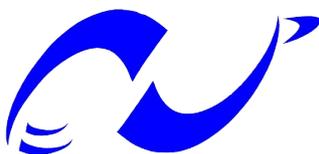


REPUBLIQUE ALGERIENNE DEMOCRATIQUE ET POPULAIRE

**MINISTERE DE L'ENSEIGNEMENT SUPERIEUR ET DE LA
RECHERCHE SCIENTIFIQUE**

ECOLE NATIONALE POLYTECHNIQUE



المدرسة الوطنية المتعددة التقنيات
Ecole Nationale Polytechnique

**Département Génie Electrique
Spécialité Automatique**

Mémoire

**Pour l'obtention du diplôme
De Master en Automatique**

Thème

**Gestion de l'occultation sur la base de
l'accélération constante**

Réalisé par : Mlle GABOUR Nour El Houda

Encadré par :

Pr. CHEKIREB Hachemi

Examineurs :

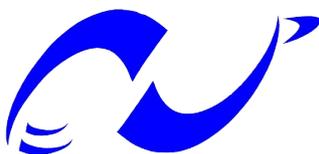
**Pr. BERKOUK El Madjid
Mr. ABDLEOUEL Lahcen**

Juin 2016

REPUBLIQUE ALGERIENNE DEMOCRATIQUE ET POPULAIRE

**MINISTERE DE L'ENSEIGNEMENT SUPERIEUR ET DE LA
RECHERCHE SCIENTIFIQUE**

ECOLE NATIONALE POLYTECHNIQUE



المدرسة الوطنية المتعددة التقنيات
Ecole Nationale Polytechnique

**Département Génie Electrique
Spécialité Automatique**

Mémoire

**Pour l'obtention du diplôme
De Master en Automatique**

Thème

**Gestion de l'occultation sur la base de
l'accélération constante**

Réalisé par : Mlle GABOUR Nour El Houda

Encadré par :

Pr. CHEKIREB Hachemi

Examineurs :

**Pr. BERKOUK El Madjid
Mr. ABDLEOUEL Lahcen**

Juin 2016

ملخص

العمل المقدم في هذه الأطروحة يتمحور حول الروبوتات المتنقلة في الأماكن المغلقة، في إطار ابحار روبوت دليل في محيط داخلي مزدحم، عن طريق المراقبة البصرية لتتبع هدف متحرك بغاية تحسين استقلالية المنصة المتحركة (الروبوت) من خلال تعزيز تفاعله مع البيئة المحيطة به عن طريق مختلف أجهزة الاستشعار الموجودة على متنه.

كلمات البحث: الروبوتات، المراقبة البصرية، يعتم.

Abstract

The work presented in this thesis deals with the indoor mobile robotics, it is placed under the navigation of an inner guide robot in a cluttered environment by performing a visual servoing for tracking a dynamic target in the purpose of improving the performance of a mobile platform by enhancing its interaction with the environment surrounding it, through its various on-board sensors.

Key words : robotics, visual servoing, occlusion.

Résumé

Le travail présenté dans ce mémoire traite de la robotique mobile d'intérieur, il est placé dans le cadre de la navigation d'un robot guide d'intérieur dans un environnement encombré en effectuant un asservissement visuel pour le suivi d'une cible dynamique dans le but de l'amélioration de l'autonomie d'une plateforme mobile en améliorant son interaction avec le milieu l'entourant à travers ses différents capteurs embarqués.

Mots clés : robotique, asservissement visuel, occultation.

Remerciements

Tous d'abord on remercie Dieu de nous avoir accordé la connaissance, donner le courage, la patience et la santé pour réaliser mémoire de master.

Nos remerciements les plus vifs s'adressent tout particulièrement à Mr CHEKIREB Hachemi professeur à l'École Nationale Polytechnique ENP pour m'avoir aidé, conseillé et encouragé tout au long de mes recherches.

Mes remerciements s'adressent également aux membres du jury qui ont accepté de lire et d'évaluer ce mémoire.

Nous tenons aussi à remercier

Enfin, je remercie

Toute personne ayant contribué de près ou de loin à la réalisation de ce travail.

Sommaire

Remerciement

Résumé

Liste des figures

CHAPITRE 1 : Généralités sur les robots mobiles et leurs asservissements Visuels

Introduction générale.....	8
1.1 Les robots mobiles	8
1.1.1 Généralités sur les robots mobiles	8
1.1.2 Intérêts des robots mobiles	9
1.1.3 Classification des robots mobiles selon leur degré d'autonomie	9
1.1.4 Classification des robots mobile à roue	10
a) Robots omnidirectionnels	10
b) Robots non holonomes	10
1.2 Commande visuelle	10
1.2.1 Commande référencée capteurs.....	10
1.2.2 Commande référencée vision	11
1.3 Gestion des occultations	12

CHAPITRE 2 : Gestion de l'occultation

Introduction	15
2.1 Présentation du robot B21r	16
2.1.1 Modélisation du système robotique	17
2.2 Le système de vision	21
2.2.1 Description matérielle de la Kinect de Microsoft	21
2.2.2 Calibrage d'un capteur de vision.....	21
2.2.3 Généralités sue la Reconnaissance de Personnes dans une Scène	21

2.2.4	Reconnaissance physique basée sur la Kinect.....	21
2.3	Gestion de l'occultation	24
2.3.1	Généralités et hypothèses de travail	24
2.3.2	Prédiction de la trajectoire	26
2.3.2.1	Prédiction à vitesse constante	27
2.3.2.2	Prédiction à accélération constante.....	28
	Conclusion.....	31
	Bibliographie.....	32

Liste des figures

Fig.1.1 : roue d'un robot holonome

Fig. 2.1 : Robot mobile B21r

Fig. 2.2 : Repères du système Robot-Kinect

Fig.2.3 : Kinect sans cache

Fig.2.4 Squelettes « trackés » et « non-trackés »

Fig.2.5 Représentation d'un squelette tracké

Fig 2.6 Représentation des deux repères et des différentes projections

CHAPITRE 1

Généralités sur les robots mobiles et leurs asservissements Visuels

• Introduction générale

La robotique, apparue dans les années 1960, consiste à concevoir et à étudier des machines intelligentes, c'est-à-dire des machines capables de réaliser des tâches avec un certain degré d'autonomie. Les disciplines de recherche associées à cette thématique sont articulées autour de la perception (traitement du signal, reconnaissance de formes), de la décision (intelligence artificielle) et de l'action (automatique). Dans un premier temps, les chercheurs se sont intéressés à la robotique de manipulation et à ses applications notamment dans l'industrie manufacturière. A partir des années 1980, en raison de l'augmentation de la puissance de calcul embarquée et de l'amélioration des capteurs, la robotique mobile devient un thème de recherche important avec comme objectif la création de robots capables d'explorer des zones dangereuses pour l'homme en ce qui concerne la robotique d'extérieur, en ce qui concerne la robotique d'intérieur, l'objectif est la création de robots accompagnateurs capables d'interagir avec des humains : aide aux personnes âgées, robots guide dans une institution (musée, hôpital...).

