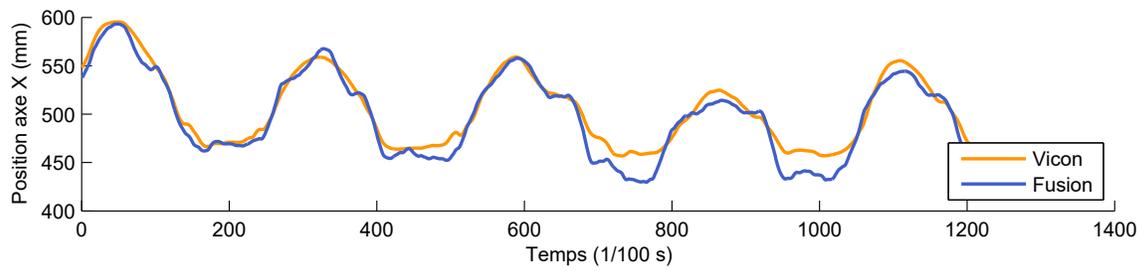
On observe une fiabilité des résultats de mesure de position et d'angle considérablement améliorée relativement à la fiabilité des données initiales des sous-systèmes individuels.

## Erreur angulaire et erreur de position

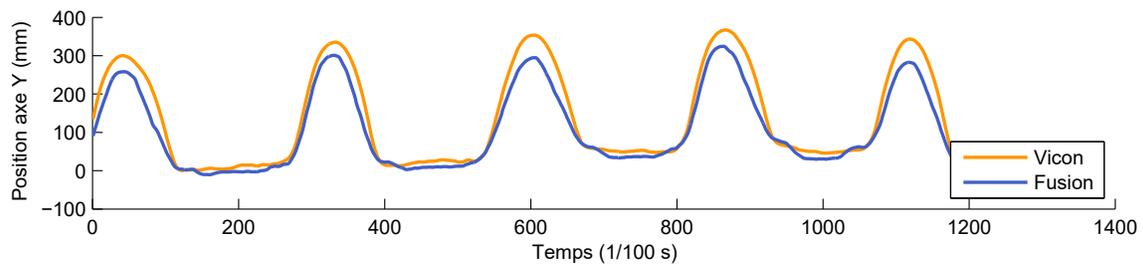
Les données suivantes représentent l'erreur en distance absolue (distance euclidienne) de suivi des joints entre les observations du système développé basé sur la fusion par filtre particulaire et les observations du système de validation Vicon. La [Table 3.11](#) contient les erreurs en distances moyennes absolues en cm pour chaque joint et pour chaque exercice. La [Figure 3.14](#) présente un exemple de suivi de position de joint dans le temps. Notez que les échelles sont ajustées individuellement pour chaque dimension afin de mieux percevoir les erreurs de mouvement dans l'espace.

Table 3.11 – Distance euclidienne entre les joints observés par les deux systèmes (cm)

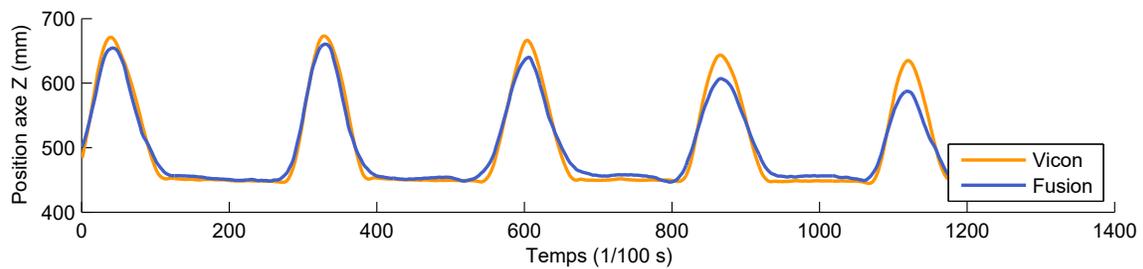
	CG	GG	HG	CD	GD	HD	D	Moyenne
Flexion	6,08	3,64	8,21	6,34	4,21	8,85	15,34	7,52
Manutention	6,13	4,33	5,98	7,39	4,80	6,51	9,31	6,35
Manutention +	11,30	6,32	9,16	6,55	5,37	7,06	9,67	7,92
Marche	10,28	6,29	4,12	8,33	5,48	3,77	4,04	6,04
Marche Statique	3,67	3,54	4,07	4,23	3,25	2,54	2,91	3,46
Pose Fixe	1,70	0,95	1,66	1,97	1,71	1,11	1,18	1,47
Rotation D	21,93	15,35	7,00	7,91	7,45	8,95	20,86	12,78
Rotation G	9,12	6,44	11,99	7,09	5,89	6,84	17,38	9,25
Squat	5,97	3,64	7,29	6,94	3,74	6,89	8,75	6,17
Squat +	15,06	6,76	16,28	22,25	19,29	15,36	13,81	15,54
Écart-type	5,95	3,85	4,22	5,39	4,89	3,97	6,50	4,97
Moyenne	9,12	5,73	7,57	7,90	6,12	6,79	10,33	7,65



(a)



(b)



(c)

Figure 3.14 – Suivi d'un joint en position dans le temps au cours d'un exercice. Mesures de position du joint du genou gauche sur l'axe global X, Y, et Z (dans le même ordre (a), (b) et (c)) pour l'exercice marche statique.

On observe une variation de l'erreur moyenne de position en fonction du joint observé et de l'exercice effectué. Les valeurs sont généralement sous les 10 cm et varient entre 1.47 cm et 12.78 cm selon les exercices et entre 6.12 cm et 10.33 cm selon les joints. La moyenne totale est de 7.65 cm. Ces erreurs sont en partie causées par la différence dans la définition de la position des joints malgré la correction des modèles qui est effectuée selon une pause fixe.

