

# Système d'Analyse de Mouvements Humains Acrobatiques

## A Human Acrobatic Motion Analysis System

R. Cassel<sup>1</sup>

C. Collet<sup>2</sup>

R. Gherbi<sup>1</sup>

<sup>1</sup> LIMSI-CNRS, Université de Paris XI  
BP133 91403 ORSAY Cedex France

<sup>2</sup> IRIT, Université Paul Sabatier  
118 route de Narbonne 31062 TOULOUSE Cedex France

{ryan.cassel, rachid.gherbi}@limsi.fr  
christophe.collet@irit.fr

### Résumé

*Dans le domaine de la vision par ordinateur, les techniques d'analyse du mouvement sont de plus en plus robustes et commencent à sortir des laboratoires. Les machines ayant leur puissance de calcul grandissante, les traitements temps réel (ou proche du temps réel) sont plus fréquents. Le domaine sportif est un secteur fortement demandeur en analyse du mouvement et plus particulièrement les sports mettant en action de l'acrobatie. L'acrobatie mêle des rotations complexes suivant des axes variés combinées avec des postures variées. Cet article présente un système d'analyse de gestes acrobatiques qui se base sur une caractérisation du mouvement initialement utilisée en compétition de trampoline. L'analyse est réalisée à partir de mesures globales extraites des séquences vidéo ou d'un flux vidéo directe. Les informations extraites des vidéos sont utilisées soit pour l'amélioration de la performance sportive, soit pour des premiers stades de reconnaissance de mouvements.*

### Mots Clef

Filtrage, suivi, analyse de mouvement, reconnaissance, geste sportif.

### Abstract

*In the field of computer vision, the motion analysis techniques are increasingly robust and beginning to have an impact beyond laboratories. Computers are able nowadays to perform image processing in real time. Indeed, they could be used to build non-intrusive systems for gesture analysis. In particular, gesture and motion analysis is in strong demand in the athletics field. This is especially true for sports dealing with acrobatics, as acrobatics mix complex spatial rotations over multiple axes and may be combined with various postures. Thus, any sport using acrobatics*

*requires a fine analysis. This paper presents a vision system focused on the analysis of acrobatic gestures from various sports. Instead of classical systems requiring modelizing human bodies, our system is based on the characterization of basic acrobatics movements. The system was successively tested first on trampoline sport movements, but also on other sports (gymnastics, diving, etc.). Within the system, the gestures analysis is mainly carried out by using global measurements, extracted from recorded movies or live video.*

### Keywords

Filtering, tracking, motion analysis, recognition, sport gesture.

## 1 Introduction

Dans le domaine de la vision par ordinateur, les techniques d'analyse du mouvement sont de plus en plus robustes et commencent à sortir des laboratoires. Les machines ayant leur puissance de calcul grandissante, les traitements temps réel (ou proche du temps réel) sont plus fréquents. Ils sont désormais exploités pour évaluer la performance des gestes.

Dans le domaine de l'analyse du geste, de nombreuses études se consacrent à la langue des signes. Cette tâche s'avère difficile tant pour le suivi et la reconnaissance de la main que pour l'analyse et la reconnaissance des mots des signeurs [3, 4]. Cette difficulté vient du fait que le modèle utilisé pour la reconnaissance n'est pas toujours adapté. Les phrases en langue des signes (LS) ne sont pas construites de la même façon que celles de la langue parlée. Le contexte sémantique ajoute de la complexité dans la tâche de reconnaissance [10].

Un autre axe de recherche concerne la reconnaissance de l'activité humaine [1]. Le domaine sportif est un sec-

teur fortement demandeur en analyse du mouvement et l'avènement de la vidéo offre de nouvelles possibilités. Cependant il n'existe que très peu d'outils offrant une analyse automatique et en temps réel. C'est en partie expliqué par le fait qu'on trouve peu d'articles sur les systèmes automatiques d'analyse du geste sportif. M. Yamamoto dans [5] présente une étude qualitative sur le mouvement des skieurs. Cette étude discrimine les mouvements d'un skieur novice de ceux d'un expert. Dans une autre étude, Gopal Pingali dans [9] présente un système pour suivre et analyser la trajectoire d'une balle de tennis en temps réel. Ce système a été utilisé pour l'US Open 2000 de tennis et montre ainsi l'intérêt de tels systèmes. Mais les concevoir est un vrai challenge. Le contexte sportif est extrêmement contraignant (du point de vue environnemental et gestuel).

