Perception visuelle du geste de préhension.

Pablo Augusto Negri, Xavier Clady, Maurice Milgram
LISIF - PARC, UMPC (Paris 6)
3 rue Galilee 94200 Ivry-sur-Seine, FRANCE
pablo.negri@lisif.jussieu.fr,
clady@ccr.jussieu.fr,
maum@ccr.jussieu.fr

Résumé – Nous présentons dans cet article, un nouvel algorithme de traitement d'images, la "Transformée Chinoise", permettant d'estimer la localisation des doigts d'une main. Cette approche utilise une technique inspirée de la Transformée de Hough qui prend en compte la disposition des pixels de contour ainsi que l'orientation du gradient en ces pixels. Elle a été intégrée dans un système d'acquisition visuelle monoculaire des gestes humains de préhension.

Abstract – This article describe a new algorithm, the Chinese Transform, for the localization of the fingers. This approach is inspired in the Hough Transform utilizing the position and the orientation of the gradient from the image edge's pixels. A vision system using this technique is proposed, requiring only one camera and a computer.

1 Introduction

Dans la taxonomie des gestes [9], ceux de préhension se classent dans la catégorie des gestes techniques. Ils font l'objet de nombreuses études dans les communautés cognitivistes et médicales, dont [4, 5] font office de travaux fondateurs. Ces études sont souvent réalisées dans l'objectif de déterminer les influences de maladies motrices ou psychomotrices (Parkinson [1], lésions cérébrales [3], etc.) sur la coordination du geste de préhension. Pour la plupart, les expérimentations consistent à disposer sur un plan plusieurs objets, généralement cylindriques, de différentes tailles et à différentes positions. Les cobayes doivent prendre les objets suivant un protocole défini ; tout en gardant la main parallèle au plan. Des émetteurs infrarouges sont installés sur le pouce et l'index, ainsi que sur la paume. Un système de vision dédié, par exemple l'Optotrack, traque ces marqueurs.

Dans ce papier, nous allons proposer un système non intrusif et peu onéreux, ne nécessitant qu'une seule caméra numérique grand public. D'autres applications seraient possibles telles qu'une Interface Homme-Machine naturelle [14] pour la téléopération de robots manipulateurs [13] ou des jeux en Réalité Virtuelle (par exemple, un jeu d'échecs où les pièces seraient virtuelles).

Nous disposons d'une mire, dite Plateforme d'Evolution (PE), composée d'amers circulaires de couleurs différentes et de géométrie connue. La main, formant une pince avec le pouce et l'index, se déplace dans un plan horizontal (parallèle à la plateforme) sans variation de la hauteur (composante Z, estimée à 5cm). Une caméra fixe est placée en contre-plongée, à une distance relativement grande par rapport à la PE, de sorte que la variabilité de profondeur des points de la main puisse être considérée comme négligeable. La position 3D des doigts est calculée à partir de ces hypothèses et en utilisant aussi une méthode itérative d'estimation de la pose de la caméra (cf. fig. 1).

Dans la prochaine section, nous décrivons les procédures

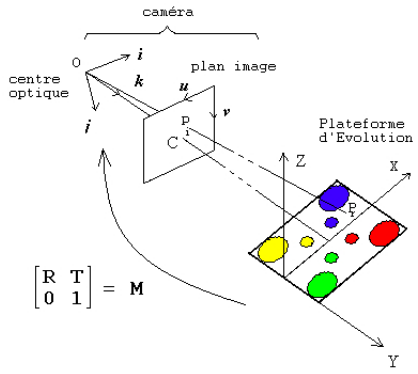

Fig. 1 – Relation "scène-caméra".

pour extraire les positions des doigts dans l'image. Tout d'abord, un algorithme de soustraction de fond nous permet de déterminer une zone d'intérêt dans l'image. Dans cette zone, une image de distance à la couleur de peau est calculée pour l'extraction des bords orientés de la main. Ces bords sont employés dans un algorithme original, la Transformée Chinoise (TC), pour la segmentation et la localisation des doigts. Cette approche est inspirée de la Transformée de Hough. Cet algorithme permet l'extraction des segments des doigts. La section 3 concerne le suivi du geste. Nous utilisons un filtre de Kalman adapté au suivi des segments. Un modèle simplifié permet la reconstruction de la main à partir de ces segments. La section 4 présente quelques résultats obtenus avec notre système. Nous concluons cet article avec quelques perspectives.