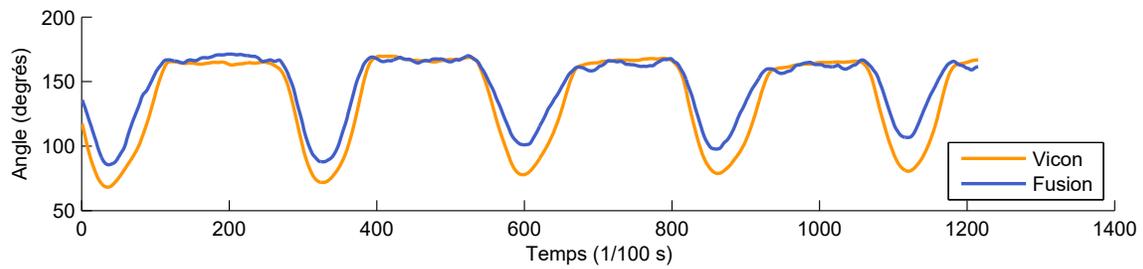
La [Table 3.12](#) et la [Table 3.13](#) présentent l'erreur d'angle absolue (distance angulaire) de suivi de la pose des jambes et du dos entre les observations du système développé et du système de validation. Ces données sont représentées selon la moyenne des erreurs par exercice et par segment. Afin d'évaluer les performances du système développé dans des conditions normales, les données angulaires mesurées et présentées sont avant la correction des modèles et présentent un certain biais dans le positionnement des joints sur le squelette. La [Figure 3.15](#) présente un exemple de suivi de l'angle des deux jambes dans le temps. La [Figure 3.16](#) présente un exemple de suivi de l'angle du dos dans le temps.

Table 3.12 – Erreur angulaire jambes (degrés)

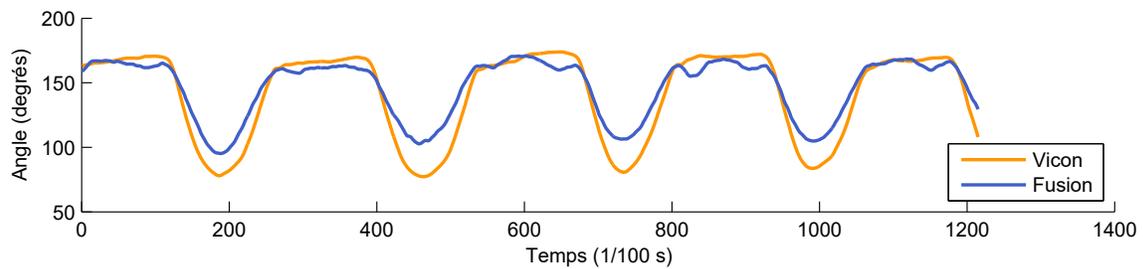
	JG	JD	Moyenne
Flexion	6,69	8,93	7,81
Manutention	11,81	10,00	10,90
Manutention +	16,73	13,73	15,23
Marche	9,56	10,93	10,24
Marche Statique	11,00	10,18	10,59
Pose Fixe	3,60	7,53	5,56
Rotation D	10,36	11,26	10,81
Rotation G	9,06	5,07	7,06
Squat	17,88	9,11	13,50
Squat +	10,15	18,32	14,24
Écart-type	4,22	3,58	3,90
Moyenne	10,68	10,51	10,59

Table 3.13 – Erreur angulaire dos (degrés)

Flexion	10,31
Manutention	5,11
Manutention +	5,50
Marche	1,95
Marche Statique	2,02
Pose Fixe	1,82
Rotation D	9,89
Rotation G	10,17
Squat	4,93
Squat +	7,43
Écart-type	3,42
Moyenne	5,91



(a)



(b)

Figure 3.15 – Suivi de l'angle des jambes dans le temps. Les deux graphes représentent respectivement la jambe gauche (a) et la jambe droite (b) observées selon les deux systèmes. Mesures d'angles des jambes pour l'exercice marche statique.

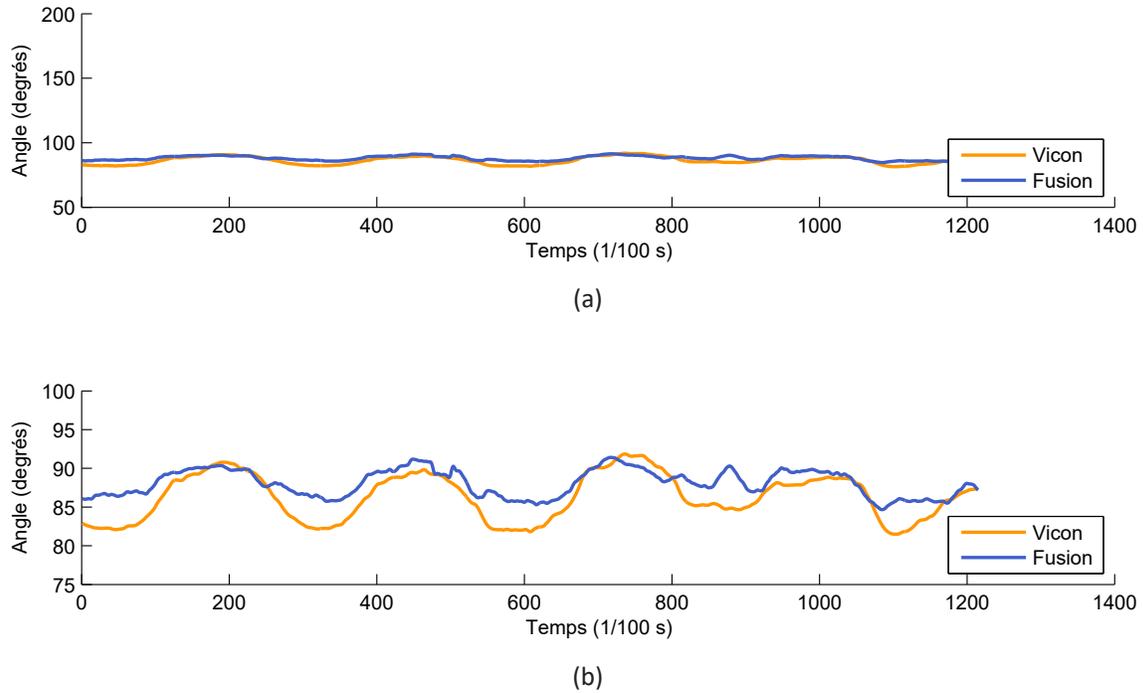


Figure 3.16 – Suivi de l'angle du dos dans le temps au cours d'un exercice. Les graphes représentent l'angle du dos observé individuellement par les deux systèmes selon la même échelle que la [Figure 3.15](#) (a) et selon une échelle étirée (b). Mesures d'angles du dos par rapport au sol pour l'exercice marche statique.