Cet article présente un système d'analyse de mouvement pour le geste acrobatique. L'acrobatie génère des séquences extrêmement complexes (à réaliser et à analyser). Tout sport mettant en scène de l'acrobatie nécessite une fine analyse pour améliorer ou juger de la qualité du geste. Et cette analyse n'est pas triviale. Il est fréquent qu'un expert manque de reconnaître certains éléments importants tant les nuances sont fines. La mise en place de système comme celui présenté ici se veut être un complément pour l'entraînement et une aide pour les juges en compétition. Il extrait de nombreuses informations concernant l'acrobatie. Les données relatives aux déplacements, rotations et positions corporelles sont accessibles aussi bien pour leurs coordonnées que pour leur vitesse et leur accélération. Les rotations de type longitudinal (voir plus loin) ne sont pas encore extraites. Ce type de rotation est délicat à traiter. Le système a été spécifié suivant les conditions de pratique. De ce fait l'athlète ne sera jamais contraint par le système.

D'autres études qui travaillent sur l'analyse du geste, utilisent des capteurs sur le corps pour extraire de manière efficace ses coordonnées. De tels systèmes sont dits intrusifs, ils gênent le mouvement les rendant moins naturels. Les mouvements acrobatiques rendent l'utilisation de capteurs délicate. De plus, ces dispositifs sont coûteux, encombrants et contraignants. Ils sont cependant terriblement précis comparé aux techniques de traitement vidéo. Notre choix a été de préserver au maximum la naturalité du geste. Notre étude n'utilise aucun capteur pouvant interférer avec le sportif. Les mouvements sont captés par un système de vision monoculaire (caméra fixe au format PAL 720x576 en 25 images par secondes).

L'analyse du geste se base sur une caractérisation du mouvement initialement utilisée en compétition de trampoline. Ce modèle du mouvement a été modifié et est présenté brièvement en partie 2. L'architecture du système ainsi que ses algorithmes sont décrits en partie 3. La partie 4 évalue la robustesse du système et la partie 5 présente l'utilisation du système dans des situations réelles d'entraînement. Les perspectives et la conclusion sont développées dans les deux dernières parties.

## 2 Modèle du mouvement

La conception du système d'analyse s'appuie sur un modèle du mouvement établi et utilisé par les trampolinistes (pour plus de détails se référer au Code de Pointage officiel de la Fédération Internationale de Gymnastique [11]). Il se base essentiellement sur une décomposition chronologique et axiale du mouvement. La notation du geste est divisée en trois parties. La première partie concerne la trame temporelle des rotations transversales<sup>1</sup> du corps. La deuxième concerne les rotations longitudinales<sup>2</sup> du corps distribuées par la trame temporelle des rotations transversales (voir exemple plus loin). La dernière concerne la configuration spatiale du corps. Ce modèle du mouvement abouti à une notation alpha numérique décrite en détail dans [2].

Ainsi chaque saut en trampoline est identifié par un nom ou encore une valeur. Cette valeur est appelée notation numérique. C'est un code alphanumérique construit à l'aide d'expressions régulières de la façon suivante :

$$s \ q \ v_1 \dots v_n \ p$$

$$s = [b|f]$$

$$q = [0 - 9]^*$$

$$v_i = [0 - 9]^*$$

$$p = [o < | /]$$

où  $s$  renseigne sur le sens de la figure ( $f$ , en avant ou  $b$ , en arrière),  $q$  représente le nombre de quart de rotation du corps suivant l'axe de rotation transversal,  $v_1 \dots v_n$  représente la distribution des demi vrilles par rotation complète ( $360^\circ$  de rotation) entamée et enfin  $p$  représente la position du corps (groupée ou tuck =  $o$ , carpée ou pike =  $<$  ou tendue ou layout =  $/$  figure 1). Ainsi,  $b \ 8 \ 0 \ 0 \ o$  représente un double salto ( $720^\circ$  de rotation) en arrière groupé sans vrille alors que  $f \ 4 \ 1 \ /$  représente un salto ( $360^\circ$  de rotation) en avant tendu avec une demi vrille.

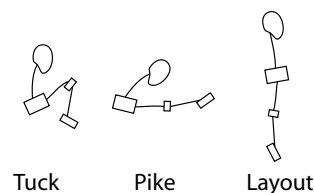


FIG. 1 – Positions du corps humain en trampoline.

<sup>1</sup>Les rotations transversales correspondent à la rotation du corps suivant son axe transversal. Axe passant par le centre de gravité et parallèle à la droite passant par la tête fémoral gauche et droite du gymnaste quand il se trouve en position debout, jambes serrées. Ces rotations sont souvent appelées rotations salto.

<sup>2</sup>Les rotations longitudinales correspondent à la rotation du corps suivant son axe longitudinal. Axe passant par le centre de gravité et parallèle à la droite passant par la tête et le bassin du gymnaste quand il se trouve en position debout, jambes serrées. Ces rotations sont souvent appelées rotations vrille.

Cette notation est utilisable dans toutes disciplines acrobatiques car elle permet d'identifier efficacement n'importe quel mouvement acrobatique. La reconnaissance de chaque partie de cette notation nous renseigne sur la qualité de réalisation. Elle est à la base de notre système d'analyse. La section 5 montre quelques exemples d'utilisation d'un tel système en entraînement de gymnastique (barre fixe et trampoline).

