

Perception visuelle du geste de préhension : application à la robotique manipulatrice.

Pablo Negri, Xavier Clady, Maurice Milgram
LISIF - PARC, UMPC (Paris 6)
3 rue Galilée 94200 Ivry-sur-Seine
Pablo.Negri@lis.jussieu.fr

Résumé – Cet article décrit un système d’acquisition visuelle de gestes humains de préhension. L’analyse et la reconnaissance de ce type de gestes est destinée à la commande de robots manipulateurs.

Abstract – This article deals with a visual system to capture human grasping gestures. The recognition of these gestures could be used to control manipulating robots.

1 Introduction

En téléopération et notamment en robotique manipulatrice, l’Interface Homme Machine (IHM) est un élément essentiel. Il permet de traduire les volontés de l’opérateur en consignes pour le robot. Depuis plusieurs années, les interfaces graphiques ou *graphical user interfaces* (GUIs) utilisant un protocole dit WIMPs (*windows, icons, menus, pointing*) sont les plus usitées. Cependant, celles-ci ont été conçues pour des applications de type bureautique et ne sont donc pas adaptées aux nouveaux défis de la robotique.

En effet, le besoin croissant d’interventions à distance hors la présence de l’homme nécessite le développement de robots manipulateurs capables d’effectuer des tâches avec une grande dextérité et une certaine autonomie. Pour cela, il est souvent nécessaire de transmettre au robot des informations de type 3D, pour lesquelles ni la souris ni le clavier ne sont appropriés.

Aussi de nouvelles interfaces dites naturelles (*Perceptual User Interfaces* ou PUIs) voient le jour. Le lecteur pourra se référer à [10] pour une bonne introduction dans ce domaine. Il s’agit de créer des interfaces capables de comprendre la façon dont l’homme interagit naturellement avec ses congénères ou son environnement. Ainsi l’intégration d’une IHM naturelle permet au système d’analyser et de comprendre les ordres émis spontanément par l’opérateur. De telles interfaces peuvent faire appel à diverses technologies qui peuvent être ou non combinées : interfaces haptiques, OCR, reconnaissance de la parole et la vision artificielle.

Dans cet article, nous décrivons les premiers travaux que nous avons entrepris afin de constituer un système intégrant une IHM naturelle destinée à la téléopération d’un robot manipulateur, de type bras et main articulés. Dans ce cadre, nous pensons que l’analyse du geste humain par vision peut s’avérer fort utile. En effet, sa perception et son interprétation permettront la commande intuitive de la saisie et la manipulation d’objets par le système robotique : l’opérateur mime un geste devant une caméra, et

peut contrôler l’exécution de ce geste dans un environnement en Réalité Augmentée intégrant les données acquises par la caméra et les données perçues sur le robot et son environnement (cf. figure 1).

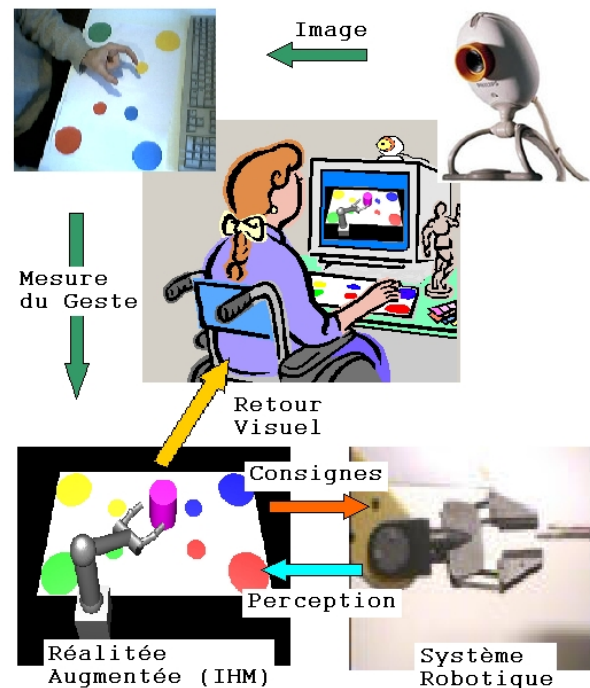


FIG. 1 – Illustration d’un système de téléopération d’une main robotique via une IHM naturelle par vision artificielle.

