

Modélisation et suivi de la main

Hocine OUHADDI Patrick HORAIN Krystian MIKOLAJCZYK

Institut National des Télécommunications
Département Signal et Image
9, rue Charles Fourier, 91011 Evry Cedex

<http://www-sim.int-evry.fr>

Prénom.nom@int-evry.fr

1. INTRODUCTION

Les gestes sont un moyen naturel de communication entre personnes. Dans la vie courante, ils viennent ponctuer ou renforcer l'expression orale. Avec la langue des signes, ils constituent un moyen de communication pour les malentendants. Bien des machines, à travers des interfaces mécaniques ou électroniques, sont conçues pour être pilotées par des gestes. Citons par exemple un volant de voiture, une commande de robot télémanipulateur, ou simplement une souris d'ordinateur.

Les possibilités de communications par le geste avec les machines se sont enrichies avec l'apparition de gants instrumentés, munis de nombreux capteurs mesurant les angles des articulations et la position de la main dans l'espace. Ceux-ci ont été utilisés pour commander des acteurs virtuels, pour désigner et manipuler des objets graphiques sur des écrans d'ordinateurs [1, 2] ou même pour la reconnaissance de la langue des signes [5]. Malheureusement, de tels gants restent chers et fragiles, et leurs câbles de liaison constituent une entrave. C'est pourquoi de plus en plus de travaux de recherche s'intéressent à la vision artificielle pour l'acquisition des gestes. Nous les classons en 3 grandes familles.

- *L'analyse 2D* des gestes utilise la silhouette de la main dans l'image ou des points caractéristiques tels que les bouts des doigts pour la reconnaissance de postures. L'analyse est facilitée généralement par l'utilisation de gants ou de marqueurs colorés sur la main de l'opérateur. Une telle approche ne permet pas un traitement simple et général des auto-occultations.
- La reconstruction par *stéréovision* fournit la forme 3D de la main ou du bras à partir de plusieurs images.
- Méthodes par *ajustement et recalage d'un modèle générique 3D* à partir d'une ou plusieurs vues.

Ces deux dernières approches fournissent une interprétation du geste plus riche par estimation de paramètres 3D. L'utilisation d'un modèle générique 3D peut permettre de n'utiliser qu'une seule vue pour déterminer la posture de la main en 3D. C'est pourquoi nous nous intéressons à cette approche.

Dans la première partie de cet article, nous présentons le modèle générique 3D articulé que nous proposons ainsi que son ajustement sur la morphologie de la main. La partie

suivante présente nos travaux en cours pour le recalage et le suivi de la main dans une séquence d'images. La dernière partie présente les résultats obtenus.

2. MODELE DE LA MAIN

Pour décrire les gestes à partir des angles des articulations, nous avons représenté la main par un modèle de type **volumique articulé**.

Celui-ci, conformément au paramétrage MPEG-4 [11], peut être animé suivant les 6 degrés de liberté pour le positionnement global, et 5 degrés de liberté pour chaque doigt. La contrainte dynamique à laquelle obéit le pouce, et celles portant sur les flexions interphalangiennes des 4 autres doigts réduisent le nombre de degrés de liberté de 31 à 26 [10].

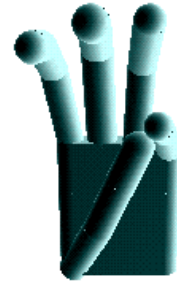


Figure 1. Modèle à base de primitives simples.

3. AJUSTEMENT DU MODELE

Le modèle générique 3D ne correspond pas en général à la morphologie particulière de la main d'un opérateur. L'ajustement préalable des paramètres de morphologie du modèle est nécessaire pour la robustesse de la procédure de suivi de la main dans des séquences. Cette opération est effectuée à partir d'une seule image de la main ouverte (doigts écartés).

La silhouette de la main est obtenue à partir d'un seuillage sur la teinte de la peau et les doigts sont séparés par filtrage morphologique. Les paramètres de la morphologie des doigts (rayon et longueur) et de la paume (longueur et largeur) sont calculés à partir des dimensions des régions correspondantes [9].

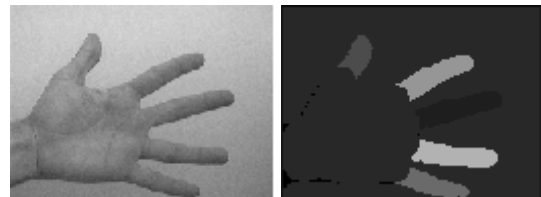


Figure 2. Segmentation de la main par seuillage sur la couleur et séparation des doigts.

