Reconnaissance 3D des Gestes pour l'Interaction Naturelle Homme Robot

Hajar Hiyadi, Fakhr-Eddine Ababsa, El Houssine Bouyakhf, Christophe Montagne, Fakhita Regragui

Reconnaissance 3D des Gestes pour l'Interaction Naturelle Homme Robot
3D Gesture Recognition For Natural Human Robot Interaction

H. HIYADI[12] F. ABABSA[1] E. H. BOUYAKHF[2] C. MONTAGNE[1] F. REGRAGUI[2]

[1] Université d'Evry Val-d'Essonne, Evry, France
[2] Université Mohamed V, Rabat, Maroc

Hajar.Hiyadi@ufrst.univ-evry.fr
Fakhr-Eddine.Ababsa@ufrst.univ-evry.fr
Christophe.Montagne@ufrst.univ-evry.fr
bouyakhf@fsr.ac.ma
regragui@fsr.ac.ma

Résumé

Dans ce papier, nous proposons une méthode de reconnaissance 3D des gestes pour l'interaction homme robot (HRI) basée sur l'information de profondeur fournie par la Kinect. Le suivi du corps est réalisé avec l'algorithme Skeleton fourni par le Kinect SDK. L'idée de ce travail est de calculer les angles des articulations de la partie supérieure du corps durant l'exécution du geste. Les variations de ces angles seront les entrées des Modèles de Markov Cachés afin de reconnaître les gestes dynamiques. Les résultats montrent que notre méthode est très robuste ; elle nécessite peu de prétraitements et n'est pas influencée par les conditions de l'environnement comme les changements d'éclairage et la complexité de la scène.

Mots-Clef

Reconnaissance des gestes 3D, Suivi des gestes, Image de profondeur, Modèles de Markov Cachés.

Abstract

In this paper we propose a 3D dynamic gesture recognition method for human robot interaction (HRI) based on depth information provided by a Kinect. The body is tracked using the skeleton algorithm provided by the Kinect SDK. The main idea of this work is to compute the angles of the upper body articulations which are active when executing gesture. The variation of these angles will be used as inputs for a Hidden Markov Models (HMMs) in order to recognize the dynamic gestures. Results demonstrate that our method is very robust ; it does not need much preprocessing steps and is not influenced by environment conditions like illumination changes and scene complexity.

Keywords

3D gesture recognition, Gesture tracking, Depth image, Hidden Markov Models.

1 Introduction

1.1 Motivation

Un geste est un mouvement du corps qui a une signification. Par conséquent, la reconnaissance des gestes a des applications dans de nombreux domaines de recherche. Parmis eux on s'intéresse à l'interaction homme robot (HRI). Le but des applications visées pour l'interaction homme-robot est de parvenir à bien simuler l'interaction homme à homme afin de permettre aux robots de réussir et d'effectuer naturellement l'interaction. Comme dans la communication homme à homme, les gestes sont aussi très utilisées dans la communication homme robot. On distingue deux ensembles d'approches utilisées dans ce domaine : Les méthodes basées sur les dispositifs et celles basées sur la vision. Les méthodes basées sur les dispositifs utilisent généralement des gants pour suivre les mains ou des marqueurs pour suivre la silhouette. Ces méthodes emploient des capteurs mécaniques ou optiques attachés à un gant ou des points marqueurs qui transforment les flexions des membres en signaux électriques pour déterminer la posture. En utilisant ces méthodes, on peut acquérir différentes informations à savoir les angles, les articulations de la main..., etc. Cependant, ces méthodes nécessitent que la personne porte un dispositif pouvant être encombrant et lourd, ce qui freine l'interaction naturelle. Par contre, les méthodes basées sur la vision automatique ne sont pas contraignantes et permettent le suivi et la reconnaissance des gestes sans intermédiaire entre l'humain et le robot. Les capteurs à base de vision comprennent les capteurs 2D et 3D. Cependant, il existe certaines limites dans la reconnaissance des gestes en fonction des images 2D. Tout d'abord, les images peuvent pas être sous un niveau déclairage différent. Deuxièmement, les éléments de l'arrière plan peuvent rendre la reconnaissance plus difficile. Avec l'émergence du capteur Kinect [25], l'information de profondeur ob-

tenue rend possible et pratique d'obtenir non seulement des informations de position, mais aussi des informations d'orientation. Dans cet article notre objectif est d'utiliser uniquement l'information 3D pour construire un système de reconnaissance gestuelle pour l'interaction homme robot.

