

Conception, ajustement et recalage d'un modèle 3D articulé pour le suivi de la main dans des séquences d'images

Design, adjustment and registration of a 3D articulated model
for hand tracking in image sequences

Hocine OUHADDI, Patrick HORAIN

Institut National des Télécommunications
Département Signal et Image
9, rue Charles Fourier, 91011 Evry Cedex
<http://www-sim.int-evry.fr>
prenom.nom@int-evry.fr

Résumé

Après un état de l'art des méthodes de modélisation du geste par recalage d'un modèle en vision artificielle, nous présentons nos travaux en cours sur la modélisation 3D et le suivi de la main dans des séquences d'images monoscopiques, sans recourir à des marqueurs. Un modèle générique 3D est ajusté à la morphologie de l'opérateur sur une image de sa main ouverte. La silhouette de la main est extraite par segmentation sur un critère de teinte de la chair, puis les doigts sont séparés par filtrage morphologique et le modèle est ajusté à partir de points caractéristiques et de moments d'inertie. A partir de cette position initiale, le modèle est animé pour reproduire les mouvements observés dans la séquence. Les angles des articulations ainsi que l'attitude sont ajustés pour rapprocher les projections des contours occultants du modèle des contours extraits de l'image. Des résultats sont présentés pour des mouvements de translation et d'abduction/adduction. A terme, un tel système de modélisation du geste par la vision artificielle pourrait être utilisé pour des interfaces homme-machine par le geste, pour la désignation ou la manipulation d'objets dans l'espace, le codage à très bas débit du geste pour la télé-présence, ou la reconnaissance de la langue des signes.

Mots clés : Modèle articulé, anthropométrie, biomécanique, ajustement morphologique, recalage, suivi, interaction gestuelle

Abstract

We first present a state of the art of gesture recognition by model registration and artificial vision, and then our current work on 3D modeling and tracking of the hand in monoscopic images sequences, without markers. A 3D generic model is adjusted to the operator's morphology in an image of his open hand. The hand silhouette is segmented using a skin hue criteria. Then, fingers are separated by morphological filtering and the model is adjusted using characteristic points and inertia moments. The model is animated, from this initial pose to reproduce movements observed in a video sequence. Joint angles and hand pose are adjusted to match the projected model contours with the contours extracted from the image. Results are presented for translation and abduction/adduction movements. Gesture modelisation by artificial vision could be a valuable help for applications such as human-machine interaction by gesture for object designation or manipulation in space, low bit rate coding of gesture for telepresence, or sign language recognition.

Keywords : Articulated model, anthropometry, biomechanics, morphological adjustment, registration, tracking, gesture interaction

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 communication 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 [Baudel93, Bellik96] ou même pour la reconnaissance de la langue des signes [Braffort96]. 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 [Iwai96, Hienz96]. Une telle approche ne permet pas un traitement simple et général des auto-occultations (figure 1).

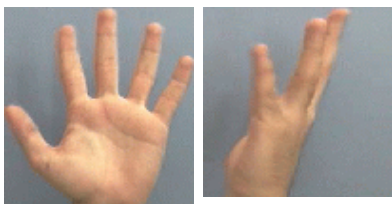


Figure 1: Images de la main sans et avec auto-occultations.

- La *reconstruction par stéréovision* fournit la forme 3D de la main ou du bras à partir de plusieurs images [Delamarre98, Fua97, Ishibuchi92, Fukumuto94, Kuno94, Utsu96].
- Méthodes par *ajustement et recalage d'un modèle générique 3D* à partir d'une ou plusieurs vues de geste [Davis94, Dorner94, Heap96, Kush95, Lee95, Millar96, Mochimaru94, Regh94, Shimada95, Shimada96].

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 dernière approche.

Dans la première partie de cet article nous présentons un état de l'art des méthodes d'analyse de gestes par recalage d'un modèle 3D. Nous présentons ensuite quelques rappels sur l'anatomie, l'anthropométrie et la biomécanique de la main. Le modèle générique 3D articulé que nous proposons ainsi que son ajustement sur la morphologie de la main est ensuite décrit. Enfin, nous décrivons nos travaux en cours pour le recalage et le suivi de la main dans une séquence d'images, et nous présentons quelques résultats.

2 La reconnaissance de gestes par recalage d'un modèle 3D

Les méthodes d'analyse de gestes par recalage de modèles génériques 3D prennent en considération la nature tridimensionnelle des gestes en utilisant un modèle 3D pour représenter la main. Différents types de modèles peuvent être utilisés : squelettiques, surfaciques ou volumiques.

2.1 Modèles squelettiques

Les modèles squelettiques représentent la main par un ensemble de segments et de repères locaux associés aux différentes articulations.

Dorner [Dorner94] a proposé une méthode de reconnaissance de la langue de signes en utilisant un modèle squelettique dont les degrés de liberté sont estimés par minimisation itérative des distances entre les points représentant les projections des articulations du modèle et leurs correspondants dans l'image. La détermination de ces derniers est facilitée par l'utilisation d'un gant coloré.

Une méthode similaire a été proposée par Lee et Kunii [Lee95] pour l'identification de postures. Les points caractéristiques sont les bouts de doigts. A la différence de la méthode précédente, l'appariement n'est pas effectué dans le plan image mais les positions 3D des points caractéristiques (marqués) sont calculées par stéréovision. Les degrés de liberté sont estimés ensuite en utilisant la cinématique inverse. Les contraintes biomécaniques permettent de choisir parmi des solutions multiples.

