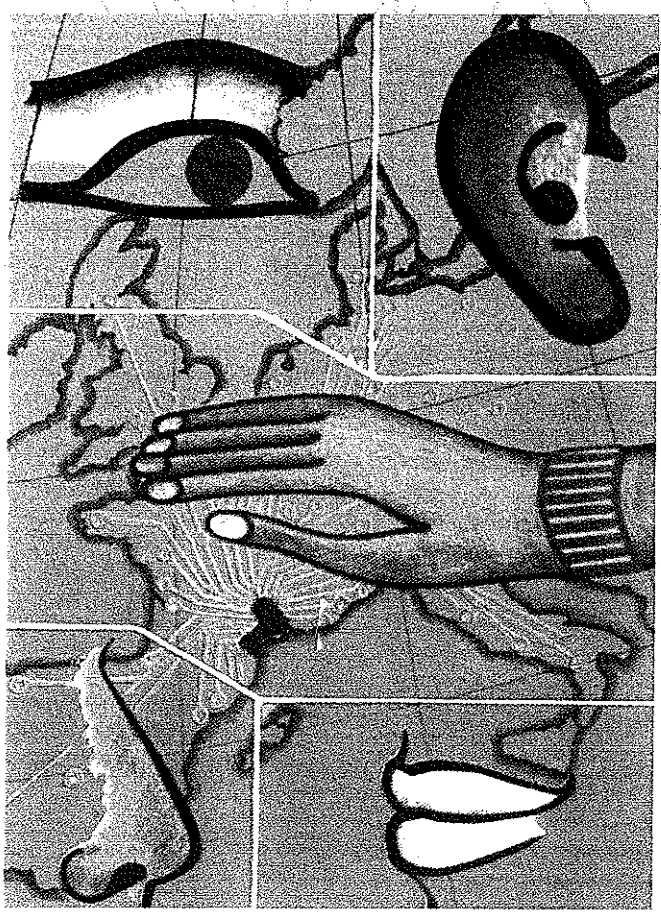
Lettre mensuelle de veille sur les technologies de l'information et les outils de productivité intellectuelle

Numéro 123

MAI 1997

La Lettre de l'IA

Interfaces 97



actes

proceedings

FUSION DE DONNÉES ÉLECTROMAGNÉTIQUES ET VIDÉO POUR LE SUIVI DE MOUVEMENTS EN TEMPS RÉEL

VINCENT RODIN, FLORENT GAUDILLAT & JACQUES TISSEAU

École Nationale d'Ingénieurs de Brest

Technopôle Brest-Iroise, CP 15
29608 Brest Cedex, France
Phone : 02 98 05 66 26
Fax : 02 98 05 66 29
e-mail : rodin@enib.fr

Résumé : Cet article décrit un système de suivi de gestes de la main en temps réel fonctionnant sur une station de travail Indy de Silicon Graphics. Il porte plus particulièrement sur la fusion de données, issues d'un capteur de position et d'une caméra vidéo couleur, en vue d'accélérer les traitements. Les variations de la forme de la main (déformations locales) sont prises en compte à l'aide d'un contour actif évoluant dans les images et les déplacements de la main dans l'espace (déformations globales) sont détectés par le capteur de position. Nous décrivons une méthode permettant de trouver les déformations locales par segmentation, filtrage et évolution d'un contour actif. Nous détaillons également l'utilisation des déformations globales permettant de suivre des mouvements rapides de l'utilisateur. Ce dernier point implique une calibration de l'ensemble caméra-capteur pour connaître à chaque instant la position du capteur dans l'image.

Mots clés : Capteur de position, contour actif, fusion de données, suivi de la main, traitement d'images.

Abstract: This paper describes a real-time movement tracking system running on a Silicon Graphics Indy station. It deals more particularly with the data fusion from a position sensor and a standard color camera so as to speed up processing. Variations of the shape of the hand (local deformations) are taken into account using an active contour model and the motion of the hand (global deformations) is detected by the position sensor. We describe a method for finding local deformations by the segmentation, filtering and moving of an active contour. We also take into account details for the use of global deformations allowing to follow the user's quick movements. The latter involves a calibration of the camera-sensor ensemble which enables knowing the exact location of the sensor in the image at all times.

Keywords: Active contour, data fusion, hand tracking, image processing, position sensor.

