Modélisation 3D du geste par la vision

Hocine OUHADDI, Patrick HORAIN

Unité de projet ARTEMIS, Département Signal et Image

Institut National des Télécommunications,

9 rue Charles Fourier, 91011 Evry cedex

{*prénom.nom*}@int-evry.fr
http://www-sim.int-evry.fr/Artemis

Résumé

Nous présentons une méthode de modélisation 3D du geste par recalage d'un modèle articulé de la main sur une séquence d'images monoscopique, sans marqueur. Nous décrivons l'ajustement du modèle sur la morphologie de l'opérateur, puis nous comparons différentes fonctionnelles et méthodes d'optimisation pour le recalage d'un modèle de la main. Une originalité de notre approche réside dans l'intégration des contraintes biomécaniques dans la procédure de recalage, ce qui permet de réduire considérablement l'espace de recherche et évite l'obtention de configurations irréalistes. Les applications potentielles d'un tel système sont, entre autres, l'animation d'acteurs virtuels, le codage SNHC et la transmission à très bas débit des gestes humains ou leur reconnaissance.

Mots clés : Analyse de gestes, modèle articulé, suivi de la main, recalage, fonctionnelle de coût, optimisation, simplexe, contraintes biomécaniques.

1. Introduction

Moyen de communication dans la vie courante, les gestes peuvent aussi être utilisés pour communiquer avec des machines, par exemple au moyen de gants instrumentés munis de nombreux capteurs. Ceci permet par exemple de commander des acteurs virtuels ou de désigner et manipuler des objets graphiques sur des écrans d'ordinateurs [1]. Les gants instrumentés demeurent toutefois une technologie onéreuse et contraignante en terme d'ergonomie du système.

Nous nous intéressons dans cet article à la modélisation 3D de la main et des gestes par vision artificielle. L'acquisition est effectuée par une seule caméra et sans marqueur. Notre approche consiste à ajuster un modèle générique 3D sur une image de la main ouverte de l'opérateur en appliquant des contraintes morphologiques, puis à recaler le modèle articulé sur la main en mouvement tout en respectant des contraintes biomécaniques.

2. Modélisation biomécanique et anthropométrique de la main

La main possède 6 degrés de liberté au niveau du poignet : 3 rotations et 3 translations. Deux degrés de liberté, l'abduction/adduction, et la flexion/extension, sont associés aux articulations métacarpo-phalangienne de chaque doigt, et 1 autre (flexion/extension) au niveau de chaque articulation interphalangienne. Notre modèle possède par conséquent 26 degrés de liberté.

Nous avons développé un modèle 3D articulé de la main intégrant tous ces degrés de liberté et respectant le paramétrage pour l'animation du corps défini par la norme MPEG4 [2] (figure 1).

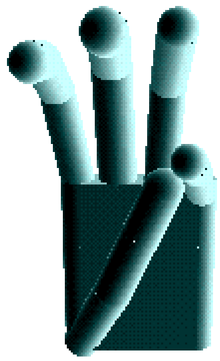


Figure 1: Modèle de la main

Le modèle générique est d'abord ajusté à la morphologie particulière de la personne observée par une technique non supervisée utilisant une seule image de la main en configuration ouverte (doigts écartés). La silhouette de la main est extraite par une classification sur la teinte chair. Les doigts et la paume sont ensuite segmentés au moyen d'opérateurs de filtrage et de reconstruction de la morphologie mathématique [3]. Cela permet d'estimer, à un facteur d'échelle près, les paramètres de la morphologie des doigts et de la paume. Une première originalité de notre contribution réside dans l'utilisation de contraintes anthropométriques pour valider les mesures obtenues et déterminer les mesures manquantes (longueurs des phalanges) [4].

3. Recalage du modèle 3D de main sur des séquences vidéo

Nous avons étudié le recalage du modèle pour des positions quelconques de la main dans des séquences d'images. Notre approche consiste à estimer les paramètres de position globale et les angles des articulations qui minimisent l'écart entre la projection du modèle et des primitives 2D extraites de l'image.

3.1. Etat de l'art

Mochimaru et Yamazaki [5] puis Kuch et Huang [6] ont minimisé l'écart entre la silhouette du modèle projeté et celle extraite de l'image par une procédure de "perturbation locale de degrés de liberté". Celle-ci consiste à modifier successivement la valeur de chaque degré de liberté d'une petite quantité. Si la fonctionnelle diminue, la modification est répétée dans la même direction, sinon la direction de perturbation est inversée. Les méthodes à base de perturbation locale de degrés de liberté semblent assez efficaces en pratique. Toutefois, les différents paramètres doivent être choisis de manière empirique.