Dans la section 2, nous explicitons le contexte applicatif de notre IHM. Il est aussi l’occasion de positionner nos travaux par rapport à l’analyse de geste humain par vision. Les sections suivantes sont consacrées à la description du démonstrateur et des outils que nous développons dans ce cadre. Enfin, la dernière section sera l’occasion de discuter des perspectives que nous donnons à ces travaux.

2 Contexte applicatif

Nous proposons d'étudier et de mettre en oeuvre un système de perception visuelle pour une IHM naturelle, où c'est au système d'analyser et de comprendre les ordres émis spontanément par l'opérateur et non à l'opérateur d'apprendre un langage artificiel et des contraintes particulières pour interagir avec lui.

Nous désirons particulièrement nous intéresser au geste de préhension. Sa perception et son interprétation permettront la commande intuitive de l'atteinte et la saisie d'un objet virtuel ou réel par le système, l'opérateur n'ayant qu'à mimer le geste qu'il voudrait voir se réaliser. L'utilisation de la vision pour cette tâche rendra le système peu coûteux et peu encombrant : il pourra n'être constitué que d'une caméra numérique et d'un ordinateur portable.

2.1 Les applications visées

Une des principales applications industrielles que nous visons est la téléopération de robots manipulateurs (ou bras robotiques). Les applications immédiates se trouvent bien sûr dans les milieux hostiles tels que les fonds sous-marins, l'espace ou les environnements radioactifs, mais également dans des domaines plus proches du grand public : les tâches manufacturières, les opérations chirurgicales et surtout l'aide aux handicapés. Ces robots sont appelés à sortir des sites spécialisés pour opérer dans des environnements qui ne leur sont pas adaptés, encombrés d'obstacles fixes ou même mobiles. Aussi, pour de nombreuses tâches, leur téléopération humaine sera recommandée. De plus, ils devront être téléopérés par des personnes non-spécialistes, telles qu'une personne alitée ou hospitalisée. Une IHM naturelle telle que celle que nous proposons, trouve donc un emploi évident.

Dans le même esprit que l'IHM pour les robots, on peut imaginer des jeux en Réalité Augmentée. Par exemple, un jeu d'échecs virtuel où une partie ou l'ensemble des pièces seront virtuelles.

En dehors du domaine de la robotique, d'autres applications sont possibles, telles que l'aide au diagnostic médical. Il s'agirait alors de noter les différences entre les trajectoires prévues, dites "normales", pour un humain et celles réalisées par l'opérateur pour la préhension de l'objet. Et ainsi reconnaître de manière automatique les symptômes de certaines maladies motrices ou psychomotrices affectant la coordination du geste.

2.2 Le geste humain par vision

L'analyse de mouvements humains est un domaine de la vision par ordinateur encore jeune. En effet, percevoir précisément et analyser le mouvement du corps humain posent de grandes difficultés intrinsèques par la nature de celui-ci. Cependant, les progrès importants de l'analyse de séquences d'images, couplés à l'évolution des puissances de calcul, la miniaturisation et le bas coût des capteurs rendent à présent possible la réalisation d'applications tels que l'analyse de mouvement sportif (ex : aide durant l'entraînement ou analyse du jeu adverse), la télésurveillance

(ex : prévention sur sites dangereux, statistiques sur la circulation des clients dans un point de vente).

Au niveau du geste humain, la plupart des travaux de recherche ont été menés autour des langages gestuels [17]. Le langage des signes est celui le plus connu du grand public. Son interprétation par vision présente de nombreuses applications potentielles (ex : traduction automatique du langage des malentendants). Dans certains milieux industriels, des langages gestuels, composés de sémaphores, sont aussi souvent utilisés. Leur analyse automatique par vision peut permettre de constituer des IHM (ex : conduite d'un pont polaire de levage [13]). Cependant, les langages gestuels nécessitent un apprentissage : un opérateur désirant les utiliser doit suivre une formation longue et coûteuse.