1.2 Travaux antérieurs

Un système de reconnaissance des gestes contient plusieurs étapes : la détection de la personne, le suivi, l'extraction des gestes et la classification. Le suivi de la main peut être réalisé en se basant sur la couleur de la peau. Pour accomplir cette tâche, la classification de la couleur (sur un espace couleur) est utilisée. Dans [18] la couleur de la peau est utilisée afin d'extraire la main puis suivre le centre de la région qui correspond à la main détectée. La surface extraite représente une ellipse dans chaque espace de chrominance. Cette observation a permis de penser à un modèle statistique de couleur pour la détection de la couleur de la peau, nommé modèle de contour elliptique. L'idée a été étendue dans [6] dans le but de détecter et localiser la tête et les mains. La segmentation est une étape très importante pour le suivi. Ce processus permet de laisser la personne cible et d'éliminer tout ce qui l'entoure. La segmentation basée sur le *clustering* est largement utilisée pour la détection de la peau. On peut citer deux méthodes à savoir les K-means et Expectation-Maximisation dont les avantages ont été combinés dans [19] afin de proposer une nouvelle technique plus robuste (KEM). D'autres méthodes basées sur le template matching 2D/3D ont été développées dans [2][16][11]. En revanche, les approches basées sur la couleur de la peau sont très influencées par le changement de la luminosité et la complexité de l'arrière plan. Par conséquent, les études récentes ont tendance à utiliser une nouvelle information qui est la profondeur. En effet, l'information de profondeur fournie par les capteurs de profondeur augmente les performances des systèmes de reconnaissances des gestes. Il existe beaucoup de travaux qui combinent l'information de profondeur avec la couleur pour la segmentation ainsi que le suivi [1][12][22][8]. Alors qu'il y a d'autres travaux où la parole est ajoutée à ces deux indices [5]. Dans l'article [12], les auteurs utilisent une technique basée sur un modèle de silhouette afin de segmenter le corps humain ; ensuite ils combinent les mouvements et les coordonnées 3D pour suivre l'humain dans la scène. Les approches basées sur les filtres sont de même utilisées pour le suivi comme l'Unscented Kalman Filter [3]. Il existe d'autres méthodes basées sur les points d'intérêt [7] qui ont plus de contraintes sur la fonction d'intensité ce qui les rend plus fiables que les méthodes basées sur les contours. D'ailleurs, les points d'intérêt sont robustes aux occultations et existent dans la majorité des images. Le geste dynamique est caractérisé par sa variabilité spatio-temporelle. En fait, le même geste peut différer dans sa forme, sa vitesse et sa durée. Ces caractéristiques rendent la reconnaissance des gestes dynamiques plus difficile par rapport aux gestes statiques [23]. Concernant la prédiction des gestes, l'article [28] propose un système génère une prédiction d'occupation d'espace de travail humaine en calculant le volume balayé de trajectoires tirées des mouvements humains. Comme dans la reconnaissance de la parole, des caractères et de l'écriture manuscrite [9][10], les Modèles de Markov Cachés (HMMs) ont été utilisés avec succès dans la reconnaissance des gestes [13][20][15]. En fait, les HMMs sont capables de modéliser des séries spatio-temporelles tout en gardant l'identité du geste. Les auteurs de l'article [26] ont développé un système de reconnaissance des gestes dynamiques en se basant sur les orientations de deux articulations du bras gauche fournies par la Kinect. Afin de modéliser et reconnaitre les séquences d'un geste, d'autres modèles mathématiques sont aussi utilisés ; on peut citer : Input-Output Hidden Markov Model (IOHMM) [24], Hidden Conditional Random Fields (HCRF) [21] et Dynamic Time Warping [4].

Dans notre travail, nous nous intéressons à la reconnaissance 3D des gestes dynamiques caractérisés par les mouvements des membres de la partie supérieure du corps. D'abord, l'algorithme Skeleton fourni par la Kinect est utilisé pour le suivi de la partie supérieur du corps.Ensuite, les coordonnées 3D des joints sont extraites et utilisées pour calculer des nouvelles informations pertinentes qui sont les angles entre les membres. Enfin, les modèles de Markov Cachés de typologie Left-Right Banded sont utilisés pour modéliser et classifier les gestes. Nous reprenons à peu près les mêmes gestes utilisés dans l'article [26]. Cependant nous les modélisent avec les angles intérieurs entre chaque deux membres tandis qu'ils utilisent les orientations des joints autour des trois axes (roulis, lacet et tengage). En plus, contrairement à leur base de données entraînée avec une seule personne, la nÃt'tre est entraînée avec 20 personnes ce qui donne un intervalle large de variabilité. Finalement et afin de reconnaître des gestes plus naturels, nous ne fixant pas la durée des gestes. Les résultats donnent des taux de reconnaîssance intéressants par rapport à l'approche [26] et peut atteindre 100% pour certains gestes. Le reste du papier est organisé comme suit : la section 2 décrit notre approche de reconnaissance 3D des gestes dynamiques et les données que nous avons utilisés. La section 3 donne des résultats expérimentaux. Enfin, une conclusion et perspectives sont présentés dans la section 4.