### 3 Architecture du système

Le système se décompose en plusieurs modules connectés de manière hiérarchique. Une première couche bas niveau permet d'extraire les informations pertinentes (typiquement, les pixels de l'acrobate). Une deuxième couche haut niveau transforme ces informations en données interprétables et les analyse. Nous présentons ici ces différentes couches.

#### 3.1 Couche bas niveau

La couche bas niveau extrait les pixels de l'acrobate. Une approche classique est d'utiliser un modèle statistique d'image de fond. Puis d'utiliser une différence d'image pour faire ressortir ce qui est en mouvement. Vient ensuite une étape de filtrage pour éliminer les données parasites.

**Modèle d'image de fond** Dans cet article, le terme d'image de fond fait référence aux pixels de l'image qui ne sont pas en mouvement. En effet, le système a comme connaissance a priori qu'il n'y a qu'une personne dans le champ de la caméra et cette dernière est fixe. L'athlète est toujours en mouvement tandis que le fond d'image est censé être constant. Toutefois, le fond d'image n'est jamais vraiment fixe dans des conditions d'entraînement ou en compétition. Il y a des variations de lumière, et il est fréquent que des personnes passent dans l'arrière plan. Pour s'adapter à ces conditions nous avons mis en œuvre une génération d'image de fond adaptative. Cette génération se base sur la moyenne de luminance des  $N$  dernières images.

La moyenne est calculée de la façon suivante :  $m_{(x,y)} = \frac{S_{(x,y)}}{N}$  où  $S_{(x,y)}$  est la somme de la valeur des pixels aux coordonnées  $(x, y)$  et  $N$  est le nombre des dernières images collectées.

La différence entre l'image courante et l'image de fond permet de mettre en évidence les parties de l'image en mouvement. Bien souvent il y a plus que les pixels du gymnaste qui sont en mouvement. Il est nécessaire de filtrer cette image.

**Filtrage par bloc** La mise en place d'un tel filtrage a pour but d'éliminer les zones en mouvements qui sont petites et isolées. Ce filtrage élimine un certain nombre de pixel provenant du bruit de l'image (nous incluons dans le bruit, tout pixel en mouvement qui n'appartient pas au gymnaste). En se basant sur la connaissance des conditions de tournage, on sait que la personne filmée est au premier plan. La différence d'image présentée précédemment,

aboutit à une image binaire dans laquelle on retrouve le sujet observé plus du bruit sous forme de pixels aléatoire. La personne en premier plan est la forme binaire la plus grosse. Le filtrage par bloc permet de garder certains éléments calibrés par la taille des blocs. Ici, le filtrage élimine les formes binaires les plus petites.

Le principe du filtrage par bloc est de sous échantillonner l'image et de pondérer chacun des pixels en tenant compte du poids des pixels voisins. Ainsi, certains pixels avec un poids élevé mais isolé seront éliminés alors que d'autres avec un poids plus faible mais entourés de pixels d'un poids plus fort seront conservés. Pour cela, l'image est divisée en bloc de taille  $(n \times m)$ . On effectue un parcours d'image bloc par bloc en les marquant comme valide, invalide ou voisin. Au terme du parcours de l'image on ne conserve que les blocs valides et voisins. Un bloc est valide s'il a une proportion de pixels binaires suffisant. Dans le cas contraire il est invalide. Un bloc est voisin si l'un de ses voisins est valide (en 8-connexité). En fixant la taille des blocs plus grands que le bruit (personnes en arrière plan, bruit de l'image) et en fixant une proportion de pixels binaire d'un bloc supérieur aux éléments du bruit, nous éliminons un grand nombre de pixels ne provenant pas du sujet analysé.

La figure 2 montre le résultat de l'algorithme. L'image du haut est l'image originale. L'image du milieu est le résultat de la binarisation par la technique présentée précédemment (cette image fait ressortir le mouvement de l'image). On constate que le public génère du mouvement. Dans des situations de compétition, le public est le plus gros problème. Cependant il reste plus petit que le gymnaste car celui-ci est en premier plan. On applique ensuite la technique de filtrage par bloc qui donne l'image du bas.

L'algorithme de filtrage par bloc élimine les éléments binaires de petite taille (en fonction de la taille des blocs et de leur pourcentage de validité). Il permet d'extraire une forme binaire qui coïncide avec les pixels de l'acrobate.

Cet algorithme est efficace quand le public est loin du premier plan et quand il n'est pas trop agité. En effet, un mouvement collectif du public de type "holà" perturbe fortement le filtrage. Dans la suite de l'étude, le corpus vidéo a été trié pour éviter de traiter ces cas particuliers.

**Filtre de Kalman** Pour être efficace en terme de temps de calcul, il est important de ne pas réaliser les traitements sur l'intégralité des images. Seul la forme binaire est pertinente. Il est nécessaire de suivre de cette forme et l'utilisation d'un filtre de Kalman [6] permet d'estimer et de prédire ses positions.