Aussi plusieurs chercheurs travaillent sur la mise en place d'interfaces gestuelles plus facilement mémorisables à court et long terme. Cela passe par la définition de classes de gestes communément utilisés ou ayant une symbolique intelligible. Dans un premier temps, l'essentiel de ces travaux ont porté sur des gestes de pointage [7] du fait de la simplicité du modèle statique de la main (main refermée avec un seul doigt, l'index, pointant dans une direction). D'autres [8, 9] définissent des dictionnaires réduits de configurations d'une main, chaque configuration correspondant à une action donnée : dans [9], les actions sont associées au nombre de doigts tendus détectés sur la main. Plus récemment, certains travaux [1] portent sur la reconnaissance de 16 gestes 3D bi-manuels considérés comme naturels (par exemple, le signe STOP est une main levée au niveau du visage paume ouverte, le signe VOLER correspond au geste mimant le vol de l'oiseau,...).

Dans la taxonomie des gestes [16], les gestes décrits précédemment sont dits de communication. Ceux qui intéressent nos travaux se classent dans la catégorie des gestes techniques. Dans ce domaine particulier, les seuls travaux en vision sont, à notre connaissance, ceux de *Triesch et al* [14]. A l'aide d'un capteur stéréo, ils détectent et suivent deux doigts opposés (le pouce et l'index) indiquant une trajectoire pour aller prendre un objet. Ensuite un robot muni d'une pince va suivre cette trajectoire pour atteindre l'objet.

Dans l'idée, le système décrit dans la suite s'inspire dans un premier temps de ces derniers travaux. Il s'agit de concevoir un démonstrateur nous permettant d'acquérir la position de deux doigts lors de la phase d'approche et de saisie. Les données acquises sont ensuite transmises à un environnement de Réalité Virtuelle afin de commander simultanément, et non pas en différé comme dans [14], un robot virtuel muni d'une pince.

3 Description du démonstrateur

Le système complet (cf. fig. 1) est composé, pour la partie matérielle :

1. d'une mire constituée d'une plateforme constituée d'une feuille de taille A3. Huit cercles de quatre couleurs différentes y sont disposés afin de servir d'amers (pour le calcul de la pose de la caméra et la localisation de la main). Cette plateforme est nom-

mée plateforme d'évolution : tous les gestes sont réalisés au dessus.

2. d'une caméra de bas coût, Philips PCVC740K qui est placée en hauteur par rapport à la scène (à environ 60cm) et fournit de la scène des images en contre-plongée, de taille 640x480 pixels dans l'espace couleur YUV
3. et bien sûr d'un ordinateur de type PC, cadencé à 1,6 GHz.

La partie logicielle est organisée en trois modules :

1. le module dit d'acquisition de la scène (localisation de plateforme d'évolution, de la main),
2. un module spécialisée dans l'acquisition et l'analyse du geste,
3. le dernier module doit permettre un retour visuel du geste à l'utilisateur, dans un environnement intégrant le geste et les objets à saisir (ceci pouvant être réalisé en Réalité Augmentée).

Les trois sections suivantes décrivent successivement les travaux réalisés pour constituer ces trois modules.

4 Acquisition et localisation de la scène

L'objectif de ce module est de localiser la main par rapport à la scène. Il s'agit donc de reconstruire les positions 3D de points de la main à partir de leurs positions 2D dans l'image. Pour ce faire, trois principales hypothèses sont posées :

- la plateforme d'évolution est de géométrie connue : les positions des amers de couleur par rapport à un repère lié à la plateforme sont connues.
- la main se déplace dans un plan horizontal (parallèle à la plateforme) sans variation de la composante Z. La distance Z est approximée à 5 cm en moyenne.
- la caméra est statique et est placée à une relative grande distance de la plateforme d'évolution. De sorte que les variations de profondeur des points de la main soient considérées comme négligeables.