2 L'approche proposée

Dans le cadre de l'interaction homme-robot, notre but est de proposer une nouvelle approche de reconnaissance des gestes dynamiques. En fait, nous nous intéressons aux gestes déictiques notamment les gestes de contrôle et de pointage. Le travail consiste à reconnaître en trois dimensions cinq gestes dynamiques en se basant sur l'information de profondeur. Les cinq gestes que nous voulons reconnaître sont : *viens, recule, stop, pointage à droite* et *pointage à gauche*. La figure 1 représente l'exécution de

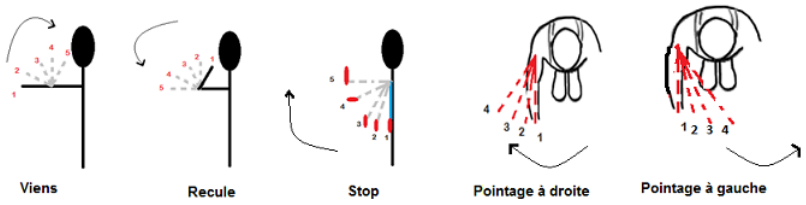

FIGURE 1 – Les cinq gestes à reconnaître.

chacun de ces gestes.

2.1 Suivi et acquisition des données

Avant de passer à la reconnaissance des gestes, il est indispensable d'effectuer d'abord un suivi. Comme mentionné dans l'état de l'art, la plupart des méthodes existantes dans les applications interaction homme robot sont basées sur l'information de la couleur. Or, la couleur n'est pas stable et fortement influencée par de différents facteurs à savoir le changement de la luminosité et les occultations. Par conséquent, les méthodes basées sur la couleur ne parviennent pas toujours à donner une posture 3D de l'humain. Dans notre travail, nous avons choisit d'utiliser le capteur Kinect qui fournit la profondeur. La profondeur est la distance entre le capteur et un point dans la scène. La figure 2 représente le système des coordonnées du capteur utilisé. Les

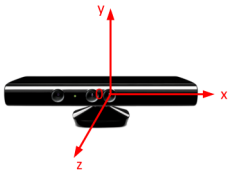

FIGURE 2 – Kinect system coordinate.

coordonnées x, y et z désignent respectivement les positions x, y et la profondeur. En fait, nous effectuons le suivi en utilisant la méthode de Skeleton fournie par la Kinect SDK [1]. Cette méthode projete un squelette sur l'image du corps humain de sorte que chaque articulation du corps est liée à une articulation du squelette projetté et lui associé un *identifiant*. Ainsi, on obtient un ensemble de 20 articulations pour chaque personne détectée. La figure 3 montre la nature de l'information utilisée dans notre approche : l'image de profondeur (b) et le suivi du Skeleton (c). L'idée principale de notre approche est d'estimer en temps réel les variations des angles actifs lors de l'exécution de chaque geste. Les angles considérés sont : coude α, épaule β and aisselle γ, comme montrés dans la figure 4. Chaque angle est calculé à partir des coordonnées 3D des 3 articulations qui le composent comme suit :

– α (l'angle coude) est calculé à partir des coordonnées 3D des articulations : poignet, coude et épaule.

1. http ://msdn.microsoft.com/en-us/library/jj131025.aspx

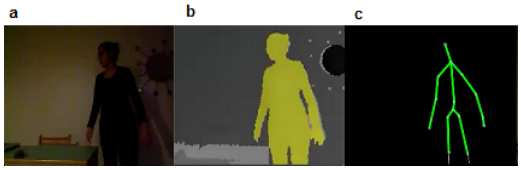

FIGURE 3 – (a) image RGB, (b) image de profondeur, (c) suivi du Skeleton.

– β (l'angle épaule) est calculé à partir des coordonnées 3D des articulations : poignet, épaule droite et épaule gauche.
– γ (l'angle aisselle) est calculé à partir des coordonnées 3D des articulations : coude, épaule et hanche.

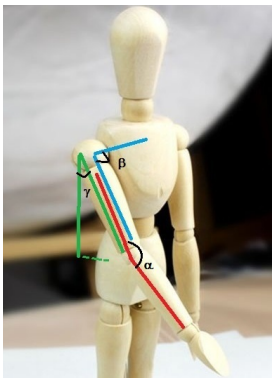

FIGURE 4 – Les angles α, β et γ.

En exécutant un geste, on enregistre les valeurs des angles dans des vecteurs :

$$V_\alpha = [\alpha_1, \alpha_2, ..., \alpha_T] \quad (1)$$

$$V_\beta = [\beta_1, \beta_2, ..., \beta_T] \quad (2)$$

$$V_\gamma = [\gamma_1, \gamma_2, ..., \gamma_T] \quad (3)$$

Avec T la longueur de la séquence du geste. Le vecteur d'entrée pour notre système de reconnaissance sera donc écrit de la manière suivante :

$$V_\alpha = [\alpha_1, \alpha_2, ..., \alpha_T, \beta_1, \beta_2, ..., \beta_T, \gamma_1, \gamma_2, ..., \gamma_T] \quad (4)$$

Cette description basée sur les variations des angles permet une distinction entre les gestes humains. En effet pour

TABLE 1 – Les variations de l'angle principal dans chaque geste.