Millar et Crawford [Millar96] déterminent directement (sans itération) les valeurs des degrés de liberté du modèle à partir des positions 3D des bouts des doigts et du poignet désignées interactivement dans des paires d'images stéréo.

Les méthodes de recalage des modèles squelettiques sur les images recourent à des marqueurs sur les points caractéristiques de la main ou une intervention de l'opérateur.

2.2 Modèles surfaciques

Les modèles surfaciques décrivent l'enveloppe de la main plutôt que son squelette.

Dans le modèle de Kush et Huang [Kush95], la surface de la main est représentée par des β -splines. La reconnaissance de gestes est effectuée par ajustement du modèle puis suivi de la main. L'étape d'ajustement a pour but d'obtenir un modèle dont les mesures morphologiques sont similaires à celles de la main de l'opérateur. L'ajustement est effectué de manière supervisée à partir de 3 vues de la main. Les degrés de liberté des articulations sont estimés par une procédure itérative de minimisation de l'écart entre la silhouette de la main dans l'image et celle du modèle après projection.

Heap et Hogg [Heap96] utilisent un modèle surfacique à base de maillage simplexe. La méthode de suivi proposée procède par déformation du maillage et se limite à l'estimation des paramètres globaux de mouvement du modèle (rotations, translations et facteur d'échelle) par appariement des contours occultants du modèle et ceux extraits de l'image.

Les modèles surfaciques fournissent des contours occultants qu'il est possible d'apparier avec les contours dans l'image. Toutefois la gestion des déformations du maillage est plus complexe que celle des articulations d'un modèle volumique.

2.3 Modèles volumiques articulés

La première méthode utilisant un modèle 3D volumique pour le suivi de la main dans une séquence d'images a été proposée par Regh et Kanade [Regh94]. Le modèle est composé de cylindres. Les degrés de liberté des articulations sont estimés par minimisation des distances entre les projections des bouts des doigts du modèle et les points correspondants mesurés dans l'image, ainsi que l'inclinaison des axes

des phalanges estimés par rapport à ceux mesurés à partir de l'image. Deux caméras sont utilisées pour traiter les auto-occultations.

La méthode proposée par Davis et Shah [Davis94] construit un modèle de type cylindre généralisé pour chaque phalange après extraction des doigts (de la main ouverte). La procédure de suivi mise en oeuvre se limite à l'estimation des paramètres globaux de mouvement (translations et rotations) en se basant sur le calcul du flot optique.

Mochimaru et Yamazaki [Mochimaru94] utilisent une procédure similaire à celle proposée par Kush et Huang, à la différence que le modèle est composé de cônes plutôt que de β -splines.

Shimada *et al.* utilisent une seule caméra, un modèle volumique à base de parallélépipèdes et estiment tous les degrés de liberté du modèle. Leur méthode consiste à sélectionner pour chaque image les postures possibles du modèle parmi un ensemble fini, à calculer leurs projections et à prendre la posture qui recouvre au mieux la silhouette extraite de l'image. Le nombre de postures candidates est réduit par l'examen des postures dans les images précédentes dans la séquence [Shimada95]. Cette méthode ne fournit qu'une estimation grossière des degrés de liberté. L'utilisation d'un filtre de Kalman étendu et l'ajout de contraintes biomécaniques permettent d'améliorer la précision de l'estimation des paramètres et de traiter les auto-occultations [Shimada96].

3 La main

3.1 Anatomie de la main

Le squelette de la main comporte les os du poignet, les métacarpes et les os des doigts. L'articulation de chaque doigt sur son métacarpe est appelée métacarpo-phalangienne (MCP). Les articulations entre phalanges sont appelées interphalangiennes proximales (IPP) et interphalangiennes distales (IPD), ou, pour le pouce, simplement interphalangiennes (IP). L'articulation située à la base du pouce est dite carpo-métacarpienne (CMC) (figure 2).

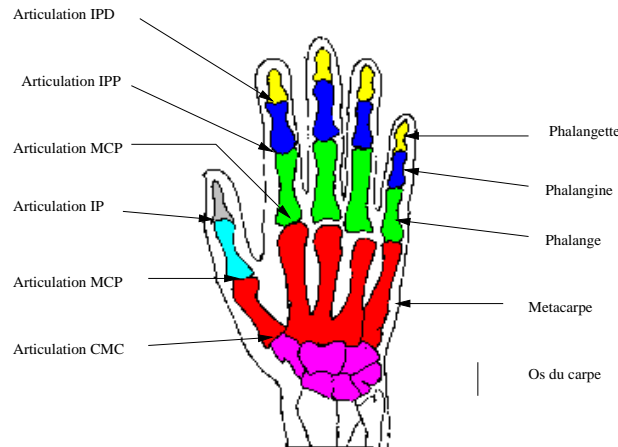


Figure 2: Squelette de la main.

3.2 Anthropométrie de la main

L'anthropométrie est l'étude comparative des mesures du corps humain. Une des études les plus élaborées, à notre connaissance, est celle de Wagner [Wagner88], réalisée sur un échantillon de 127 hommes âgés de 17 à 63 ans et 111 femmes âgées de 16 à 64 ans, qui repose sur 48 mesures

morphologiques. A partir de ces mesures, un certain nombre de contraintes anthropométriques sont formulées sous forme de relations entre mesures et d'intervalles de valeurs possibles. Par exemple, la longueur de la première phalange du majeur est égale à $0,266 \pm 0,003$ fois la longueur de la main.