Les première et troisième hypothèses nous permettent de calculer la pose de la caméra par rapport à la plateforme d'évolution. Ce calcul est réalisé en utilisant une calibration préalable de la distance focale et du centre image de la caméra (via la méthode de Toscani-Faugeras [11] selon l'implémentation de Jean-Yves Bouguet) et l'algorithme itératif POSIT de Dementhon adapté à des points coplanaires [5, 4]. Ce dernier algorithme permet d'obtenir la matrice homogène de transformation M, composée de la matrice R et du vecteur de translation T, qui permet de passer du repère lié à la plateforme à celui de la caméra. Il utilise en entrée les positions 2D et 3D des centres des

amers couleurs de la plateforme. Les positions 2D sont calculées en modélisant la projection de ces amers comme des ellipses¹. Les paramètres de ces ellipses sont déterminés via une approximation des moindres carrés robustes utilisant les points de contour [2]. Les différentes couleurs des ellipses servent à associer les points 2D (centres des ellipses) avec les points 3D (centres des amers).

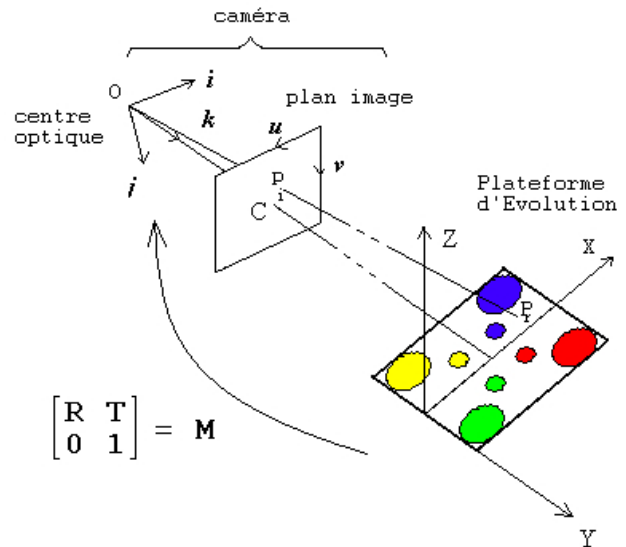


FIG. 2 – Systèmes de repère de la caméra et de celui lié à la plateforme

La seconde hypothèse va nous permettre de localiser dans le repère lié à la plateforme les points appartenant à la main. Connaissant les matrices de transformation entre le plan image et la caméra (via la calibration) et entre la caméra et la plateforme (via POSIT), nous sommes en mesure de calculer les coordonnées de la droite réalisant la projection perspective d'un point 3D, $P_r(x, y, z)$, sur le plan image au point $P_i(u_p, v_p)$. L'intersection de cette droite avec le plan $Z=5\text{cm}$ nous donnent la position 3D d'un point détecté comme appartenant à la main.

Dans cette section, nous avons vu comment localiser un point appartenant à la scène ou à la main dans un repère lié à la scène à partir de la vue de la caméra. Dans la section suivante, nous verrons comment sont segmentés les pixels appartenant à la main, et comment nous analysons ceux-ci afin de déterminer la position des doigts (formant une pince) dans l'image.

5 Acquisition du geste

Le module d'acquisition et d'analyse du geste se limite pour l'instant à l'aspect statique du geste : il s'agit de déterminer la position de la main dans l'image, puis de celles des doigts. Tout d'abord nous réalisons plusieurs pré-traitements qui nous permettent de détecter le mouvement dans l'image par suppression de l'arrière plan, et

¹Les coniques ne sont pas conservées par la projection perspective. Cependant les conditions du démonstrateur (caméra placée en hauteur et relativement loin) s'approche d'une projection affine, ce qui permet de négliger les erreurs dues à notre modélisation.

de détecter la teinte chair. Ensuite, une méthode de vote appelée Transformation Chinoise est utilisée afin de déterminer la position des doigts et leurs orientations.

5.1 Détection du mouvement

Nous utilisons la détection du mouvement pour localiser une région dans l'image où se trouve la main.

Nous appliquons une méthode adaptative de suppression d'arrière plan [6, 3]. La méthode de Stauffer modélise la couleur de chaque pixel avec un nombre maximum K de Gaussiennes. L'algorithme initialise le modèle de l'arrière plan à partir des premières images prises. Il peut avoir une seule Gaussienne dans le modèle d'arrière plan pour des arrière plans fixes, ou plusieurs s'il y a des variations répétitives (ombres,...). Les valeurs du pixel qui ne sont pas comprises dans les Gaussiennes sont considérées comme avant-plan. En permettant l'apprentissage de nouvelles Gaussiennes, le système s'adapte aux changements de lumière, à de nouveaux éléments fixes introduits ou enlevés dans la scène,...