	α	β	γ
Viens	180° →30°	-	-
Recule	30° →180°	-	-
Pointage à D.	-	90° →150°	-
Pointage à G.	-	90° →40°	-
Stop	-	-	30° →80°

chaque geste on constate qu'il y a un angle principal qui change le plus et prend différentes valeurs tandis que les deux autres angles ne changent que légèrement. Considérons les cinq gestes définis auparavant ; l'angle variable est α dans *viens* et *recule*, γ dans *stop*, et β dans les deux gestes de pointage. Les variations de l'angle principal dans chaque geste sont représentées dans le Tableau 1. Dans ce travail, nous proposons d'utiliser les séquences des variations des angles comme des vecteurs d'entrées de notre système de reconnaissance des gestes présenté dans la section suivante.

2.2 Méthode de classification des gestes

Notre méthode de reconnaissance des gestes est basée sur les Modèles de Markov Cachés (MMCs). Les MMCs sont très utilisés dans la reconnaissance de la parole, l'écriture manuscrite et les modèles temporels. Ils sont capables de modéliser les séries spatio-temporelles quand la forme et la durée du même signale peut changer.

Modèles de Markov Cachés. Un MMC peut être exprimé par le triplet des paramétres $\lambda = (A, B, \pi)$ et décrit par les éléments :

a) Un ensemble de N états $S = \{s_1, s_2, \ldots, s_N\}$.
b) Une distribution de la probabilité initiale pour chaque état $\Pi = \{\pi_j\}$, $j = \{1, 2, \ldots, N\}$, avec π_j = P(S_j à $t = 1$).
c) Une matrice de transition d'ordre N, $A = \{a_{ij}\}$, avec a_{ij} la probabilité de transition de s_i vers s_j ; $1 \leq i,j \leq N$. La somme des éléments de chaque ligne de la matrice doit être égale à 1 parce qu'elle correspond à la somme des probabilités de faire une transition d'un état donné à chacun des autres états.
d) Un ensemble d'observations $O = \{o_1, o_2, ..., o_t\}$, $t = \{1, 2, ..., T\}$.
e) Un ensemble de m symboles discrets $U = \{u_1, u_2, ..., u_m\}$.
f) Une matrice d'observation d'ordre N, $B = \{b_{im}\}$, avec b_{im} la probabilité de génération du symbole u_m par l'état s_i. La somme des éléments de chaque ligne de la matrice doit être égale à 1 pour la même raison citée auparavant.

Un état est une partie du geste. Un geste peut être divisé en N états selon sa complexité ; si le geste est simple le N est petit et vice versa. Chaque état génére un où plusieurs symboles (dits aussi observations). Dans notre application, une observation correspond à un élément du vecteur d'entrée (c'est à dire une valeur d'angle) et un état correspond à une suite d'observations. Il y a trois problèmes majeurs dans les MMCs : l'entrainement, l'évaluation et le décodage qui sont résolus en utilisant, respectivement, les algorithmes Baum Welch, Forward et Viterbi. En outre, les MMCs ont trois topologies : le modèle Ergodique où tous les états sont liés entre eux, le modèle Left-Right où chaque état est lié avec tous les états qui se produisent après et troisièmement le modèle Left-Right Banded où chaque état et lié seulement avec l'état qui le suit. Nous avons choisis le modèle Left-Right Banded parce qu'il est simple, et modélise bien les séries temporelles dont les propriétés changent dans le temps.

Initialisation des paramètres pour le modèle LRB. Nous avons réalisé cinq MMCs, un MMC pour chaque geste. Tout d'abord, il faut initialiser les paramètres de chaque MMC. On commence d'abord par la probabilité initiale. Pour un MMC de 8 états, le premier paramètre qui est le vecteur des probabilités initiales sera donc désigné par :

$$\Pi = (1\,0\,0\,0\,0\,0\,0\,0) \tag{5}$$

Le premier élément vaut 1 afin de s'assurer que le MMC commence par le premier état. Le deuxième paramètre est la matrice de transition A :

$$A = \begin{matrix} a_{ii} & 1-a_{ii} & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & a_{ii} & 1-a_{ii} & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & a_{ii} & 1-a_{ii} & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & a_{ii} & 1-a_{ii} & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & a_{ii} & 1-a_{ii} & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & a_{ii} & 1-a_{ii} & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & a_{ii} & 1-a_{ii} \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & a_{ii} \end{matrix} \tag{6}$$

avec a_{ii} la probabilité de transition entre deux états et initialisée par une valeur aléatoire. Le dernier paramètre est la matrice d'émission B déterminée par :

$$B = \{b_{im}\} \tag{7}$$

avec b_{im} la probabilité de génération d'un symbole par un état et initialisée aléatoirement.