3.3 Biomécanique de la main

La main possède 6 degrés de liberté au niveau du poignet : 3 rotations et 3 translations. Deux degrés de liberté (abduction/adduction (figure 3) et flexion/extension (figure 4) sont associées aux articulations MCP de chaque doigt, 1 autre (flexion/extension) au niveau de chaque articulation interphalangienne. D'après [Buchholz92], l'articulation CMC du pouce possède 3 degrés de liberté : abduction/adduction, flexion/extension et une pseudo-rotation due à l'incongruité entre les os du carpe et la base du métacarpe du pouce, et au relâchement des ligaments les reliant. Au total, la main possède donc 28 degrés de liberté.

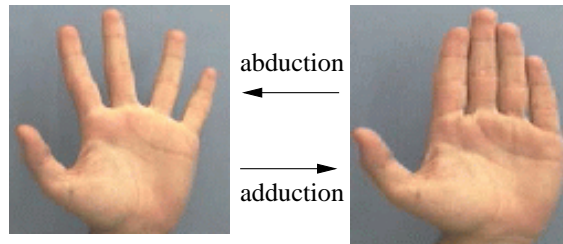


Figure 3: Mouvements d'adduction et d'abduction.

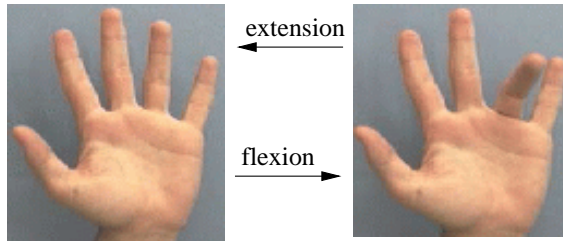


Figure 4: Mouvements de flexion et d'extension.

Les mouvements des doigts de la main sont régis par des contraintes biomécaniques qui rendent certaines 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. Par exemple, l'angle de flexion/extension des premières phalanges des quatre doigts est compris entre -110° et 15° .

Les contraintes dynamiques représentent les relations entre les degrés de liberté des articulations des doigts. Pour les doigts autres que le pouce, ces contraintes sont [Kush95, Lee95, Millar96, Teo94, Buchholz92, Rijpkema91] :

- la relation entre les flexions de la phalangette et de la phalange :

$$\theta_{f/e}^{\text{IPD}} = \frac{2}{3} \theta_{f/e}^{\text{IPP}} \quad (1)$$

- l'interdépendance entre la flexion/extension et l'abduction/adduction de chaque phalange. En effet, plus l'angle de flexion/extension est grand, plus l'abduction ou l'adduction des doigts est

limitée. La valeur maximale de l'angle d'abduction/adduction de la phalange dans le cas où l'angle de flexion/extension n'est pas nul est donnée par :

$$|\theta_{a/a}^{\text{MCP}}| \leq \left(1 - \frac{1}{\theta_{f/e \text{ max}}^{\text{MCP}}}\right) \cdot \theta_{f/e}^{\text{MCP}} \cdot \theta_{a/a}^{\text{MCP}} \quad (2)$$

où $\theta_{f/e \text{ max}}^{\text{MCP}}$ est la valeur maximale de l'angle de flexion/extension de la phalange.

- la relation entre les flexions/extensions des phalanges de doigts voisins ; par exemple, la flexion de la phalange de l'index ou de l'annulaire entraîne la flexion de la phalange de majeur.

$$\theta_{f/e}^{\text{MCP Majeur}} \geq \max \left(\begin{array}{l} \theta_{f/e}^{\text{MCP Index} - 25^\circ} \\ \theta_{f/e}^{\text{MCP Annulaire} - 45^\circ} \\ \theta_{f/e \text{ min}}^{\text{MCP Majeur}} \end{array} \right) \quad (3)$$

où $\theta_{f/e \text{ min}}^{\text{MCP Majeur}}$ est la valeur minimale de l'angle de flexion/extension de la phalange de majeur.

A cause de la morphologie particulière du pouce, ses mouvements sont soumis à un ensemble de contraintes différentes. Ainsi, la flexion du pouce au niveau de l'articulation CMC est comprise entre -80° et 25° . De plus deux contraintes dynamiques relient les flexions/extensions et les abductions/adductions des deux premières articulations du pouce :

$$\theta_{f/e}^{\text{MCP}} = 2\left(\theta_{f/e}^{\text{CMC}} - \frac{\pi}{6}\right) \quad (4)$$

$$\theta_{a/a}^{\text{MCP}} = \frac{7}{5}\theta_{a/a}^{\text{CMC}} \quad (5)$$

4 Modèle de la main

Pour représenter les gestes, nous avons choisi de décrire la main par un modèle de type volumique articulé. Ce modèle est constitué de 2 parties. La première décrit la structure articulée du modèle par un graphe représentant les translations et les rotations associées aux différents degrés de liberté de la main (degrés de liberté globaux, abductions/adductions, flexions/extensions et torsions). La seconde partie représente la forme des éléments articulés composant le modèle.