À noter que contrairement à [3], nous n'utilisons pas l'espace RGB, mais uniquement les chrominances de l'espace couleur YCbCr. Cela nous permet d'ignorer les variations de la composant Y qui sont essentiellement dues aux ombres générées par la main ou l'environnement.

À partir de la détection du mouvement, nous définissons une fenêtre de recherche. L'avantage est bien entendu essentiellement la réduction de l'espace de travail, avec pour conséquence un gain en temps de calcul appréciable pour la suite.

5.2 Détection de la teinte chair

Nous désirons obtenir une image en niveaux de gris où les pixels qui ont une chrominance proche à celle de la teinte chair seront nettement distingués. La méthode proposée consiste à calculer une image de distance à la teinte chair dans l'espace des chrominances.

Soit S la séquence des images et I une image de cette séquence dans l'espace RGB. Nous appelons I_{YCbCr} le résultat de la conversion de l'image I , de l'espace RGB à l'espace YCbCr. Premièrement, nous calculons l'inverse à la distance de la peau en soustrayant les valeurs des chrominances b et r de I_{YCbCr} avec b_{chair} et r_{chair} , qui ont été obtenues à partir de la moyenne d'échantillons de teinte chair avec un panel incluant plusieurs différentes personnes :

$$\bar{I}_b = I_{YCbCr}(:, 2) - b_{chair}$$

et

$$\bar{I}_r = I_{YCbCr}(:, 3) - r_{chair}$$

Puis nous réalisons l'addition des valeurs de \bar{I}_b et \bar{I}_r , puis nous renforçons les extrêmes avec la fonction racine carrée :

$$\bar{I}_{br} = \sqrt{\bar{I}_b + \bar{I}_r}$$

Après une normalisation et une inversion de \bar{I}_{br} ,

$$I_{br} = 1 - \frac{\bar{I}_{br}}{\max(\bar{I}_{br})}$$

nous obtenons l'image I_{TC} en niveaux des gris avec les pixels de teinte chair proches à 255. La figure 3 illustre le type de résultats obtenus.

$$I_{TC} = 255 * \frac{I_{br}}{\max(I_{br})}$$

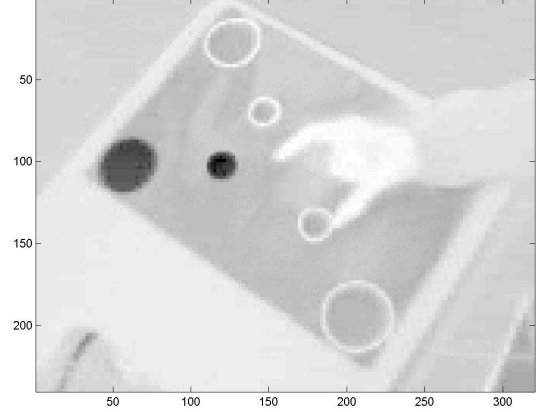


FIG. 3 – Illustration de l'image I_{TC} , représentant la distance à la couleur chair.

5.3 La Transformation Chinoise

La Transformation Chinoise (TC) tient son nom de *Shong-guo*, "l'Empire du Milieu" : c'est ainsi que les Chinois nomment leur pays [12]. Il s'agit d'une méthode de vote : le vote est attribué au centre ou *milieu* de deux points ayant des directions de gradients opposées.