1. INTRODUCTION

En dépit de nombreux travaux sur la reconnaissance et l'interprétation de gestes [1, 2], le suivi de mouvements de la main en temps réel, est toujours un problème difficile en vision par ordinateur. La plupart des systèmes de suivi de gestes actuels ne fonctionnent pas en temps réel [3, 4] ou bien, utilisent de puissantes machines [5] ou des cartes de traitement d'images spécialisées [6]. Revenons sur deux articles qui nous semblent refléter les principales approches en suivi de geste.

Les modèles stochastiques permettent la modélisation et la segmentation d'objets bidimensionnels déformables en mouvement dans un séquence d'images. Dans [3], les auteurs proposent d'utiliser un tel modèle en y incorporant des connaissances *a priori* sur la structure et les variations de l'objet à suivre. Une configuration particulière de la forme de la main est dérivée du modèle original en spécifiant deux sortes de transformations :

- Les déformations globales correspondant à la translation, la rotation, le facteur d'échelle et à n modes de variations principales associées à la transformation de Karhunen-Loeve des déformations observées sur une population représentative.
- Les déformations locales modélisées comme un processus aléatoire qui modifie localement la position des points appartenant au modèle déformable global.

L'avantage de cette méthode est sa robustesse au bruit et aux occlusions. L'inconvénient est que la recherche du contour dans une image est relativement longue.

L'utilisation d'un modèle 2D déformable, décrit dans [5], est une autre approche pour le suivi de gestes. En effet, ce type de modèle est attiré par les contours de l'image et les propriétés extraites d'une image sont utilisées pour initialiser l'image suivante. Le modèle est constitué de n points de contrôle. Il se présente sous la forme :

$$X = \overline{X} + wV$$

où : $\overline{X}$ représente la forme moyenne de l'objet.

V est la matrice des vecteurs des variations les plus significatives.

w correspond à un vecteur de poids pour les variations individuelles.

Chaque point va se déplacer vers le contour le plus fort de son voisinage. Le déplacement de chaque point est étudié afin d'en déduire les paramètres de translation, de rotation et de facteur d'échelle du modèle. Les mouvements résiduels sont accommodés en déformant le modèle sous certaines contraintes (ajustement des paramètres w_i). Le modèle est utilisé en combinaison avec un algorithme génétique pour accomplir la recherche de l'image initiale globale. La reconnaissance de gestes est aisée, car le modèle utilise peu de paramètres et fonctionne même avec des fonds chargés et des variations d'intensité lumineuse. Il travaille en temps réel. L'inconvénient de ce système est qu'il est seulement capable de détecter les contours d'une main ouverte et nécessite une machine puissante (DEC Alpha).

Pour notre part, nous avons réalisé un système de suivi de gestes en temps réel fonctionnant sur une station de travail Indy (100 Mhz) de Silicon Graphics équipée d'une caméra couleur standard. Notre système utilise également un capteur de position à champ électromagnétique pulsé [7] (Flock of Birds) qui permet d'accélérer les traitements en fournissant à chaque instant la position de la main dans l'image.

Afin de retrouver les contours de la main dans une image, nous utilisons un modèle de contour actif [8] qui évolue sur une image de régions. Nous combinons ce modèle avec les données du capteur de position pour initialiser la recherche du contour et éviter les décrochages lors de mouvements rapides.

Dans notre système, nous considérons la main comme un objet déformable caractérisé par des déformations locales (variations de la forme de la main) et des déformations globales (rotation, translation, facteur d'échelle). Les déformations locales sont prises en compte par le modèle de contour actif tandis que le capteur de position fournit des informations sur les déformations globales.

Dans la suite de cet article, nous présentons une méthode pour trouver les déformations locales par segmentation, filtrage et évolution d'un contour actif. Nous détaillons également l'utilisation des déformations globales qui permettent de suivre des mouvements rapides de l'utilisateur. Ce dernier point implique une calibration de l'ensemble caméra-capteur pour connaître à tout instant la position du capteur dans l'image.

2. Les déformations locales

Afin de retrouver les contours de la main, nous avons utilisé un modèle de contour actif. Ce modèle, introduit par Kass, Witkin et Terzopoulos [8], a été mis en œuvre dans la plupart des systèmes de suivi de mouvements. Un contour actif se présente sous la forme d'une courbe (fermée ou non) définie par un certain nombre de points de contrôle. L'initialisation du contour actif doit être située à proximité du contour recherché et son évolution s'effectue selon un processus itératif de déformations contrôlé par un test de convergence [9]. Dans le cas le plus élémentaire, l'initialisation s'effectue par interaction avec l'utilisateur. La convergence du contour actif est généralement vue comme une condition de stabilité du modèle en terme de contraintes physiques (raideur, élasticité...) attachées aux données. Ce modèle de contour actif a été largement employé en suivi de mouvement car le contour détecté dans une image peut servir à initialiser le modèle pour trouver le contour dans l'image suivante.