2 Détection des doigts

2.1 Localisation de la scène et prétraitements

Les centres des amers colorés nous permettent d'obtenir la pose de la caméra dans la matrice homogène de

transformation M [8]. Cette matrice sert à reconstruire les positions de 3D de points de la main dans les repères liés à la PE, à partir de leurs positions 2D dans l'image. Il existe des solutions plus élégantes [12], permettant de résoudre notre problème sans calcul explicite de la pose ; cependant cette connaissance pourrait s'avérer utile dans des travaux futurs. Dans une évolution future de notre système, nous pourrions ajouter une seconde caméra, liée mécaniquement à la première, qui observerait le visage de l'opérateur, afin de déterminer son champ de vision de la scène. D'où l'importance de localiser la plateforme par rapport au bâti instrumenté des deux caméras.

Pour restreindre la recherche à une fenêtre dans l'image, avec pour conséquence un gain de calcul, nous utilisons une méthode de suppression du fond. La méthode dite de Stauffer et Grimson [11, 2] utilise des mixtures de Gaussiennes pour modéliser la couleur des pixels. Dans notre approche, nous n'utilisons pas l'espace RGB, mais uniquement les chrominances de l'espace couleur YC_bC_r. Cela nous permet de minimiser l'influence des variations de la composante Y (essentiellement dues aux ombres générées par la main ou l'environnement).

De la fenêtre obtenue, nous extrayons une image en niveaux de gris I_{tc}, où les pixels qui ont une chrominance proche à celle de la teinte chair seront nettement distingués. La méthode développée consiste à calculer une image de distance normalisée à la teinte chair dans l'espace des chrominances. L'image I transformé de l'espace RGB à l'espace YC_bC_r est appelée I_{ybr}. Ensuite, nous calculons la soustraction sur chaque composante, rouge (r), et bleue (b), du I_{ybr} avec les valeurs moyennes expérimentales de la teinte chair, b_{tc} and r_{tc} respectivement.

$$\begin{aligned} \overline{I_b} &= |I_{ybr}(b) - b_{tc}| \\ \overline{I_r} &= |I_{ybr}(r) - r_{tc}| \\ \overline{I_{br}} &= \sqrt{\overline{I_b} + \overline{I_r}} \end{aligned}$$

Enfin, nous obtenons l'image de distance à la teinte chair (cf. fig. 4) définie par :

$$I_{tc} = 1 - \frac{\overline{I_{br}}}{max(\overline{I_{br}})}$$

2.2 Transformation Chinoise

La Transformation Chinoise (TC) tient son nom du mot Shongguo, *l'Empire du Milieu* : c'est ainsi que les Chinois nomment leur pays. Il s'agit d'une méthode de votes : chaque vote est attribué au centre ou milieu de deux points ayant des directions de gradients opposées. Cette méthode a le même principe que le travail de Reisfeld [10].

Dans l'exemple de la fig. 2, deux points du contour e_1 et e_2, obtenus de l'image d'une ellipse I_e, sont définis à partir des paramètres suivants : le vecteur normal, n_i, et sa position dans l'image, $p_i(x, y)$, avec $i = \{1, 2\}$. Chaque vecteur normal représente l'orientation du gradient dans ce point. p_{12} est le segment que lie les deux points avec p_v son point milieu.

En superposant e_1 sur e_2, nous pouvons comparer ces orientations. Nous affirmons que e_1 et e_2 ont des orientations opposées si l'angle α_{12} formé entre n_1 et $-n_2$ accomplit la condition :

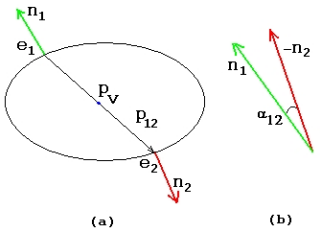

FIG. 2 – (a) montre deux points de l'image et ses vecteurs normals, (b) montre les deux vecteurs normals superposés qui forment un angle α.

