

# Estimation mono-vue de la posture du bras en utilisant un modèle biomécanique.

Frederick GIANNI<sup>1</sup>

Boris LENSEIGNE<sup>1</sup>

Patrice DALLE<sup>1</sup>

<sup>1</sup> IRIT, Université Paul Sabatier

118, route de Narbonne  
31062 TOULOUSE cedex 4

gianni@irit.fr, lenseign@irit.fr, dalle@irit.fr

## Résumé

L'étude du mouvement, et plus particulièrement des gestes d'une personne humaine à partir de séquences d'images, est un domaine du traitement d'images en pleine expansion. D'autre part l'étude des gestes et des mouvements corporels a conduit à l'émergence de la biomécanique qui a pour but de modéliser la mécanique de ces gestes et des composants permettant de les effectuer (squelette, muscles, etc ...). Nous associons ces deux disciplines pour proposer un nouvel algorithme de reconstruction de la posture du bras à partir d'une image du locuteur prise par une caméra calibrée. La mise en œuvre de cette technique se fait en deux étapes : d'une part la validation du principe et des modèles à partir d'une scène et d'un dispositif d'acquisition virtuels, d'autre part, l'étude sur des images réelles. La première étape nous permet de dégager les contraintes et tolérances sur les opérateurs de traitement d'images à mettre en œuvre, la seconde nécessite l'emploi d'opérateurs de traitement d'images dont les performances concordent avec les contraintes définies précédemment. Cet article concerne la première étape, c'est-à-dire la définition des connaissances préalables sur la scène et sur le dispositif d'acquisition, la résolution purement géométrique du problème. L'application de cet algorithme sur des séquences d'images et l'étude de la sensibilité de l'algorithme au bruit. Enfin nous proposons un modèle biomécanique du bras permettant le passage des coordonnées cartésiennes des articulations dans l'espace à leurs coordonnées angulaires. Les coordonnées angulaires sont finalement utilisées pour éliminer les positions ou trajectoires impossibles des ensembles de candidats.

## Mots Clés

Vision par ordinateur, analyse du geste, espace atteignable, paramètres de Denavit Hartenberg

## Abstract

*Motion and especially human gesture analysis is a new growing field of interest in Computer Vision research. On the other hand, human gesture and body motion have been widely studied in biomechanics. The aim of biomechanic is to build models of the mechanical structure (bones, muscles,*

*etc...) used to produce these gestures. We use methods combining these two disciplines, to propose a new simple technique to build the posture of the arm from a calibrated single-view video sequence. We implement this method in two stages : in a first time, the method is geometrically validated on a virtual scene watched with a virtual camera. This stage allows us to define the constraints and tolerances on the image processing operators we need. The second stage concerns the implementation of the method on real images, which needs the use of image processing tools whose performances agree with the constraints defined previously. This paper only concerns the first stage. We define some preconditions about the scene and the acquisition system and we present a geometric method which reduces the possible configurations for the arm to only four of them. We show how to use this algorithm on image sequences and we study its sensitivity to noise. Finally, we propose a biomechanical model defining the transformation from Cartesian coordinates of each joint in space to the angular coordinates of those joints. Angular coordinates are used to eliminate impossible configurations from the set of possible configurations.*

## Keywords

*Computer Vision, gesture analysis, reachable space, Denavit Hartenberg parameters*

## 1 Problématique

Notre travail se situe dans le cadre de l'étude des composantes langagières du geste. Plus exactement, nous nous intéressons aux cas où le geste est produit avec une intention de communication, tels que les langues des Signes ou encore certains gestes coverbaux.

### 1.1 Étude des langues des Signes

Les langues des Signes sont les langues naturelles des sourds à travers le monde. L'étude de ces langues au moyen de l'outil informatique a jusqu'ici porté essentiellement sur des outils de transcription/notation, des dictionnaires multimédia permettant la recherche dans des bases de données vidéo et des outils de traduction

automatique. Des outils de transcription/notation existent également depuis un certain temps<sup>1234</sup>. Ces outils fournissent une interface permettant d’annoter manuellement des séquences vidéo, mais certaines recherches récentes visent à automatiser partiellement cette tâche en intégrant des opérateurs de traitement d’images dans les logiciels de transcription [4]. Dans le domaine de la traduction automatique, diverses techniques de reconnaissance, voisines de celles utilisées en reconnaissance de la parole, sont utilisées [2] [19] [20] [14] [5]. Les primitives fournies à ces systèmes de reconnaissance proviennent, soit de périphériques spécifiques tels que des gants numériques, soit de systèmes de Vision. Dans le premier cas, l’information disponible concerne directement les valeurs articulaires des doigts ainsi que la position 3D de la main dans l’espace. Parmi les systèmes se basant sur la Vision par ordinateur, certains systèmes de reconnaissance fonctionnent directement dans l’espace 2D ou 2D + t de l’image [17], [19]; mais ces systèmes ne reconnaissent que des configurations ou des signes standards appris préalablement par le système. Nous nous intéressons pour notre part à la langue des Signes “naturelle” dans le but de concevoir un système d’analyse d’images dédié à l’étude de cette langue. Comme les langues des Signes font appel, en plus du vocabulaire standard, à des structures de grande iconicité [6] (construction de l’espace de signation, pointages, transferts), l’utilisation d’informations tridimensionnelles semble indispensable. En l’absence de capteurs spécifiques, il devient nécessaire d’exploiter des techniques de Vision artificielle, en particulier les techniques d’analyse du mouvement (Cf. 1.3).

## 1.2 Analyse des gestes coverbaux

Les gestes coverbaux accompagnent un acte de communication orale, souvent sans même que le locuteur en ait conscience. Ils complètent - ou remplacent - l’information orale. À la différence des langues des Signes, il n’existe pas de vocabulaire standard ; les gestes coverbaux sont le plus souvent dépendants du contexte du discours et de l’environnement du locuteur. De plus, leur compréhension demande de la part des deux interlocuteurs une certaine culture commune, permettant par exemple d’attribuer le même sens à une catégorie de gestes. En revanche, à la différence des applications mettant en œuvre de la commande gestuelle dans leur interface, la précision requise est moindre puisque ces gestes sont toujours effectués dans un contexte donné : les informations tridimensionnelles requises visent essentiellement à résoudre des

dilemmes du type “devant/derrière” ou à déterminer quel objet parmi plusieurs est l’objet du discours.