Entrainement et évaluation. Après avoir initialisé les paramètres des MMCs, on passe à l'étape de l'entrainement et l'évaluation. Nous utilisons l'algorithme Baum-Welch pour faire un entrainement complet des paramètres initialisés des MMCs $\lambda = (\Pi, A, B)$. Notre système est entrainé en variant le nombre d'état entre 3 et 12. Après l'entrainement on obtient des nouveaux paramètres $\lambda' = (\Pi', A', B')$ pour chaque geste. Ces nouveaux paramètres sont les entrées des algorithmes Forward et Viterbi pour le teste. Pour un vecteur d'entrée discret l'algorithme Forward calcule la probabilité de son appartenance à chaque MMC en changeant le nombre des états. Ainsi le chemin de Viterbi est reconnu correspondant à la probabilité maximale entre les cinq MMCs.

3 Les résultats expérimentaux

3.1 La base des données

Nous avons construit notre base de données à partir d'une population de 20 individus. Chaque geste est exécuté 5 fois par personne. En somme, on obtient 500 séquences ; 250 séquences sont utilisées pour l'entrainement et 250 séquences pour le test. La division de la base est faite par geste. Chaque MMC est entrainé avec 50 séquences par geste pour l'entrainement et 50 séquences par geste pour le test. Pour la phase de validation, on a demandé à un sujet qui n'a pas l'entrainement d'exécuter chaque geste 10 fois.

3.2 Le protocole expérimental

Avant de commencer les expériences, les sujets doivent connaître le type du geste à exècuter et son déroulement à savoir le début et la fin du geste. La durée de l'exècution d'un geste n'est pas fixée. Le sujet peut execute un geste lentement, comme il peut le faire rapidement. La distance entre la Kinect et le sujet doit être entre 80 cm et 3 m afin de bien détecter la personne tandis que la caméra reste fixe. La figure 5 montre quelques cas où le Kinect n'arrive pas Ãă dÃl'tecter totalement le corps. L'environnement est plus ou moins encombré avec l'absence d'obstacles entre le sujet et la Kinect. Tout au long du geste le sujet est debout et en face de la Kinect.

3.3 Résultats de la reconnaissance

Les variations des angles au cours du temps d'exécution de chaque geste sont présentées sur les figures de 6 à 10 dont chaque figure correspond à l'execution d'un geste. D'après ces figures, on peut constater que chaque geste est distingué par un angle qui varie le plus. Pour déterminer le nombre final des états exigés pour chaque MMC, nous l'avons varié et, d'après les résultats des expérimentations, nous avons retenu et fixé pour chaque MMC le nombre d'état qui fournit le meilleure taux de reconnaissance. Nous avons trouvé que le taux de reconnaissance est meilleur quand le nombre d'états est 11 pour les gestes *viens* et *recule* ainsi que *pointage à droite*, 12 pour le geste *pointage à gauche*, et 8 pour le geste *stop* (figure 11). Par conséquent, nous utilisons ces paramètres dans les prochaines expérimentations. Etant donnée une séquence d'un geste, le geste est reconnu dans un temps moyen de 0,1508 seconde.
Les résultats de reconnaissance sont présentés sur les tableaux 2 et 3. Le tableau 2 correspond à la matrice de confusion deux à deux entre les gestes. Pour le premier geste, il n' y a aucune mauvaise classification tandis que pour les restes on trouve 1 mauvaise reconnaissance pour le geste 3, 2 pour le geste 4 et 4 pour le geste 5. On peut bien remarquer que la méthode proposée donne de très bons résultats de reconnaissance et notamment pour les gestes opposés comme *viens* et *recule*, *pointage à droite* et *à gauche*. Ceci revient au fait que, dans les deux premiers gestes, l'angle qui change est le même mais il varie dans deux sens différents ; il décroît dans le geste *viens* et accroît dans *recule*. Le même raisonnement peut être donné