1.1 Les robots mobiles

1.1.1 Généralités sur les robots mobiles

Un robot est un système alimenté en énergie qui évolue dans un environnement statique ou dynamique, il est formé d'un microcontrôleur ainsi que d'un ou plusieurs capteurs et actionneurs. La conception d'un robot se base sur son cahier des charges. Elle comprend l'analyse du comportement souhaité pour le robot et sa synthèse théorique, à l'aide notamment des théories d'asservissement, ainsi que l'implémentation logicielle et matérielle du robot.

L'ensemble des problèmes particuliers liés à la conception de tels robots sont :

- La conception mécanique liée à la mobilité,
- La détermination de la position et de la latitude (orientation),
- La détermination du chemin optimal pour atteindre le lieu de la tâche.

La structure d'un robot est contrôlée de manière à effectuer un ensemble de tâche. Ce contrôle inclut trois phases distinctes qui se répètent en boucle : la perception, le traitement et l'action. Un robot fonctionne par l'exécution continue d'un programme informatique constitué d'algorithmes. Ce programme est écrit dans un langage de programmation dont la nature est choisie par le constructeur.

La phase de perception est assurée par l'utilisation de capteurs. Les capteurs donnent une information à propos de l'environnement ou des composants internes (ex : position d'un moteur ou d'un vérin, état d'une LED, etc.). Cette information est utilisée pour calculer l'ordre approprié à envoyer aux actionneurs.

La phase de traitement est assurée par un microcontrôleur ou un ordinateur embarqué elle peut varier en complexité. À un niveau réactif, un robot peut traduire l'information brute d'un capteur directement en commande d'un actionneur (ex : un arrêt d'urgence ; si un obstacle est détecté alors arrêt des moteurs). Avec des tâches plus sophistiquées, il faut utiliser des algorithmes.

La phase d'action est réalisée à l'aide d'actionneurs.

1.1.2 Intérêts des robots mobiles

Les robots mobiles ont une place particulière en robotique. Leur intérêt réside dans leur mobilité qui ouvre des applications dans de nombreux domaines. Comme les robots manipulateurs, ils sont destinés à assister l'homme dans les tâches pénibles (transport de charges lourdes), monotones ou en ambiance hostile (nucléaire, marine, spatiale, lutte contre l'incendie, surveillance...). L'aspect particulier de la mobilité impose une complexité technologique et méthodologique qui s'ajoute en général aux problèmes rencontrés par les robots manipulateurs. La résolution de ces problèmes passe par l'emploi de toutes les ressources disponibles tant au niveau technologique (capteurs, motricité, énergie) qu'à celui du traitement des informations par l'utilisation des techniques de l'intelligence artificielle ou de processeurs particuliers. L'autonomie du robot mobile est une faculté qui lui permet de s'adapter ou de prendre une décision dans le but de réaliser une tâche malgré des informations préliminaires absentes ou éventuellement erronées.

1.1.3 Classification des robots mobiles selon leur degré d'autonomie

Une classification est proposée dans la littérature se basant sur le degré d'autonomie du robot mobile. On distingue :

- Véhicule télécommandé par un opérateur lui imposant chaque tâche élémentaire à réaliser.
- Véhicule télécommandé par un opérateur lui imposant des tâches plus complexes.
- Véhicule semi-autonome réalisant sans l'aide de l'opérateur des tâches prédéfinies.

- Véhicule autonome qui réalise des tâches semi-définies. Ce type de véhicule pose des problèmes d'un niveau de complexité élevé de représentation des connaissances, de capacité décisionnelle et de génération de plans qui sont résolus à bord dans la mesure du possible.

1.1.4 Classification des robots mobile à roue

On trouve généralement deux grandes catégories de robots mobiles à roues :

a) Robots omnidirectionnels

Dit aussi Holonome, ce terme se réfère à la relation entre les degrés commandabilité et le degré de liberté d'un robot. Si le degré de commandabilité et de liberté sont égaux, alors le robot est dit holonome. Un robot construit sur roues ou Omni-roues est un bon exemple de robot holonomes car il peut se déplacer librement dans toutes les directions. L'image montre une roue pivotante qui peut tourner dans les deux axes X et Y.

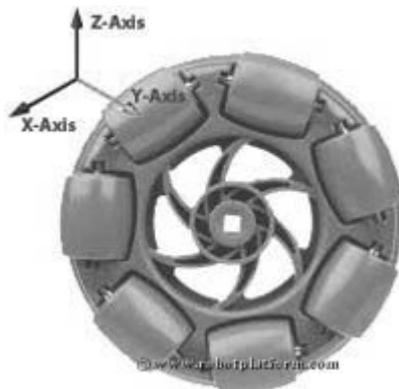


Fig.1.1 : roue d'un robot holonome

b) Robots non holonomes

Si le degré de commandabilité est inférieur au degré de liberté, alors le robot est considéré comme non-holonome. Une voiture a trois degrés de liberté ; sa position est calculée dans deux axes et son orientation. Cependant, il y a seulement deux degrés de liberté contrôlables qui sont l'accélération (ou le freinage) et l'angle de braquage du volant. Cela rend difficile pour le conducteur de tourner la voiture dans toutes les directions.

1.2 Commande visuelle

1.2.1 Commande référencée capteurs

La commande référencée capteurs consiste à exprimer les tâches robotiques à réaliser, non plus dans l'espace des configurations, mais directement dans l'espace du capteur sous la forme d'une relation locale entre le robot et son environnement. On cherche ainsi à définir des commandes en boucle fermée sur la base des informations provenant des capteurs et non plus sur la base de la configuration du robot. De ce fait, les lois de commande généralement synthétisées dans le cadre de cette approche sont de type *retour de sortie* et permettent de réaliser les tâches désirées de manière plus précise. Ainsi, toute tâche pouvant s'exprimer sous la forme d'une relation entre le robot et l'environnement sera réalisable par le biais de la commande référencée capteurs. Toutefois, cette approche requiert l'utilisation de capteurs performants, capables de fournir les mesures à une fréquence compatible avec la fréquence d'échantillonnage de l'asservissement.

1.2.2 Commande référencée vision

La commande référencée vision consiste à contrôler les mouvements d'un système robotique en utilisant des informations visuelles issues d'une ou plusieurs caméras (ou plus généralement d'un capteur de vision) embarquées ou non sur le système. De nombreux travaux sont basés sur l'exploitation de ces données pour réaliser différents objectifs tels que le positionnement face à un objet, son suivi, sa saisie, etc. La première utilisation de la vision en boucle fermée est due à Shirai et Inoue qui décrivent comment un capteur de vision pouvait augmenter la précision du positionnement [Shi 73]. On parlait alors de retour par vision (visual feedback). Mais c'est à Hill et Park que l'on doit l'apparition du terme asservissement visuel (visual servoing) [Hil 79]. Plusieurs approches ont depuis vu le jour : Sanderson et Weiss ont proposé ainsi deux grandes classes selon l'utilisation des informations visuelles donc on a l'asservissement visuel 3D (ou position-based control) et l'asservissement visuel 2D (ou image-based control) [San 80]. Notons qu'il existe aussi des approches intermédiaires plus récentes telles que l'asservissement visuel 2D $\frac{1}{2}$ ou d2D/dt.