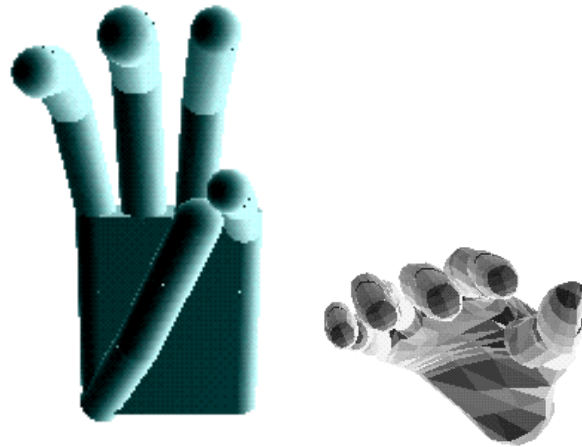
Cette dissociation de la structure et de la morphologie du modèle permet de passer facilement du modèle constitué de primitives géométriques simples (cube, cylindres et sphères) à un modèle composé de maillages triangulaires (figure 5).

Elle permet aussi de réaliser l'ajustement du modèle et son animation sous forme d'actions limitées aux seuls nœuds morphologiques et de degrés de liberté de l'arbre représentant le modèle.

Notre modèle, conforme au paramétrage MPEG-4 [Preteux98], peut être animé suivant les 6 degrés de liberté pour le positionnement global, et 5 degrés de liberté pour chaque doigt (1 flexion pour l'articulation IP du pouce, 1 flexion pour chaque articulation IPD, 3 rotations pour l'articulation CMC du pouce et pour chaque articulation MCP). Il ne prend pas en compte l'abduction de l'articulation MCP du pouce mentionnée dans [Buchholz92].

La contrainte dynamique de flexion à laquelle obéit le pouce, et celle portant sur les flexions interphalangiennes des 4 autres doigts réduisent le nombre de degrés de liberté de 31 à 26.

Les degrés de liberté en torsion des doigts permettent de modéliser l'effet des forces externes pour les tâches de préhension. Pour la reconnaissance des gestes en l'absence de force externe, les angles



(a) Modèle à base de primitives simples.

(b) Modèle à base de primitives maillées.

Figure 5: Modèle de la main.

de torsion des doigts peuvent être considérés comme nuls. Ainsi notre modèle possède 26 degrés de liberté à ajuster en respectant les contraintes biomécaniques.

La cohérence des mesures morphologiques est assurée par l'intégration des contraintes anthropométriques. Vingt-et-une contraintes portant sur les intervalles des valeurs possibles pour les longueurs des doigts et des phalanges ainsi que les dimensions de la paume ont été retenues [Wagner88].

5 Animation du modèle

Pour animer le modèle, nous avons choisi une technique explicite consistant à spécifier entièrement les paramètres de chaque plan de la séquence, tout comme l'application de suivi devra les fournir. Des séquences d'animation ont été créées en spécifiant les valeurs des paramètres des articulations de façon à reproduire les gestes représentant quelques lettres de l'alphabet de la langue des signes (figure 6).

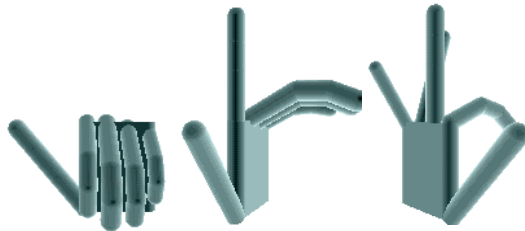


Figure 6: Synthèse de signes de l'alphabet de la langue des signes (A, D, F respectivement.)

Après ajustement, le modèle articulé est destiné à être animé pour reproduire les mouvements de la main dans des séquences vidéo. Les méthodes d'animation procédurale et déclarative [Teo94] peuvent utiliser la dynamique inverse [Rijpkema91], l'analyse par éléments finis ou l'interpolation de plans [Gourret89].

Le modèle de la main ainsi que les animations ont été développés en VRML sur station Silicon Graphics au moyen du logiciel Open Inventor.

6 Ajustement du modèle

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 du suivi de la main dans des séquences. Notre approche, non supervisée, utilise une seule image de la main ouverte (doigts écartés).

6.1 Segmentation de l'image

Pour extraire de l'image la silhouette de la main, nous avons utilisé l'information de couleur. Un apprentissage de la distribution des couleurs de la peau est effectué à partir de l'histogramme des teintes et des saturations (peu sensible aux conditions d'éclairage de la scène). Une région de cet espace est définie par des seuils. La silhouette de la main est considérée comme étant la plus grande composante connexe correspondant aux couleurs sélectionnées. Les trous de la silhouette dus aux bruits sont remplis (figure 7).

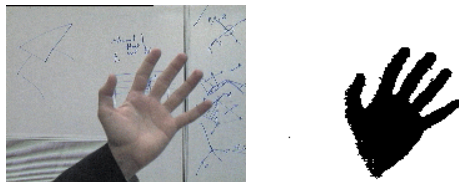


Figure 7: Segmentation de la main par seuillage sur la couleur.

L'estimation des paramètres morphologiques du modèle requiert l'identification des doigts et de la paume dans l'image. Les doigts sont extraits par un opérateur de morphologie mathématique : un chapeau haut-de-forme [Serra94] dont l'élément structurant est un disque de diamètre légèrement supérieur à la largeur des doigts. Les régions représentant les doigts sont les plus grandes composantes connexes obtenues (figure 8). Elles sont mises en correspondance avec les doigts par un balayage perpendiculairement à l'axe d'inertie de la silhouette.