## 1.3 Analyse et représentation des gestes

L’étude du geste en tant qu’acte de communication repose sur l’étude du mouvement, plus particulièrement sur celui des membres supérieurs. La description de ces mouvements a fait l’objet d’un grand nombre de travaux. Un état de l’art de ces travaux peut être trouvé dans [10]. Plus récemment dans [8] les auteurs utilisent plusieurs caméras non calibrées pour estimer la posture du corps complet. D’autres applications n’utilisent qu’une seule caméra, calibrée pour reconstruire le mouvement d’un bras [7]. On peut également noter les travaux de [21] qui estime la conformation d’une chaîne articulée à partir d’une caméra non calibrée, en estimant simultanément une matrice de projection perspective. Dans [3] les auteurs modélisent les mouvements d’un corps articulé (une chaîne cinématique) afin de calculer dans une image les valeurs angulaires des articulations à partir de celles connues dans l’image précédente. Enfin [18] fournit un modèle du bras humain à quatre degrés de liberté et l’utilise pour estimer la pose d’un bras à partir d’images fixes. Nous utilisons pour notre part une caméra calibrée associée à un modèle mécanique du bras. Ce modèle nous permet de lier les coordonnées articulaires du bras et sa position dans l’espace et de prendre en compte les valeurs limites sur les articulations.

Un modèle proche du notre associé à des butées articulaires est également utilisé pour la reconstruction et la prédiction de posture de bras manipulateurs dans [13], qui reprend les travaux d’Abdel-Malek sur le calcul d’espace de travail de bras articulés [12]. Enfin, un modèle du bras humain basé sur les paramètres de Denavit-Hartenberg (utilisé en robotique) est exploité dans [15] pour piloter un bras robot anthropomorphe. L’auteur donne également les valeurs des butées pour chaque articulation. Nous reprenons ces résultats pour mener à bien le filtrage des solutions fournies par l’algorithme de Vision (Cf. section 3).

## 1.4 Définition du problème

Nous proposons pour notre part un algorithme simple de Vision par ordinateur qui nous permet de définir un ensemble des configurations possibles (au maximum quatre) pour le bras étudié. Cet algorithme est basé sur l’utilisation d’une seule caméra calibrée, associée à des informations a priori sur l’espace atteignable par les différentes articulations. Un modèle biomécanique, inspiré de ceux utilisés en robotique pour les bras manipulateurs, est utilisé ensuite pour réduire encore cet espace. Ce modèle nous donne également accès à une représentation de la posture et, dans le cas de séquences d’images, de la trajectoire de chaque ar-

<sup>1</sup><http://www.sign-lang.uni-hamburg.de/Projects/HamNoSys.html>

<sup>2</sup><http://www.signwriting.org/>

<sup>3</sup><http://www.sign-lang.uni-hamburg.de/Projects/syncWRITER.html>

<sup>4</sup><http://www.bu.edu/asllrp/SignStream/>

tication, à la fois dans l'espace objet cartésien et dans l'espace des configurations angulaires de chaque articulation. Les aspects géométriques de cette démarche ont été, dans un premier temps évalués sur une scène de synthèse, en utilisant les images générées par une caméra virtuelle. Nous utilisons ensuite le modèle biomécanique pour vérifier la validité des solutions trouvées de façon géométrique. De nombreux problèmes de traitement d'images restent bien évidemment ouverts à ce stade de l'étude (Cf.2.1), mais cette étape de validation du principe est toutefois un préalable nécessaire à toute mise en œuvre sur des images réelles. Elle nous permettra en outre, au travers de l'étude de la sensibilité de l'algorithme au bruit (Cf. 2.5), de définir les tolérances supportables dans l'estimation des paramètres de position des articulations pour le traitement des images réelles.

## 2 Vers une solution géométrique

Nous décrivons dans cette section comment reconstruire, à partir d'une image d'un locuteur faisant des gestes devant une caméra calibrée, la posture 3D de l'un de ses bras. Nous appelons posture la configuration (fixe) du bras à un instant donné, cette configuration étant définie, soit par les valeurs angulaires des articulations (épaule, coude) du bras, soit par les positions des extrémités de chaque segment (épaule, coude, poignet) dans l'espace 3D. Nous montrons comment nous pouvons ramener l'espace des postures possibles à quatre configurations seulement à partir d'un modèle de la scène et du dispositif d'acquisition, ceci moyennant certaines hypothèses. Nous démontrons la validité de cette solution sur des images de synthèse.

### 2.1 Pré requis

Notre technique repose sur un certain nombre d'hypothèses et de connaissances préalables, parfois contraignantes dans le cadre d'une exploitation en environnement non contrôlé. Elles peuvent toutefois être en partie levées si nous nous limitons à une estimation à un facteur d'échelle près. Il faut de plus noter qu'une telle reconstruction reste tout à fait utilisable dans la majorité des applications d'analyse des gestes, notamment en ce qui concerne la résolution de problèmes tels que l'estimation de la position relative.

**Dispositif d'acquisition :** Le dispositif d'acquisition est constitué d'une seule caméra, calibrée de façon à pouvoir calculer l'équation d'un rayon de projection passant par un pixel donné. Ceci suppose la connaissance de la matrice de projection perspective  $\mathbf{C}$ . De nombreuses techniques de calibrage ont été proposées dans la littérature, par exemple par O. Faugeras dans [9], Z. Zhang dans [22] ou P. Gurdjos [16].

**Suivi des articulations :** Nous faisons également l'hypothèse que nous sommes capables d'identifier

dans une image les positions des trois articulations du bras. En l'absence de marqueurs, le problème peut s'avérer d'autant plus difficile à résoudre que les points recherchés correspondent en réalité aux axes de rotation de chaque articulation. Leur position devra alors probablement être déduite du résultat de la recherche de primitives plus facilement détectables dans l'images, par exemple les différents segments du bras. Enfin le problème des occultations (bras tendu vers l'avant par exemple) devra également être pris en compte. De façon générale, le choix d'un algorithme de détection de ces points devra donc être guidé par les contraintes d'exécution de l'application et par la précision des résultats souhaités.

**Posture du bras :** Nous ne traitons ici que de la reconstruction de la posture d'un bras, sans tenir compte de la main, soit si ces coordonnées sont exprimées dans le repère caméra :

- pour l'épaule :  $P_1 = (X_1, Y_1, Z_1)^T$
- pour le coude :  $P_2 = (X_2, Y_2, Z_2)^T$
- pour le poignet :  $P_3 = (X_3, Y_3, Z_3)^T$

Cette représentation permet de ramener le problème de l'estimation de la posture à celui de l'estimation de la position de trois points dans l'espace.