1.3 Gestion des occultations

Le mot occultation, est emprunté du latin *occultatio* : action de cacher ; en astronomie, il signifie la disparition passagère d'un astre caché par un autre.

Habituellement, l'occultation se définit comme étant un phénomène de recouvrement d'un élément par un autre. En asservissement visuel, cela se traduit par la perte des informations sensorielles généralement liée au masquage du motif visuel par un élément de l'environnement. Dans le cadre de notre étude, le fait que la cible est susceptible de sortir du champ de vision de la caméra, est également assimilé à une occultation.

Le suivi de personnes est devenu essentiel pour le service robotique puisque le robot est supposé interagir et naviguer dans un environnement encombré de personnes. La vision référencée capteurs est basée sur l'asservissement visuel, il opère dans l'espace image en se basant sur la caméra pour accomplir différentes tâches.

Les approches de l'asservissement visuel ont été initialement appliquées à un bras manipulateur à 6 degrés de liberté. L'application a été étendue aux robots mobiles depuis 1990. Dans le contexte de la navigation des robots mobiles, trois tâches critiques doivent être prises en considération et plus particulièrement pour le cas des robots non holonomes :

- Les tâches de positionnement tout en prenant en compte les propriétés non holonomes du robot.
- La combinaison entre l'asservissement visuel dans l'image et d'autres tâches tel que l'évitement d'obstacles.
- La visibilité des primitives de l'image pendant l'asservissement plus particulièrement lorsque le capteur de vision est fixé sur la plateforme mobile.

Pour la troisième tâche, assurer la visibilité reste un continuel défi. La visibilité nécessite qu'un nombre minimum d'informations visuelles reste dans le champ de vision de la caméra pendant l'asservissement. Si ces informations viennent à être perdues, l'asservissement visuel échoue. La perte des informations est causée par l'occlusion de la cible ou par l'application d'une loi de commande pour une tâche spécifique telle que l'évitement d'obstacles. Des méthodes ont été développées pour résoudre cette question à travers l'évitement de l'occultation de la cible et ainsi garder les informations visuelles dans le dit champ de vision pendant l'asservissement. Ces méthodes ont été dédiées aux bras manipulateurs avec 6 degrés de liberté et tirent avantage des redondances. Dans le cas de la robotique mobile Chirubini et al. [Chi 08] proposent une approche basée sur la redondance en rajoutant des degrés de liberté pour la caméra. Cette approche permet de garder la cible dans le champ de vision pendant l'évitement d'obstacles. Folio et al proposent une méthode pour prendre en compte la visibilité et les collisions [Fol 05].

Contrairement à la préservation de tous les indices visuels, une première approche permet le changement de ces derniers durant l'asservissement, par contre cette méthode est limitée aux changements partiels de la visibilité, mais encore si le nombre des indices visuels qui apparaissent est plus faible que celui requis, la méthode échoue également [Gar 05].

Dans le but de traiter la question de perte totale des indices visuels, une méthode est introduite, elle applique un algorithme d'estimation des informations visuelles. Elle utilise l'information précédente et l'entrée de commande afin d'estimer la prochaine information nécessaire lorsque celle-ci est perdue (la cible est sortie du champ de vision). Cependant la méthode est sujette aux erreurs d'estimation [Intelligent Robots and Systems, 2008].

Dans le cadre de nos travaux, nous nous intéressons à l'estimation des indices visuels en se basant sur un estimateur qui va gérer les occultations, tout en utilisant les informations envoyées précédemment par la caméra (Kinect).

Le travail présenté dans ce mémoire traite de la robotique mobile d'intérieur, il est placé dans le cadre de l'amélioration de l'autonomie de la plateforme mobile b21r grâce à une meilleure interaction avec le milieu l'entourant à travers ses capteurs embarqués.

Chapitre 2

Gestion de l'occultation

• Introduction

Un défi quotidien auquel notre cerveau doit constamment faire face consiste à s'adapter à l'environnement l'entourant, qui est le plus souvent caractérisé comme étant ambigu et manquant d'informations sensorielles.

Etant le système sensoriel principal, nous comptons sur lui pour recueillir des indices sur le monde extérieur.

Il n'est pas rare que des objets que nous suivant du regard viennent à être caché par d'autres, en d'autres termes, ils peuvent disparaître temporairement de la vue, occlus par des obstacles visuels au premier plan.

Néanmoins, nous sommes souvent en mesure de garder notre regard sur eux tout au long de l'occlusion. Cela implique que le cerveau peut combler les lacunes de l'information sensorielle manquant en extrapolant le déplacement de l'objet à travers l'occlusion.

Dans le cadre de la robotique de service, le suivi de personnes est devenu essentiel puisque le robot est supposé interagir et naviguer dans un environnement encombré. La vision référencée capteurs est basée sur l'asservissement visuel, il opère dans l'espace image en se basant sur la caméra pour accomplir différentes tâches.

Durant cet asservissement, la cible considérée va avoir une certaine trajectoire (qui est la plupart du temps imprévisible) donc, tout en se déplaçant, elle peut être cachée par des objets de l'environnement, ou quitter la scène ce qui causerait la perte de la cible du champ de vision du robot. On dira à ce moment que durant le suivi, la cible vient d'être occultée.

Pour l'estimation de la trajectoire de la cible, deux variantes de procédure existent :

- Une estimation basée sur une vitesse constante de la cible,
- Une estimation basée sur une accélération constante de la cible.

Dans le cadre du mémoire de fin d'étude nous avons traité le cas de l'estimation de la trajectoire de la cible qui se déplace à vitesse constante durant l'occultation.

Dans le cadre du mémoire du master, on s'intéresse à la seconde variante ; à savoir : l'estimation de la trajectoire de la cible qui se déplace avec une accélération constante.

2.1 Présentation du Robot mobile B21r

Le robot mobile *B21r* est une plate-forme expérimentale construite par la société *iRobot* pouvant se déplacer sur un terrain non accidenté ayant comme type de traction, la traction synchrone. Il dispose de quatre roues décentrées orientables tournant selon deux axes : une rotation selon l'axe y pour engendrer la translation, et une rotation selon l'axe verticale au sol pour engendrer une rotation sur lui-même. Ce robot mobile est muni de deux ceintures de capteurs à ultrasons, une ceinture de capteurs infrarouges, un laser, des capteurs tactiles placés le long de ses parois, et une caméra CCD N/B.