Dans notre application, nous faisons évoluer un contour actif discret (non approximé par une B-spline à partir des points de contrôle) sur une séquence d'images. Chaque image de la séquence est traitée par segmentation couleur afin de faire ressortir une région uniforme sur laquelle le contour va s'adapter (la région qui correspond à la main).

2.1 Segmentation couleur

Afin de détecter plus facilement la main dans l'image, nous avons réalisé une première expérience où l'utilisateur porte un gant de couleur verte. L'extraction des régions vertes de l'image est alors très simple, car un pixel appartiendra à la couleur verte s'il vérifie l'équation : Vert–Bleu–Rouge≥0.

Un filtre médian [10] est ensuite appliqué afin de supprimer le bruit résiduel dans l'image des régions (voir figure 1).

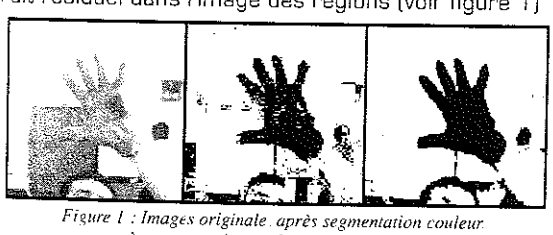


Figure 1 : Images originale, après segmentation couleur, après segmentation couleur et filtrage médian

L'inconvénient principal de cette méthode est que les couleurs sombres peuvent être interprètées comme une forme particulière de vert.

Afin de résoudre ce problème, nous avons effectué une deuxième expérience permettant de détecter directement la couleur de la peau. Cette nouvelle approche consiste à présenter au système la main de l'utilisateur afin de créer une table de couleurs par échantillons. La calibration du système de couleur doit se faire sur plusieurs plans, car la couleur de la main dépend du plan sur lequel elle se trouve. En effet, plus elle est proche de la caméra, plus elle est claire.

Ainsi, au cours de cette deuxième expérience, plusieurs tables de couleurs sont obtenues par échantillonnage en des plans connus (via le capteur de position). Lors du suivi de la main dans une séquence d'images, la table de couleurs à utiliser pour chacune des images est facilement déterminée. En effet, comme nous le verrons dans la suite de cet article, le capteur de position permet de connaître à chaque instant la distance entre la main et l'objectif de la caméra et donc de connaître la table de couleurs à utiliser.

Il est à noter que, lors de l'utilisation de la couleur de la peau, la présence du visage et d'autres parties du corps peuvent poser des problèmes lorsqu'ils se trouvent superposés dans l'image.

2.2 Évolution du contour actif

À l'aide de l'image des régions obtenue précédemment, nous faisons évoluer un contour actif qui va converger vers les frontières de la région de la main. Pour cela, nous utilisons un modèle de contour actif fermé et discret constitué d'un certain nombre de points de contrôle. Le nombre de ces derniers varie entre 50 et 200 en fonction de la position et de la taille de la main dans l'image. C'est-à-dire, lorsque la distance entre deux points de contrôle devient trop importante, on crée un nouveau point entre les deux. Réciproquement, si la distance entre deux points de contrôle devient trop petite, on supprime un des deux points. Cela traduit la notion d'élasticité du modèle.

Chaque point de contrôle se déplace dans la direction perpendiculaire à la direction définie par ses voisins (voir figure 2-a). Chaque point peut se déplacer soit vers l'intérieur, soit vers l'extérieur du modèle. Le sens du déplacement est défini par un test d'appartenance ou de non appartenance à la région de la main.

Le critère d'arrêt des points de contrôle est le changement de région du point considéré (voir figure 2-b). Le contour actif évoluant sur une image binaire, il lui est facile de détecter les frontières de la main. Signalons que nous n'avons pas de stratégie particulière pour choisir le point de contrôle à faire évoluer. En effet, les points de contrôle évoluent dans l'ordre dans lequel ils apparaissent dans le contour actif.