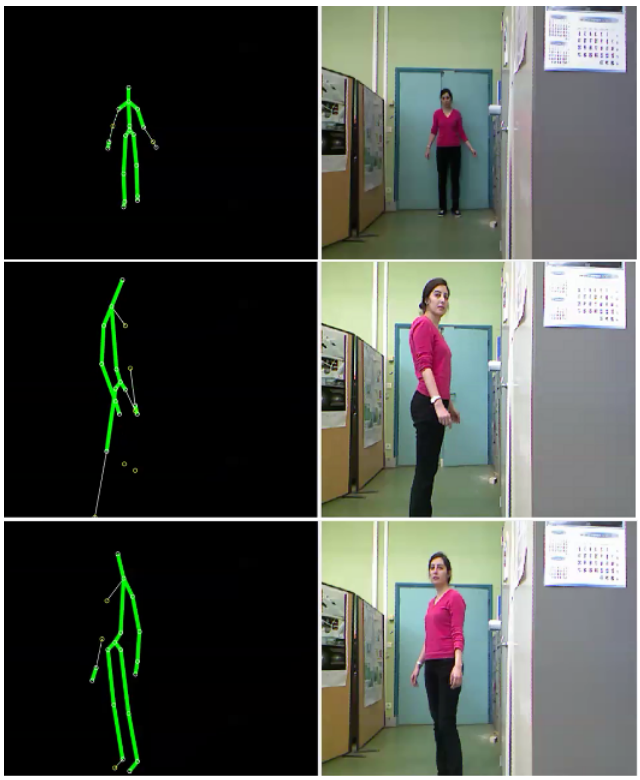

FIGURE 5 – Les cas d'échec de la dètection par le Kinect ; la première image : la distance est supérieure à $3m$, la deuxième et la troisième image : le sujet n'est pas en face du Kinect.

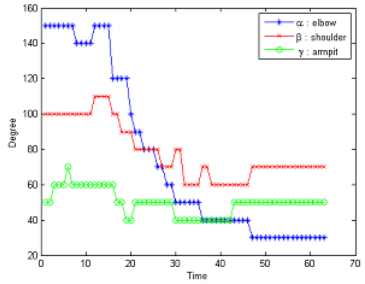

FIGURE 6 – Les variations des angles pour le geste *come*.

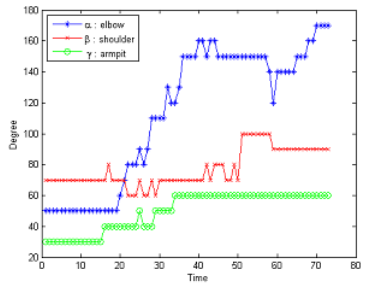

FIGURE 7 – Les variations des angles pour le geste *recede*.

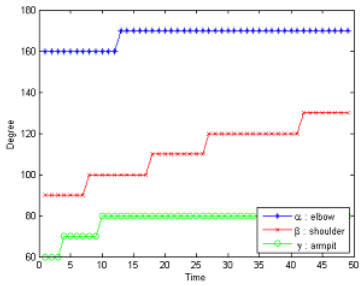

FIGURE 8 – Les variations des angles pour le geste *pointing to the right*.

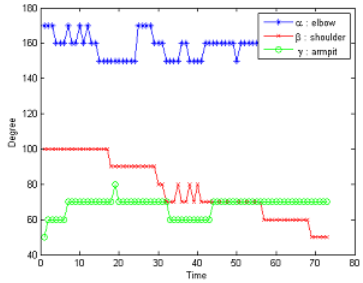

FIGURE 9 – Les variations des angles pour le geste *pointing to the left*.

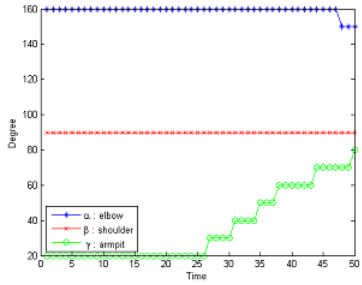

FIGURE 10 – Les variations des angles pour le geste *stop*.

TABLE 2 – La matrice de confusion et la précision des différents gestes avec entrainement (1, 2, 3, 4, 5 désignent resp. *viens, recules, pointage à droite, pointage à gauche, stop*).

	1	2	3	4	5	Précision
1	50	0	0	0	0	100%
2	0	50	0	0	0	100%
3	0	0	49	0	1	98%
4	0	0	0	48	2	96%
5	1	0	0	3	46	92%
Précision moyenne 97,2%						

pour les deux autres gestes. Le tableau 3 représente une comparaison entre notre méthode et celle présentée dans l'article [26]. Les auteurs de cet article utilisent l'orientation des articulations *coude* et *épaule* du bras gauche pour caractériser les gestes. Leur base de données contient cinq gestes. Ils ont entraîné leur base de données avec une seule personne et la testent avec deux personnes. La durée des gestes est fixée au préalable. Le taux de reconnaissance en mode hors ligne et avec les participants à l'entrainement est de 85% pour leur méthode et 97,2% pour notre méthode. Le taux de reconnaissance pour les non participants à l'entrainement est de 73% pour leur méthode et 82% pour la nôtre.

Les gestes que nous avons définit pour l'interaction homme robot sont naturels. Ils représentent presque les gestes utilisés dans nos quotidiens. Par contre, la plupart des méthodes dans l'état de l'art sont basées sur des gestes contraignants en utilisant des signes ce qui rend l'interaction moins naturelle. La méthode de reconnaissance de geste que nous avons proposé est basée uniquement sur l'information de la profondeur, ce qui la rend robuste vis-à-vis le changement de luminosité, les occultations et la complexité de l'arrière plan.