4. LE SUIVI DE LA MAIN

La procédure de suivi consiste, après une étape d'initialisation, à répéter sur chaque image les étapes suivantes :

- extraction et projection des contours occultants du modèle,
- extraction des contours de la main dans l'image, puis
- recalage itératif de ces contours à partir de l'état précédant ou initial de ces contours par carte de distances du modèle 3D sur l'image.

4.1 INITIALISATION

Le suivi est initialisé à partir d'une image de la main ouverte, après ajustement morphologique du modèle. Les angles des articulations de la main sont estimés en supposant que toutes les flexions sont nulles et que l'axe optique de la caméra est perpendiculaire au plan de la main. Ceci permet d'identifier l'orientation globale

du modèle dans le plan image à l'orientation de l'axe d'inertie de la silhouette de la main. Les angles d'abduction / adduction des articulations sont estimés à partir des axes d'inertie des doigts.

Les translations dans le plan image sont calculées en ramenant le poignet du modèle de la main à la position du poignet estimée à partir de l'image (calculée durant l'étape d'ajustement morphologique). Le facteur d'échelle est estimé ensuite à partir des positions des bouts des doigts. Cette initialisation peut être affinée en appliquant dès la première image le processus de recalage sur les contours décrit ci-après pour le suivi.

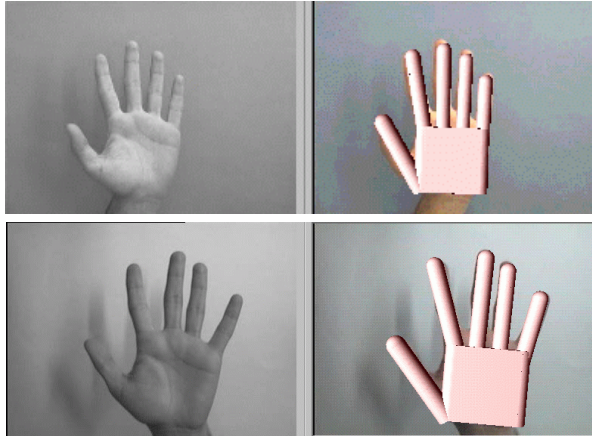


Figure 3. Exemples d'initialisation de suivi sur la main ouverte.

4.2 EXTRACTION ET PROJECTION DES CONTOURS OCCULTANTS

Les contours occultants du modèle (bords des cylindres dans le cas d'un modèle à primitives cylindriques) sont des segments repérés par leurs points extrémités. La position dans l'espace d'un point du modèle articulé est calculée par multiplication des matrices en coordonnées homogènes [7] associées aux degrés de liberté regroupés dans le vecteur p des paramètres du modèle.

La matrice de transformation globale est :

$$M_{globale} = T \cdot R \cdot S$$

où T représente une translation dans l'espace, R une rotation autour d'un axe et S une homothétie.

Dans le cas d'objets articulés, les degrés de liberté internes doivent être pris en compte. Par exemple, la transformation $M(p)$ pour la phalange d'un doigt peut être calculée à partir de :

- M_{doigt} la matrice de transformation de la base du doigt,
- $M_{a/a}^{phal1}$ et $M_{f/e}^{phal1}$ les matrices de transformation de l'abduction / adduction et flexion / extension de la phalange, et
- $M_{f/e}^{phal2}$ la matrice de transformation de la flexion / extension de la phalange,

par la relation :

$$M(p) = M_{globale} \cdot M_{doigt} \cdot M_{a/a}^{phal1} \cdot M_{f/e}^{phal1} \cdot M_{f/e}^{phal2}$$

Désignant par \underline{X} les coordonnées d'un point de la phalange dans le repère lié à celle-ci, les coordonnées dans l'espace de ce point du modèle articulé sont :

$$\underline{X}(p) = M(p) \cdot \underline{X}$$

Les matrices associées aux articulation ainsi que leurs dérivées partielles sont précalculées. Les points de projection des segments de droites représentant les limbes sont ensuite obtenus par l'algorithme de Bresenham [7]. Pour simplifier, nous avons modélisé la caméra par une projection orthographique suivant l'axe z . Dans le cas d'un modèle à primitives maillées, les limbes peuvent être simplement obtenus par un algorithme de Z-buffer [3].

4.3 EXTRACTION DES CONTOURS DANS L'IMAGE

Les contours occultant de la main sont des limbes, c'est-à-dire que, à l'approche de ces contours, la normale à la surface devient progressivement perpendiculaire à la direction d'observation. Dans le cas où les doigts sont jointifs ou cachent une autre partie de la main, les contours occultants apparaissent comme des lignes sombres dans l'image, c'est-à-dire des minimums locaux des niveaux de gris.