Fig. 2.1 : Robot mobile B21r

2.1.1 Modélisation du système robotique

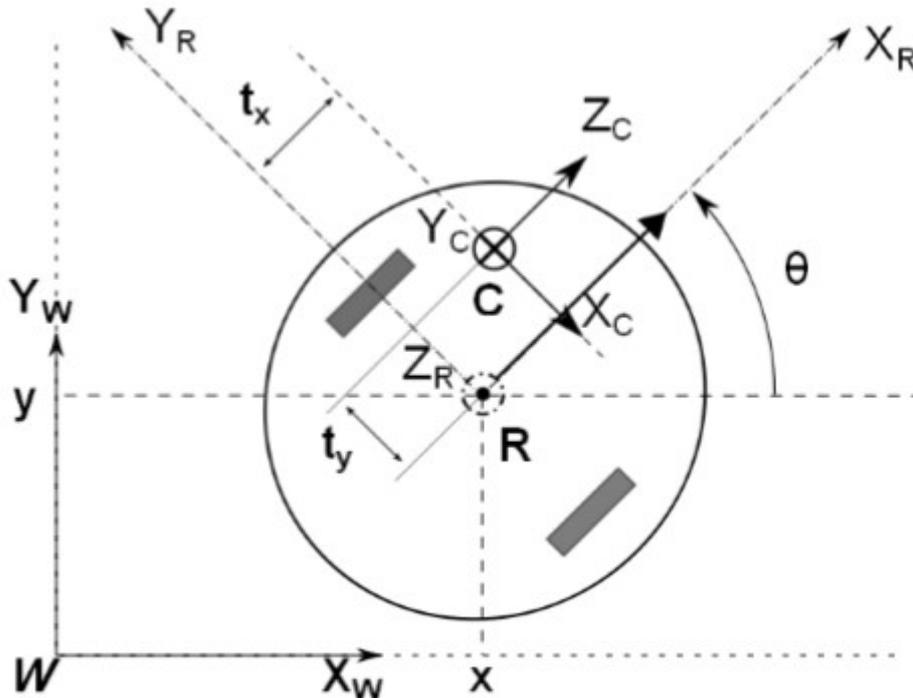


Fig. 2.2 : Repères du système Robot-Kinect

La détermination de la commande d'un système donné, nécessite le plus souvent la connaissance d'un modèle du système. Cette modélisation doit être menée avec rigueur, en vue d'obtenir le modèle le plus proche de la réalité afin de garantir par la suite de meilleures performances en termes de stabilité et de précision.

En effet, nous allons présenter une modélisation explicite et détaillée de la plateforme robotique utilisée. Pour ce faire nous définissons tout d'abord les repères suivants :

- le repère lié à la scène $R_W(W, X_W, Y_W, Z_W)$;
- le repère lié à la base mobile $R_R(R, X_R, Y_R, Z_R)$;
- et le repère lié à la Kinect $R_C(C, X_C, Y_C, Z_C)$.

De plus, nous notons par (t_x, t_y) les coordonnées de l'origine du repère lié à la Kinect dans le repère mobile.

La position du robot mobile est définie par les coordonnées (x, y) du point R dans le repère R_W , par contre son orientation est repérée par l'angle ϑ entre l'axe X_W et l'axe X_R . En outre la position relative de la camera dans la base mobile est définie par le vecteur :

$$\overline{RC} = \begin{pmatrix} t_x \\ t_y \\ 0 \end{pmatrix} \quad (2.1)$$

Le vecteur configuration du robot est défini comme suit :

$$q = \begin{pmatrix} l \\ \vartheta \end{pmatrix} \quad (2.2)$$

Où l représente l'abscisse curviligne du robot dans le repère de la scène ainsi la dérivée par rapport au temps du vecteur configuration représente le vecteur de commande.

Dans la mesure où notre système robotique est commandable en vitesse, il y a lieu de définir la relation entre le torseur cinématique du robot noté V_{Cin} et le vecteur commande noté \dot{q} . Par définition, ce torseur cinématique est constitué des composantes des vitesses de translation et de rotation ; il est alors comme suit :

$$V_{Cin} = [V_C \ \omega_C]^T \quad (2.3)$$

Avec :

$$V_C = (V_{X_C} \ V_{Y_C} \ V_{Z_C})^T \quad (2.4)$$

$$\omega_C = (\omega_{X_C} \ \omega_{Y_C} \ \omega_{Z_C})^T \quad (2.5)$$

La vitesse ${}^C V_{C,W}$ est déduite en exploitant la loi de composition des vitesses :

$${}^C V_{W,C} = {}^C V_{R,C} + {}^C V_{W,R} + {}^C \omega_R \wedge {}^C \overline{RC} \quad (2.6)$$

La Kinect est repérée par le vecteur ${}^C \overline{RC} = (-t_y \ 0 \ t_x)^T$ et comme elle est fixée sur le robot par conséquent sa vitesse, par rapport au repère lié au robot, est nulle et donc :

$${}^C V_{R,C} = \frac{d}{dt} {}^C \overline{RC} = 0 \quad (2.7)$$

En outre, le robot est repéré dans la scène par le vecteur \overline{WR} et sa vitesse par rapport à la scène est alors donnée par :

$${}^cV_{W,R} = \frac{d{}^c\overline{WR}}{dt} \quad (2.8)$$

Donc :

$${}^cV_{W,R} = \begin{pmatrix} 0 \\ 0 \\ v \end{pmatrix} \quad (2.9)$$

Sachant que ${}^c\overline{RC} = (0 \ 0 \ l)^T$ et $v = \frac{dl}{dt}$ le calcul du terme ${}^c\omega_R$ conduit à

$${}^c\omega_R = \begin{pmatrix} 0 \\ \dot{\vartheta} \\ 0 \end{pmatrix} \quad (2.10)$$

En remplaçant les termes dans l'équation on obtient le vecteur suivant :

$${}^cV_{W,R} = \begin{pmatrix} \dot{\vartheta}tx \\ 0 \\ v + \dot{\vartheta}ty \end{pmatrix} \quad (2.11)$$

Par conséquent, le torseur cinématique de la Kinect, par rapport à Roet projeté dans le repère Kinect, est exprimé par :

$$\begin{pmatrix} v_{X_C} \\ v_{Y_C} \\ v_{Z_C} \\ \omega_{X_C} \\ \omega_{Y_C} \\ \omega_{Z_C} \end{pmatrix} = \begin{pmatrix} 0 & tx \\ 0 & 0 \\ 1 & ty \\ 0 & 0 \\ 0 & 1 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} v \\ \dot{\vartheta} \end{pmatrix} \quad (2.12)$$

Donc la matrice Jacobienne est telle que :

$$J = \begin{pmatrix} 0 & tx \\ 0 & 0 \\ 1 & ty \\ 0 & 0 \\ 0 & 1 \\ 0 & 0 \end{pmatrix} \quad (2.13)$$

La forme réduite du Jacobien est obtenue en ne considérant que les mouvements réalisables par la Kinect donc ne sont prises en compte que la translation le long de X_k et celle le long de l'axe Z_k en plus de la rotation autour de l'axe Y_k d'où on obtient :

$$J_r = \begin{pmatrix} 0 & tx \\ 1 & ty \\ 0 & 1 \end{pmatrix} \quad (2.14)$$

2.2 Système de vision

2.2.1 Description matérielle de la Kinect de Microsoft

Le mot « Kinect » est issu des mots anglais « *kinetic* » (qu'on peut traduire par « cinétique ») et « *connect* » (qu'on peut traduire par « connecter »).