4 Conclusion et perspectives

Nous avons présenté une méthode efficace de la reconnaissance 3D des gestes naturels et dynamiques pour l'interaction homme robot. Nous avons décrit et identifié cinq gestes déictiques qui ont été reconnus en utilisant uniquement l'information de la profondeur. L'idée est d'extraire les coordonnées 3D des articulations qui se trouvent sur la partie supérieure du corps humain et ensuite calculer les angles correspondant à ces articulations. La variation de ces angles est utilisée comme entrée des Modèles de Markov Cachés. Nous avons proposé un modèle MMC pour chaque geste. Voici les caractéristiques de notre système de reconnaissance des geste. Premièrement, la phase d'entrainement est simple, il suffit d'enregistrer le geste lors de son exécution. deuxièmement, Le système peut reconnaître les gestes même si la distance ou l'emplacement des personnes changent. Troisièmement, bien que la vitesse

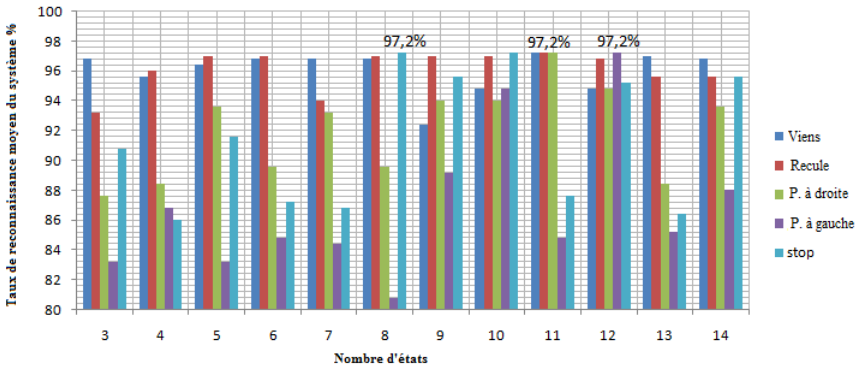

FIGURE 11 – Le taux de reconnaissance moyen du système de reconnaissance en variant le nombre d'état des cinq MMCs de 3 à 14 états.

TABLE 3 – La comparaison des performances de notre méthode avec la méthode de l'article [26] (N. de P. : Nombre de personnes.

Méthodes	Ye et Ha [26]	Notre méthode
Nature des gestes	Dynamique	Dynamique
Info. utilisées	Angles des articulations par rapport au torse	Angles intérieurs des articulations
N. des gestes	5	5
N. d'articulations	2	5
Données utilisées	Ségmentées	Brutes
Classification	MMC	MMC
La base d'entraînement	75	500
N. de p. pour le test	2	20
Durée de geste	Fixe	Variable
Taux de recon.	73%	97,2%

des gestes peut varier d'une personne à autre, le système reste capable de reconnaître le geste. Finalement, le changement de la durée d'un geste d'une personne à autre n'influence pas la reconnaissance. Les résultats expérimentaux montrent que notre méthode donne de meilleurs résultats par rapport à l'état de l'art. En effet, le taux de reconnaissance peut atteindre 100% pour certains gestes. Actuellement, nous avons pu reconnaître juste cinq geste séparés. Dans les travaux futurs, nous voulons étendre notre base de données afin de reconnaitre plusieurs gestes dans la même séquence. Nous voulons aussi combiner l'information de profondeur avec la reconnaissance de la parole pour pouvoir automatiser la détection du début et la fin d'un geste et rendre la reconnaissance plus robuste.

Références

[1] A. Bleiweiss and M.l Werman, Fusion time-of-flight depth and color for realtime segmentation and tracking *DAGM Symposium for Pattern Recognition,* Jena, Germany, 2009, pp 58-69.

[2] A. Barczak and F. Dadgostar, Real-time hand tracking using a set of cooperative classifiers based on haar-like features, *Research Letters in the Information and Mathematical Sciences*, 2005, pp 29-42.

[3] A. Boesen Lindbo Larsen, S. Hauberg, and K. Steenstrup Pedersen, Unscented kalman filtering for articulated human tracking, *17th Scandinavian Conference, SCIA*, Ystad, Sweden, 2011, pp 228-237.

[4] A. Corradini, Dynamic time warping for off-line recognition of a small gesture vocabulary, *ICCV Workshop on Recognition, Analysis, and Tracking of Faces and Gestures in Real-Time Systems*, IEEE (2001), pp 82-89.

[5] C. Matuszek, L. Bo, L. Zettlemoyer and D. Fox, Learning from Unscripted Deictic Gesture and Language for Human-Robot Interactions, *I. J. Robotic*, 2014, pp 581-599.