Les opérateurs de différentiation, qui détectent les maximums de la pente des niveaux de gris, vont dans ce cas donner deux réponses de part et d'autre de ces lignes sombres. Nous avons préféré rechercher les contours internes de la main comme étant des thalwegs de l'image, c'est-à-dire les vallées des niveaux des gris considérés comme un relief [8]. Les contours externes sont ceux de la silhouette qui est obtenue comme à l'étape d'ajustement du modèle.

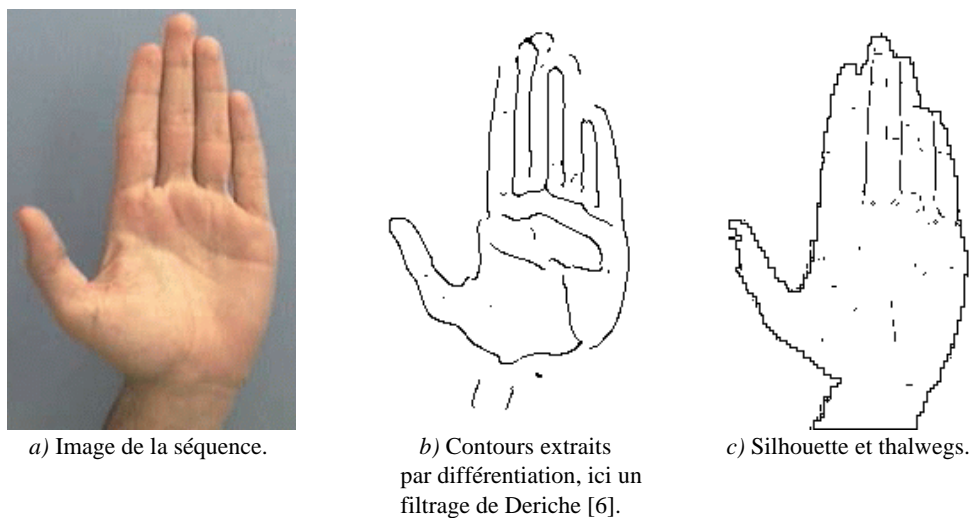


Figure 4. Estimation des contours dans l'image.

4.4 RECALAGE SUR LES CONTOURS

Le recalage du modèle consiste à rapprocher les limbes projetés des contours extraits de l'image. Cette correspondance fournit les positions et les orientations des objets 3D par rapport à la caméra. Lorsque l'on dispose d'une estimation initiale proche de la solution optimale, l'utilisation d'une carte de distances évite d'avoir à rechercher un appariement explicite des primitives. Une transformation en distances convertit l'image

binaire des contours en une image qui associe à chaque pixel sa distance au pixel de contour le plus proche [4]. Elle réalise un appariement implicite des points du modèle projeté avec les pixels des contours de l'image (fig. 5).

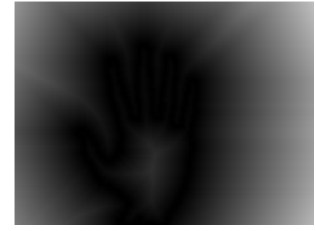


Figure 5. Contours de la main et carte de distances.

La fonctionnelle à minimiser est la somme des distances pour les pixels estimés sur les projections des arêtes du modèle. Son expression est donnée par :

$$F = \min_p \sum_{n_e} \sum_{n_p} d(\text{Proj}(X(p)))$$

où n_e est le nombre des arêtes du modèle, n_p est le nombre de pixels de la projection de l'arête et $d(\text{Proj}(X(p)))$ est la distance (lue dans la carte de distances) du pixel représentant le point X du modèle projeté dans l'image pour le vecteur de paramètres p .

L'erreur d'appariement est minimisée par une méthode d'optimisation non linéaire qui s'apparente à une descente de gradient de type Levenberg-Marquardt [9]. Les paramètres correspondant à cet optimum constituent l'estimation de l'attitude du modèle. Cette méthode nécessite la connaissance des dérivées partielles par rapport aux paramètres de position et d'orientation à estimer. Ceux-ci sont calculés en combinant les dérivées partielles des positions des points projetés des arêtes du modèle avec le gradient de la carte de distance.