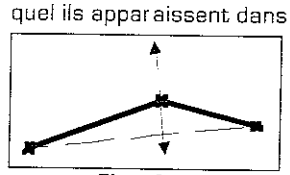


Figure 2-a :
Déplacement d'un point de contrôle

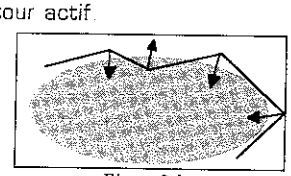


Figure 2-b :
Évolution des points de contrôle

Pour des raisons de rapidité d'exécution, nous avons supprimé le calcul de la fonction d'énergie généralement associée à un contour actif. Malheureusement, ceci engendre un pro-

blème. En effet, lorsque l'image des régions est imparfaite (visible par la présence de trous dus au bruit), le contour a tendance à effectuer des boucles infinies. Il faut donc supprimer toutes les boucles se formant durant l'évolution. Une méthode efficace consiste à calculer la distance entre le point à déplacer et les segments composant le contour (voir figure 3). Si cette distance est petite, tous les points de contrôle qui composent la boucle sont supprimés.

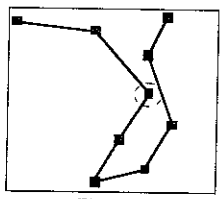


Figure 3 :
Élimination d'une bouche présente dans un contour actif

Si nous supposons que les variations de la position du contour de la main d'une image à l'autre sont faibles, le contour détecté dans une image peut alors être utilisé pour initialiser la recherche du contour dans l'image suivante.

Nous devons donc effectuer une première recherche du contour pour initialiser le modèle. L'initialisation du modèle nécessite la connaissance d'un point image appartenant à la main. Le contour actif est alors centré sur ce point et évolue en s'écartant vers les frontières de la région de la main. Le temps mis pour trouver le premier contour est de l'ordre de trois secondes pendant lesquelles la main doit rester immobile. Le point image par lequel commence l'initialisation du modèle peut être déterminé soit en plaçant la main au centre de l'image, soit en utilisant le capteur de position. En effet, le capteur de position nous permet de connaître la position de la main dans l'image (cf. section 4).

L'initialisation du contour actif peut être suivie sur la figure 4. Nous avons placé un carré blanc pour simuler une imperfection (un trou) dans la région de la main. Nous observons que le contour actif a tendance à contourner le trou des deux côtés. Lorsque ceux-ci se rejoignent et que le trou a été franchi, la boucle est éliminée.

Le contour obtenu est ensuite projeté d'une image à l'autre et doit à chaque fois s'adapter au nouveau contour de la main.

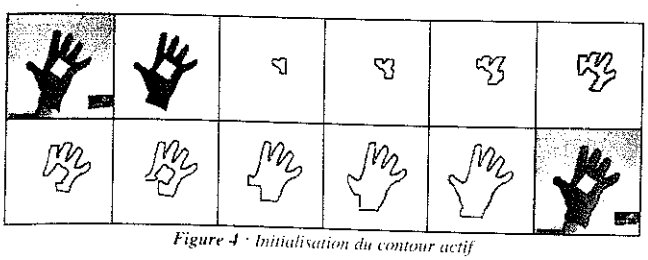


Figure 4 : Initialisation du contour actif

Cette méthode est efficace et permet de suivre des mouvements lents en temps réel à l'aide uniquement d'une caméra vidéo couleur. Malheureusement, pour des mouvements rapides de la main, des difficultés apparaissent si l'on n'utilise que la caméra. En effet, la convergence du contour actif est plus longue car l'écart entre le contour projeté et le nouveau contour devient important. Nous avons constaté expérimentalement que, lors d'un écart d'une dizaine de pixels, le système commençait à ne plus suivre et finalement décrochait. Cet écart dépend bien évidemment de la puissance de la machine sur laquelle s'exécute le suivi de geste.

3. Les déformations globales

Afin de minimiser, lors de mouvements rapides, l'écart entre le contour projeté et le nouveau contour de la main, nous introduisons un nouvel élément qui est un capteur de position électromagnétique. Ce capteur est capable de fournir son orientation et sa position dans l'espace (déformations globales) par rapport à un émetteur de champ magnétique continu pulsé [7] (Flock of Birds). L'émetteur est constitué d'une antenne fixe créant un champ tournant dans l'espace au moyen de trois solénoïdes dont les axes sont mutuellement perpendiculaires. Le capteur (ou récepteur) est lui aussi constitué de trois solénoïdes perpendiculaires. L'orientation du capteur est déduite des différentes phases entre le champ émis et les courant induits. La position du capteur est évaluée à l'aide de l'intensité des courants.