### 2.2 Modèle géométrique du bras

Toujours dans le cadre de la résolution géométrique, nous modélisons le bras par le système articulaire reliant l'épaule au poignet. Ce système est constitué d'articulations (liaison pivot pour l'épaule et rotule pour le coude) reliant des segments rigides (bras et avant bras) notés  $l_i$ , le segment  $l_i$  reliant les articulations  $P_i$  et  $P_{i+1}$ . La position de l'organe terminal correspond à celle du poignet, c'est-à-dire à  $P_3$ , extrémité du segment  $l_2$ . Ces articulations ne permettant que des rotations pures du segment  $l_i$  autour de l'articulation  $P_i$ , nous pouvons définir l'ensemble des positions atteignables par l'articulation  $P_j$  ( $j = 2, 3$ ) comme étant une sphère, sphère centrée sur l'articulation précédente  $P_{j-1}$  et dont le rayon est  $\|l_i\|^5$  (cf. figure 1).

Grâce à cette modélisation, l'espace de recherche pour chaque articulation se trouve ramené à une sphère dont les paramètres sont connus dès lors que l'on a déterminé la position de l'articulation précédente et la longueur de chaque segment du bras. Ce qui nécessite donc de connaître dès le début la position 3D de l'épaule. Il faut toutefois noter que, dans la mesure où nous utilisons un dispositif d'acquisition calibré, le fait de connaître la position de l'épaule dans l'image et une de ses coordonnées 3D, permet de calculer de façon simple les coordonnées manquantes. De la même

<sup>5</sup> $\|l_i\|$  est la norme du segment  $l_i$

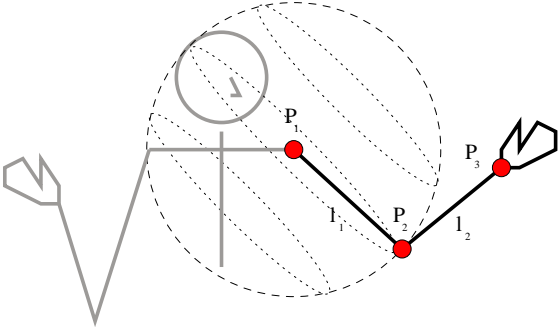


Figure 1: Modèle du système articulaire du bras. La sphère représente l'ensemble des positions possibles pour le coude.

façon, la position de l'épaule pourra être initialisée à un facteur d'échelle près si l'on fixe à l'avance une de ses coordonnées (par exemple la distance à la caméra)

### 2.3 Algorithme

La méthode que nous proposons exploite un modèle géométrique simple de la scène et spécialement de la structure - le bras - dont nous souhaitons déterminer les paramètres (Cf. 2.2). Nous supposons connues les coordonnées, dans l'image, des points correspondants aux articulations, notées en coordonnées homogènes :

- pour l'épaule :  $\tilde{\mathbf{p}}_1 = (u_1, v_1, 1)^T$
- pour le coude :  $\tilde{\mathbf{p}}_2 = (u_2, v_2, 1)^T$
- pour le poignet :  $\tilde{\mathbf{p}}_3 = (u_3, v_3, 1)^T$

Grâce au calibrage de la caméra, nous pouvons alors associer à chaque point image un rayon de projection, c'est-à-dire la droite (passant par le centre optique et le point image considéré) sur laquelle se trouve l'homologue 3D de ce point.

**Espace des configurations possibles pour le coude :** Étant donné  $P_1$  la position dans l'espace de l'épaule (éventuellement arbitraire), nous pouvons définir l'espace des positions possibles pour le coude comme étant la sphère  $S_1$  centrée sur l'épaule et dont le rayon est la longueur  $\|l_1\|$  du bras. L'équation cartésienne de cette sphère est la suivante :

$$(X_1 - x)^2 + (Y_1 - y)^2 + (Z_1 - z)^2 - \|l_1\|^2 = 0 \quad (1)$$

**Équation du rayon de projection :** Soit  $\tilde{\mathbf{p}}_1$  la position de l'épaule dans l'image, exprimée en coordonnées homogènes. Le calibrage de la caméra nous donne la matrice de projection perspective  $C$  définissant le passage du repère 3D de la caméra. L'inverse de cette matrice nous donne pour un point de l'image  $\tilde{\mathbf{p}}_i$  l'équation du rayon de projection associé. Le point 3D recherché se trouve nécessairement sur cette droite :

$$R_i(\lambda) = \lambda \tilde{\mathbf{p}}_i \quad (2)$$

Il s'agit d'une équation paramétrique où  $\lambda$  est un simple coefficient multiplicateur et où  $\tilde{\mathbf{p}}_i$  représente les coordonnées du point image considéré exprimées dans le repère caméra par :

$$\tilde{\mathbf{p}}_i = C^{-1} \tilde{\mathbf{p}}_i \text{ avec } C^{-1} = \begin{bmatrix} \frac{1}{fk_u} & 0 & \frac{u_0}{fk_u} \\ 0 & \frac{1}{fk_v} & \frac{v_0}{fk_v} \\ 0 & 0 & 1 \end{bmatrix} \quad (3)$$

$$\text{Soit : } \tilde{\mathbf{p}}_i = \begin{bmatrix} \frac{(u_i - u_0)}{fk_u} \\ \frac{(v_i - v_0)}{fk_v} \\ 1 \end{bmatrix} \quad (4)$$

La position 3D recherchée se trouve alors à l'intersection de la surface de la sphère  $S_1$  définissant l'ensemble des configurations possibles pour le coude, et du rayon de projection  $R_i(\lambda)$ . Le calcul de ces intersections dans le repère caméra revient donc à déterminer des valeurs de  $\lambda$  telles que :

$$(X_1 - \lambda \frac{(u_i - u_0)}{fk_u})^2 + (Y_1 - \lambda \frac{(v_i - v_0)}{fk_v})^2 + (Z_1 - \lambda)^2 - \|l_1\|^2 = 0 \quad (5)$$

Cela revient à calculer les solutions du polynôme du second degré  $a\lambda^2 + b\lambda + c = 0$  dont les coefficients sont :

$$\begin{aligned} a &= \left(\frac{(u_i - u_0)}{fk_u}\right)^2 + \left(\frac{(v_i - v_0)}{fk_v}\right)^2 + 1 \\ b &= 2\left[\left(\frac{(u_i - u_0)}{fk_u}\right)(-X_1) + \left(\frac{(v_i - v_0)}{fk_v}\right)(-Y_1) - Z_1\right] \\ c &= X_1^2 + Y_1^2 + Z_1^2 - l_1^2 \end{aligned} \quad (6)$$

La résolution de ce polynôme nous donne donc deux valeurs possibles pour  $\lambda$  (éventuellement une seule), dont nous pouvons aisément déduire des positions  $\hat{P}_{2,j}$  ( $j = 1, 2$ ) possibles pour le coude puisque  $R_i(\lambda) = \lambda \tilde{\mathbf{p}}_i$ .