Les distances angulaires absolues sont directement dépendantes de la position des joints sur lesquels les vecteurs sont calculés. L'erreur varie en fonction du segment et l'angle du dos présente une erreur considérablement plus basse que celle associée aux jambes. Dans les deux cas, cette erreur varie en fonction de l'amplitude de l'angle du segment. Comme pour l'erreur de position, cette erreur est en partie explicable par la différence dans la définition de la position des joints entre les deux modèles de squelette. Les valeurs d'erreurs angulaires sont semblables pour les deux jambes (10.68 et 10.51 degrés) et ne semblent pas grandement affectées par la marche (voir l'erreur associée aux exercices *Marche* et *Marche* statique ainsi que *Manutention* et *Flexion*). Par contre, certaines poses comme le *Squat* semblent plus difficiles à estimer correctement par l'algorithme d'estimation de pose utilisé.

La méthode de fusion proposée est robuste aux données erronées et aux occlusions par l'approche de fusion particulière et de données multivues. Nous observons en général une très forte corrélation entre les deux ensembles de données. Les erreurs en distance euclidienne et angulaire sont explicables en partie par la définition de la position biométrique des joints

(partiellement corrigée par la calibration des modèles). Nous avons évalué la performance en utilisant un montage uniforme afin d'obtenir des résultats non biaisés selon les exercices, il sera possible d'optimiser le montage afin d'effectuer des analyses sur des exercices spécifiques. Nous avons utilisé un des algorithmes d'estimation de pose les plus étudiés (voir section 3.2) pour l'extraction de squelettes utilisés comme données initiale de notre approche de fusion, la méthode développée s'adapte sans problème à n'importe quel autre algorithme proposant un modèle de squelette différent.

# Conclusion

Nous avons présenté un système de vision complet permettant la réadaptation en environnement virtuel avec un suivi de la pose humaine dans un espace de volume adaptable en plus d'un retour haptique et visuel ainsi qu'une analyse en temps réel et en post-traitement. Le projet multidisciplinaire profite de l'expertise combinée de chercheurs en robotique, en vision artificielle et en recherche clinique et réadaptation. Cette pluralité de connaissances a permis d'accomplir le développement de trois grandes composantes, le mécanisme à câbles permettant un retour haptique, le système de vision artificielle permettant le suivi du mouvement en environnement virtuel et la recherche clinique en réadaptation pour l'analyse des données de pose.

La méthode développée et couverte par cette thèse est flexible et s'adapte à différents projets sans devoir être nécessairement utilisée en réadaptation et n'est pas dépendante des composantes mécaniques comme le mécanisme à câbles. La plupart des éléments présentés dans les chapitres précédents sont réutilisables individuellement. Pour en énumérer quelques un: l'approche de calibration des capteurs et d'alignement des nuages de points, le prétraitement et le transfert des données en parallèle par les sous-systèmes, le suivi du casque de réalité virtuelle et l'affichage de l'avatar en temps réel, l'estimation de pose d'un objet rigide par triangulation ainsi que la méthode de fusion des données de positions 3D en temps réel.

L'analyse de la première itération de données de suivi du mouvement enregistrées est intéressante sur beaucoup d'aspects et servira de base à l'optimisation du système. La méthode de fusion développée présente une amélioration considérable de la fiabilité et de la robustesse des données individuelles d'estimation des poses en plus de permettre une adaptation de l'espace observé selon une approche modulaire divisée en sous-systèmes.

## Suivi et travaux futurs

Le système est présentement installé au Centre interdisciplinaire de recherche en réadaptation et intégration sociale (CIRRIIS) et un deuxième ensemble de sous-systèmes NUC et capteurs Kinects est prêt à l'installation au Laboratoire de Vision et Système Numérique (LVSN) et pourra être utilisé pour différents projets utilisant certaines méthodes développées

lors de notre projet. Nous avons développé un logiciel de calibration de nuages de points simple à utiliser (Multicam3D) comportant un affichage en temps réel. Nous avons aussi développé le logiciel plus spécifique (MulticamVR) utilisé pour la réadaptation en environnement basé sur la fusion de données.

Le système est présentement utilisé par des chercheurs au CIRRIS afin d'évaluer les possibilités d'en faire un outil clinique économique permettant la capture du mouvement sans marqueurs. Le système pourrait éventuellement être répliqué et installé dans différents centres de réadaptation pour une fraction du prix des systèmes de capture de mouvement populaires dans ce domaine. De plus, notre système présente différents avantages comme l'intégration en environnement virtuel, la possibilité de retour haptique et la facilité d'utilisation par l'approche de suivi sans marqueurs. Le projet est maintenant bien intégré au CIRRIS et nous maintenons le suivi afin de nous assurer de la continuité du développement.

# Annexe A

## Spécifications des caméras Bonita Vicon

# Bonita

Size. Performance.  
Reliability.



## Specifications

### All Bonita Range

Interface	Gigabit Ethernet, RJ45
System latency	2ms
Strobe	68 high powered NIR LEDs @ 780nm
Power	802.3af PoE
Compatibility	T-Series T10, T20S, T40S, T160 Bonita Video

### Bonita 3

Frame Rate	240 fps
Resolution	0.3 megapixel (640 x 480)
Lens Operating Range	Up to 12m
Angle of View Wide (4mm)	82.7° x 66.85°
Angle of View Narrow (12mm)	32.7° x 24.81°

### Bonita 10

Frame Rate	250 fps
Resolution	1 megapixel (1024 x 1024)
Lens Operating Range	Up to 13m
Angle of View Wide (4mm)	70.29° x 70.29°
Angle of View Narrow (12mm)	26.41° x 26.41°

Welcome to Bonita. Vicon's small optical camera family reliably streams the precise motion of markers, whether they are attached to people, animals or machines. Building on over 30 years of innovation, Bonita's speed, flexibility and affordability redefine the possibilities of motion capture.

### Solution

Ergonomics. Training. Injury prevention. Synthetic environments. Static facial capture. Immersive reality. Virtual prototyping. Simulation. Visualization. Animation. Gait analysis. Sports biomechanics. Animal studies. Rehabilitation. Virtual camera tracking.