Ainsi l'algorithme de la transformation chinoise utilise la direction et l'intensité du gradient. Dans le cas 2D (cf. figure 3), Ce vecteur gradient G est défini à partir de l'image I_{TC} de tel sorte qu'au point (x, y) , on ait :

$$\vec{G}(x, y) = \begin{bmatrix} \frac{\partial I_{TC}(x, y)}{\partial x} \\ \frac{\partial I_{TC}(x, y)}{\partial y} \end{bmatrix} = \begin{bmatrix} I_x(x, y) \\ I_y(x, y) \end{bmatrix}$$

Le calcul des dérivées peut être réalisé en faisant la convolution de l'image avec deux masques, une pour la direction x , M_x , et l'autre pour la direction y , M_y . Le module du gradient est alors défini :

$$G(x, y) = |M_x * I_{TC}(x, y)| + |M_y * I_{TC}(x, y)|$$

et son orientation :

$$\phi(x, y) = \tan^{-1}\left(\frac{M_x * I_{TC}(x, y)}{M_y * I_{TC}(x, y)}\right)$$

avec :

$$M_x = \begin{bmatrix} -1 & 1 \end{bmatrix} \text{ et } M_y = \begin{bmatrix} -1 \\ 1 \end{bmatrix}$$

Nous déterminons les points dont le module du gradient est supérieur à un seuil (déterminé de manière automatique selon le maximum du module). Les valeurs de la direction du gradient, $\phi(x, y)$, vont de 0 à 2π . Nous échantillons ces valeurs en N directions (fig. 4). N dépend de la résolution angulaire désirée.

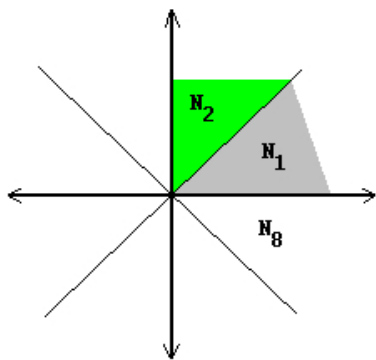


FIG. 4 – Directions du gradient avec $N = 8$.

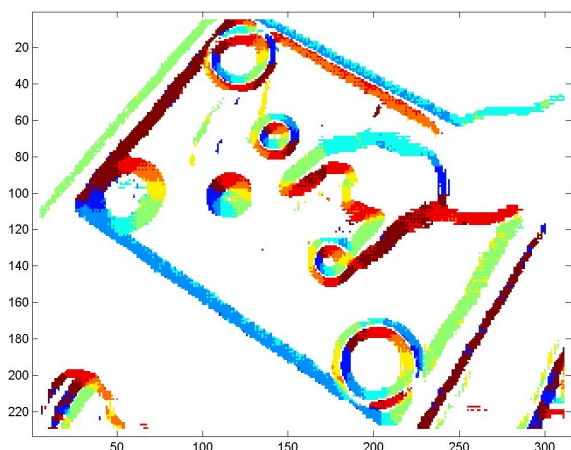


FIG. 5 – Orientation du gradient aux contours de l'image. Les valeurs du gradient ont été codées avec $N = 8$. Chaque orientation est représentée par une couleur différente dans l'image.

Les points de contours et leurs directions obtenues, il reste à déterminer ceux qui sont de directions opposées. Soit M_1 un point du contour, maxima local du module du gradient. Notre but est de chercher des points M_2 du contour (cf. figure 6) tels que :

- La direction du gradient de M_2 est opposée à celle de M_1 ,
- La distance entre M_1 et M_2 est inférieure au seuil s prédéfini : $|\overline{M_2M_1}| < s$.

L'algorithme vote pour le point P_v , point milieu du segment $\overline{M_2M_1}$. Cette opération est répétée pour tous les couples M_1 et M_2 du contour de l'image. Nous pouvons voir les zones d'accumulation dans la figure 6.

Nous profitons de la forme tubulaire de l'index et du pouce (les deux doigts formant la pince). Leurs contours parallèles accomplissent les conditions de distance et des gradients opposés. Les zones d'accumulation trouvées (fig. 7) permettent de définir les régions intérieures des doigts. L'application d'une transformée de Hough sur l'image d'accumulation permet de déterminer les segments correspondant aux doigts.