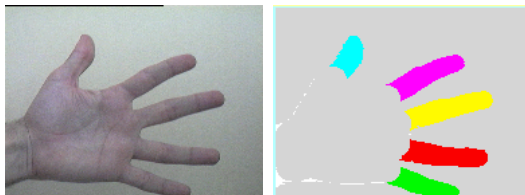


Figure 8: Détection des doigts.

Le poignet est localisé approximativement comme étant le centre de gravité des points d'intersection de l'axe d'inertie de la main et ceux des 4 grands doigts. La paume est obtenue par soustraction des régions correspondant aux doigts et à l'avant-bras.

6.2 Ajustement morphologique

L'étape de segmentation fournit la silhouette des différentes parties de la main. 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 à un facteur d'échelle près. L'épaisseur de la paume est supposée égale au diamètre du pouce. Les contraintes anthropométriques sont utilisées pour valider les mesures et déterminer les longueurs des phalanges du modèle.

7 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 précédent (ou initial) de ces contours par carte de distances du modèle 3D sur l'image.

7.1 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 9).

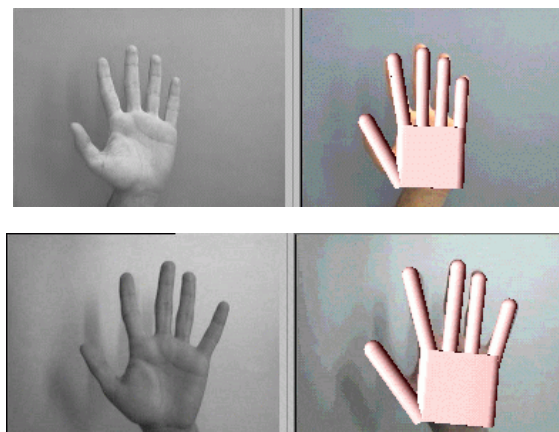


Figure 9: Exemples d'initialisation de suivi sur la main ouverte.

7.2 Extraction et projection des contours occultants du modèle

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 associées aux degrés de liberté regroupés dans le vecteur p des paramètres du modèle. La matrice de transformation globale est :

$$M_{globale} = T.R.S \quad (6)$$

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

Dans le cas d'objets articulés, les degrés de liberté internes doivent être pris en compte. Par exemple, la transformation 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 :

$$M(p) = M_{globale} \cdot M_{doigt} \cdot M_{a/a}^{phal1} \cdot M_{f/e}^{phal1} \cdot M_{f/e}^{phal2} \quad (7)$$

Désignant par 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 :

$$X(p) = M(p).X \quad (8)$$

L'ensemble des matrices associées aux articulations ainsi que leurs dérivées partielles sont pré-calculées. Les points de projection des segments de droites représentant les limbes sont ensuite obtenus par l'algorithme de Bresenham. Pour simplifier, nous avons modélisé la caméra par une projection orthographique. Dans le cas d'un modèle à primitives maillées, les limbes peuvent être obtenus par un algorithme de Z-buffer [Bourges97].

7.3 Extraction des contours dans l'image

Les contours occultants de la main sont des limbes, c'est-à-dire qu'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 occultent 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 comme étant des thalwegs de l'image, c'est-à-dire des vallées des niveaux de gris considérés comme un relief [Horain89] (figure 10). Les contours externes de la main sont ceux de la silhouette qui est obtenue comme précédemment à l'étape d'ajustement du modèle.

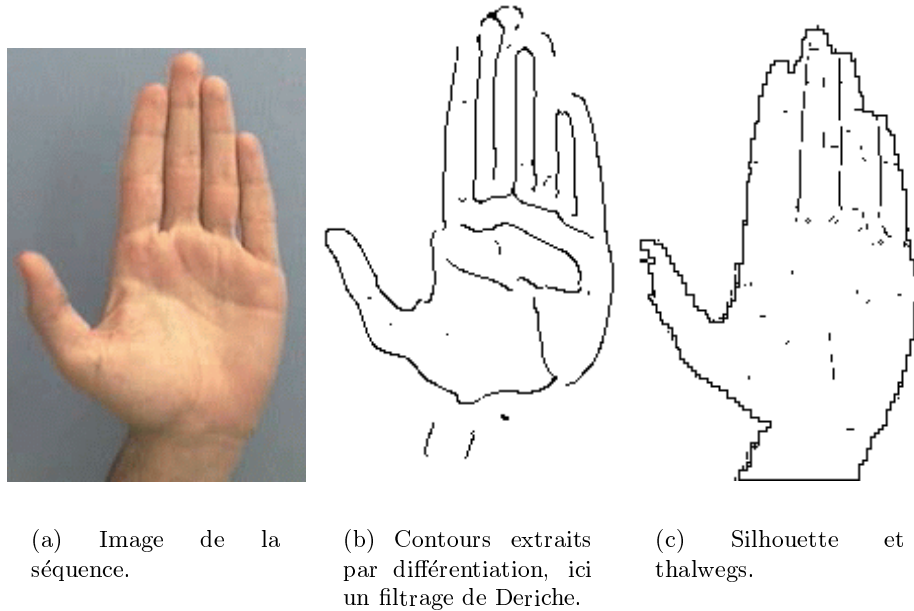


Figure 10: Estimation des contours dans l'image.

7.4 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 [Borgefors86]. Elle réalise un appariement implicite des points du modèle projeté avec les pixels de contour de l'image (figure 11).