L'utilisateur porte le capteur au niveau de la main. Ceci permet de connaître à tout instant la position de la main dans l'image. Le principal avantage de l'utilisation de ce capteur est de pouvoir projeter le contour actif trouvé dans une image à une position très proche du nouveau contour à extraire dans l'image suivante (voir figure 5).

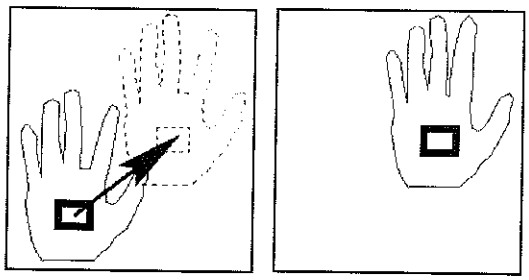


Figure 5 : Projection du contour actif d'une image à l'autre à l'aide du capteur de position

L'initialisation de la recherche du contour dans l'image suivante est de la forme :

$$\vec{X}_i = \vec{X}_{i-1} + \vec{T}_i$$

où : $\vec{X}_{i-1}$ est le modèle extrait de l'image précédente

$\vec{X}_i$ est le modèle projeté dans l'image courante

$\vec{T}_i$ est la translation calculée à l'aide du capteur de position

Il est à noter que, pour des raisons de rapidité et de simplicité des calculs, nous ne prenons en compte ici que de la translation de la main. Dans l'avenir, nous pensons utiliser l'orientation de la main fournie par le capteur. En effet, celle-ci devrait faciliter et accélérer la reconnaissance de gestes simples en temps réel.

4. La calibration et la détection du capteur dans l'image

Le calcul de la position du capteur dans l'image nécessite une calibration du système caméra-capteur. La méthode de calibration retenue passe tout d'abord par une modélisation de la caméra à l'aide du modèle pin-hole [11] (figure 6). Il est à noter que d'autres modèles plus complexes existent (modèle lentille mince, modèle lentille épaisse) et que l'on peut se ramener, pour chacun de ces modèles, à un modèle pin-hole équivalent.

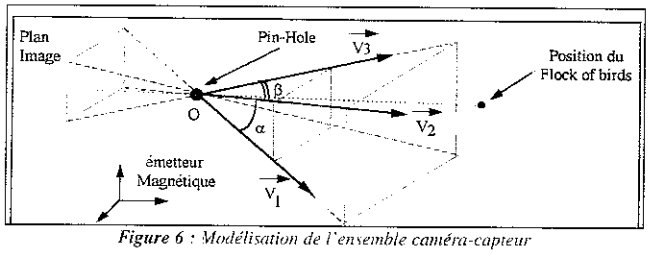


Figure 6 : Modélisation de l'ensemble caméra-capteur

Le but de la calibration est de déterminer les angles α et β correspondant aux angles de prise de vue de la caméra (figure 6) Le calcul de la position du capteur dans l image consistera à calculer les angles α_M et β_M qui représentent les coordonnées du capteur dans le cône de la caméra (figure 7) À partir de ces angles il est très simple de trouver la position du capteur dans l'image

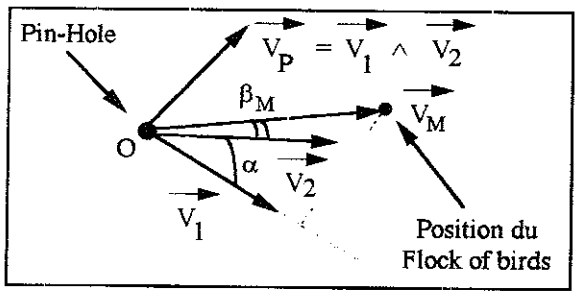


Figure 7 Détermination des α_M et β_M

4.1. Détermination des angles α et β

Afin de calculer les angles α et β nous déterminons les vecteurs $\vec{V}_1$ $\vec{V}_2$ et $\vec{V}_3$ (figure 6) à partir de 4 mesures de la position du capteur (capteur situé sur l'objectif de la caméra et dans 3 coins différents de l image) Les angles α et β sont alors définis par les produits scalaires :

$$\vec{V}_1 \cdot \vec{V}_2 = \left\|\vec{V}_1\right\| \left\|\vec{V}_2\right\| \cos\alpha \quad \text{et} \quad \vec{V}_2 \cdot \vec{V}_3 = \left\|\vec{V}_2\right\| \left\|\vec{V}_3\right\| \cos\beta$$