La Kinect est une caméra utilisant des techniques d'interaction par commande vocale, reconnaissance de mouvement et d'image.

Ce périphérique de Microsoft est constitué d'une barre horizontale connectée à sa base via un petit moteur. La barre horizontale constitue l'élément principal de la technologie Kinect. Elle est équipée d'une série de multi-microphones, d'une caméra RGB, et enfin d'un capteur de profondeur. L'exploitation de ces trois dispositifs permet la reconnaissance vocale et faciale de l'utilisateur, ainsi que la « capture » de l'ensemble de la pièce (et donc du corps) en 3D.

Caractéristiques en détails

- Capteur :
 - Lentilles détectant la couleur et la profondeur
 - Micro à reconnaissance vocale
 - Capteur motorisé pour suivre les déplacements
- Champ de vision :
 - Champ de vision horizontal : 57 degrés
 - Champ de vision vertical : 43 degrés
 - Marge de déplacement du capteur : ± 27 degrés
 - Portée du capteur : 1,2 m – 3,5 m (à partir de 50 cm pour la version Kinect pour Windows)
- Flux de données :
 - 320×240 en couleur 16 bits à 30 images par seconde
 - 640×480 en couleur 32 bits à 30 images par seconde

- Audio 16 bits à 16 kHz
- Système de reconnaissance physique :
 - Jusqu'à 6 personnes et 2 joueurs actifs (4 joueurs actifs avec le SDK 1.0)
 - 20 articulations par squelette
 - Application des mouvements des joueurs sur leurs avatars Xbox Live
- Audio :
 - Chat vocal Xbox Live et chat vocal dans les jeux
 - Suppression de l'écho
 - Reconnaissance vocale multilingue

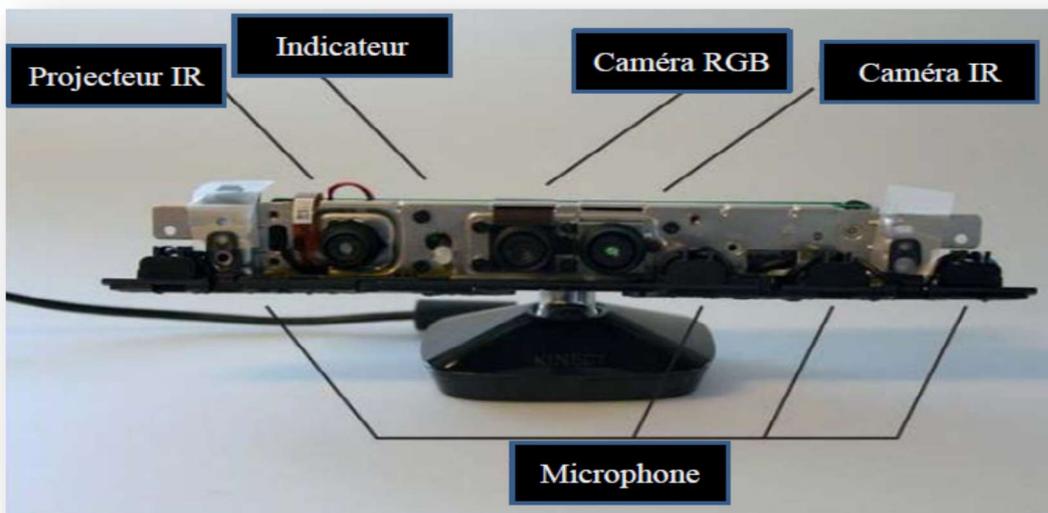


Fig.2.3 : Kinect sans cache

2.2.2 Calibrage d'un capteur de vision stéréoscopique

L'opération de calibration d'une caméra revient à modéliser le processus de formation des images, il consiste à déterminer la relation mathématique existant entre les coordonnées des points 3D de la scène observée et les coordonnées 2D de leur projection dans l'image (points-image). Cette étape de calibrage constitue le point initial pour plusieurs applications de la vision artificielle, comme par exemple la reconnaissance et la localisation d'objets, le contrôle dimensionnel de pièces, la reconstruction de l'environnement pour la navigation d'un robot mobile, etc.

Plusieurs modèles décrivant le processus de formation des images existent. Le plus simple est le modèle du sténopé ou modèle du trou d'épingle (*pinhole* dans la littérature anglo-saxonne). Ce dernier est couramment utilisé en traitement d'image. Ce modèle est développé en détail dans [pfe annexe B].

2.2.3 Généralités sur la Reconnaissance de Personnes dans une Scène

La vision par ordinateur (aussi appelée vision artificielle ou vision numérique) est une branche de l'intelligence artificielle dont le principal but est de permettre à une machine d'analyser, traiter et comprendre une ou plusieurs images prises par un système d'acquisition (par exemple : caméras).

La détection de personnes est un domaine de la vision par ordinateur consistant à détecter un humain dans une image numérique. C'est un cas particulier de détection d'objet, où l'on cherche à détecter la présence et la localisation précise, dans une image, d'une ou plusieurs personnes, en général dans une posture proche de celle de la situation debout ou démarche. On parle également de détection de piéton, en raison de l'importance des applications en vidéosurveillance et pour les systèmes de vision embarqués dans des véhicules.

Étudiée à partir de la fin des années 1990, la détection de personne s'est révélée être un sujet assez difficile, en raison de la grande variété d'apparences des personnes, de l'articulation du corps humain et des phénomènes d'occultations. Bénéficiant des progrès méthodologiques réalisés en détection de visage, la détection de personne a inspiré des méthodes spécifiques, comme les histogrammes de gradient orienté, particulièrement performants. Les méthodes les plus efficaces construisent des modèles statistiques par apprentissage supervisé, à partir de caractéristiques de forme ou d'apparence, calculées sur de nombreux exemples d'images de personnes.

2.2.4 Reconnaissance physique basée sur la Kinect

L'exécution de la tâche de suivi d'une cible par asservissement visuel passe par l'étape de la reconnaissance des points décrivant la cible parmi le nuage 3D observé par la Kinect. Il est à noter que la Kinect détecte un squelette (ensemble de points particuliers liés à une personne) en guise de personne. La Kinect, mise à notre disposition, peut détecter au maximum jusqu'à six squelettes de personnes assises ou debout. Cette caractéristique représente l'une des principales limites actuelles du périphérique. Parmi ces six squelettes, deux seulement peuvent être

reconnus totalement : on parle alors de squelettes « trackés » et les quatre autres sont dits squelettes « non-trackés ». L'image de la **Figure 2.3** illustre cette fonctionnalité.

De la Figure 3, nous constatons que les squelettes « non-trackés » sont gérés différemment par rapport à ceux « trackés ». Dans le premier cas, la Kinect reconnaît un seul point correspondant à la position courante du squelette dans son repère alors que dans le second cas, elle détecte 20 points représentant le squelette de la personne détectée : c'est pourquoi nous parlons de squelettes et non de personnes.