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 :

$$F = \min_p \sum_{n_e} \sum_{n_p} d(Proj(X(p))) \quad (9)$$

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

La méthode de Levenberg-Marquardt 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 des arêtes du modèle avec le gradient de la carte de distance.

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

$$\nabla_p d(X(p)) = \sum_{n_e} \sum_{n_p} \nabla_p d(X(p)) \quad (10)$$

$$\nabla_p d(X(p)) = \nabla_{X(p)} d \cdot \nabla_p X(p) \quad (11)$$

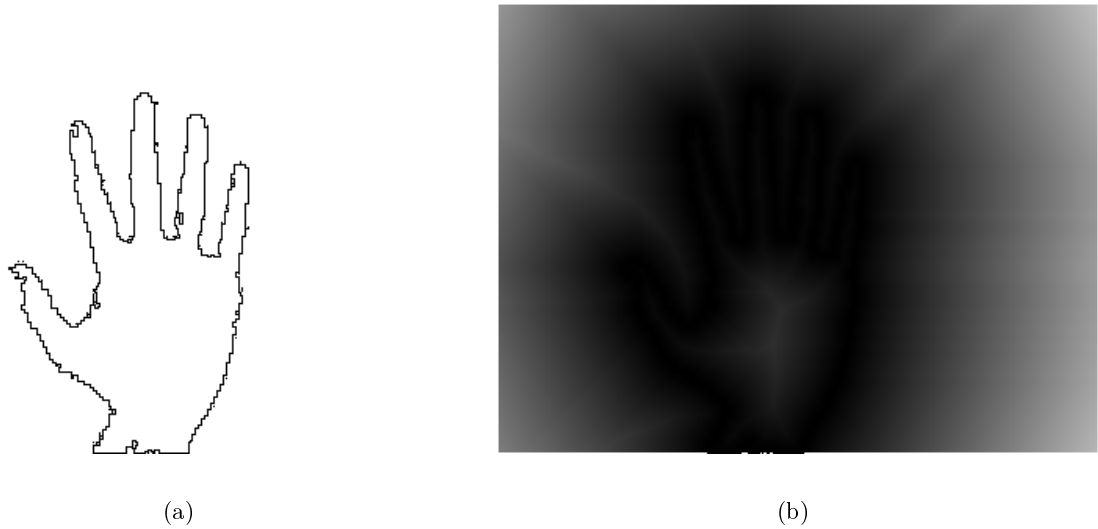


Figure 11: Contours de la main et carte de distances.

où $\nabla_X d$ 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 de 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 que 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 incorrects sont ramenées dans l'intervalle permis (figure 12).

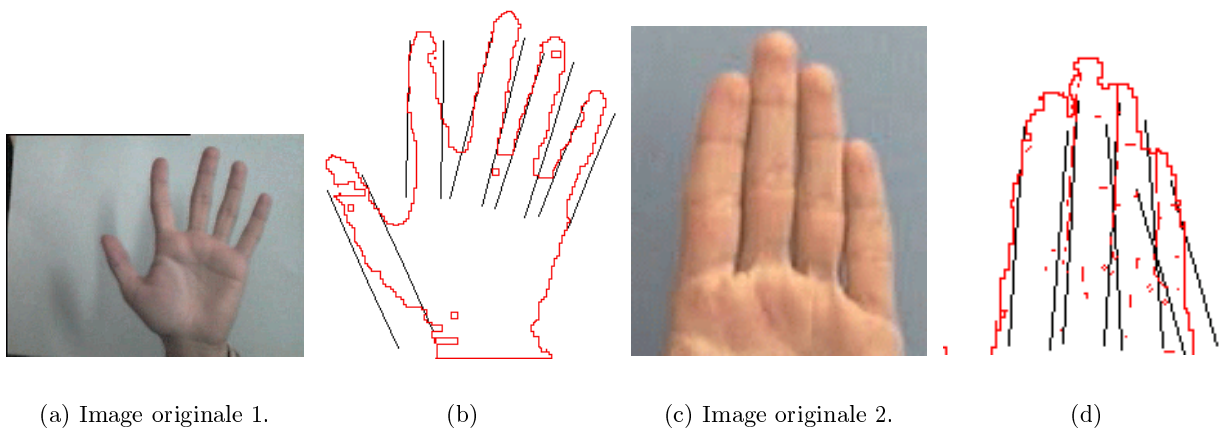


Figure 12: Recalage sur la silhouette, et éventuellement, les thalwegs.

7.5 Suivi dans les séquences

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 du suivi). Cette procédure est stabilisée par un filtre de Kalman étendu. Le vecteur d'état à estimer est constitué des 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 recalées 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 (figure 13). Nous envisageons de lever cette limitation en utilisant un filtre adaptatif combiné

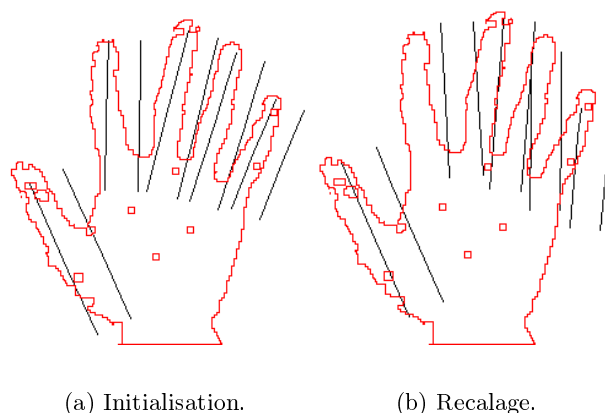


Figure 13: Recalage incorrect, lorsque l'initialisation est trop éloignée de la solution en cas de changement brusque du mouvement.