4.2. Détermination des angles α_M et β_M

Soit le vecteur $\vec{V}_M$ défini par la position du capteur dans l espace et par la position de l'objectif déterminée lors de la calibration (figure 7)

L angle β_M correspond à l angle entre la droite définie par (O $\vec{V}_M$) et le plan P défini par ($\vec{V}_1$ $\vec{V}_2$)

Soit $\vec{V}_P = \vec{V}_2 \wedge \vec{V}_1$ un vecteur normal au plan P. D après la figure 7 on peut en déduire l'angle β_M par la relation suivante :

$$\vec{V}_M \cdot \vec{V}_P = \left\|\vec{V}_M\right\| \left\|\vec{V}_P\right\| \cos\left(\frac{\pi}{2} - \beta_M\right) = \left\|\vec{V}_M\right\| \left\|\vec{V}_P\right\| \sin\beta_M$$

α_M peut bien entendu être évalué par la même méthode

Connaissant maintenant les angles α_M et β_M la position (i j) du capteur dans l'image peut être déterminée par :

$$i = \frac{\alpha_M}{\alpha} \text{ nbLignes} \quad \text{et} \quad j = \frac{\beta_M}{\beta} \text{ nbColonnes}$$

5. Conclusion

Les travaux présentés sont consacrés à l étude d un système de suivi de gestes Cette étude porte plus particulièrement sur la fusion de données provenant d un capteur de position et d une caméra couleur en vue d accélérer les traitements Les variations de la forme de la main sont prises en compte à l aide d un contour actif et les déplacements de la main sont détectés par le capteur de position

Le système fonctionne sur une station de travail Indy (SGI) ayant un processeur à 100 Mhz et est capable de traiter de 15 à 28 images par seconde (suivant la taille de la main dans l'image et suivant la précision du tracé souhaitée). Il fonctionne bien avec des fonds chargés Cependant les reflets et le contre-jour au niveau de la main sont sources d erreurs de segmentation Le système présente malgré tout une bonne immunité au bruit.

Différentes évolutions doivent être envisagées pour améliorer notre système Nous pensons tout d abord que l utilisation d'un modèle de la main serait très utile pour résoudre le problème des occlusions partielles lors d'un suivi De plus il serait intéressant de combiner avec le traitement de la couleur, une contrainte sur le gradient temporel Ceci devrait permettre de séparer la main des autres parties du corps (visage .) Nous pensons également que l'orientation de la main fournie par le capteur électromagnétique devrait faciliter et accélérer la reconnaissance de gestes

Références

[1] J Lee et T. L. Kunii : *Model-based analysis of hand posture* ; IEEE Computer Graphics and Applications vol. 15, n° 5 septembre 1995.

[2] S Gibet A Braffort C Collet F. Forest R. Gherbi et T. Lebourque : *Gesture in human-machine communication capture analysis-synthesis recognition semantics* Proceedings of gesture workshop 96 (York GB). Progress in Gestural interaction, Springer-Verlag 1996

[3] C Kervrann et F Heitz : *Robust tracking of stochastic deformable model in long image sequence* ; IEEE International Conference on Image Processing novembre 1994 Austin États-Unis

[4] T Heap et D Hogg : *3D deformable hand models* ; Proceedings of gesture workshop 96 (York, GB). Progress in gestural interaction Springer-Verlag 1996

[5] T Heap : *Real-time hand tracking and gesture recognition using smart snakes* ; Actes des Journées l'Interface des Mondes Réels et Virtuels (Éditions EC2) Montpellier 1995.

[6] A Sutherland : *Real-time video-based recognition of sign language gestures using template matching* ; Proceedings of gesture workshop'96 (York GB) Progress in gestural interaction Springer-Verlag, 1996

[7] Ascension Technology Corporation PO Box 527, Burlington Vermont 05402 États-Unis *The Flock of Birds™ Position and orientation measurement system*, 1995

[8] M Kass A Witkin et D Terzopoulos : *Active contour models* ; International Journal of Computer Vision vol 1 pp 312-331 1988

[9] J.-P. Cocquerez et S Philipp : *Analyse d images filtrage et segmentation* ; Masson 1995

[10] J W Tukey : *Exploratory data analysis* ; Addison-Wesley 1977

[11] B K Horn : *Robot Vision* ; The MIT Press 1986