Ces points sont appelés « joints ». Ils correspondent pour la plupart d'entre eux aux articulations de l'homme et constituent le squelette de la personne.

La **Figure 2.4** détaille les 20 joints d'un squelette tracké. Dans le cas d'un squelette non-tracké, un seul joint est pris en compte : HIP_CENTER. Chaque joint, associé à un squelette, est identifié par sa position en X, Y et Z dans le repère 3D ainsi que son orientation par les quaternions dans ce même repère 3D.

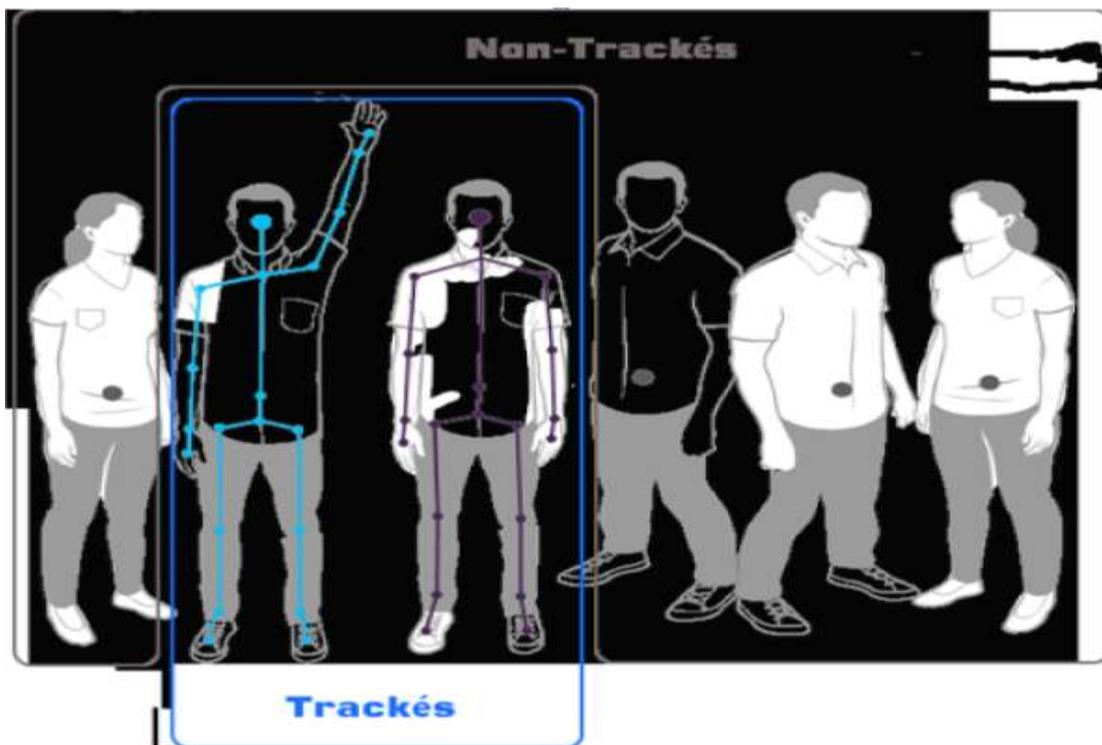


Fig.2.4 Squelettes « trackés » et « non-trackés »

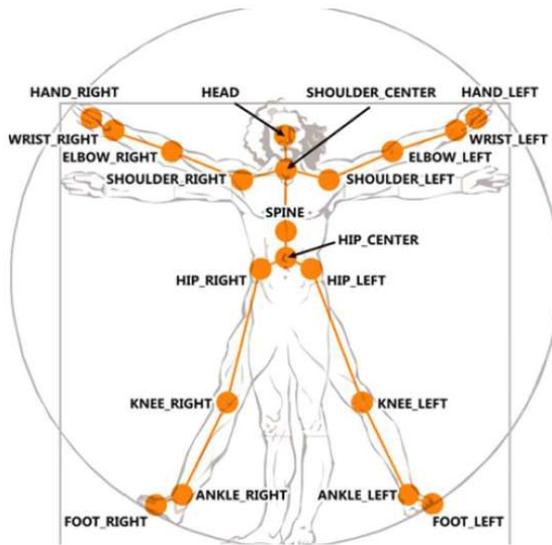


Fig.2.5 Représentation d'un squelette tracké

2.3. Gestion de l'occultation - prédiction de la trajectoire

2.3.1 Généralités et hypothèses de travail

Un des objectifs de notre projet est de gérer les occultations, c'est-à-dire, la disparition de la cible du champ de vision. Nous nous plaçons ainsi dans un scénario où, durant la poursuite de la cible, cette dernière est cachée par un objet (ou une personne) durant un intervalle de temps T_{oc} . Durant T_{oc} , la Kinect donne la dernière position (avant occultation) de la cible durant le laps de temps T_{oc1} , puis n'envoie plus d'informations une fois ce temps dépassé, la cible apparaît dans le champ de vision après une durée T_{oc2} , enfin, le programme de perception repérera la cible après un intervalle de temps T_{oc3} , depuis sa réapparition. Le robot aura alors pour référence une mauvaise position de la cible qu'il poursuivra pendant T_{oc1} , puis s'arrêtera dès que la Kinect n'émet plus d'information pendant $T_{oc2}+T_{oc3}$ pour finalement redémarrer et poursuivre la cible (s'il arrive à retrouver cette dernière).

$$T = T_{oc1} + T_{oc2} + T_{oc3}$$



Cette situation présente deux inconvénients majeurs :

- Le robot arrête de poursuivre la cible durant toute la période T, ce qui n'est pas permis par le cahier de charge.
- Si le robot s'arrête durant l'occultation, la cible réapparaîtra à un endroit éloigné de la position actuelle du robot et risque de ne pas être repérée.

Notre objectif est que le robot continue de poursuivre la cible même si cette dernière est occultée, et ce, jusqu'à ce que la Kinect détecte à nouveau la cible. Pour ce faire, nous effectuons une prédiction de la trajectoire de la cible réelle durant l'occultation. La précision de cette prédiction dépendra de deux paramètres essentiels :

- la durée de l'occultation. En effet, plus la durée d'occultation de la cible est importante, plus les erreurs de prédictions s'accumulent. Ceci conduit le robot à dévier de la trajectoire réelle de la cible et donc il aura moins de chances de la retrouver dans le voisinage de la position prédite quand celle-ci apparaîtra, ce qui a pour conséquence la perte définitive de la cible.
- la prédictibilité de la trajectoire. En outre, plus la trajectoire de la cible est aléatoire (accélération soudaine durant l'occultation, demi-tour...) plus faible sera la chance de retrouver la cible une fois celle-ci réapparue. Le cas idéal serait qu'elle garde la même direction et vitesse qu'avant l'occultation.