### Precise

With its high resolution, quality optics and sophisticated algorithms, Bonita's new 1 megapixel camera, the Bonita 10, captures with precision, down to 0.5mm in a 4m x 4m volume.

### Affordable

You don't have to choose between performance and price with the new

Bonita family. With cameras starting at \$3,000 / £2,000, Bonita could change your idea of motion capture.

### Fast

Capture speeds of up to 250 fps with the Bonita 10 and 240 fps with the Bonita 3, means you can easily and accurately capture the movements of fast moving objects.

### Flexible

The beauty of the compact Bonita camera is you can mount them anywhere. With its Variable focal length lens and single cable ethernet connection, it's easy to optimize your set up. From the footwell of a car, to a complex virtual reality cave environment, you won't lose performance with Bonita.

### Plug and play

Some motion capture systems have a reputation for being hard to set up. Not with Bonita. Just like its big brother T-series, Bonita uses a single cable to provide everything it needs – power, data and synchronisation. Calibrate and start streaming data in less than 2 minutes.

## Discover More



facebook.com/vicon  
twitter.com/vicon  
youtube.com/vicon100

[vicon.com/system/bonita](http://vicon.com/system/bonita)

Denver  
T: +1 303.799.8686  
Los Angeles  
T: +1 303.799.8686

[info@vicon.com](mailto:info@vicon.com)

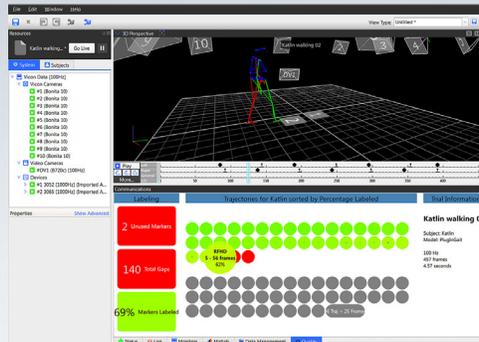
Oxford  
T: +44 (0) 1865 261800  
Singapore  
T: +65 6400 3500

# Annexe B

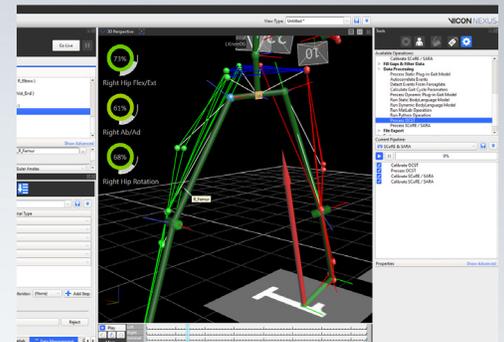
Nexus: logiciel de capture de données Vicon

# Nexus 2

Track. Analyze. Discover.



Automatic Quality Assessment Panel



Real-time subject calibration feedback

Welcome to Nexus 2, Vicon's data capture software developed for our Life Science customers. With a host of new automated features, intelligent processing and flexible controls, Nexus 2 sets a new standard for motion capture.

## Under the hood

The Nexus 2 data processing engine has been completely re-engineered to give you significant speed and accuracy improvements when processing your data.

Labeling in Nexus 2 is intelligent. Not only is it faster and more accurate, but real-time labeling is also significantly improved. The new data processing engine also allows real-time subject calibration feedback and automatically initializes labeling of your subject – no need to label manually.

## Powerful modeling

As the only optical motion capture provider on the MATHWORKS partners program, Vicon has developed native integration with MATLAB in Nexus 2. Run your MATLAB scripts from Nexus with one click.

Build your MATLAB processing scripts and see your results immediately with the Nexus-MATLAB dynamic connection. Run scripts from within the Nexus MATLAB pane or save these directly to your Nexus pipeline.

Not a MATLAB user? Nexus also comes with Python pre-installed. Access all your trial data with the Vicon SDK.

## Automatic quality assessment

Save time with Nexus 2's Automatic Quality Assessment. By automatically detecting gaps and labeling quality in your data, Nexus 2 presents an informative, visual overview of your trial. Improve your results with the expanded set of data fill options including the new Rigid Body and Kinematic Fills with Intelligent Automatic Donor Selection.

## Customize your workflow

Creating simple, universal lab protocols has never been easier. Replacing the previous 'Protocols' concept, the Biomechanics Workflow builder creates a series of steps to combine data collection and offline processing. It also makes it simple to get started with the new SCoRE and SARA Functional Calibration, providing real-time

visual feedback in the Nexus 2 workspace to assist with subject calibration.

## Manage your data

ProEclipse is the new data management system for Vicon Life Science software users. With improved database navigation and intelligent search functionality, ProEclipse will enable you find your data with ease.

## Seamless video

Nexus offers support for the Bonita Video camera which was developed from the ground up for Vicon Life Science customers. Bonita Video achieves full frame synchronization up to 350Hz or 1280x720 HD from a single point of connection for both power and data. Spend less time setting up your system and more time on research thanks to the all-new one step dynamic video and optical calibration.

Discover More

[vicon.com/software/nexus](http://vicon.com/software/nexus)

[info@vicon.com](mailto:info@vicon.com)