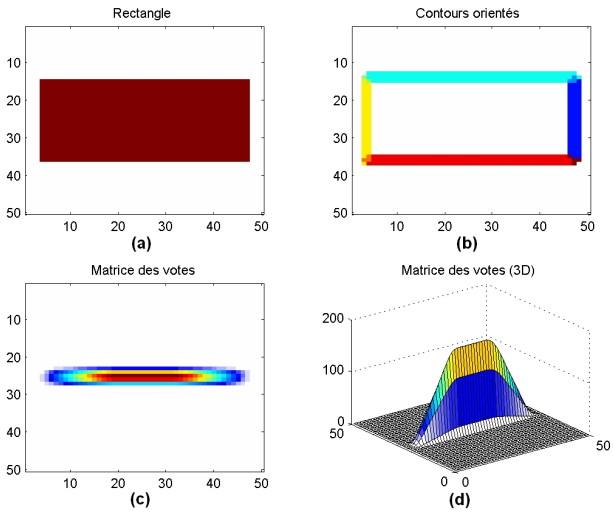

FIG. 3 – Exemple de la TC pour un rectangle. (a) image originale, (b) contours orientés, (c) et (d) matrices des votes en 2D et 3D.

$$\alpha_{12} < \alpha_{threshold} \quad (1)$$

Ensuite, la TC vote pour p_v, le point milieu de p_{12}, si :

$$|p_{12}| < d \quad (2)$$

Nous créons et incrémentons un accumulateur (matrice de votes) avec tous les couples qui accomplissent les conditions (1) et (2).

La fig. 3 est un exemple de l'algorithme de la TC appliqué à un rectangle. La fig. 3.a représente l'image originale et la fig. 3.b montre ces contours orientés en différentes couleurs. Dans la pratique, nous échantillons les orientations du gradient en $N = 8$ directions ; cette opération fixe une valeur pratique pour $\alpha_{threshold}$. La matrice de votes (voir la fig. 3.c et 3.d) est le résultat de l'application de la TC pour tous les points de contour de la fig. 3.a avec $d =$ 35 pixels.

Dans notre application, nous profitons de la forme tubulaire de l'index et du pouce (les deux doigts formant la pince). Leurs bords parallèles satisfont les conditions de direction, de distance et de gradient. Les zones d'accumulation trouvées permettent de définir les régions intérieures des doigts (cf. fig. 4.d). L'application d'une Transformée de Hough sur la matrice de votes permet de déterminer les segments correspondant aux doigts.

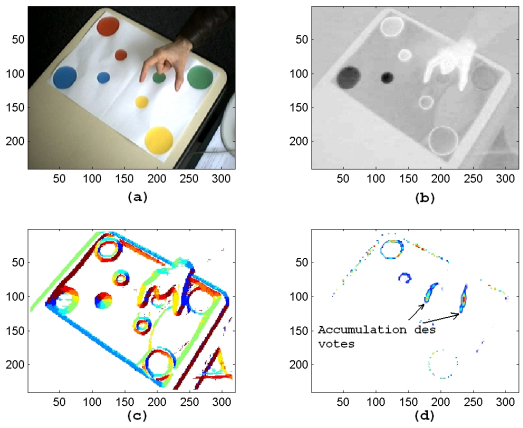

FIG. 4 – Illustration de la TC. (a) Image originale, (b) Distance à la couleur chair I_{tc}, (c) Orientation du gradient aux contour de l'image : chaque orientation est répresentée par une couleur différente, (d) Matrice d'accumulation de votes.

Les résultats de la TC peuvent se comparer à un algorithme de squelettisation morphologique. Cependant, la squelettisation utilise les régions comme primitive et est soumise à ses défauts caractéristiques : les trous, les lacunes et les irrégularités des bords.

3 Suivi du geste et répresentations

Un filtre de Kalman [6] adapté aux segments nous permet de filtrer les pistes, pour ne retenir que celles liées aux doigts formant la "pince" (constituée du pouce et de l'index), et d'éliminer les fauses alarmes.

3.1 Suivi des segments

L'objectif est le suivi des segments appartenant aux doigts dans une séquence d'images. Ceux-ci ont été obtenus à partir de la matrice de votes de la TC et de l'application de la Transformée de Hough. Les paramètres identifiant chaque segment sont (voir fig. 5) : $P_m(x_m, y_m)$, coordonnées du point milieu, l et θ, respectivement la longueur et l'angle du segment.

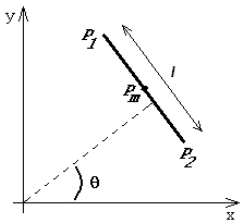

FIG. 5 – Modèle du segment.

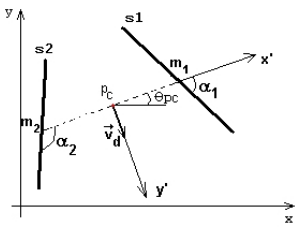

FIG. 6 – Modèle de la pince.