Nous pouvons ensuite calculer les positions  $\hat{P}_{3,j}$  ( $j = 1, \dots, 4$ ) du poignet, dans le repère caméra. Nous considérons pour cela les deux sphères dont le centre est donné par chaque position estimée du coude et le rayon par la longueur de l'avant-bras. Suivant le même principe, nous pouvons alors calculer pour chaque valeur de la position du coude deux positions possibles pour le poignet, soit en tout quatre configurations possibles pour la posture du bras. Cet algorithme permet donc de ramener l'espace des configurations possibles pour un bras à quatre possibilités. L'étude de séquences d'images exploitant des hypothèses sur la continuité du mouvement (Cf. 2.4) et l'apport du modèle biomécanique (Cf. 3.1), nous permettent de déterminer parmi ces quatre postures, la plus probable.

### 2.4 Analyse de séquences d'images

Dans le cas d'une séquence d'images, nous calculons à partir de chaque image l'ensemble des points 3D candidats. Nous constituons ensuite les trajectoires. Il y a à chaque étape (sauf configuration particulière) deux points à répartir sur deux trajectoires. Toutefois, dans la mesure où il n'est pas possible de savoir

a priori quel point rajouter à quelle trajectoire, nous devons faire appel à un critère de linéarité afin de les reconstruire. Nous calculons pour cela l'angle  $\alpha$  entre les vecteurs  $\vec{V}_{i,j,k}$  et  $\vec{V}_{i,j,k+1}$ , où  $\vec{V}_{i,j,k}$  est défini par les points  $\hat{P}_{i,j,k-1}$  et  $\hat{P}_{i,j,k}$ , et  $\vec{V}_{i,j,k+1}$  par  $\hat{P}_{i,j,k}$  et  $\hat{P}_{i,j,k+1}$ .  $\hat{P}_{i,j,k}$  est la  $j^{eme}$  ( $j = 1, 2$ ) estimée des coordonnées spatiales de l'articulation  $i$  à l'étape  $k$  (figure 2).

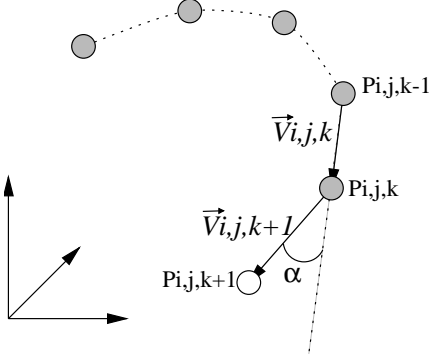


Figure 2: Calcul du produit vectoriel entre le dernier vecteur directeur de la trajectoire considérée et un nouveau vecteur directeur construit à partir du point candidat à affecter (en blanc). Le critère de linéarité des trajectoires utilisé consiste à affecter chaque point candidat à la trajectoire adéquate de façon à minimiser ces produits vectoriels.

Dans la pratique, cela revient à calculer la norme du produit vectoriel  $\|\vec{V}_{i,j,k} \wedge \vec{V}_{i,j,k+1}\| = \|\vec{V}_{i,j,k}\| \|\vec{V}_{i,j,k+1}\| \sin(\alpha)$ . Le point candidat pour lequel  $\sin(\alpha)$  est le plus faible est alors affecté à la trajectoire pour laquelle cette valeur minimale a été calculée. Le second point de la branche est alors affecté à la trajectoire restante. Enfin, dans le cas où le bras reste immobile d'une image à l'autre,  $\sin(\alpha)$  est considéré comme nul.

**Configurations particulières :** La construction des trajectoires décrite ci-dessus s'effectue correctement dans le cas général où l'algorithme donne deux intersections du rayon de projection avec chaque sphère. Il existe cependant des configurations où cela ne sera pas le cas. Ces configurations doivent être prises en compte dans l'algorithme ; elles peuvent également servir de détecteur pour des mouvements particuliers.

Il existe deux catégories de configurations particulières :

1. Le polynôme (6) n'a qu'une seule solution. C'est le cas où le segment considéré (bras ou avant-bras) se situe dans un plan parallèle au plan image. Dans ce cas, le rayon de projection arrive de façon tangentielle à la sphère. Il n'y aura donc qu'une seule "intersection". Ce point est alors concaténé aux deux trajectoires considérées, dans la mesure où cela correspond effectivement à un cas où ces deux trajectoires se croiseront.

2. Le polynôme (6) n'a pas de solution. En l'absence de bruit, ce cas ne peut se produire que pour le poignet. En effet, après calcul des deux positions possibles pour le coude, nous définissons les deux sphères qui forment l'ensemble des positions possibles du poignet. Il peut alors se produire des cas où, lorsque l'on se base sur la "mauvaise" position du coude, la sphère ainsi définie n'a pas d'intersection avec le rayon de projection. De telles configurations permettent directement d'éliminer une branche complète de l'arbre des solutions.

## 2.5 Prise en compte de données bruitées

Lors de la mise en œuvre de cet algorithme sur des images réelles, la principale source de bruit proviendra de la détection des articulations dans l'image. Nous cherchons donc à étudier de quelle façon un bruit  $\epsilon_i$  sur la position de chaque point  $p_i$  se répercute sur l'ensemble des positions 3D calculées. De façon intuitive, une erreur sur la position du point  $\tilde{p}_i$  entraînera une déviation du rayon de projection, et donc un décalage de l'intersection avec la sphère d'autant plus important que l'angle incident est éloigné de la normale à la surface. Dans les cas extrêmes, c'est-à-dire lorsque le rayon de projection devient tangent à la sphère, le rayon peut même "manquer" éventuellement la sphère. Nous proposons dans cette section deux approches différentes permettant d'appréhender ce problème :

1. Propagation du bruit : dans un premier temps, nous étudions dans quelle mesure les erreurs sur la détection de l'articulation  $\tilde{p}_i$  influenceront le calcul de la position de l'articulation  $P_{i+1}$ .
2. Définition d'un ensemble de solutions valides : certaines tâches d'interprétation ne nécessitent qu'une estimation approximative de la position de chaque articulation dans l'espace. Nous étendons alors le raisonnement précédent (Cf. 2.3) de façon à prendre en compte directement la présence de bruit sur la détection des points. La résolution du problème posé de cette façon nous permet de définir le volume décrit par l'ensemble des positions possibles de chaque articulation, ou encore de dire de façon simple si un point de l'espace donné appartient à cet ensemble.