Ohya et Kishino [7] utilisent un algorithme génétique pour recaler un modèle représentant la partie supérieure du corps humain sur des paires d'images stéréo (images de synthèse), par minimisation de l'écart entre les silhouettes projetée et extraite de l'image. Les résultats présentés sont obtenus à partir d'une population d'environ 1000 solutions initiales et nécessite plusieurs centaines d'itérations, ce qui est très coûteux en temps de calcul.

Gavrila et Davis [8] recalent un modèle du corps humain sur des paires d'images stéréo par minimisation d'une fonctionnelle de distance entre les contours du modèle projeté et ceux extraits des images. La minimisation est réalisée après discrétisation de l'espace de paramètres et recherche hiérarchique dans un arbre représentant l'ensemble des configurations possibles.

Shimada *et al.* [9] recalent un modèle de la main sur une séquence d'images monoscopiques par génération et sélection d'un ensemble discret de postures candidates en fonction d'une part du taux de recouvrement entre la silhouette de modèle et celle extraite de l'image, et d'autre part des distances entre les contours des doigts dans l'image et les contours du modèle projeté. Le nombre de postures générées est réduit en fonction des postures dans les images précédentes en utilisant une loi de dépendance statistique entre configurations successives. La discrétisation de l'espace des paramètres nécessite un compromis entre la précision et le coût de recalage.

3.2. Choix d'une méthode de recalage

Nous avons comparé deux fonctionnelles d'évaluation du recalage de notre modèle articulé sur les images. L'une est la moyenne des distances entre contours occultants, et l'autre est le taux de non recouvrement des silhouettes (figure 2) :

$$F = \frac{\mathrm{Surface}(S_i - S_m) + \mathrm{Fonction}(S_m - S_i)}{\mathrm{Surface}(S_i) + \mathrm{Surface}(S_m)}$$

S_i et S_m sont respectivement la silhouette de la main extraite de l'image (par seuillage sur la teinte) et la silhouette du modèle projeté. Notons que notre modèle actuel ne comporte pas d'avant-bras ; pour ne pas introduire de biais dans l'évaluation du recalage. Nous avons supprimé interactivement l'avant-bras dans les images (la localisation du poignet pourrait être automatisée [10]).

Ces deux fonctionnelles ont été tour à tour minimisées suivant trois méthodes classiques : Levenberg-Marquardt, descente de simplexe et Powell. La méthode de Levenberg-Marquardt [11] est une descente de gradient utilisant les dérivées partielles de la fonctionnelle par rapport aux paramètres à optimiser. La méthode de descente du simplexe [12] consiste à appliquer, dans l'espace des paramètres, des transformations géométriques à un simplexe construit autour d'un vecteur initial, de façon à adapter sa forme à la topologie locale de la fonctionnelle. La méthode de Powell [13] effectue une succession de minimisations monodimensionnelles pour chaque paramètre, jusqu'à convergence.

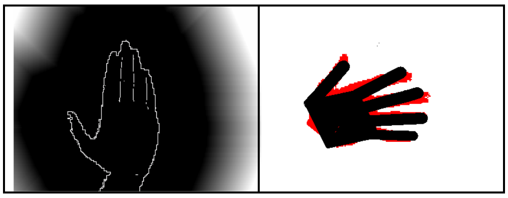


Figure 2 : Fonctionnelles retenues

Nous avons comparé 5 combinaisons de ces 2 fonctionnelles avec les 3 méthodes d'optimisation (le gradient de la surface de non recouvrement étant difficile à estimer, nous n'avons pas mis en œuvre la méthode de Levenberg-Marquardt pour cette fonctionnelle).

Méthodes d'optimisation/ Fonctionnelles d'évaluation	Levenberg-Marquardt	Descente du simplexe	Powell
Surfaces de non recouvrement		✔	✔
Distances aux contours	✔	✔	✔

Tableau des combinaisons des méthodes d'optimisation et des fonctionnelles d'évaluation que nous comparons (en gras : celle retenue).