Notre système est composé de trois filtres de Kalman indépendants. Deux filtres scalaires pour la longueur et pour l'orientation, et un filtre vectoriel pour la position. Si nous considérons la vitesse constante, les vecteurs d'état sont :

$$X^{P_m} = \begin{pmatrix} x_m \\ \dot{x_m} \\ y_m \\ \dot{y_m} \end{pmatrix} \qquad X^{l} = \begin{pmatrix} l \\ \dot{l} \end{pmatrix} \qquad X^{\theta} = \begin{pmatrix} \theta \\ \dot{\theta} \end{pmatrix}$$

Nous suivons tous les segments de chaque image de la séquence. Il subsiste des fausses alarmes, qui vont disparaître dans les images successives.

3.2 Suivi de la pince

A partir de deux segments, nous modélisons la pince (cf. fig. 6) avec les paramètres suivants :

- p_c, point milieu du segment m_1m_2,
- Θ_{pc}, angle qui défine l'inclination de la pince par rapport à l'axe x,
- v_d, vecteur unitaire directeur qui définit l'orientation de la pince,
- l_{12}, longueur du segment m_1m_2,

Nous ajoutons deux autres paramètres : les orientations α_1 et α_2 des segments s_1 et s_2, calculées après un changement de repères, de (x, y) au (p_c, x', y'), l'axe lié à la pince, (voir fig. 6).

Les nouveaux vecteurs d'état pour le suivi de la pince sont : X^{p_c}, $X^{\Theta_{pc}}$, X^{α_1}, X^{α_2} et $X^{l_{12}}$.

Tous les couples de segments ne forment pas forcément une pince. Des contraintes sur la longueur l_{12} et les angles α_1 et α_2 nous permettent de filtrer les "fausses pinces". Cette représentation a été inspirée des études réalisées sur le geste de préhension [4, 5, 3, 1]. Elle permet d'observer facilement les principales grandeurs utilisées dans ces études : distance inter-doigts (ouverture de la pince), position et orientation de la main par rapport à la scène (et aux objets). Par ailleurs, elle nous a permis de définir un modèle articulatoire de la main.

4 Résultats

Nous appliquons la TC dans une séquence vidéo composée de trois étapes. La première étape montre une main allant saisir un objet imaginaire dans un coin de la PE (cf. la fig. 7). Dans la prochaine étape, le sujet "pose" l'objet imaginaire dans le coin opposé de la PE. L'étape finale montre la main retournant à la position initiale. Nous présentons les résultats obtenus pour l'étape 1.

Dans la fig. 8.a, nous pouvons voir tous les segments enregistrés durant la première étape. La fig. 8.b montre la trajectoire du point p_c. Les courbes d'évolution de l'ouverture des doigts et de la vitesse de la main pour l'étape 1 sont présentées respectivement dans la fig. 8.c et 8.d. D'après Jeannerod [4, 5], l'acte de préhension d'un sujet normal est divisé en 2 phases : une phase à grande vitesse correspondante au 75% du mouvement d'approche vers l'objet et une phase finale à plus faible vitesse. Dans la fig. 8.d, nous remarquons que la première phase jusqu'au frame 15 est caractérisée par une grande accélération et une augmentation de la distance inter-doigts (fig. 8.c). Ensuite, il y a une désaccélération de la main en s'approchant vers l'objet pour atteindre la distance inter-doigts finale.

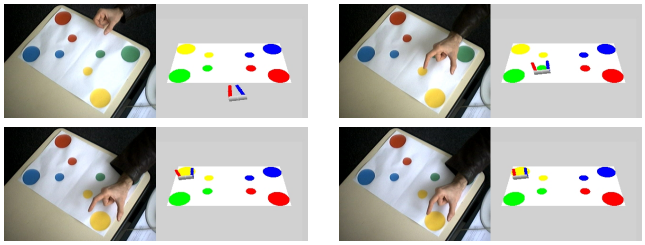

FIG. 7 – Cette figure montre des images de l'étape 1. Sur l'image de gauche, nous pouvons voir le point de vue de la caméra et sur la droite, l'environnement OpenGL avec une pince virtuelle reproduisant simultanément le geste.