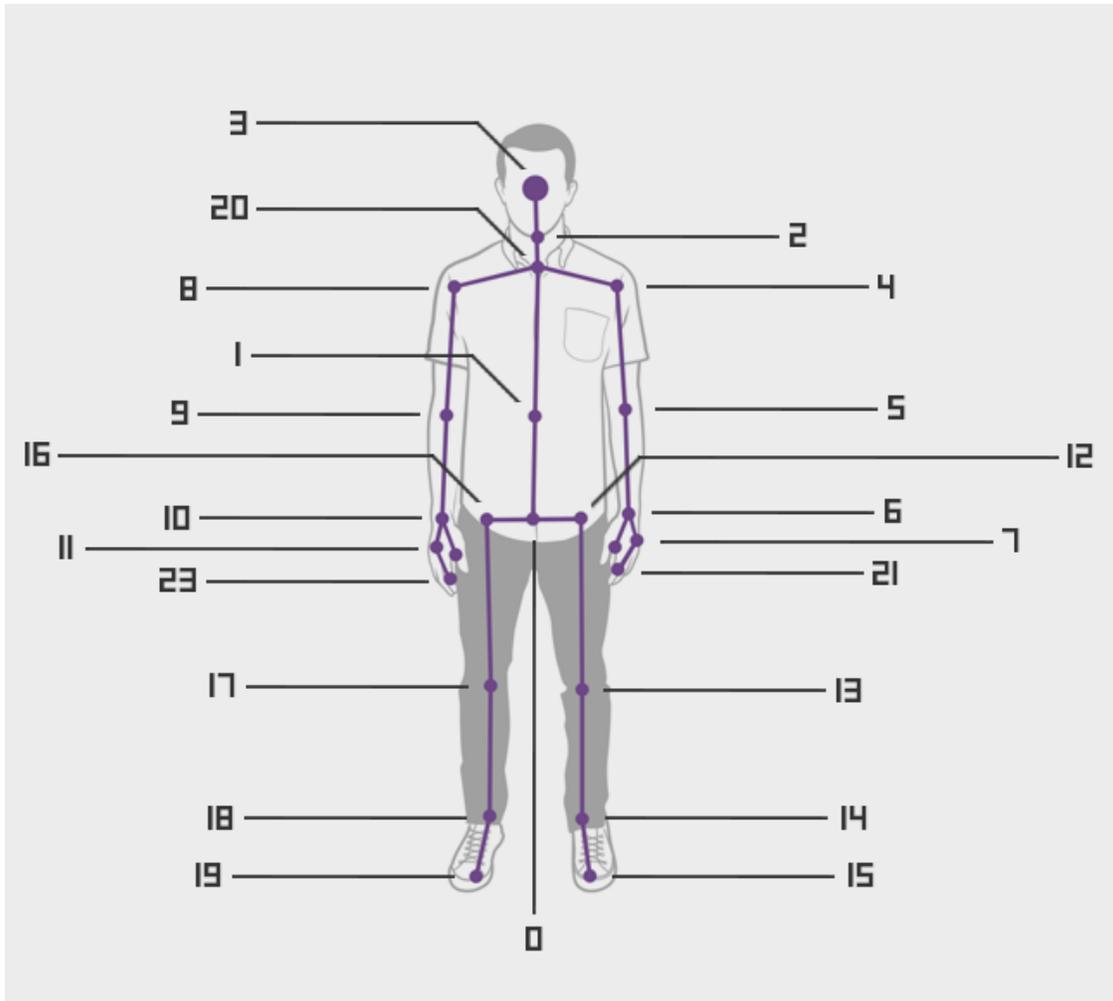
[facebook.com/vicon](https://facebook.com/vicon)  
[twitter.com/Vicon](https://twitter.com/Vicon)  
[youtube.com/vicon100](https://youtube.com/vicon100)

Denver  
 T: +1 303.799.8686  
 Los Angeles  
 T: +1 303.799.8686

Oxford  
 T: +44 (0) 1865 261800  
 Singapore  
 T: +65 6400 3500

# Annexe C

Modèle du Squelette de l'algorithme d'estimation de pose de Microsoft



[60]

# KINECT V2 JOINT ID MAP

25 JOINTS

6 BODIES

JOINTTYPE_SPINEBASE	= 0.
JOINTTYPE_SPINEMID	= 1.
JOINTTYPE_NECK	= 2.
JOINTTYPE_HEAD	= 3.
JOINTTYPE_SHOULDERLEFT	= 4.
JOINTTYPE_ELBOWLEFT	= 5.
JOINTTYPE_WRISTLEFT	= 6.
JOINTTYPE_HANDLEFT	= 7.
JOINTTYPE_SHOULDERRIGHT	= 8.
JOINTTYPE_ELBOWRIGHT	= 9.
JOINTTYPE_WRISTRIGHT	= 10.
JOINTTYPE_HANDRIGHT	= 11.
JOINTTYPE_HIPLEFT	= 12.
JOINTTYPE_KNEELEFT	= 13.
JOINTTYPE_ANKLELEFT	= 14.
JOINTTYPE_FOOTLEFT	= 15.
JOINTTYPE_HIPRIGHT	= 16.
JOINTTYPE_KNEERIGHT	= 17.
JOINTTYPE_ANKLERIGHT	= 18.
JOINTTYPE_FOOTRIGHT	= 19.
JOINTTYPE_SPINESHOULDER	= 20.
JOINTTYPE_HANDTIPLEFT	= 21.
JOINTTYPE_THUMBLEFT	= 22.
JOINTTYPE_HANDTIPRIGHT	= 23.
JOINTTYPE_THUMBRIGHT	= 24.

[60]