[6] D. Xu, Y.L. Chen, X. Wu and Y. Xu, Integrated approach of skincolor detection and depth information

for hand and face localization, *IEEE International Conference on Robotics and Biomimetics - ROBIO*, 2011, pp 952 - 956.

[7] D. Koller, S. Thrun, C. PlagemannVarun and V. Ganapathi, Realtime identi

cation and localization of body parts from depth images, *IEEE International Conference on Robotics and Automation (ICRA)*, Anchorage, AK, 2010, pp 3108-3113.

[8] D. Xu, X. Wu, Y. Chen, Y. Xu, Online Dynamic Gesture Recognition for Human Robot Interaction, *IEEE Journal of Intelligent Robotic Systems*, 2014.

[9] G. Saon and J. T. Chien, Bayesian sensing hidden markov models, *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 20, 2012, pp 43-54.

[10] G. Saon and J. T. Chien, Viewing sea level by a one-dimensional random function with long memory, *Mathematical Problems in Engineering*, vol. 2011, Article ID 654284, 2011.

[11] J. Xu, Y. Wu and A. Katsaggelos, Part-based initialization for hand tracking, *The 17th IEEE International Conference on Image Processing (ICIP)*, 2010, pp 3257-3260.

[12] Lu Xia, Chia-Chih Chen and J.K. Agarwal, Human detection using depth information by kinect, *Computer Society Conference on Computer Vision and Pattern Recognition - CVPR*, Colorado Springs, 2011, 15 - 22.

[13] M. Elmezain, A. Al-Hamadi and B.Michaelis, Real-time capable system for handgesture recognition using hidden markov models in stereo color image sequences, *Journal of WSCG*, vol. 16, 2008, pp 65-72.

[14] M. Elmezain, A. Al-Hamadi, J. Appenrodt and B. Michaelis, A Hidden Markov Model-Based Isolated and Meaningful Hand Gesture Recognition, *International Journal of Electrical and Electronics Engineering*, 2009.

[15] N. D. Binh and T. Ejima, Real-time hand gesture recognition using pseudo 3-d Hidden Markov Model, *Proceedings of the 5th IEEE International Conference on Cognitive Informatics (ICCI âĂŹ06)*, 2002, pp 820-824.

[16] Q. Chen, N. Georganas and E. Petriu, Hand gesture recognition using haar-like features and a stochastic context-free grammar, *IEEE Transactions on Instrumentation and Measurement*, 2008, pp 1562-1571.

[17] R. R. Lawrence, A Tutorial on Hidden Markov Models and Selected Applications in Speech Recognition, *Proceeding of the IEEE*, Vol. 77(2), 1989, pp 257-286.

[18] S. Swarup Rautaray, A. Agrawal, A real time hand tracking system for interactive applications, *International journal of computer Applications*, 2011.

[19] S.E. Ghobadi, O.E. Leopprich, K. Hartmann and O. Loffeld, Hand segmentation using 2d/3d images, *Proceeding of image and Vision Computiong* , Hamilton, New Zealand, 2007, pp 64-69.

[20] S. Eickeler, A. Kosmala and G. Rigoll, Hidden markov model based continuous online gesture recognition, *Proceedings of 14th International Conference on Pattern Recognition*, vol. 2, 1998, pp. 1206-1208.

[21] S. Wang, A. Quattoni, L. Morency, D. Demirdjian, T. Darrell, Hidden conditional random fields for gesture recognition, *IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR)*, vol. 2, 2006, pp 1521-1527.

[22] S. Qin, X. Zhu, Y. Yang, Y. Jiang, Real-time Hand Gesture Recognition from Depth Images Using Convex Shape Decomposition Method, *Journal of Signal Processing Systems*, Vol. 74, 2014, pp 47-58.

[23] X. Wang, M. Xia, H. Cai, Y. Gao and C. Cattani, Hidden-Markov-Models-Based Dynamic Hand Gesture Recognition, *Mathematical Problems in Engineering*, vol. 2012, 2012, 11 pages.

[24] Y. Bengio and P. Frasconi, Input-output HMMs for sequence processing, *IEEE Transactions on Neural Networks*, 1996, pp 1231-1249.

[25] Z. Zhang, Microsoft Kinect Sensor and its effect, *Multi Media*, vol. 19, no. 2, pp. 4-10, feb. 2012.

[26] Y. Gu, H. Do, Y. Ou, W. Sheng, Human Gesture Recognition through a Kinect Sensor, *IEEE International Conference on Robotics and Biomimetics*, Guangzhou, Chine, 2012.

[27] O. Lason, K. Nikolaos, A. Argyros, Efficient Model-based 3D Tracking of Hand Articulations using Kinect *Proceedings of the British Machine Vision Conference*, 2011.

[28] J. Mainprice, D. Berenson, Human-Robot Collaborative Manipulation Planning Using Early Prediction of Human Motion *IROS, IEEE*, 2013.