Les dérivées partielles sont estimées à partir du gradient de la fonctionnelle calculé comme suit :

$$\nabla_p d(X(p)) = \sum_{n_e} \sum_{n_p} \nabla_p d(X(p))$$

$$\nabla_p d(X(p)) = \nabla_{X(p)} d \cdot \nabla_p X(p)$$

où $\nabla_{X(p)} d$ est le gradient numérique de la carte de distance au point X et $\nabla_p X(p)$ est le gradient de $X(p)$ par rapport au vecteur des paramètres p .

A l'approche de l'optimum, il peut arriver que la valeur de la fonctionnelle ne décroisse plus et que l'ajustement soit prématurément bloqué. Ce problème provient de la nature discrète de la carte de distance, lorsque les distances sont lues au plus proche voisin et que le pas de descente du gradient devient petit. Il est alors nécessaire d'échantillonner de plus en plus finement la carte de distance par interpolation, jusqu'à convergence de la procédure d'optimisation.

Sur les séquences d'abduction / adduction que nous avons traitées, nous nous sommes contentés d'utiliser les contours des 2 dernières phalanges de chaque doigt. On pourra utiliser des primitives supplémentaires pour suivre des mouvements plus complexes. La conformité des angles des articulations aux contraintes statiques est contrôlée et les valeurs incorrectes sont ramenées dans l'intervalle permis.

4.5 SUIVI DANS LES SEQUENCES

La procédure de recalage pour chaque image de la séquence est initialisée à partir de l'estimation obtenue pour l'image précédente (ou à l'initialisation de suivi). Cette procédure est stabilisée par un filtre de Kalman étendu. Le vecteur d'état à estimer est constitué de paramètres globaux de mouvement de la main, des angles des articulations, ainsi que leurs vitesses respectives. Le vecteur des observations est constitué des coordonnées des milieux, des longueurs et des orientations des projections des limbes recalés sur les contours par la procédure d'optimisation.

En cas de changement brusque du mouvement, il peut arriver que la procédure de recalage soit initialisée trop loin de la solution souhaitée et que la procédure d'optimisation fournisse un ajustement erroné (fig. 7). Nous envisageons de lever cette limitation en utilisant un filtre adaptatif combiné à une approche probabiliste.

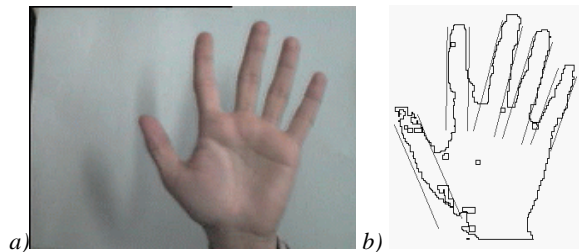


Figure 6. Recalage sur les contours
a), c) Images originales
b), d) Recalage sur la silhouette et, éventuellement, les thalwegs.

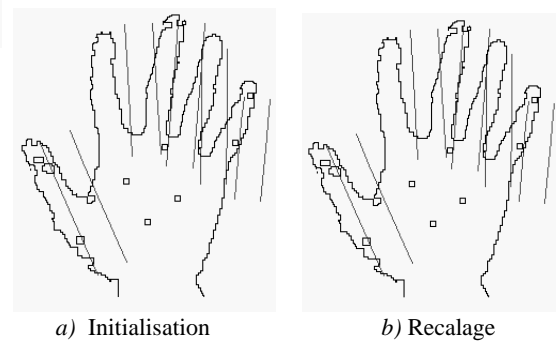


Figure 7. Exemple de recalage incorrect lorsque l'initialisation est trop éloignée de la solution en cas de changement brusque du mouvement.

5. RESULTATS

Nous avons utilisé cette procédure de recalage sur une séquence de 28 images présentant des mouvements d'abduction / adduction de la main. Nos premiers résultats montrent que l'algorithme de recalage avec les cartes de distance permet d'estimer les paramètres globaux du mouvement de la main dans un plan perpendiculaire à l'axe optique de la caméra ainsi que les angles d'abduction / adduction des différents doigts (fig. 8).