Le vecteur d'état à estimer est constitué de paramètres liés aux déplacements. Il contient les valeurs de position, vitesse et accélération du centre de gravité du gymnaste.

$$X_k = (x_k, \dot{x}_k, \ddot{x}_k, y_k, \dot{y}_k, \ddot{y}_k)$$

Nous en déduisons une matrice de transition  $A$  vérifiant l'équation de commande :



FIG. 2 – Filtrage par bloc.

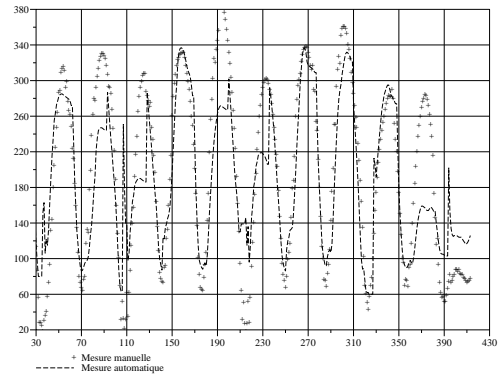


FIG. 3 – Suivi sans prédiction des positions verticales.

$$x_{k+\delta t} = Ax_k + w_k$$

avec

$$A = \begin{pmatrix} 1 & \delta t & \frac{\delta t^2}{2} & 0 & 0 & 0 \\ 0 & 1 & \delta t & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & \delta t & \frac{\delta t^2}{2} \\ 0 & 0 & 0 & 0 & 1 & \delta t \\ 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}$$

et  $w_k$  le bruit de processus.

Notre mesure est la position du centre de gravité du gymnaste. On mesure le vecteur  $z_k = (x_k, y_k)$  vérifiant l'équation de mesure :

$$z_k = Hx_k + v_k$$

avec

$$H = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \end{pmatrix}$$

et  $v_k$  le bruit de mesure.

Le filtre de Kalman donne une estimation d'un processus en utilisant un système de rétroaction : le filtre estime le processus à un temps  $k$  et obtient alors un retour sous la forme d'une mesure bruitée. En tant que tels, les équations du filtre de Kalman sont séparées en deux groupes : les équations de mise à jour temporelle et les équations de mises à jour par la mesure. Les équations de mises à jour temporelles sont assimilées dans cet article à des équations de prédiction et les équations de mises à jour par la mesure sont vues comme des équations de correction.

Dans cette étude, le filtre de Kalman permet de limiter la zone de traitement. Nous utilisons la prédiction pour estimer les positions futures de l'athlète. La position de la zone de traitement est estimée en fonction de la prédiction. Le fait de restreindre la zone de traitement à plusieurs effets. Tout d'abord, le temps de calcul est réduit car les

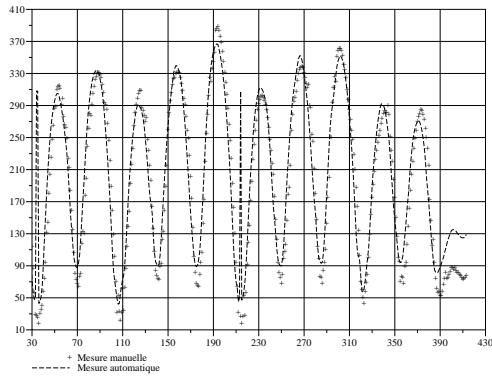


FIG. 4 – Suivi avec prédiction des positions verticales.

traitements ne sont pas fait sur l'intégralité des images. D'autre part, cela permet de faire abstraction des nouveaux éléments de l'image pouvant perturber l'analyse (le plus souvent, des personnes en image de fond traversant le champs de la caméra).

On peut utiliser le filtre de Kalman pour filtrer les positions et/ou pour les prédire. Ici la prédiction est essentielle. En effet, le geste sportif est très souvent rapide. Lorsque l'on effectue une mesure sur l'image, on localise les zones importantes. A l'image suivante, on n'a que l'information a posteriori pour effectuer le calcul. Or, on est certain que les zones d'intérêts se sont déplacées. Pour éviter de faire un prétraitement de pour localiser les zones d'information, nous utilisons la prédiction. Sans cette prédiction les performances chutent considérablement comme nous le montre les figures 3 et 4. Nous avons évalué notre système avec et sans prédiction par rapport au suivi des positions du centre de gravité. Nous avons comparé les positions du centre de gravité extraites à la main avec celles extraite par notre système. Les évaluations montrent un écart type de 41 pixel sans utiliser la prédiction contre 26 avec la prédiction. On constate sur la figure 3 que le système se trompe souvent quand il n'utilise pas la prédiction. Avec la prédiction, la figure 4 montre que le suivi est correcte.