Dans le mémoire de fin d'étude nous avons émis les hypothèses de travail suivantes :

- L'occultation est de courte durée (moins de 10 secondes) ;
- Durant le temps d'occultation, la cible ne change pas soudainement de direction ou de vitesse.

2.3.2 Prédiction de la trajectoire

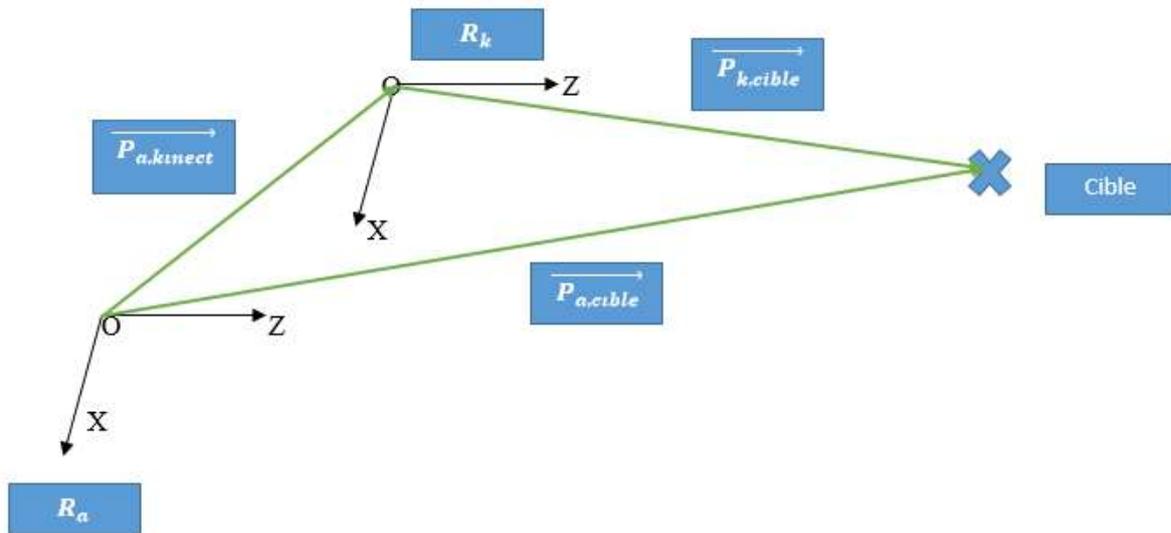


Fig 2.6 Représentation des deux repères et des différentes projections

Soit (R_k) le repère lié à la Kinect à l'instant $t = k\Delta t$ et soit (R_a) le repère lié à la Kinect à $t=0$, appelé repère absolu. Ces deux repères sont bidimensionnels, sur un plan parallèle au sol, il n'y a donc pas de composantes pour exprimer la hauteur bien que la Kinect puisse le faire. Ce choix est justifié par le fait que le robot ne peut pas faire de mouvement vertical, et donc la composante " y_k " (qui est la composante verticale pour la Kinect) est superflue.

Soit ${}^kP_{k,cible}$ et ${}^aP_{k,cible}$ les positions de la cible par rapport au repère (R_k) exprimées respectivement dans les repères (R_k) et (R_a) .

Soit ${}^aP_{a,cible}$ et ${}^aP_{a,kinect}$ les positions respectives de la cible et de la Kinect, par rapport à (R_a) et exprimées dans (R_a) .

Nous noterons aussi ${}^aP_{a,kinect}$ par $P_{kinect} = (x_k, z_k)$.

Soit θ l'orientation de la Kinect (ou du robot) par rapport à (R_a) , qui représente aussi l'orientation de (R_k) par rapport à (R_a) .

${}^kP_{k,cible}$ est le vecteur position donné par la Kinect, la loi de commande utilise ce vecteur pour calculer les vitesses du robot. Ainsi, le rôle de la gestion de l'occultation est de fournir un vecteur ${}^kP_{k,cible}$ à la loi de commande quand la Kinect ne donne plus d'information.

2.3.2.1 Prédiction à vitesse constante

Dans le mémoire de fin d'étude, notre stratégie était de prédire à chaque itération la position de la cible en se basant sur la vitesse précédente calculée de celle-ci et en utilisant un algorithme qui gère les occultations.

Nous préférons la vitesse de la cible dans le repère absolu ${}^aV_{a,cible} = ({}^aV_{x cible}, {}^aV_{y cible})$ car celle exprimée dans le repère de la Kinect ${}^kV_{k,cible}$ n'est pas la vitesse réelle de la cible, cette vitesse dépend aussi de celle du robot. Ainsi, si le robot se dirige vers une cible immobile, ${}^kV_{k,cible}$ prédite serait non nulle, et nous prédirions une cible qui se dirige constamment vers le robot, ce qui le poussera à reculer constamment.

Nous utilisons donc pour la prédiction :

$${}^aV_{a,cible}(k\Delta t) = ({}^aP_{a,cible}(k\Delta t) - {}^aP_{a,cible}((k-1)\Delta t))/\Delta t \quad (2.15)$$

En notant $(k\Delta t)$ par $[k]$; on obtient :

$${}^aV_{a,cible}[k] = ({}^aP_{a,cible}[k] - {}^aP_{a,cible}[k-1])/\Delta t \quad (2.16)$$

On pose

$$H = \begin{pmatrix} {}^ax_{a,cible} \\ {}^ay_{a,cible} \\ {}^aV_{x cible} \\ {}^aV_{y cible} \end{pmatrix} \quad (2.17)$$

$$A = \begin{pmatrix} 1 & 0 & \Delta t & 0 \\ 0 & 1 & 0 & \Delta t \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \quad (2.18)$$

L'étape de prédiction dans le repère absolu est alors :

$$\mathbf{H}_p[\mathbf{k} + 1] = \mathbf{A} \cdot \mathbf{H}[\mathbf{k}] \quad (2.19)$$

Où $\mathbf{H}_p[\mathbf{k} + 1]$ représente le vecteur prédit à l'étape $\mathbf{k}+1$ à partir du vecteur \mathbf{H} donné par la Kinect à l'étape \mathbf{k} .

$\mathbf{H}[\mathbf{k}]$ se renouvelle tant que la Kinect donne encore des informations viables. Dès que la cible est perdue \mathbf{H}_p est calculé à partir du vecteur prédit à l'étape précédente :

$$\mathbf{H}_p[\mathbf{k} + 1] = \mathbf{A} \cdot \mathbf{H}_p[\mathbf{k}] \quad (2.20)$$

2.3.2.2 Prédiction à accélération constante

La stratégie précédente d'estimation ne permet pas de prendre en compte le cas où la vitesse de la cible varie durant l'occultation.

Comme la cible à suivre est un être humain, et que dans la plupart des cas ce dernier à une trajectoire quelconque, la variation de vitesse est très probable à cause des intentions de la cible et des aléas de l'environnement.

Notre stratégie est de prédire à chaque itération la position de la cible en se basant sur la vitesse et l'accélération précédentes calculées.