Nous avons obtenu les résultats suivants :

- La minimisation par descente du simplexe de la fonctionnelle de non recouvrement permet de traiter des mouvements rigides, des mouvements d'abduction/adduction et de flexion. Si la procédure d'optimisation converge avec un taux de non recouvrement résiduel important (supérieur à un seuil que nous avons fixé empiriquement à 8 %) alors elle a convergé vers un minimum local qui n'est pas satisfaisant. Ce problème est généralement résolu en réinitialisant la procédure d'optimisation avec un simplexe plus grand.

- Le recalage avec une fonctionnelle utilisant les distances entre contours semble très sensible à l'initialisation et au bruit dans l'image. Ceci est dû au fait que la transformation distance effectue en chaque point un appariement implicite avec le contour le plus proche. Si la configuration initiale est très éloignée de l'optimum recherché, ou si les arêtes extraites de l'image sont bruitées, de faux appariements peuvent conduire à un mauvais recalage du modèle.

- Les résultats obtenus par la méthode de Powell sont en général équivalents à ceux de descente du simplexe. Toutefois, le nombre d'évaluations de la fonctionnelle est plus grand. En effet, la méthode de Powell fait appel à une procédure de minimisation mono-dimensionelle qui effectue plusieurs évaluations de la fonctionnelle afin de déterminer un intervalle contenant un minimum de la fonctionnelle dans une direction donnée.

Ces résultats nous ont conduit à retenir la fonctionnelle utilisant les surfaces de non recouvrement et la méthode d'optimisation avec descente du simplexe.

4. Descente de simplexe avec contraintes biomécaniques

Afin de prendre en compte la connaissance *a priori* sur les configurations possibles de la main, nous avons introduit des contraintes biomécaniques dans l'algorithme d'optimisation afin d'interdire les postures irréalisables. Ces contraintes sont statiques ou dynamiques. Les contraintes statiques traduisent les limites des angles d'abduction/adduction ou de flexion/extension possibles des différentes articulations.

Les contraintes dynamiques représentent les relations entre les degrés de liberté des articulations des doigts. Par exemple, les flexion/extension et l'abduction/adduction de chaque doigt se limitent mutuellement (figure 3).

Ces contraintes (linéaires) définissent des demi-espaces dans l’espace des paramètres, dont l’intersection est un domaine convexe. Nous imposons alors à chaque transformation du simplexe de générer un simplexe contenu dans ce domaine. Le domaine des contraintes étant convexe, toute contraction d'un simplexe à l'intérieur de cette région crée un nouveau simplexe qui respecte les contraintes. Par conséquent, il suffit de tester les contraintes seulement après les symétries et les expansions. Si un sommet du simplexe est déplacé à l'extérieur de cette région, alors celui-ci est remplacé par un nouveau sommet situé à l'intersection de l'hyperplan de contrainte et de la droite qui relie ce nouveau sommet à l'ancien (figure 4). Cette stratégie originale permet de respecter les contraintes biomécaniques tout en réduisant l'espace de recherche.

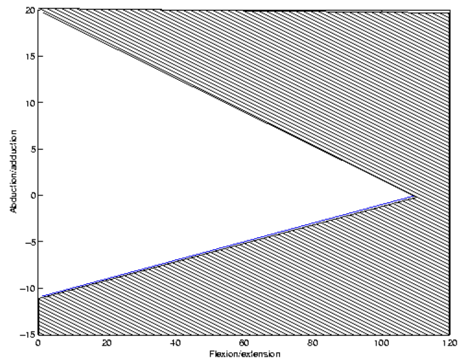
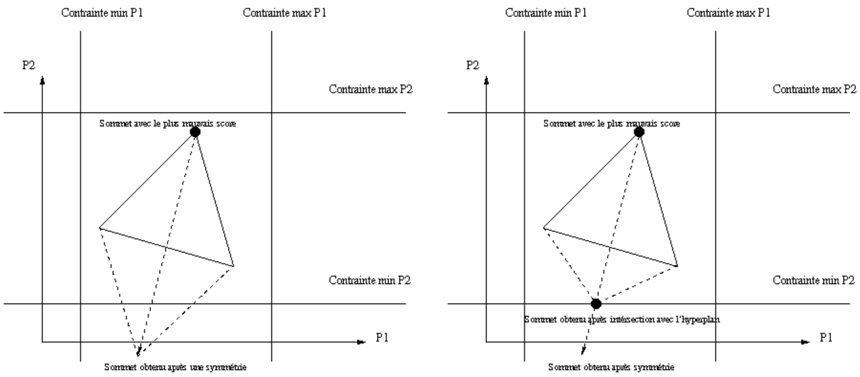


Figure 3 : Exemple de contrainte biomécanique : interdépendance entre abduction/adduction et flexion/extension (angles en degrés).