**Propagation du bruit :** Afin d'étudier l'effet de l'accumulation du bruit lors du calcul des positions successives de chaque articulation, nous supposons pour cela la position 3D de l'épaule connue avec précision. Nous faisons ensuite l'hypothèse que la position  $\tilde{p}_i$  de chaque point est entaché d'un bruit  $\epsilon_i$  uniforme dans toutes les directions et de faible amplitude par rapport à la taille de l'image, soit  $\tilde{p}_i^* = \tilde{p}_i + \epsilon_i$  la

position bruitée du point  $\tilde{p}_i$ . Si nous reprenons les équations données en section 2.3, nous pouvons poser l'équation du rayon de projection :

$$R(\lambda) = \lambda \tilde{\mathbf{p}}_i^* \quad (7)$$

L'équation (5) définissant a priori l'ensemble des configurations possibles du coude reste inchangée. Nous pouvons alors récrire les coefficients du polynôme (6) donnant les valeurs possibles de  $\lambda$  :

$$\begin{aligned} a &= \left(\frac{(u_i + \epsilon_i) - u_0}{fk_u}\right)^2 + \left(\frac{(v_i + \epsilon_i) - v_0}{fk_v}\right)^2 + 1 \\ b &= 2\left[\left(\frac{(u_i + \epsilon_i) - u_0}{fk_u}\right)(-X_1) + \left(\frac{(v_i + \epsilon_i) - v_0}{fk_v}\right)(-Y_1) - Z_1\right] \\ c &= X_1^2 + Y_1^2 + Z_1^2 - l_1^2 \end{aligned} \quad (8)$$

En développant à partir de (8) l'expression du discriminant, nous obtenons :

$$\begin{aligned} \Delta &= 8\left[\left(\frac{(u_i + \epsilon_i) - u_0}{fk_u}\right)\left(\frac{(v_i + \epsilon_i) - v_0}{fk_v}\right)X_1Y_1 \right. \\ &\quad \left. + \left(\frac{(u_i + \epsilon_i) - u_0}{fk_u}\right)X_1Z_1 + \left(\frac{(v_i + \epsilon_i) - v_0}{fk_v}\right)Y_1Z_1\right] \\ &\quad - 4\left[\left(\frac{(u_i + \epsilon_i) - u_0}{fk_u}\right)^2(Y_1^2 + Z_1^2 - l_1^2) \right. \\ &\quad \left. + \left(\frac{(v_i + \epsilon_i) - v_0}{fk_v}\right)^2(X_1^2 + Z_1^2 - l_1^2) \right. \\ &\quad \left. + (X_1^2 + Y_1^2 - l_1^2)\right] \end{aligned} \quad (9)$$

Dans le cas général, cette expression confirme les résultats expérimentaux (figure 3) : les solutions générales  $\lambda = \frac{-b \pm \sqrt{\Delta}}{2a}$  varient suffisamment peu en fonction du bruit pour que les erreurs commises lors de la détection du coude ne rendent pas impossible l'estimation de la position du poignet. D'autre part, lorsque  $\Delta$  devient nul, la solution double s'écrit :

$$\lambda_{1,2} = \frac{\left(\frac{(u_i + \epsilon_i) - u_0}{fk_u}\right)X_1 + \left(\frac{(v_i + \epsilon_i) - v_0}{fk_v}\right)Y_1 + Z_1}{\left(\frac{(u_i + \epsilon_i) - u_0}{fk_u}\right)^2 + \left(\frac{(v_i + \epsilon_i) - v_0}{fk_v}\right)^2 + 1} \quad (10)$$

À ce moment là, la présence dans cette expression d'un terme constant additif au dénominateur nous garantit que les valeurs trouvées pour  $\lambda$  resteront bornées. D'autre part, les cas où  $\Delta$  devient nul correspondant à des configuration particulières (Cf. 2.4), ces dernières peuvent être détectées à l'avance et les postures correspondantes déduites de celles calculées précédemment. Enfin, des cas où le polynôme n'a pas de solution peuvent se produire ne présence de bruit, il s'agit de cas où le rayon de projection est dévié au point de "manquer" la sphère. Comme par construction, il ne s'agit que de rayons initialement proches de la tangente à la sphere, une approximation valable consiste à se ramener au cas où  $\Delta = 0$ .

**Définition d'un ensemble de solutions valides**  
Nous proposons maintenant une modélisation géométrique prenant directement en compte le bruit sur les positions des articulations dans l'image. Grâce à cette modélisation, nous allons pouvoir étudier

directement les volumes définissant l'ensemble des positions possibles de chaque articulation. Cette modélisation ne nous permet toutefois pas à ce stade de l'étude de définir analytiquement les volumes et surfaces impliqués ; elle pose cependant les premières bases d'un raisonnement permettant de les calculer de façon itérative ou de définir un prédicat permettant de déterminer si un point de l'espace appartient ou pas à l'ensemble des positions possibles.

Si nous supposons que nous connaissons la valeur maximale de l'erreur dont est entachée la position détectée pour chaque  $\tilde{p}_i$ , le point à détecter dans l'image deviendra un disque de rayon  $\|\epsilon_i\|$  qui contient nécessairement le point réel. Le rayon de projection associé devient un cône de sommet 0 et dont l'angle au sommet est donné par  $\arctan\left(\frac{\|\epsilon_i\|}{f}\right)$ . Pour le calcul de la position du coude, la frontière de chaque intersection (entrante et sortante) du cône avec la sphère peut être définie de façon analytique ce qui n'est par contre pas le cas de l'ensemble a priori des positions possibles du poignet. Cet ensemble est décrit par l'ensemble des sphères de rayon  $\|l_2\|$  dont le centre se trouve sur une des deux portions de surface définissant la position du coude. Celles du poignet seront alors données par les intersections du second cône de projection avec ce volume (figure. 3).

Ce raisonnement nous assure de l'existence de ces ensembles de positions possibles même si le bras approche d'une configuration singulière (Cf. section 2.4). En effet, même si le rayon lancé à partir du point détecté n'intersecte pas la sphère, il y aura forcément intersection entre le cône de projection et le volume définissant l'ensemble des configurations possibles dans la mesure où le disque considéré dans l'image contient bien le point réel.

### 3 Filtrage des solutions

Nous disposons d'un arbre binaire représentant les 4 configurations possibles pour le bras. Ces configurations sont données dans l'espace objet  $3D$ . Toutefois, lors du calcul de ces configurations, nous ne prenons pas en compte les possibilités réelles de mouvement d'un bras humain. En effet, l'espace de travail défini précédemment par des sphères est en réalité réduit à des portions seulement de ces sphères par les butées présentes sur les articulations. Nous aurons donc, parmi les configurations candidates définies géométriquement, un certain nombre de configurations aberrantes, au sens des valeurs angulaires sur les articulations. D'autre part, en plus des positions aberrantes, nous aurons également un certain nombre de configurations plausibles mais peu probables du fait de leur inconfort. Cette notion est appelée "zone de travail adroite" par A. Malek dans [12]. Afin d'exploiter ces informations, nous complétons notre modélisation par un second modèle du bras, similaire à ceux utilisés

$i$	$\theta_i$	$d_i$	$\alpha_i$	$a_i$
1	$\theta_1$	0	$-\pi/2$	0
2	$\theta_2$	0	$\pi/2$	0
3	$\theta_3$	$-l_1$	$-\pi/2$	0
4	$\theta_4$	0	$\pi/2$	0
5	0	$-l_2$	0	0

Table 1: Paramètres de DH définissant le modèle du bras utilisé

en robotique.