Les résultats peuvent se comparer à une squelettisation des doigts. Cependant, nous évitons avec la méthode TC de travailler avec des régions. Cela permet d'éviter les pro-

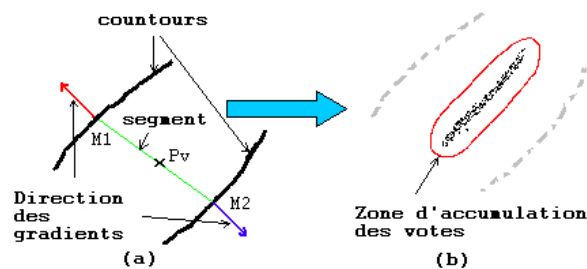


FIG. 6 – Points des votes pour la transformation chinoise. Dans l'image (a), deux points M_1 et M_2 du contour, de directions des gradients opposées. L'algorithme vote pour le point P_v , milieu du segment $\overline{M_2M_1}$. Dans l'image (b), le résultat de l'algorithme est un nuage de points de votes correspondant au traitement de tous les points du contour.

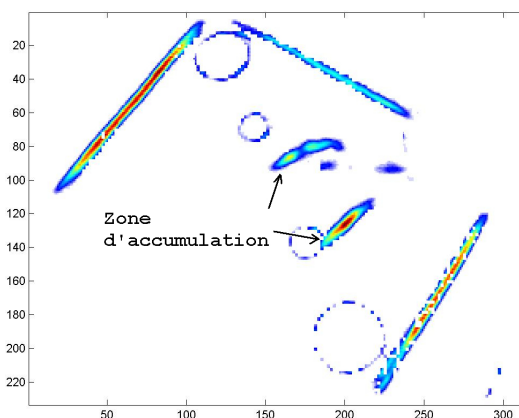


FIG. 7 – Zone d'accumulation des votes de l'image. Les deux nuages qui correspondent aux régions des doigts peuvent être modélisés par deux segments.

blèmes caractéristiques de la squelettisation : sensibilité aux trous, aux lacunes et aux irrégularités des bords.

6 Retour visuel

La localisation de la main et des doigts dans l'image, puis rapport à la scène, nous permet de déterminer la position de l'effecteur (de type pince) d'un robot virtuel (cf. figure 8). Ceci est réalisé simplement avec le modèle géométrique inverse du robot, sans tenir compte de la dynamique du robot.

Actuellement, nous travaillons avec OPENGL. Dans le futur, nous envisageons d'utiliser Open Dynamics Engine (ODE). ODE est une bibliothèque de software gratuit qui peut s'exécuter sous Windows ou Linux. Les avantages principales de cette bibliothèque est la simulation de la dynamique des corps rigides et la détermination des collisions entre objets.

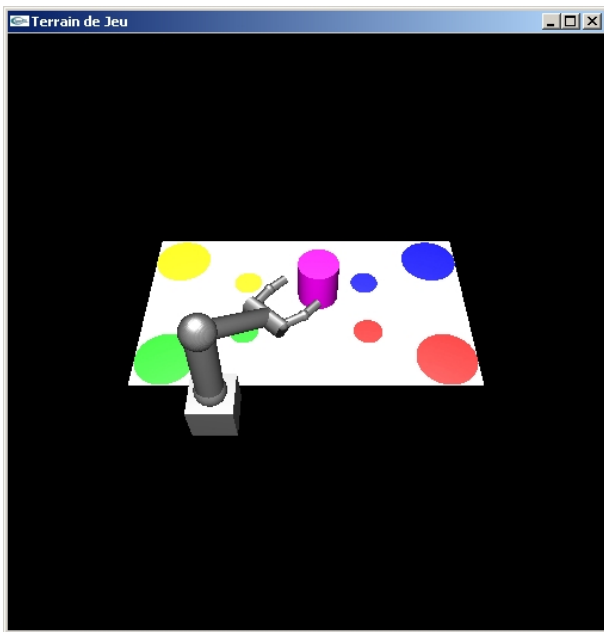


FIG. 8 – Bras robot virtuel

7 Conclusion et perspectives

Les travaux exposés dans cet article représentent une première étape dans la conception d'un système expérimental pour l'acquisition et l'analyse de gestes humains de préhension.

La réalisation de ce système nous a permis l'identification des composantes nécessaires pour la constitution d'une IHM naturelle pour la commande de robots manipulateurs. Ces composantes sont les suivantes :

- Acquisition de la scène.
- Segmentation de la main et reconnaissance du geste.
- Commande du robot virtuel et/ou réel.