(a) Symétrie non contrainte

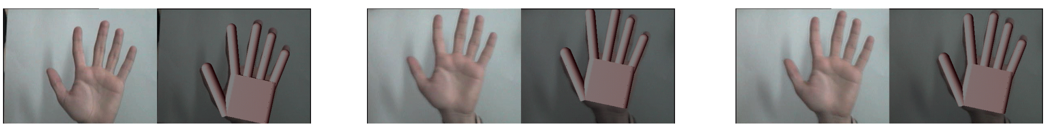
(b) Symétrie contrainte

Figure 4 : Transformation contrainte du simplexe.

5. Résultats

Nous avons testé notre méthode sur des séquences naturelles présentant divers mouvements (translations, abductions et flexions).

Parmi les 21 degrés de liberté du modèle, seuls la position du modèle, les abductions et les flexions des doigts ont été d'abord retenus, soient 12 paramètres. Les résultats obtenus sont satisfaisants (figures 5, 6, 7, 8, 9).

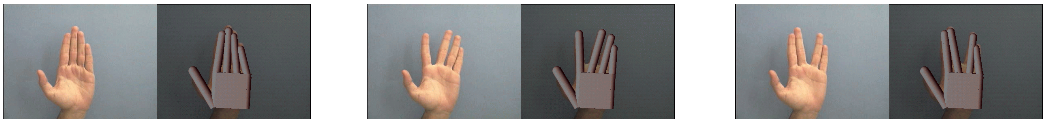


(a) image 7

(c) image 38

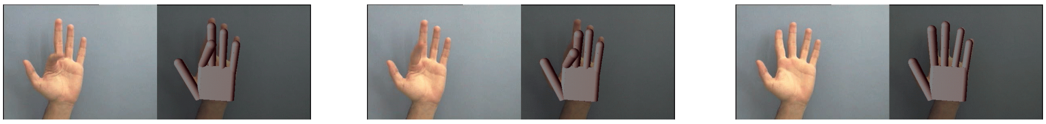
(d) image 4

Figure 5 : Translation de la main.



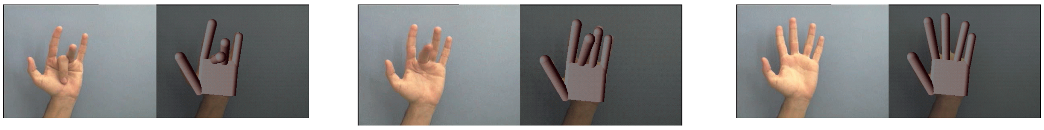
(a) image 18 (b) image 55 (c) image 67

Figure 6: Abduction/adduction des doigts.



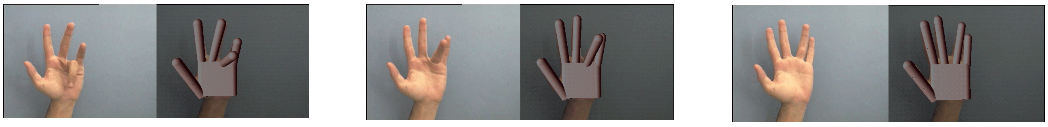
(a) image 13 (b) image 34 (c) image 44

Figure 7: Flexion de l'index.



(a) image 27 (b) image 35 (c) image 41

Figure 8 : Flexion de majeur.



(a) image 26 (b) image 44 (c) image 50

Figure 9 : Flexion de l'annulaire.

Dans le cas de la flexion, il peut arriver que la troisième et la seconde phalange soient fléchies devant la paume, et apparaissent à l'intérieur de la silhouette. La fonctionnelle de non recouvrement ne fournit alors aucune information sur cette partie de la main, et le recalage de ces phalanges ne peut être correctement réalisé. Nous envisageons d'utiliser les contours internes de l'image pour traiter les configurations où des doigts sont devant la paume.

Nous avons également mis en oeuvre une technique hiérarchique consistant à recaler d'abord la position globale seulement, puis à rajouter successivement le facteur d'échelle, les degrés de liberté de rotation, d'abduction et enfin de flexion des doigts (rigides). Avec ces 16 degrés de liberté, le recalage est moins précis. Nous envisageons d'intégrer l'information de mouvement dominant apparent (flot optique) dans la procédure de recalage pour choisir l'ordre du recalage hiérarchique.