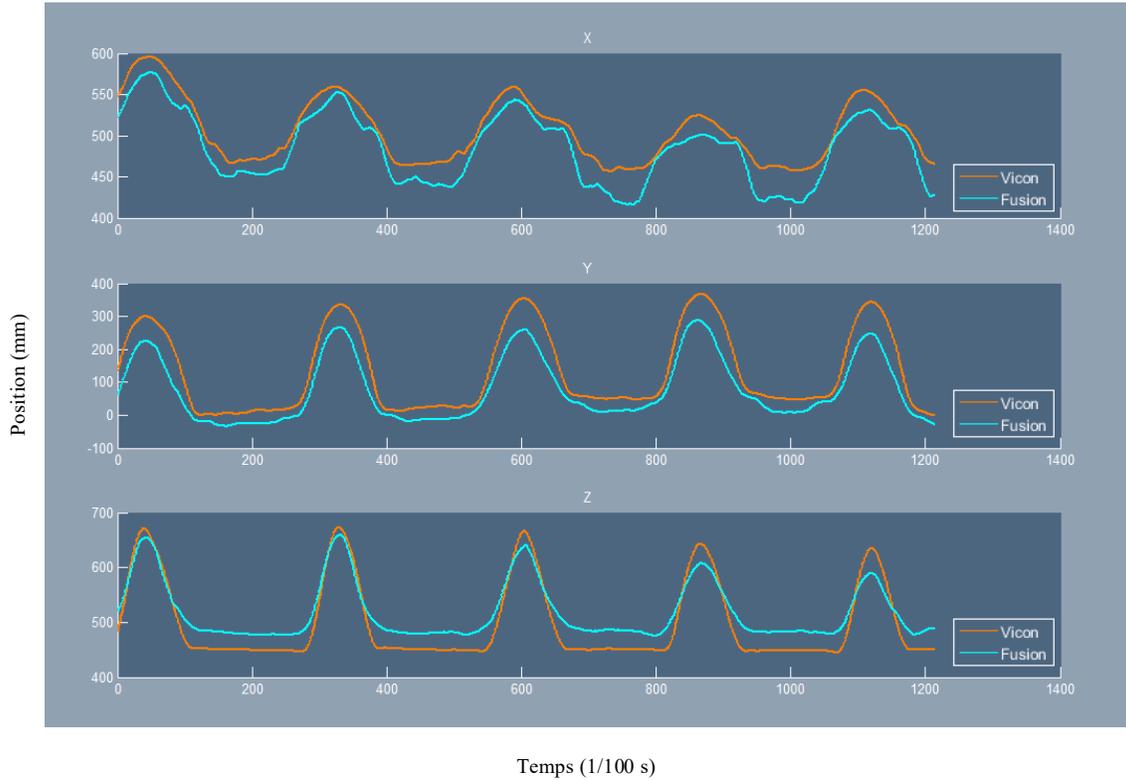
# Annexe D

## Données supplémentaires

**Avant correction du biais:** Distance euclidienne entre les joints observés par les deux systèmes (cm)

	CG	GG	HG	CD	GD	HD	D	Moyenne
Flexion	10,456	4,270	12,594	10,665	5,679	7,623	21,967	10,465
Manutention	10,288	6,873	11,191	9,148	6,948	11,994	20,610	11,008
Manutention +	15,837	8,516	13,479	9,853	7,898	11,971	20,709	12,609
Marche	15,419	7,257	10,730	13,886	7,073	11,613	20,003	12,283
Marche Statique	8,433	6,990	8,810	9,122	6,250	10,162	18,722	9,784
Pose Fixe	9,800	4,094	10,939	10,681	5,966	11,148	18,829	10,208
Rotation D	25,222	15,051	15,028	8,221	8,406	13,596	25,059	15,798
Rotation G	7,609	6,504	13,425	6,999	5,706	15,350	24,379	11,425
Squat	9,738	6,425	10,185	10,346	7,365	11,063	19,910	10,719
Squat +	18,205	7,263	17,651	16,922	21,203	18,234	25,645	17,875
Écart-type	5,527	3,031	2,601	2,874	4,644	2,909	2,567	3,450
Moyenne	13,101	7,324	12,403	10,584	8,249	12,276	21,583	12,217

Joint: Genou gauche X, Y et Z  
Exercice: Marche statique



**Avant correction du biais:** Suivi d'un joint en position dans le temps au cours d'un exercice. Les trois graphes représentent chacun la position des joints selon le référentiel global pour une des trois dimensions (X, Y et Z).

**Avant fusion:** Erreur angulaire jambes (degrés)

	KA		KB		KC		KD		
	JG	JD	JG	JD	JG	JD	JG	JD	Moyenne
Flexion	57,55	12,86	4,15	5,12	15,73	77,87	31,63	22,20	28,39
Manutention	15,28	14,22	11,90	10,98	16,83	36,91	62,05	8,02	22,02
Manutention +	21,48	13,15	13,27	18,88	17,77	27,10	53,45	29,59	24,34
Marche	14,21	14,68	12,72	12,11	18,67	12,53	12,27	13,89	13,89
Marche Statique	8,54	8,60	11,16	13,16	24,47	24,46	20,08	21,78	16,53
Pose Fixe	9,88	12,61	6,80	4,00	15,95	42,75	2,70	4,03	12,34
Rotation D	27,30	8,54	6,29	10,50	8,63	13,99	31,50	6,93	14,21
Rotation G	36,56	12,14	4,09	5,98	28,19	24,76	14,09	16,72	17,82
Squat	14,94	21,06	22,68	7,23	18,91	25,30	56,45	6,73	21,66
Squat +	15,20	18,25	10,74	13,86	26,18	30,46	37,08	5,77	19,69
Écart-type	15,03	3,85	5,54	4,63	5,77	18,64	20,27	8,71	10,30
Moyenne	22,09	13,61	10,38	10,18	19,13	31,61	32,13	13,57	19,09

**Avant fusion:** Erreur angulaire dos (degrés)

	KA	KB	KC	KD	Moyenne
Flexion	10,50	10,93	14,14	10,01	11,40
Manutention	5,77	7,82	5,49	4,64	5,93
Manutention +	5,32	8,24	5,34	4,99	5,97
Marche	1,58	2,54	1,82	3,31	2,31
Marche Statique	2,05	5,47	2,10	3,11	3,18
Pose Fixe	0,96	4,23	2,89	2,84	2,73
Rotation D	10,47	12,34	9,16	2,95	8,73
Rotation G	15,32	9,04	11,45	10,34	11,54
Squat	1,43	5,13	5,19	6,52	4,57
Squat +	9,69	8,17	1,59	5,90	6,34
Écart-type	4,96	3,05	4,33	2,79	3,78
Moyenne	6,31	7,39	5,92	5,46	6,27

# Références

- [1] Moons T, Van Gool L, Vergauwen M. 3D Reconstruction from multiple images, part 1: Principles. Now Publishers Inc; 2009.
  
- [2] Fankhauser P, Bloesch M, Rodriguez D, Kaestner R, Hutter M, Siegwart R. Kinect v2 for mobile robot navigation: evaluation and modeling. In Advanced Robotics (ICAR), 2015 International Conference on 2015 Jul 27 (pp. 388-394). IEEE.
  
- [3] Andersen MR, Jensen T, Lisouski P, Mortensen AK, Hansen MK, Gregersen T, Ahrendt P. Kinect depth sensor evaluation for computer vision applications. Technical Report Electronics and Computer Engineering. 2015 Jul 2;1(6).
  
- [4] Microsoft. Kinect for Windows Sensor Components and Specifications. <https://msdn.microsoft.com/enus/library/jj131033>
  
- [5] Microsoft Kinect for Windows (SDK 1.5, 1.6, 1.7, 1.8) NUI Constants description. <https://msdn.microsoft.com/enus/library/hh855368>
  
- [6] França JG, Gazziro MA, Ide AN, Saito JH. A 3D scanning system based on laser triangulation and variable field of view. In IEEE International Conference on Image Processing 2005 2005 Sep 11 (Vol. 1, pp. I-425). IEEE.
  