à une approche probabiliste.

8 Résultats

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 (figure 14).

9 Conclusion 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 du modèle sont ajustées automatiquement sur la main ouverte de l'opérateur dans la première 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 les abductions/adductions des doigts. Ils sont calculés par une procédure de recalage 3D/2D.

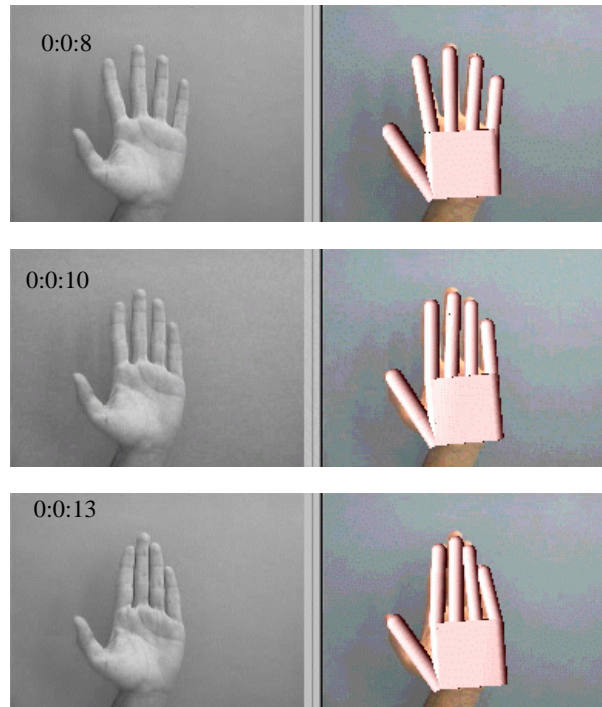


Figure 14: Recalage sur quelques images de la séquence abduction/adduction(28 images).

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 au moyen d'une transformation 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 le filtre de Kalman étendu.

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

References

- [Baudel93] T. BAUDEL and M. BEAUDOUIN-LAFON, "CHARADE : Remote Control of objects using Free-Hand Gestures", *Communications of the ACM*, Vol. 36, no 7, 1993, p. 28-35.
- [Bellik96] 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>
- [Buchholz92] B.BUCHHOLZ and T.J.ARMSTONG, "A Kinematic model of the human hand to evaluate its prehensile capabilities", *J. Biomechanics*, Vol. 25, no 2, 1992, p. 149-162.
- [Borgefors86] G, BORGEFORS, "Distance Transformations in Digital Images". *Computer Vision, Graphics and Image Processing*, Vol. 34, 1986, p. 344-371.

- [Bourges97] M. BOURGES-SEVENIER, P. HORAIN, F. PRETEUX, P. LERAY, “Recalage d’un modèle générique sur une séquence d’images 2D”, Actes *CORESA’97: 3^{ème} Journées d’études et d’échanges Compression et Représentation des Signaux Audiovisuels*, Issy-les-Moulineaux, France, 26-27 March 1997, p. 163-171.
- [Braffort96] A. BRAFFORT, *Reconnaissance et Compréhension de gestes, application à la langue des signes*, thèse de l’université de Paris XI, spécialité informatique, juin 1996.
- [Davis94] J. DAVIS, J. and M. SHAH, “Determining 3-D Hand Motion”, *Proc. 28th Asimolar Conf. Signals, Systems and Computers*, 1994.
- [Delamarre98] Q. DELAMARRE and O. FAUGERAS, “Finding pose of hand in video images: a stereo-based approach”, *Proceedings of FG’98*, April 14-16, 1998 Nara, Japan.
- [Deriche87] R. DERICHE, “Using Canny’s Criteria to Derive a Recursively Implemented Optimal Edge Detector”, *International Journal of Computer Vision*, Vol. 1, no. 2, 1987, p. 167-187
- [Dorner94] B. DORNER, *Chasing the color glove : visual hand tracking*, Master Thesis, Simon Fraser University, 1994.
- [Fua97] P. FUA, A. GRUEN, R. PLANKERS, N. APPUZO, D. THALMANN, “Human Body Modeling and Motion Analysis from Video Sequences”. Int. Symp. on Real Time Imaging and Dynamic Analysis, June 1998, Haddokate, Japon.
<http://ligwww.epfl.ch/~thalmann/FNRS-Fua.html>
- [Fukumuto94] M. FUKUMOTO, K. MASE and Y. SUENAGA, “Finger-Pointer: Pointing Interfaces by Image Processing”, *Computer and Graphics*, Vol. 18, no. 5, 1994, p. 633-642.
- [Gourret89] J. GOURRET, N. MAGNENAT-THALMANN and D. THALMANN, “Simulation of Object and Human Skin in a Grasping Task”, *Comp. Graphics*, Vol. 23, no. 3, Siggraph’89, 1989, p. 21-30.
- [Hienz96] H. HIENZ, K. GROBEL and G. OFFNER, “Real-Time Hand Arm Motion Analysing using a Single Video Camera”, *Proc. of Gesture Workshop’96*, Univ. of York, p. 323-327.
- [Heap96] T. HEAP and D. Hogg, “3D Deformable Hand Models, Progress in Gestural Interaction”, *Proc. of Gesture Workshop’96*, March 19th, 1996, University of York, UK, Philip A. Harling and Alstair D. N. Edwards (Eds), p. 131-139.
- [Horain89] P. HORAIN, “Extraction des arêtes dans des images de distance”, *Actes du 7^{ème} congrès RFIA*, AFCET-INRIA, Paris, 1989.
- [Iwai96] Y. IWAI, Y. YAGI and M. YACHIDA, “Gesture Recognition using Colored Gloves”, *Proc. of ICPR’96*, p. 662-666.
- [Ishibuchi92] ISHIBUCHI, H. TAKEMURA and F. KISHINO, “Real Time Hand Shape Recognition for Man-Machine Interfaces”, *IEEE. Int. Conf On Neural Networks*, 1992, p. 1407-1412.