6. Conclusion et perspectives

Nous avons présenté une méthode pour le suivi de la main dans une séquence d'images à partir d'une seule vue et sans utilisation de marqueurs. La main est représentée par un modèle 3D articulé dont les mesures morphologiques sont ajustées automatiquement sur une image de la main ouverte de l'opérateur. Nous avons comparé ensuite 2 fonctionnelles de coût et 3 méthodes d'optimisation pour le recalage du modèle 3D. Minimiser le taux de non recouvrement par descente du simplexe apparaît comme étant la méthode la plus efficace parmi celles testées.

Une originalité de notre travail réside dans l'intégration des contraintes biomécaniques dans l'algorithme de recalage. Ceci permet de réduire considérablement l'espace de recherche et évite l'obtention de configurations irréalistes. Des résultats sur des séquences naturelles montrent la validité de notre approche.

Nous étudions actuellement le traitement des scènes présentant des fonds complexes, des auto-occultations de la main (doigts fléchis devant la paume), et la prise en compte de l'information de mouvement pour le recalage du modèle articulé.

A terme, les applications potentielles incluent l'animation d'acteurs virtuels, le codage hybride synthétique / naturel (SNHC) à bas débit (transmission des seuls paramètres des gestes), et la reconnaissance des gestes (interfaces homme-machine multimodales, interprétation de la langue des signes).

Références

[1] 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`.

[2] F. Prêteux, P. Horain, H. Ouhaddi and M. Preda, "*Report on Core Experiment 3 on Hand Baps interpretation*", ISO/IEC JTC1/SC29/WG11, MPEG97/M3332, March 1998 ; `http://www-sim.int-evry.fr/Publications`.

[3] J. Serra, Image Analysis and Mathematical Morphology, Academic Press, 1982.

[4] H. Ouhaddi et P. Horain, "Conception et ajustement d'un modèle 3D articulé de la main", Actes des *6èmes Journées de Travail du GT Réalité Virtuelle*, Issy-les-Moulineaux, France, 1998, pp. 83-90 ; `http://www-sim.int-evry.fr/Publications`.

[5] M. Mochimaru and N. Yamazaki, "The three-dimensional measurement of unconstrained motion using a model-matching method", *Ergonomics*, Vol. 37, no 3, 1994, pp. 493-510.

[6] J. J. Kuch, and T. S. Huang, "Vision Based Hand Modeling and Tracking for Virtual Teleconferencing and Telecollaboration". In *Proc. of fifth International Conference on Computer Vision*, 1995, Massachusetts, pp. 666-671.

[7] J. Ohya and F. Kishino, "Human Posture Estimation from Multiple Images Using Genetic Algorithm", *Proc. of the 12th IAPR Int. Conf. On Pattern Recognition*, Vol 1, pp. 750-753.

[8] D. M. Gavrila and L. S. Davis, "3-D model-based tracking of humans in action: a multi-view approach", Proc. *IEEE Computer Vision and Pattern Recognition*, San Francisco, 1996. `http://www.umiacs.umd.edu/users/gavrila/publications.html`

[9] N. Shimada, Y. Shirai, Y. Kuno and J. Miura, "Hand Gesture Estimation and Model Refinement using Monocular Camera - Ambiguity Limitations by Inequality Constraints", *Proceedings of FG'98*, pp. 268-273, april 1998 Nara, Japan.

[10] B. Deimel and S. Schröter, "Improving Hand-Gesture Recognition via Video Based Methods for the Separation of the forearm from the Human Hand", *Gesture Workshop'99,* Orsay, France, March 1999.

[11] H. Ouhaddi, P. Horain et K. Mikolajczyk, "Modélisation et suivi de la main", Actes des *4èmes Journées d'Etudes et d'Echanges « COmpression et REprésentation des Signaux Audiovisuels » (CORESA'98)*, Lannion, 1998, pp. 109-114;

`http://www-sim.int-evry.fr/Publications`.

[12] J. A. Nelder and R. Mead, « A Simplex Method for Function Minimization », *Computer Journal*, Vol. 7, 1965, pp. 308-313.

[13] W. H. Press, B. P. Flannery, S. A. Teukolsky, W. T. Vetterling, « *Numerical Recipes in C* », Cambridge University Press, 1992.