En conclusion, la première étape consiste à mettre en évidence les zones en mouvements; puis à partir de la connaissance des conditions de tournage, filtrer ces données pour faire ressortir uniquement les pixels de l'acrobate. Un filtre de Kalman permet de suivre ces pixels. La zone des éléments à traiter est alors réduite au tour des pixels de l'acrobate. Une fois la zone d'intérêt fixée et suivie, le système fait abstraction des nouveaux éléments qui entre dans l'image. Le système est focalisé sur l'athlète qu'il est entrain de suivre, il peut alors entamer la phase d'analyse du mouvement.

### 3.2 Analyse du mouvement (couche haut niveau)

La couche haut niveau ne reçoit que de l'information pertinente. Plutôt que de chercher à reconstruire un modèle de corps humain, notre approche consiste à analyser le comportement de ces données par rapport au modèle de mouvement que nous avons défini. Nous devons déterminer des rotations et des changements de configuration corporelles 1. L'analyse du mouvement passe par le suivi de l'axe du corps et de ses variations. L'axe du corps est extrait par une méthode mathématique de décomposition en valeur propre. Les informations concernant les rotations sont extraites par l'étude des variations de l'axe du corps. Ces informations une fois segmentées permettent de déterminer les valeurs de la notation numérique [2] (les rotations salto et la position).

**Détermination des axes et de leur orientation** A partir de l'image binaire et en fonction de la zone d'intérêt, le système calcul l'axe principal de la forme binaire. Comme décrit dans [7] nous utilisons une décomposition en valeur propre pour déterminer l'orientation de l'athlète. L'aire de l'objet est définie par le moment d'ordre 0 : Find the zero<sup>th</sup> moment

$$A = \sum_x \sum_y I(x, y).$$

Le point  $(x, y)$  parcourt la zone d'intérêt et  $I(x, y)$  est la valeur du pixel dans l'image binaire (0 ou 1). Le centre de masse  $C_m = (\bar{x}, \bar{y})$  est définie par les moments d'ordre 1 :

$$M_x = \sum_x \sum_y xI(x, y),$$

$$M_y = \sum_x \sum_y yI(x, y).$$

$$\bar{x} = \frac{M_x}{A}; \quad \bar{y} = \frac{M_y}{A}.$$

L'orientation 2d de la forme binaire est déterminée par les moments d'ordre 2. Les moments d'ordre 2 sont :

$$M_{xx} = \sum_x \sum_y x^2 I(x, y),$$

$$M_{yy} = \sum_x \sum_y y^2 I(x, y),$$

$$M_{xy} = \sum_x \sum_y xy I(x, y).$$

On pose :

$$a = \frac{M_{xx}}{A} - \bar{x}^2, \quad b = 2 \left( \frac{M_{xy}}{A} - \bar{x}\bar{y} \right), \quad c = \frac{M_{yy}}{A} - \bar{y}^2.$$

L'orientation de l'objet est définie comme étant l'axe de minimum d'inertie. Cela correspond à l'axe du plus petit moment d'ordre 2. L'orientation est celle du plus grand axe :

$$\theta = \frac{\arctan\left(\frac{b}{a-c}\right)}{2}.$$

Les deux premières valeurs propres qui sont la longueur et la largeur sont calculées de la façon suivante :

$$l = \sqrt{\frac{(a+c) + \sqrt{b^2 + (a-c)^2}}{2}},$$

$$w = \sqrt{\frac{(a+c) - \sqrt{b^2 + (a-c)^2}}{2}}.$$

Ces calculs permettent de déterminer les axes principaux du corps ainsi que leur largeur comme le montre la figure 5.



FIG. 5 – Orientation du corps, longueur et largeur.

**Suivi des rotations** A partir du calcul des axes, nous déterminons une orientation du corps. Une étude sur les contraintes physiques a été réalisée sur les corpus vidéo acrobatique. Elle a permis de déterminer entre autre, les vitesses angulaires maximales engendrées par les rotations. La vitesse angulaire maximale étant de  $\omega_{max} = 22 \text{ rad.s}^{-1}$  (soit  $\omega_{max} = 0.89 \text{ rad/image}$  pour des vidéos à  $25 \text{ images/s}$ ), des variations d'ordre  $\pi \text{ rad/image}$  ou supérieures sont physiquement impossibles. On calcul la vitesse angulaire instantanée ainsi :  $\omega = \theta_{t-1} - \theta_t [\pi]$ . L'orientation correcte  $\theta'$  est calculée de la façon suivante :  $\theta'_t = \theta'_{t-1} + \omega$ . Cette orientation n'est pas absolue mais relative à l'orientation initiale.

**Détection des quarts** La reconnaissance des rotations n'est pas si évidente. Il faut comprendre que souvent, le gymnaste ne réalise pas 4 quarts de rotations. Pour un salto arrière par exemple, il peut le commencer extrêmement penché en arrière et le finir penché en avant. Le résultat

d'une mesure stricte donne bien moins que 4 quarts de rotation. Il n'y a souvent que 3 quarts réels de rotation. Pourtant, le système doit retenir les 4 quarts de rotation puisque l'acrobate part sur les pieds et arrive sur les pieds. Dans le cas d'un trampoliniste, le problème se complique quand il entre ou sort de la toile du trampoline sur le ventre ou sur le dos. Le corps disparaît en parti ne laissant apparaître que des segments verticaux. Or le corps est horizontal.