Comme pour le cas de la prédiction à vitesse constante, nous préférons la vitesse et l'accélération de la cible dans le repère absolu ${}^a\mathbf{V}_{a,cible} = ({}^a\mathbf{V}_{x\ cible}, {}^a\mathbf{V}_{y\ cible})$ et ${}^a\mathbf{a}_{a,cible} = ({}^a\mathbf{a}_{x\ cible}, {}^a\mathbf{a}_{y\ cible})$ car, pour les mêmes raisons, celles exprimées dans le repère de la Kinect ${}^k\mathbf{V}_{k,cible}$ et ${}^k\mathbf{a}_{k,cible}$ ne sont pas la vitesse et l'accélération réelle de la cible, cette vitesse ainsi que cette accélération dépendent aussi de celle du robot.

Nous utilisons donc pour la prédiction :

$${}^a\mathbf{V}_{a,cible}(k\Delta t) = ({}^a\mathbf{P}_{a,cible}(k\Delta t) - {}^a\mathbf{P}_{a,cible}((k-1)\Delta t))/\Delta t \quad (2.21)$$

$${}^a\mathbf{a}_{a,cible}(k\Delta t) = ({}^a\mathbf{V}_{a,cible}(k\Delta t) - {}^a\mathbf{V}_{a,cible}((k-1)\Delta t))/\Delta t \quad (2.22)$$

En notant $(k\Delta t)$ par $[\mathbf{k}]$; on obtient

$${}^aV_{a,cible}[k] = ({}^aP_{a,cible}[k] - {}^aP_{a,cible}[k-1])/\Delta t \quad (2.23)$$

$${}^a\mathbf{a}_{a,cible}[k] = ({}^aV_{a,cible}[k] - {}^aV_{a,cible}[k-1])/\Delta t \quad (2.24)$$

Le vecteur H est agrandi et prend en compte l'accélération.

On pose

$$H = \begin{pmatrix} {}^ax_{a,cible} \\ {}^ay_{a,cible} \\ {}^aV_{x\ cible} \\ {}^aV_{y\ cible} \\ {}^aa_{x\ cible} \\ {}^aa_{y\ cible} \end{pmatrix} \quad (2.25)$$

La matrice A s'agrandit également :

$$A = \begin{pmatrix} 1 & 0 & \Delta t & 0 & \Delta t^2/2 & 0 \\ 0 & 1 & 0 & \Delta t & 0 & \Delta t^2/2 \\ 0 & 0 & 1 & 0 & \Delta t & 0 \\ 0 & 0 & 0 & 1 & 0 & \Delta t \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix} \quad (2.26)$$

L'étape de prédiction dans le repère absolu est alors :

$$H_p[k+1] = A \cdot H[k] \quad (2.27)$$

Où $H_p[k+1]$ représente le vecteur prédit à l'étape k+1 à partir du vecteur H donné par la Kinect à l'étape k.

$H[k]$ se renouvelle tant que la Kinect donne encore des informations viables. Dès que la cible est perdue H_p est calculé à partir du vecteur prédit à l'étape précédente :

$$H_p[k+1] = A \cdot H_p[k]$$

Encore est-il que H est donné dans (R_a) , alors que la Kinect ne donne les positions que dans le repère (R_k).

Il faut donc trouver les coordonnées de ${}^a\mathbf{P}_{a,cible}$ à partir de ${}^k\mathbf{P}_{k,cible}$ pour effectuer la prédiction, puis, passer de la position prédite ${}^a\mathbf{P}_{a,cible}$ à ${}^k\mathbf{P}_{k,cible}$ utilisée par la loi de commande.

$${}^a\mathbf{P}_{a,cible} = {}^a\mathbf{P}_{a,kinect} + {}^a\mathbf{P}_{k,cible} \quad (2.28)$$

$${}^a\mathbf{P}_{a,cible} = \mathbf{P}_{kinect} + \mathbf{R}(\theta) \cdot {}^k\mathbf{P}_{k,cible} \quad (2.29)$$

Où $\mathbf{R}(\theta)$ représente la matrice de rotation du repère (\mathbf{R}_k) par rapport au repère (\mathbf{R}_a) :

$$\mathbf{R}(\theta) = \begin{pmatrix} \cos \theta & \sin \theta \\ -\sin \theta & \cos \theta \end{pmatrix} \quad (2.30)$$

Enfin, pour obtenir à nouveau ${}^k\mathbf{P}_{k,cible}$, nous utilisons la relation :

$${}^k\mathbf{P}_{k,cible} = \mathbf{R}^{-1}(\theta) \cdot ({}^a\mathbf{P}_{a,cible} - \mathbf{P}_{kinect}) \quad (2.31)$$

\mathbf{R} étant une matrice orthogonale, nous avons : $\mathbf{R}^{-1} = \mathbf{R}^T$ et donc

$$\mathbf{R}^{-1}(\theta) = \begin{pmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{pmatrix} \quad (2.32)$$

- **Conclusion**

Dans le cadre de ce master, il est présenté une variante de l'estimation de la trajectoire de la cible qui se base sur une accélération constante de cette dernière.

Suivre une cible au cours du temps signifie qu'on doit prévoir son emplacement à chaque itération si cette dernière n'est plus visible. La méthode la plus simple de prédiction est de supposer que la cible sera proche de sa dernière position connue. En d'autres termes, la détection précédente sert de prédiction suivante.

Une méthode plus sophistiquée de prévision est d'utiliser le mouvement observé précédemment de la cible. L'algorithme de gestion de l'occultation prédit l'emplacement suivant de la cible, en supposant qu'il se déplace selon un modèle de mouvement, tel que la vitesse constante ou une accélération constante.

Bibliographie

[Shi 73] : Shirai, Y. & Inoue, H. (1973), ‘Guiding a robot by visual feedback in assembling tasks’, Pattern Recognition

[Hil 79] : Hill, J. & Park, W. (1979), Real Time control of a robot with a mobile camera, in ‘Proceedings Of the 9th International Symposium On Industrial Robots’,

[Chi 08] : A. Cherubini, F. Chaumette, and G. Oriolo, “A position-based visual servoing scheme for following paths with nonholonomic mobile robots,” in IEEE/RSJ Int. Conf. on Intelligent Robots and Systems, IROS 2008, Nice, France, September 2008.

[Fol 05] : D. Folio and V. Cadenat, “A controller to avoid both occlusions and obstacles during a vision-based navigation task in a cluttered environment,” in Decision and Control, 2005 and 2005 European Control Conference. CDC-ECC ’05. 44th IEEE Conference on, Dec.

[Gar 05] : N. Garcia-Aracil, E. Malis, R. Aracil-Santonja, and C. Perez-Vidal, “Continuous visual servoing despite the changes of visibility in image features,”Robotics,IEEE Transaction son,vol.21,no.6,pp.1214–1220, 2005.

http://www.optique-ingenieur.org/fr/cours/OPI_fr_M04_C01/co/Grain_OPI_fr_M04_C01_1.html

<http://www.cdta.dz/plateforme-robotique/>