- [Kush95] J. J. KUSH, and T. S. HUANG, "Vision Based Hand Modeling and Tracking for Virtual Teleconferencing and Telecollaboration". In *Proc. of fifth International Conference on Computer Vision*, June 20-23, 1995, MIT, Cambridge, Massachusetts, p. 666-671.
- [Kuno94] Kuno Y.KUNO, M. SAKAMOTO, K. SAKATA and Y. SHIRAI, "Vision Based Human Interface with User-Centered Frame", *Proc. of IROS'94*, p. 2923-2929.
- [Lee95] J. LEE and T. KUNII, "Model-Based Analysis of Hand Posture". *IEEE Computer Graphics and Applications*, September 1995, p. 77-86.
- [Lowe92] D. G. LOWE, "Robust Model-based Motion Tracking Through the Integration of Search and Estimation". *International Journal of Computer Vision*, Vol. 8, no 2, 1992, p. 113-122.
- [Millar96] R. J. MILLAR, C. F. CRAWFORD, "A Mathematical Model for Hand-Shape Analysis", *Proc. Of Gesture Workshop'96*, March 19th 1996, University of York, UK, Philip A. Harling and Alstair D. N. Edwards (Eds), p. 235-245.
- [Mochimaru94] M. MOCHIMARU and N. YAMAZAKI, "The three-dimensional measurement of unconstrained motion using a model-matching method", *ERGONOMICS*, Vol. 37, No 3, 1994, p. 493-510.
- [Ouhaddi98] H. OUHADDI et P. HORAIN, "Conception et ajustement d'un modèle 3D articulé de la main", *Actes 6^{èmes} Journées de Travail du GT Réalité Virtuelle*, Issy-les-Moulineaux, 12-13 Mars, 1998, p. 83-90. <http://www-sim.int-evry/Publications>.
- [Preteux98] F.PRETEUX, P. HORAIN, H. OUHADDI and M. PREDA, *Report on Core Experiment 3 on Hand Baps interpretation*, ISO/IEC JTC1/WG1 MPEG97/M3332, March 1998.
- [Regh94] J.M. REGH T. KANADE, "Visual Tracking of Height DOF Articulated Structures: an Application to Human Hand Tracking", *Proc. 3^d ECCV*, Vol 2, 1994, p. 37-46.
- [Rijpkema91] H. RIJPKEMA and M. GIRARD, "Computer Animation of Knowledge-Based Human Grasping", *Computer Graphics*, Vol. 25, no. 4, 1991, p. 339-348.
- [Serra94] J. SERRA and P. SOILLE, *Mathematical morphology and its applications to image processing*, Kluwer Academic Publishers, 1994.
- [Shimada95] N. SHIMADA, Y. SHIRAI. and Y. KUNO, "Hand Gesture Recognition Using Computer Vision Based on Model-Matching Method". In *Proc. of 5 th International Conference on HCI*, p. 11-16. Elsevier, 1995.
- [Shimada96] Y. SHIMADA, and Y. SHIRAI, "3-D Hand Pose Estimation and Shape Model Refinement From a Monocular Image Sequence", 1996.
<http://www-cv.mech.eng.osaka-u.ac.jp/~shimada/research/research.html>
- [Teo94] C.G. TEO, *A Hybrid Procedural Knowledge-Based Approach to the Animation of Human Hand Grasping*, Master Thesis, Simon University, 1994.

- [Utsu96] A. UTSU, T. MIYASATO, F. KISHINO and R.NAKATSU, "Hand Gesture Recognition System using Multiple Cameras", *Proc ICPR'96*, Vol. I, p. 667-671.
- [Wagner88] C. WAGNER, "The pianist's hand anthropometry and biomechanics", *Ergonomics*, Vol. 31, no 1, 1988, p. 97-131.

Liste des figures

1	Images de la main sans et avec auto-occultations.	2
2	Squelette de la main.	4
3	Mouvements d'adduction et d'abduction.	5
4	Mouvements de flexion et d'extension.	5
5	Modèle de la main.	7
6	Synthèse de signes de l'alphabet de la langue des signes (A, D, F respectivement.)	7
7	Segmentation de la main par seuillage sur la couleur.	8
8	Détection des doigts.	8
9	Exemples d'initialisation de suivi sur la main ouverte.	9
10	Estimation des contours dans l'image.	11
11	Contours de la main et carte de distances.	12
12	Recalage sur la silhouette, et éventuellement, les thalwegs.	12
13	Recalage incorrect, lorsque l'initialisation est trop éloignée de la solution en cas de changement brusque du mouvement.	13
14	Recalage sur quelques images de la séquence abduction/adduction(28 images).	14