Actuellement, nous disposons donc d'un système permettant l'acquisition et la localisation d'une main réalisant une saisie de type pince dans un plan horizontal.

Ce système va nous permettre la mise en oeuvre de méthodes pour reconnaître le geste de préhension d'un objet. Notre ambition est de déterminer, voire d'anticiper, les points de prise d'un objet à partir des caractéristiques intrinsèques (forme, taille,...) et extrinsèques (position, orientation) de l'objet et des caractéristiques du geste de préhension [15].

Ensuite, nous envisagerons d'appliquer ces méthodes à la détermination et l'analyse de gestes 3D de préhension.

Références

- [1] O. Bernier J. E. Viallet A. Just, S. Marcel. Reconnaissance de gestes 3d bi-manuels. In *Atelier : Acquisition du geste humain par vision artificielle et applications*, Toulouse, 27 Janvier 2004.
- [2] R.B. Fisher A.W. Fitzgibbon, M. Pilu. Direct least squares fitting of ellipses. Technical Report DAIRP-794, Department of Artificial Intelligence, The University of Edinburgh, January 1996.
- [3] W.E.L. Grimson C. Stauffer. Adaptive background mixture models for real-time tracking. In *Computer Vision and Pattern Recognition*, volume 2, page 2246, Fort Collins, Colorado, June 1999.
- [4] L.S. Davis D. Oberkampf, D.F. DeMenthon. Iterative pose estimation using coplanar feature points. *Computer Vision and Image Understanding*, 63(3) :495–511, 1996.
- [5] Daniel DeMenthon and Larry S. Davis. Model-based object pose in 25 lines of code. In *European Conference on Computer Vision*, pages 335–343, 1992.
- [6] J. Eklundh E. Hayman. Statistical background subtraction for a mobile observer. In *In the Proceedings of the 9th International Conference on Computer Vision*, pages 67–74, Nice, France, 2003.
- [7] J. Coutaz et J.L. Crowley F. Bérard. Le tableau magique : un outil pour l'activité de réflexion. In *ErgoIHM'2000*, pages 3–6, Biarritz, France, Octobre 2000.
- [8] C. von der Malsburg J. Triesch. Classification of hand postures against complex backgrounds using elastic graph matching. *Image and Vision Computing*, 20(13-14) :937–943, December 2002.
- [9] T. Lindeberg L. Bretzner, I. Laptev. Hand gesture recognition using multiscale color features, hierarchical models and particles filtering. In *Proceedings Face and Gesture*, pages 423–428, Washington DC, 2002.
- [10] M. Kölsch M. Turk. *Emerging Topics in Computer Vision*, chapter Perceptual Interfaces. Prentice Hall PTR, 2005.
- [11] O. Monga R. Horaud. *Vision par Ordinateur : Outils Fondamentaux*. Editions Hermès, Paris, 2nd. edition, 1995.
- [12] Jean-Pierre Respaut. *La Chine : l'empire du Milieu*. Clio, Février 2000.
- [13] R. Marc T. Chateau, F. Jurie. Reconnaissance de gestes par vision monoculaire temps réel : application à la formation des chargés de manoeuvres pour la conduite des ponts polaires. In *RFIA - Atelier de Gestes*, Toulouse, Janvier 2004.
- [14] J. Triesch. *Vision Based Robotic Gesture Recognition*. PhD thesis, Institut für Neuroinformatik, Ruhr-Universität Bochum, Germany, Mai 1999.
- [15] I. Toni M. Jeannerod V. Frak, Y. Paulignan. El eje de oposición es la variable controlada de la prehensión. *Revista Neurológica Argentina*, (24) :14–25, 1999.
- [16] T. S. Huang V. Pavlovic, R. Sharma. Visual interpretation of hand gestures for human-computer interaction : A review. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19 :677–695, July 1997.
- [17] Stan Sclaroff Vassilis Athitsos. Estimating 3d hand pose from a cluttered image. In *IEEE Conference on Computer Vision and Pattern Recognition*, volume 2, pages II– 432–9 vol.2, June 2003.