### 3.1 Modèle biomécanique

Le modèle paramétrique du bras humain proposé est basé sur les paramètres de Denavit-Hartenberg (DH) modifiés [11]. Ce modèle décrit en fait le système par la matrice de transformation homogène permettant de passer du repère objet au repère associé à l'organe terminal - dans notre cas le poignet. Cette matrice est paramétrée par les valeurs angulaires de chaque articulation et les dimensions des segments (bras et avant-bras les reliant). Il s'agit là du modèle géométrique direct. Réciproquement, le modèle géométrique inverse permet d'associer aux coordonnées cartésiennes des articulations les valeurs angulaires correspondantes.

#### Paramètres de Denavit-Hartenberg modifiés :

Dans [11] les auteurs définissent un algorithme permettant de placer de façon judicieuse les repères  $R_i$  associés à chaque articulation  $i$  (figure 4). Les règles utilisées pour positionner ces repères sont les suivantes :

1. Les axes  $x_{i-1}, x_i$  sont les perpendiculaire communes aux axes  $z_{i-1}$  et  $z_i$ .
2. Les axes  $y_i$  sont tels que:  $y_i = z_i \wedge x_i$
3. Les paramètres définissant la configuration relative du repère  $R_i$  par rapport à  $R_{i-1}$  sont :
  - $\alpha_i$  : angle entre les axes  $z_{i-1}$  et  $z_i$  en tournant autour de  $x_{i-1}$ .
  - $d_i$  : la distance entre  $z_{i-1}$  et  $z_i$ , mesurée selon  $x_{i-1}$ .
  - $\theta_i$  : l'angle entre les axes  $x_{i-1}$  et  $x_i$  en tournant autour de  $z_i$ .
  - $a_i$  : la distance entre  $x_{i-1}$  et  $x_i$ , mesurée selon  $z_i$ .

Les repères 0 et 1 étant confondus, le modèle du bras se décrit alors par la table (1) des paramètres de DH. L'utilisation des paramètres de DH permet d'écrire la matrice de transformation homogène  ${}^{i-1}T_i$ , qui décrit le passage du repère  $R_{i-1}$  au repère  $R_i$ :

$i$	$\theta_i$	$d_i$	$\alpha_i$	$a_i$
1	$\theta_1$	0	$-\pi/2$	0
2	$\theta_2$	0	$\pi/2$	0
3	0	$-l_1$	0	0

Table 2: Paramètres de DH pour le système épaule-coude

$${}^{i-1}T_i = \begin{bmatrix} \cos \theta_i & -\cos \alpha_i \sin \theta_i & \sin \alpha_i \sin \theta_i & a \cos \theta_i \\ \sin \theta_i & \cos \alpha_i \cos \theta_i & -\sin \alpha_i \cos \theta_i & a \sin \theta_i \\ 0 & \sin \alpha_i & \cos \alpha_i & d_i \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad (11)$$

où les  $\theta_i, \alpha_i, d_i, a_i$  sont les paramètres de DH.

**Modèle géométrique direct :** Le modèle géométrique direct nous permet de passer des valeurs angulaires des articulations du bras à leurs coordonnées cartésiennes. La transformation homogène  ${}^0T_5$  permettant de passer du repère 0 de l'épaule au repère 5 du poignet est obtenue en multipliant successivement les matrices  ${}^{i-1}T_i, i = 1, \dots, 5$ . La matrice  ${}^0T_5$  (14) résultante est donnée en annexe. Ce modèle est paramétré par les valeurs  $\theta_i$  des articulations ; il définit l'espace de travail du bras. Pour effectuer le filtrage en fonction des butées sur les valeurs angulaires de chaque articulation, nous avons besoin d'un modèle permettant de passer d'une position 3D donnée de l'organe terminal à l'ensemble de valeurs angulaires correspondant.

**Modèle géométrique inverse :** Le modèle géométrique inverse permet, à partir des coordonnées cartésiennes de position de l'organe terminal (le poignet), de calculer les valeurs angulaires  $\theta_i$  correspondantes des articulations (figure 5). La première solution envisageable est de calculer l'inverse de  ${}^0T_5$  (14), mais la décomposition de la chaîne cinématique simplifie grandement les calculs. Nous calculons les valeurs angulaires de chaque articulation séparément. Nous définissons pour cela les transformations inverses permettant de passer, de la position cartésienne du poignet aux valeurs angulaires du coude  ${}^5T_3$ , et de la position du coude aux valeurs angulaires de l'épaule  ${}^3T_0$ .

En ne considérant que l'articulation de l'épaule, les paramètres de Denavit-Hartenberg sont donnés dans la table (2).

On peut alors déterminer les transformations homogènes permettant de passer de  $R_{i-1}$  à  $R_i$  pour  $i = 1, 2, 3$ , et de calculer la transformation suivante (cos  $\theta_{12}$  signifie  $\cos \theta_1 * \cos \theta_2$ ):

$${}^0T_3 = \begin{bmatrix} \cos \theta_{12} & -\sin \theta_1 & \cos \theta_{12} & -l_1 \cos \theta_1 \sin \theta_2 \\ \sin \theta_1 \cos \theta_2 & \cos \theta_1 & \sin \theta_{12} & -l_1 \sin \theta_{12} \\ -\sin \theta_2 & 0 & \cos \theta_2 & l_1 \cos \theta_2 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

Le modèle géométrique direct est donné par la quatrième colonne de  ${}^0T_3$ . Le modèle géométrique inverse

du système est alors:

$$\begin{cases} \theta_1 = \arccos(z_2 / -l_1) \\ \theta_2 = \arctan(y_2 / x_2) \end{cases} \quad (12)$$

de même pour le poignet;

$$\begin{cases} \theta_3 = \arccos(z_3 / -l_2) \\ \theta_4 = \arctan(y_3 / x_3) \end{cases} \quad (13)$$

Grâce à ce modèle, nous pouvons, pour une position calculée à l'aide de l'algorithme décrit précédemment (Cf. 2), calculer l'ensemble des valeurs angulaires correspondantes. Ces valeurs angulaires vont maintenant nous permettre de trier les trajectoires ou points candidats calculés géométriquement.