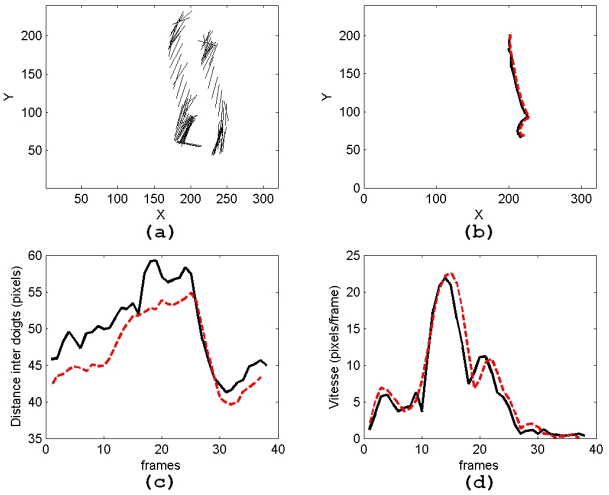

FIG. 8 – Résultats pour l'étape 1 : (a) tous les segments. (b) trajectoire du point p_c (c) distance entre les doigts (d) vitesse de la main. Dans (b), (c) et (d), les courbes en pointillés montrent la vérité terrain.

Ces informations peuvent être employées par les spécialistes afin de mesurer l'habilité du sujet et peuvent également être utiles pour un diagnostic médical.

5 Conclusion and perspectives

Cet article a présenté un système pour l'acquisition du geste humain de préhension. Il utilise une nouvelle méthode pour la détection et localisation des doigts, appelée Transformation Chinoise. Cette technique est une méthode de votes inspirée par la Transformée de Hough. Des filtres de Kalman sont utilisés pour le suivi du geste. Les résultats obtenus sont conformes aux observations des études médicales [4, 5] et peuvent être utilisés pour la détection des maladies psychomotrices.

Les prochaines étapes de nos travaux de recherche seront orientées dans la détermination, voire l'anticipation des points de prise d'un objet [7]. Pour ceci, nous devrons analyser le geste en fonction de ses propres caractéristiques, mais aussi des caractéristiques intrinsèques (forme, taille, etc.) et extrinsèques (position, orientation) de l'objet.

Références

[1] U. Castiello, K. Bennet, C. Bonfiglioli, S. Lim, and R.F. Peppard. The reach-to-grap movement in parkinson's disease : response to a simultaneous perturbation of object position and object size. *Computer Exp. Brain Res*, (125) :453–462, 1999.

[2] E. Hayman and J. Eklundh. 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.

[3] J. Hermdörfer, and Marquardt C. Ulrich, S., and and Mai N. Goldenberg, G. Prehension with the ipsilesional hand after unilateral brain damage. *Cortex*, pages 35 :139–161, 1999.

[4] M. Jeannerod. Intersegmental coordination during reaching at natural visual objects. *Attention and performance (Long J, Baddeley A, eds)*, pages 153–168, 1981.

[5] M. Jeannerod. The timing of natural prehension movements. *Journal of Motor Behavior*, pages 16 :235–254, 1984.

[6] R.E. Kalman. A new approach to linear filtering and prediction problems. *Transactions of the ASME - Journal of Basic Engineering*, (82) :35–45, 1960.

[7] C. Michel, V. Perdereau, and M. Drouin. An approach to grasp planning in constrained environments. In *Proceedings of the IEEE International Conference on Mechatronics and Robotics*, pages 761–767, Aachen, Germany, Septembre 2004.

[8] P. Negri, X. Clady, and Milgram M. Perception visuelle du geste de préhension : Application à la robotique manipulatrice. In *18ème Journée des JJCR*, Douai, France, Septembre 2004.

[9] V. Pavlovic, R. Sharma, and T. S. Huang. Visual interpretation of hand gestures for human-computer interaction : A review. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19 :677–695, July 1997.

[10] D. Reisfeld. *Generalized Symmetry Transforms : Attentional Mechanisms and Face Recognition.* PhD thesis, Tel Aviv University, Janvier 1994.

[11] C. Stauffer and W.E.L. Grimson. Adaptive background mixture models for real-time tracking. In *Computer Vision and Pattern Recognition*, volume 2, pages 22–46, Fort Collins, Colorado, Juin 1999.

[12] Peter Sturm. Algorithms for plane-based pose estimation. In *Proceedings of the Conference on Computer Vision and Pattern Recognition, Hilton Head Island, South Carolina, USA*, pages 1010–1017, June 2000.

[13] J. Triesh and C. von der Malsburg. Classification of hand postures agains complex backgrounds using elastic graph matching. *Image and Vision Computing*, 20(13-14) :937–943, Décembre 2002.

[14] M. Turk and M. Kolsch. *Emerging Topics in Computer Vision*, chapter Perceptual Interfaces. Prentice Hall PTR, 2005.