Notre solution est de calculer de trois façons différentes ces quarts de rotation. La première technique est de calculer des vitesses angulaires moyennes sur les portions de salto où cette vitesse est régulière. Une accélération angulaire nulle engendre une vitesse constante. Prenons le cas général des salti. Il y a toute d'abord une phase d'envole (appelée take off). L'athlète exerce une poussée complète sur le support sur lequel il se trouve jusqu'à atteindre une certaine hauteur. Dans cette phase, la vitesse de rotation augmente. La phase suivante consiste à accélérer la rotation pour enfin entrer dans une phase stable, là où la rotation est stabilisée. C'est dans cette phase que l'on atteint la vitesse maximale et constante. Vient ensuite la phase de réception (appelée landing) où la vitesse de rotation décroît afin de préparer la réception sur le support. En calculant la vitesse moyenne de la phase constante on arrive à avoir une bonne information du salto réalisé. En reportant la vitesse moyenne sur l'ensemble du salto, nous compensons les problèmes d'orientation du départ et de la fin du saut.

Une deuxième solution est de comparer l'angle de sortie avec l'angle d'arrivée. Ce calcul nous donne la mesure exacte de l'angle parcouru.

Enfin une dernière technique consiste à compter le nombre de cadran traversé par l'axe du corps. Les rotations sont représentées par un disque découpé en cadrans. Ce disque est divisé en quatre ( $[0; \frac{\pi}{2}[$ ,  $[\frac{\pi}{2}; \pi[$ ,  $[\pi; \frac{3\pi}{2}[$ ,  $[\frac{3\pi}{2}; 2\pi[$ , ). A chaque fois que l'axe entre dans un nouveau cadran il est comptabilisé (même s'il n'a pas été complètement traversé).

Ces trois techniques se compensent. La première technique est optimiste pour les salto rapides tandis que la deuxième est pessimiste. La troisième technique est correcte à  $\pm 1$  quart. En faisant la moyenne des trois techniques les quarts de salto sont correctement détectés (tableau 1).

**Evaluation des positions** Pour évaluer la position (configuration spatiale) du corps, le système utilise les longueurs des deux axes  $l$  et  $w$  et la surface du gymnaste. La longueur moyenne  $l_m$  des deux axes est calculée. La longueur  $l_m$  représente le diamètre du cercle  $C$  centré sur le centre de gravité de l'acrobate. Ainsi, plus l'acrobate est groupé, plus le rapport entre  $l$  et  $w$  est proche de 1, plus le rapport entre la surface de  $C$  et celle du gymnaste (inclus dans  $C$ ) est proche de 1. Cela signifie qu'une grande partie du corps se trouve inclus dans le cercle  $C$ . Inversement, plus le corps est tendu et plus le rapport entre la surface de  $C$  et celle du gymnaste est petit. Cette première méthode permet de discriminer correctement les positions tendues des positions groupées et carpées.

**Evaluation des vrilles** Les premières évaluations utilisant le flot optique pour la détection des vrilles ne sont pas concluantes. En effet, son utilisation brut dans l'image détecte bien du mouvement. Cependant l'information qui nous intéresse est bruitée par le mouvement du gymnaste et ses rotations salto. Le flot optique met en évidence les mouvements les plus rapides (les déplacements). Pour le moment, les vrilles ne sont pas encore détectées. De ce fait, la reconnaissance des éléments n'est pas complète. La détection des vrilles est la partie la plus complexe à réaliser et sera discuté dans les perspectives.

## 4 Résultats expérimentaux

Pour évaluer notre système nous avons effectué un corpus vidéo que nous avons en partie étiqueté à la main. Nous avons localisé les positions de la tête, des mains, du bassin, des genoux et des pieds. Il comporte plus de 100 séquences et plus de 1000 figures réalisées par 7 athlètes. Nous avons divisé le corpus en deux parties. Une pour le réglage des algorithmes et l'autre pour les évaluer. On rencontre différentes situations dans ce corpus. Il y a trois situations : des séquences d'entraînement (seul le gymnaste apparaît à l'image) ; des séquences d'entraînement avec l'entraîneur dans le champ de la caméra (le gymnaste apparaît à l'image ainsi que l'entraîneur) ; des séquences de compétition (le gymnaste apparaît dans l'image et dans l'arrière plan figure un public en mouvement). Les évaluations présentées ci dessous ont été effectuées sur les deux premiers types de séquences, c'est à dire sans public en arrière plan.