### 3.2 Exploitation du modèle

En utilisant le modèle géométrique inverse décrit dans la section précédente, il devient possible, connaissant les valeurs en butée de chaque articulation, de déterminer de façon instantanée les positions aberrantes parmi celles disponibles. En effet, en utilisant les butées articulaires d'un bras humain, il est possible de tester les valeurs des  $\theta_i$  et de les trier selon leurs limites. Ces valeurs sont données dans [15] :

$$40 < \theta_1 < 270, 40 < \theta_2 < 270, 0 < \theta_3 < 180, 0 < \theta_4 < 160$$

La figure (6) montre les domaines atteignables de chaque articulation après prise en compte des valeurs en butée des articulations.

D'autre part ce modèle est aisément différentiable. Le modèle cinématique inverse ainsi obtenu permet de relier les vitesses et les accélérations mesurées dans l'espace aux valeurs angulaires correspondantes. Ces informations peuvent être recoupées avec des résultats issus des neurosciences décrivant la cinématique d'un bras humain [1] afin de définir, parmi les trajectoires candidates, celle(s) dont le(s) mouvement(s) est (sont) les plus "naturels".

## 4 Conclusion, perspectives

Nous proposons une méthode simple de reconstruction de la posture d'un bras basée sur des techniques de Vision par ordinateur. Cette méthode doit nous permettre, lors de l'analyse de séquences d'images, de reconstruire un ensemble de trajectoires possibles pour le coude et le poignet. Dans le pire des cas, cet algorithme nous fournit deux trajectoires candidates pour le coude, et quatre pour le poignet. Ces nombres peuvent éventuellement être plus faibles si le bras passe par certaines configurations particulières. Nous disposons d'autre part d'un modèle biomécanique du bras qui permet d'associer à une configuration des articulations, donnée en coordonnées cartésiennes, l'ensemble des valeurs angulaires de chaque articulation correspondant. L'exploitation de ce modèle nous permet d'éliminer, parmi les trajectoires, celles contenant

des positions aberrantes, c'est-à-dire des positions impossibles au sens des valeurs prises par les articulations. Cette méthode a été validée sur une scène de synthèse dans lesquelles les coordonnées image des articulations sont accessibles directement. Sur des images réelles, ces positions devront être déterminées par le traitement d'images, seules les positions dans l'image initiale ( $t = 0$ ) pouvant être éventuellement données manuellement. Ces positions seront donc connues avec une certaine imprécision. Nous avons donc étudié l'impact du bruit sur notre algorithme et nous montrons comment ce dernier se comporte lorsque des erreurs s'accumulent lors de la détection des articulations. Nous indiquons également comment prendre en compte les cas particuliers où le bruit rend impossible l'estimation. A ce stade de l'étude, la mise en œuvre sur des images réelles n'a toutefois pas été réalisée. Cette dernière nécessite en effet une initialisation précise rendant difficilement exploitable la méthode en environnement non contrôlé. Les aspects futurs de ce travail consisteront donc à alléger ces hypothèses, notamment si l'on recherche des informations qualitatives uniquement. nous envisageons également une autre approche consistant à exploiter des résultats approximatifs fournis par la résolution géométrique pour "piloter" directement le modèle biomécanique puisque ce dernier contient implicitement les relations reliant les mouvements des différentes articulations.

## References

- [1] A. Berthoz. Le sens du mouvement. Odile Jacob, ISBN : 2738104576, Paris, 1997.
- [2] A. Braffort. *Reconnaissance et compréhension de geste, application à la langue des signes*. PhD thesis, Université Paris XI, UFR Sciences, LIMSI, juin 1996.
- [3] J. Malik C. Bregler. Video motion capture. In *SIGGRAPH 1998*, 1998.
- [4] P. Dalle C. Hudelot, B. Lenseigne. Apports d'un système d'analyse d'images pour l'étude de la langue des signes. In *Journées R-LSF 2001*, IRIT, UPS Toulouse III, novembre 2001.
- [5] D. Metaxas C. Vogler. Asl recognition based on a coupling between hmms and 3d motion analysis. In *Proceedings of the International Conference on Computer*, pages 363–369, janvier 1998.
- [6] C. Cuxac. La langue des signes française. les voies de l'iconocité. *Faits de Langue, Ophrys*, 2000.
- [7] L. Goncalves et al. Monocular tracking of the human arm in 3d. *Fifth International Conference on Computer Vision*, page 764, 2000.



- [8] R. Rosales et al. Estimating 3d body pose using uncalibrated cameras. In *IEEE Computer Vision and Pattern Recognition Conference (CVPR)*, Dec. 2001.
- [9] O. Faugeras. *Three-dimensional computer vision: a geometric viewpoint*. MIT Press, Cambridge, Massachusetts, États-Unis, 1993.
- [10] D.M. Gravila. The visual analysis of human movement: A survey. *Computer Vision and Image Understanding*, 73(1), janvier 1999.
- [11] R.S. Hartenberg J. Denavit. A kinematic notation for lower-pair mechanisms based on matrices. *Journal of Applied Mechanics*, pages 215–221, June 1955.
- [12] H. Yeh K. Abdel-Malek. Local dexterity analysis for open kinematic chains. *Mechanism and Machine Theory*, 35:131–154, 2000.
- [13] M. Jaber K. Abdel-Malek, W. Yu. Realistic posture prediction. (submitted) SAE Digital Human Modeling and Simulation.
- [14] S. Igi K. Imagawa, S. Lu. Color-based hands tracking system for sign language recognition. In *3rd IEEE International Conference on Automatic Face and Gesture Recognition*, pages 462–467. IEEE, 1998.
- [15] T. Huntsberger M. Soltys. Inverse kinematic analysis of articulated figure using the fussy self-organizing feature map, 1993.
- [16] R. Payrissat P. Gurdjos. Calibrage plan d’une caméra en mouvement à focale variable. In *RFIA 2002*, pages 859–868. AFRIF-AFIA, janvier 2002.
- [17] M. Turk R. Cutler. View-based interpretation of real-time optical flow for gesture recognition. *IEEE International Conference on Automatic Face and Gesture Recognition*, avril 1998.
- [18] E. Granum T. Moeslund. Pose estimation of human arm using kinematics constraints. In *12th Scandinavian conference on image analysis (SCIA’01)*, Bergen, Norway, June 2001.
- [19] A. Pentland T. Starner. Real-time american sign language recognition from video using hidden markov models. Technical Report TR-375, M.I.T Media Laboratory Perceptual Computing Section, 1995.
- [20] A. Pentland T. Starner, J. Weaver. Real-time american sign language recognition using desk and wearable computer based video. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 20(12):1371–1375, decembre 1998.
- [21] C.J. Taylor. Reconstruction of articulated objects from points correspondences in a single uncalibrated image. *Computer Vision and Image Understanding*, 80(3):349–363, Decembre 200.
- [22] Z. Zhang. A flexible new technique for camera calibration. Technical report, Microsoft corporation, Dec 1998 (updated, 1999).