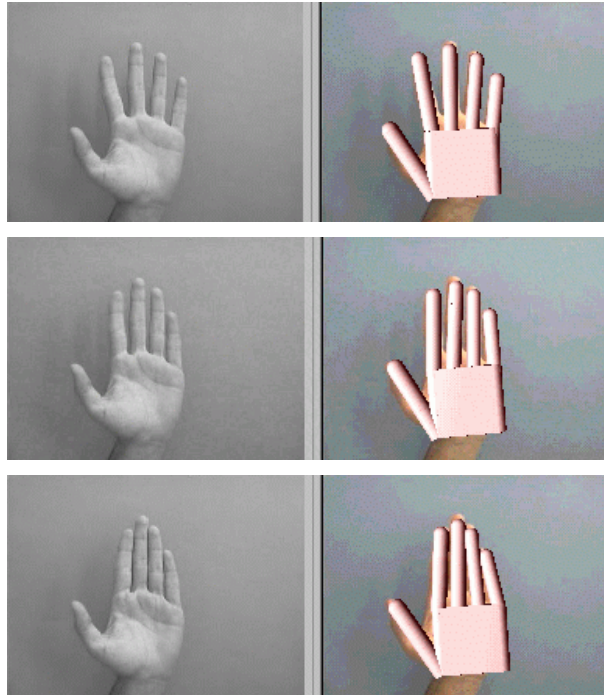


Figure 8. Recalage sur quelques images de la séquence abduction / adduction (28 images).

6. CONCLUSIONS ET PERSPECTIVES

Nous avons présenté dans cet article nos travaux en cours pour le suivi de la main dans une séquence d'images à partir d'une seule vue. La main est représentée par un modèle 3D articulé. Les mesures morphologiques de modèle sont ajustées automatiquement sur la main ouverte de l'opérateur dans la 1^{ère} image et les degrés de liberté sont initialisés pour le suivi.

Les degrés de liberté du modèle estimés sont les paramètres globaux de mouvement (translations, rotations et facteur d'échelle) ainsi que ceux correspondant abductions / adductions des doigts. Ils sont calculés par une procédure de recalage 3D/2D.

La méthode de recalage consiste à estimer les degrés de liberté par minimisation de l'écart entre les contours occultants du modèle et les contours extraits de l'image. L'écart est évalué à partir d'une carte de distance. La minimisation est réalisée par descente de gradient et interpolation de la carte de distance. Actuellement, nous n'avons pas encore traité les angles de flexion / extension. Leur estimation, éventuellement en présence d'auto-occultations, sera abordée avec un filtre de Kalman étendu pour la régularisation.

A long terme les applications envisagées concernant la transmission de geste bas débit, la télévirtualité, l'interprétation de gestes pour des interfaces homme-machine multimodales, la reconnaissance de tout ou une partie de la langue des signes.

7. BIBLIOGRAPHIE

1. T. Baudel and M. Beaudouin-Lafon, "CHARADE : Remote Control of objects using Free-Hand Gestures", *Communications of the ACM*, Vol. 36, n° 7, pp. 28-35, 1993.
2. Y. Bellik, "Modality Integration : Speech and Gesture. Survey of the State of the Art in Human Language Technology", Section 9.4, R. A. Cole ed. in chief, 1996.
<http://www.cse.orgi.edu/CSLU/HLTsurvey>.
3. M. Bourges-Sévenier, P. Horain, F. Prêteux, P. Leray, "Recalage d'un modèle générique sur une séquence d'image 2D", Actes de CORESA'97 : 3^{èmes} Journées d'études et d'échanges Compression et Représentation des Signaux audiovisuels, Issy-les-Moulineaux, France, 26-27, pp. 163-171.
<http://www-sim.int-evry/Publications>.
4. G. Borgefors, "Distance Transformations in Digital Images", *Computer Vision, Graphics and Image Processing* 34, 344-371, 1986.
5. A. Braffort, "Reconnaissance et compréhension de gestes, application à la langue des signes", thèse de l'université de Paris XI, spécialité informatique, Juin 1996.
6. R. Deriche, "Using Canny's Criteria to Derive a Recursively Implemented Optimal Edge Detector", *International Journal of Computer Vision*, 1(2) : 167-187, 1987.
7. J.D. Foley, A. Van Dam, S.K. Feiner, J.F. Hughes, *Introduction à l'infographie*, Addison-Wesley France, 1995.
8. P. Horain, "Extraction des arêtes dans des images de distance", *Actes du 7^{ème} congrès RFIA, AFCET-INRIA*, Paris, 1989.
9. D. G. Lowe, "Robust Model-based Motion Tracking Through the Integration of Search and Estimation", *International Journal of Computer Vision*, 8:2, 113-122 (1992).
10. H. Ouhaddi et P. Horain, "Conception et ajustement d'un modèle 3D articulé de la main", *Actes des 6^{èmes} Journées de Travail du GT Réalité Virtuelle*, Issy-les-Moulineaux, 12-13 mars, 1998, pp. 83-90. <http://www-sim.int-evry/Publications>.
11. F. Prêteux, P. Horain, H. Ouhaddi, and M. Preda, *Report on Core Experiment 3 on Hand Baps interpretation*. ISO/IEC JTC1/WG1 MPEG97/M3332, March 1998, <http://www-sim.int-evry/Publications>.