**Suivi du corps** A partir de nos données étiquetées, nous avons déduit le centre de masse de l'athlète (barycentre des points étiquetés). Nous avons calculé le coefficient de corrélation (corrélation linéaire simple) entre les données mesurée à la main et celles calculées automatiquement par le système. Le résultat est présenté dans le tableaux 1 sous forme de pourcentage de corrélation. L'écart type présenté est celui de la distance entre les deux mesures.

**Reconnaissance des rotations** Les quarts de rotation ont été comptés manuellement. Ils sont comparés avec ceux calculés automatiquement. Le pourcentage présenté dans le tableau représente le rapport entre les quarts comptés à la main et ceux reconnu par le système.

**Reconnaissance des positions** De la même façon, chacune des positions a été évaluée. Tout d'abord étiquetées à la main elles sont comparées à celles reconnues par le système.

Les résultats du suivi sont suffisants pour obtenir une bonne reconnaissance des salti.

Les résultats obtenus pour la détection des rotations dépendent de l'environnement. Les limites du système apparaissent clairement. En effet il n'est pour le moment pas envisageable d'utiliser ce dispositif dans des conditions de compétition avec un public frénétique en arrière plan. Dans cette situation, la reconnaissance des rotations est fortement diminuée. Un meilleur algorithme d'extraction de pri-

Type	Test	Evaluation
Tracking		
Corrélation	93 %	94 %
Ecart type	26 pixels	26 pixels
Rotations		
Reconnaissance	100 %	98 %
Positions o		
Reconnaissance	57 %	50 %
Positions <		
Reconnaissance	50 %	54 %
Positions /		
Reconnaissance	100 %	100 %

TAB. 1 – Evaluation des algorithmes de reconnaissance et de suivi.

mitives est alors de rigueur. En dehors de ces conditions difficiles, les résultats de la reconnaissance des salti sont efficaces. Il arrive que le système se trompe dans des cas extrêmes. La reconnaissance des positions n'est pas encore aboutie car il n'est pour le moment pas possible de discriminer la position groupée de la position carpée. Le système confond régulièrement les deux positions car elles sont relativement proches. Cependant les informations qu'extrait le système permettent déjà d'être exploitées. Dans des conditions d'entraînement le système démontre une très bonne robustesse. Nous verrons dans la section 5 qu'il a été utilisé en conditions réelles pour adapter l'entraînement sportif.

En terme de temps de calcul, le système fonctionne en temps réel. Après la phase d'initialisation (lors de cette phase, le système ne reconnaît rien), le traitement est achevé en moyenne en 0.019 secondes par image sur un INTEL Pentium 4 cadencé à 2.6 GHz.

## 5 Evaluation de l'entraînement

Le système proposé dans cet article n'est pas finalisé. Il est cependant capable d'extraire des informations exploitables pour l'entraînement. Dans des agrès tel que les barres (fixe & asymétriques), les anneaux, le trampoline, la poutre, il apporte de nombreuses informations qui ne sont pas évidentes pendant l'entraînement. L'entraîneur ne peut pas être vigilant à toutes les informations en même temps. Le système le seconde sur des informations difficilement remarquables. Nous présentons sommairement le type d'information que le système extrait des vidéos à travers deux exemples.

### 5.1 Gymnastique - barre fixe

La première évaluation est en condition d'entraînement. Deux exercices sont filmés et sont ensuite analysés. Un premier exercice de référence réalisé correctement, et un deuxième exercice sujet à problème.

Le mouvement étudié consiste à lâcher la barre en effec-

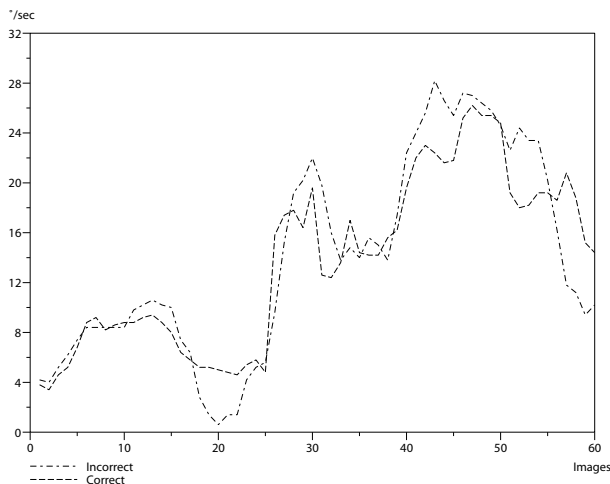


FIG. 6 – Comparaison des deux mouvements.