- [7] Zhang Z. Microsoft kinect sensor and its effect. IEEE multimedia. 2012 Feb;19(2):4-10.
  
- [8] Zennaro, S., et al. "Performance evaluation of the 1st and 2nd generation Kinect for multimedia applications." 2015 IEEE International Conference on Multimedia and Expo (ICME). IEEE, 2015.
  
- [9] <https://www.vicon.com/>

- [10] Heikkila J, Silvén O. A four-step camera calibration procedure with implicit image correction. In *Computer Vision and Pattern Recognition, 1997. Proceedings., 1997 IEEE Computer Society Conference on* 1997 Jun 17 (pp. 1106-1112). IEEE.
- [11] Zhang Z. A flexible new technique for camera calibration. *IEEE Transactions on pattern analysis and machine intelligence.* 2000 Nov;22(11):1330-4.
- [12] Z. Zhang, "Camera Calibration", Chapter 2, pages 4-43, in G. Medioni and S.B. Kang, eds., *Emerging Topics in Computer Vision*, Prentice Hall Professional Technical Reference, 2004.
- [13] Tsai R. A versatile camera calibration technique for high-accuracy 3D machine vision metrology using off-the-shelf TV cameras and lenses. *IEEE Journal on Robotics and Automation.* 1987 Aug;3(4):323-44.
- [14] Salvi J, Matabosch C, Fofi D, Forest J. A review of recent range image registration methods with accuracy evaluation. *Image and Vision computing.* 2007 May 1;25(5):578-96.
- [15] Eggert DW, Lorusso A, Fisher RB. Estimating 3-D rigid body transformations: a comparison of four major algorithms. *Machine Vision and Applications.* 1997 Mar 1;9(5-6):272-90.
- [16] Arun KS, Huang TS, Blostein SD. Least-squares fitting of two 3-D point sets. *IEEE Transactions on pattern analysis and machine intelligence.* 1987 Sep(5):698-700.
- [17] Besl PJ, McKay ND. Method for registration of 3-D shapes. In *Robotics-DL tentative* 1992 Apr 30 (pp. 586-606). International Society for Optics and Photonics.
- [18] Zhou H, Hu H. Human motion tracking for rehabilitation—A survey. *Biomedical Signal Processing and Control.* 2008 Jan 31;3(1):1-8.
- [19] Moeslund TB, Granum E. A survey of computer vision-based human motion capture. *Computer vision and image understanding.* 2001 Mar 31;81(3):231-68.
- [20] Shotton J, Sharp T, Kipman A, Fitzgibbon A, Finocchio M, Blake A, Cook M, Moore

R. Real-time human pose recognition in parts from single depth images. *Communications of the ACM*. 2013 Jan 1;56(1):116-24.

[21] Lu TW, O'connor JJ. Bone position estimation from skin marker co-ordinates using global optimisation with joint constraints. *Journal of biomechanics*. 1999 Feb 28;32(2):129-34.

[22] Obdržálek Š, Kurillo G, Ofi F, Bajcsy R, Seto E, Jimison H, Pavel M. Accuracy and robustness of Kinect pose estimation in the context of coaching of elderly population. In 2012 Annual International Conference of the IEEE Engineering in Medicine and Biology Society 2012 Aug 28 (pp. 1188-1193). IEEE.

[23] Wang Q, Kurillo G, Ofi F, Bajcsy R. Evaluation of pose tracking accuracy in the first and second generations of microsoft kinect. In *Healthcare Informatics (ICHI), 2015 International Conference on* 2015 Oct 21 (pp. 380-389). IEEE.

[24] Tao G, Archambault PS, Levin MF. Evaluation of Kinect skeletal tracking in a virtual reality rehabilitation system for upper limb hemiparesis. In *2013 International Conference on Virtual Rehabilitation (ICVR) 2013 Aug 26* (pp. 164-165). IEEE.

[25] Wiedemann LG, Planinc R, Nemeč I, Kampel M. Performance evaluation of joint angles obtained by the Kinect v2. In *Technologies for Active and Assisted Living (TechAAL), IET International Conference on* 2015 Nov 5 (pp. 1-6). IET.

[26] Khaleghi B, Khamis A, Karray FO, Razavi SN. Multisensor data fusion: A review of the state-of-the-art. *Information Fusion*. 2013 Jan 31;14(1):28-44.

[27] Del Moral P. Non-linear filtering: interacting particle resolution. *Markov processes and related fields*. 1996;2(4):555-81.

[28] Doucet A, Godsill S, Andrieu C. On sequential Monte Carlo sampling methods for Bayesian filtering. *Statistics and computing*. 2000 Jul 1;10(3):197-208.

[29] Doucet A, Johansen AM. A tutorial on particle filtering and smoothing: Fifteen years later. *Handbook of nonlinear filtering*. 2009 Dec;12(656-704):3.

- [30] Cappé O, Godsill SJ, Moulines E. An overview of existing methods and recent advances in sequential Monte Carlo. *Proceedings of the IEEE*. 2007 May;95(5):899-924.
- [31] Lo AC, Guarino PD, Richards LG, Haselkorn JK, Wittenberg GF, Federman DG, Ringer RJ, Wagner TH, Krebs HI, Volpe BT, Bever Jr CT. Robot-assisted therapy for long-term upper-limb impairment after stroke. *New England Journal of Medicine*. 2010 May 13;362(19):1772-83.
- [32] J.Y.Bouguet. MATLAB calibration tool. [http://www.vision.caltech.edu/bouguetj/calib\\_doc/](http://www.vision.caltech.edu/bouguetj/calib_doc/)
- [33] Fischler MA, Bolles RC. Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. *Communications of the ACM*. 1981 Jun 1;24(6):381-95.
- [34] Muller N, Magaia L, Herbst BM. Singular value decomposition, eigenfaces, and 3D reconstructions. *SIAM review*. 2004;46(3):518-45
- [35] Bourke P. The shortest line between two lines in 3d. Disponível na internet via WWW. URL: <http://local.wasp.uwa.edu.au/~pbourke/geometry/lineline3d>. 1998 Apr.
- [36] Darekar A, McFadyen BJ, Lamontagne A, Fung J. Efficacy of virtual reality-based intervention on balance and mobility disorders post-stroke: a scoping review. *Journal of neuroengineering and rehabilitation*. 2015 May 10;12(1):1.
- [37] Burdea GC. Virtual rehabilitation-benefits and challenges. *Methods of Information in Medicine-Methodik der Information in der Medizin*. 2003 Jan 1;42(5):519-23.
- [38] Boian RF, Deutsch JE, Lee CS, Burdea GC, Lewis J. Haptic effects for virtual reality-based post-stroke rehabilitation. In *Haptic Interfaces for Virtual Environment and Teleoperator Systems, 2003. HAPTICS 2003. Proceedings. 11th Symposium on* 2003 Mar 22 (pp. 247-253). IEEE.
- [39] Broeren J, Rydmark M, Sunnerhagen KS. Virtual reality and haptics as a training