## ANNEXES

La matrice de passage  ${}^0T_5$  du repère 0 au repère 5. ( $\cos \theta_{ij}$  signifie  $\cos \theta_i * \cos \theta_j$ ):

$$\begin{aligned}
 {}^0T_5 &= \begin{pmatrix} R_1 & R_2 & R_3 & T \\ 0 & 0 & 0 & 1 \end{pmatrix} \\
 R_1 &= \begin{pmatrix} \cos \theta_4 (\cos \theta_{123} - \sin \theta_{13}) - \sin \theta_{42} \cos \theta_1 \\ \cos \theta_4 (\cos \theta_{23} - \sin \theta_3 \cos \theta_2) - \sin \theta_{124} \\ -\cos \theta_{34} \sin \theta_2 - \cos \theta_2 \sin \theta_4 \\ -\sin \theta_3 \cos \theta_{12} - \sin \theta_1 \cos \theta_3 \end{pmatrix} \\
 R_2 &= \begin{pmatrix} -\sin \theta_3 \cos \theta_2 \sin \theta_1 + \cos \theta_3 \sin \theta_1 \\ \sin \theta_3 \sin \theta_2 \\ \sin \theta_4 (\cos \theta_{123} - \sin \theta_{13} + \cos \theta_{14} \sin \theta_2) \\ \sin \theta_4 (\cos \theta_{23} \sin \theta_1 - \sin \theta_3 \cos \theta_1) + \sin \theta_{12} \cos \theta_4 \\ -\sin \theta_{24} \cos \theta_3 + \cos \theta_{24} \end{pmatrix} \\
 R_3 &= \begin{pmatrix} -l_2 \sin \theta_4 (\cos \theta_{123} - \sin \theta_{13} + \cos \theta_{14} \sin \theta_2) - l_1 \cos \theta_1 \sin \theta_2 \\ -l_2 \sin \theta_4 (\cos \theta_{23} \sin \theta_1 - \sin \theta_3 \cos \theta_1 + \cos \theta_4 \sin \theta_{12}) - l_1 \sin \theta_{12} \\ -l_2 (-\sin \theta_{24} \cos \theta_3 + \cos \theta_{24}) - l_1 \cos \theta_2 \end{pmatrix}
 \end{aligned} \tag{14}$$

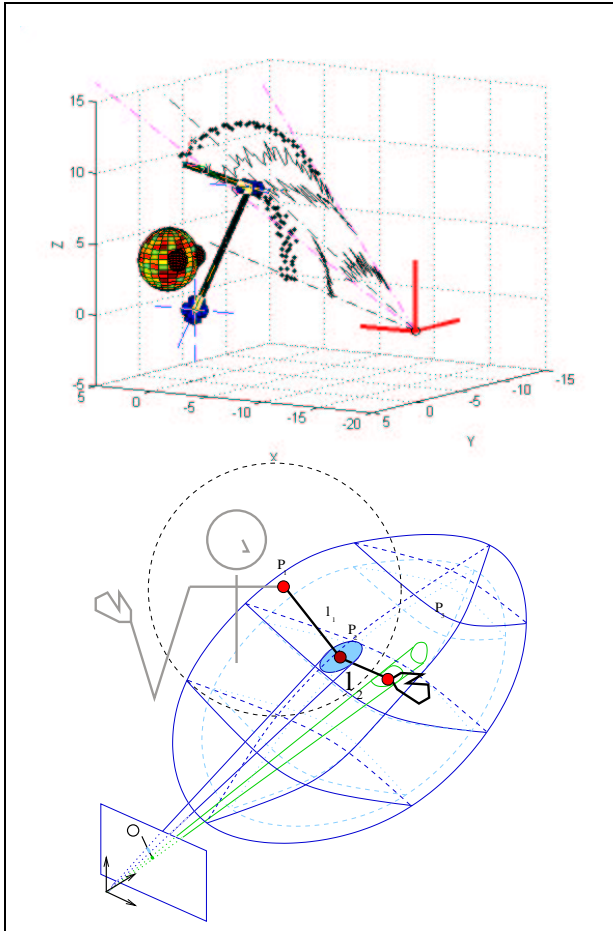


Figure 3: Deux aspects de l'étude du bruit : *en haut*, reconstruction des trajectoires dans une scène à partir d'images de synthèse. Du bruit a été rajouté sur les points image à hauteur de 5 % de la taille de l'image. Le filtrage des trajectoires a été effectué manuellement. Il est à noter que la position relative des points calculés reste inchangée. De tels résultats restent donc exploitables dans le cadre de mesures qualitatives sur les positions relatives des différents segments du bras. *En bas*, modélisation géométrique de la scène lorsque la détection de chaque point est entachée d'un bruit  $\epsilon_i$ . Nous ne montrons qu'un seul ensemble de solutions pour le coude, le volume définissant l'ensemble des configurations a priori du poignet et les deux ensembles de solutions pour le poignet.

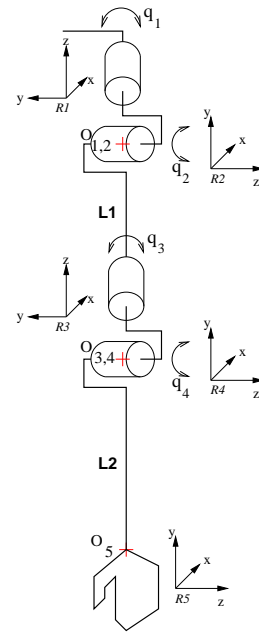


Figure 4: Modèle du bras avec les repères utilisés pour le calcul du modèle géométrique direct.

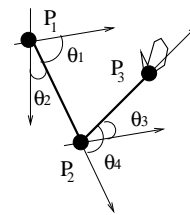


Figure 5: Le modèle inverse du bras nous donne les valeurs angulaires des articulations  $\theta_1, \theta_2, \theta_3, \theta_4$  en fonction des coordonnées spatiales de  $P_1, P_2, P_3$  (l'épaule, le coude, le poignet).

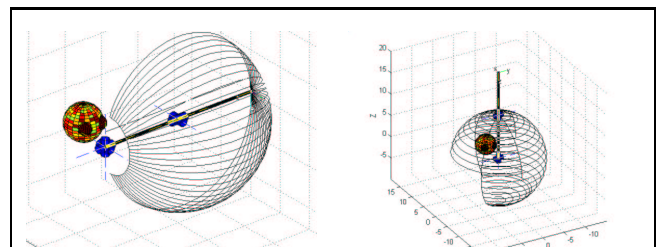


Figure 6: Domaine atteignable par les articulations du coude (à droite) et du poignet (à gauche). Dans le second cas, l'épaule reste fixe. Ces domaines prennent en compte les valeurs de butées sur les articulations.