device for movement rehabilitation after stroke: a single-case study. *Archives of physical medicine and rehabilitation*. 2004 Aug 31;85(8):1247-50.

[40] Zhang J, Cheah CC. Passivity and stability of human-robot interaction control for upper-limb rehabilitation robots. *IEEE Transactions on Robotics*. 2015 Apr;31(2):233-45.

[41] Desai PR, Desai PN, Ajmera KD, Mehta K. A review paper on oculus rift-a virtual reality headset. *arXiv preprint arXiv:1408.1173*. 2014 Aug 6.

[42] LaViola Jr JJ. A discussion of cybersickness in virtual environments. *ACM SIGCHI Bulletin*. 2000 Jan 1;32(1):47-56.

[43] Lemieux F, Laurendeau D, Branzan-Albu A, Boivin E, Drouin R, Martel H, Ouellet D, Schwartz JM. Survey of the State-of-the-Art on Synthetic Environments, Sensori-Motor Activities in Synthetic Environments, Simulation Frameworks and Real-World Abstraction Models Contract W7701-023981/QCA.

[44] Yeung KY, Kwok TH, Wang CC. Improved skeleton tracking by duplex kinects: A practical approach for real-time applications. *Journal of Computing and Information Science in Engineering*. 2013 Dec 1;13(4):041007.

[45] Arulampalam MS, Maskell S, Gordon N, Clapp T. A tutorial on particle filters for online nonlinear/non-Gaussian Bayesian tracking. *IEEE Transactions on signal processing*. 2002 Feb;50(2):174-88.

[46] Fitzgibbon AW, Fisher RB. A buyer's guide to conic fitting. *DAI Research paper*. 1996 May.

[47] Suzuki S. Topological structural analysis of digitized binary images by border following. *Computer Vision, Graphics, and Image Processing*. 1985 Apr 1;30(1):32-46.

[48] Gordon NJ, Salmond DJ, Smith AF. Novel approach to nonlinear/non-Gaus-

sian Bayesian state estimation. In *IEEE Proceedings F-Radar and Signal Processing* 1993 Apr (Vol. 140, No. 2, pp. 107-113). IET.

[49] Doucet A, De Freitas N, Gordon N. An introduction to sequential Monte Carlo methods. In *Sequential Monte Carlo methods in practice* 2001 (pp. 3-14). Springer New York.

[50] Bourke P. The shortest line between two lines in 3d. Disponível na internet via WWW. URL: <http://local.wasp.uwa.edu.au/~pbourke/geometry/lineline3d>. 1998 Apr.

[51] Clark RA, Pua YH, Fortin K, Ritchie C, Webster KE, Denehy L, Bryant AL. Validity of the Microsoft Kinect for assessment of postural control. *Gait & posture*. 2012 Jul 31;36(3):372-7.

[52] Pfister A, West AM, Bronner S, Noah JA. Comparative abilities of Microsoft Kinect and Vicon 3D motion capture for gait analysis. *Journal of medical engineering & technology*. 2014 Jul 1;38(5):274-80.

[53] Stone EE, Skubic M. Passive in-home measurement of stride-to-stride gait variability comparing vision and Kinect sensing. In *2011 Annual international conference of the IEEE engineering in medicine and biology society* 2011 Aug 30 (pp. 6491-6494). IEEE.

[54] Stone EE, Skubic M. Evaluation of an inexpensive depth camera for passive in-home fall risk assessment. In *2011 5th International Conference on Pervasive Computing Technologies for Healthcare (PervasiveHealth) and Workshops* 2011 May 23 (pp. 71-77). IEEE.

[55] Galna B, Barry G, Jackson D, Mhiripiri D, Olivier P, Rochester L. Accuracy

of the Microsoft Kinect sensor for measuring movement in people with Parkinson's disease. *Gait & posture*. 2014 Apr 30;39(4):1062-8.

[56] Dutta T. Evaluation of the Kinect™ sensor for 3-D kinematic measurement in the workplace. *Applied ergonomics*. 2012 Jul 31;43(4):645-9.

[57] Springer S, Yogev Seligmann G. Validity of the Kinect for Gait Assessment: A Focused Review. *Sensors*. 2016 Feb 4;16(2):194.

[58] Vicon Bonita <https://www.vicon.com/products/camera-systems/bonita>

[59] Kinect for Xbox One <http://www.xbox.com/en-US/xbox-one/accessories/kinect>

[60] Kinect V2 joint ID map <https://vrvv.org/documentation/kinect>

[61] Evans JD. *Straightforward statistics for the behavioral sciences*. Brooks/Cole; 1996.

[62] Ganapathi V, Plagemann C, Koller D, Thrun S. Real-time human pose tracking from range data. In *European conference on computer vision 2012 Oct 7* (pp. 738-751). Springer Berlin Heidelberg.

[63] Yub Jung H, Lee S, Seok Heo Y, Dong Yun I. Random tree walk toward instantaneous 3D human pose estimation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition 2015* (pp. 2467-2474).

[64] Girshick R, Shotton J, Kohli P, Criminisi A, Fitzgibbon A. Efficient regression of general-activity human poses from depth images. In *2011 International Conference on Computer Vision 2011 Nov 6* (pp. 415-422). IEEE.