tuant une rotation (salto vers l'arrière) complète du corps au dessus de celle-ci et de la rattraper. Quand le mouvement est correctement effectué (premier mouvement) le gymnaste rattrape la barre. Dans le deuxième mouvement le gymnaste n'arrive pas à rattraper la barre. Nous utilisons le système pour évaluer les vitesses angulaires et comparer les deux mouvements. La figure 6 illustre les variations de vitesse angulaire au cours du temps. Cette séquence de 60 images correspond au moment où, le gymnaste arrive sous la barre (images 0 - 25), effectue une accélération tout en tenant la barre (images 25 - 30), lâche la barre (image 30), le gymnaste effectue une phase ascensionnelle (images 30 - 40) avant de grouper pour tourner au dessus de la barre (images 40 - 50) et finalement lâche ses genoux (dégroupe) pour rattraper la barre (images 50 - 60). Les deux mouvements ont été calés et superposés pour mieux les comparer. On remarque sur la courbe incorrecte une forte décélération à l'image 20. Les entraîneurs interprètent cette décélération comme une anticipation du shoot des pieds pour engager la rotation. Les images 50 à 60 représentent le moment où le gymnaste lâche ses genoux pour se préparer à rattraper la barre. Le gymnaste du deuxième mouvement (mouvement incorrecte) lâche ses genoux plus tard que celui du premier. En conclusion, pour le deuxième mouvement, le gymnaste exagère le shoot des pieds, tourne très vite et dégroupé trop tard. Le premier mouvement est plus régulier, avant de lâcher la barre l'athlète maintient une vitesse angulaire constante (images 15 - 25). Le conseil à donner aux vues de ces courbes est que pour réussir, le gymnaste doit tenter de conserver une vitesse plus constante avant le lâché de barre et doit dégroupé un peu plus tôt pour se préparer à la rattraper.

L'œil humain sur les deux vidéos ne voit pas de différence. Le système, permet de mettre en évidence l'erreur du mouvement.

## 5.2 Trampoline

De la même façon nous utilisons le système pour adapter l'entraînement de trampoline. Lors de préparation aux compétitions, les entraîneurs préparent avec les gymnastes des enchaînements de 10 éléments qui devront être parfaits. Un élément est un saut (un salto par exemple). L'idéale est de réaliser ces 10 éléments à une hauteur constante. Les figures 7 et 8 illustrent les variations d'amplitude des 10 éléments d'un enchaînement. Chaque élément est décrit en utilisant la notation numérique citée précédemment. L'élaboration de ces enchaînements n'est pas simple. Chaque gymnaste a ses particularités et ses facilités sur les figures réalisées. Un mouvement peut convenir à un individu mais pas à une autre. L'exemple ci-dessous illustre une séquence classique réalisée par de nombreux gymnastes. A l'entraînement, cette séquence passe difficilement pour un individu. L'athlète éprouve des difficultés à terminer le mouvement. L'erreur classique serait de dire qu'il suffit de s'entraîner d'avantage. Cependant en regardant simplement la figure 7 on constate une première partie relativement constante (éléments de 1 à 5), un décrochement (au 6ième élément) et une remontée vers la hauteur initiale jusqu'à la fin du mouvement. La liaison 5ième élément, 6ième élément est trop éprouvante à réaliser à cet endroit de l'enchaînement. L'idée simple est de déplacer cette liaison et de la mettre par exemple à la fin de l'enchaînement. Le système évalue à nouveau l'enchaînement.

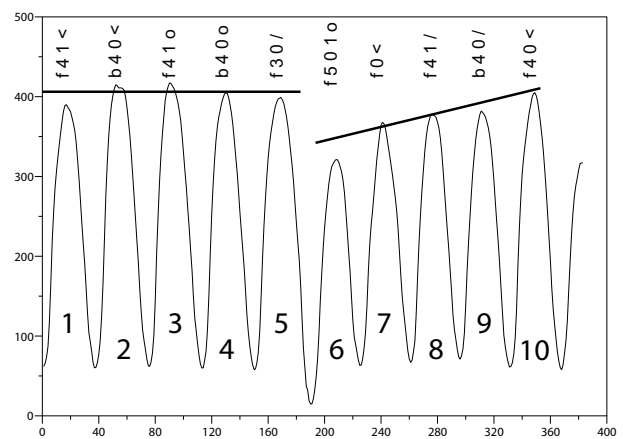


FIG. 7 – Séquence initiale (non adaptée).

Dans le mouvement corrigé (figure 8) on retrouve cette brusque perte d'amplitude (élément 8 à 9). Mais cette difficulté n'interfère plus avec les autres éléments. Il faut prendre conscience que l'athlète est acteur et ne voit pas ce qui ne va pas, c'est le rôle de l'entraîneur. Celui-ci ne peut pas être attentif à tout surtout quand il s'agit d'éléments difficilement repérables. Ces variations d'amplitude ne sont pas évidentes car on ne peut pas mettre des marqueurs aux maxima des sauts pour les visualiser. De plus si on se foca-



lise dessus, on ne fait plus attention aux éléments réalisés et on n'est même pas sûr de bien détecter les éléments incriminés dans ces pertes d'amplitude. Le système permet de mettre en évidence une perte considérable d'amplitude qui conduit à une médiocrité de l'enchaînement. L'élément incriminé est alors remplacé ou déplacé.

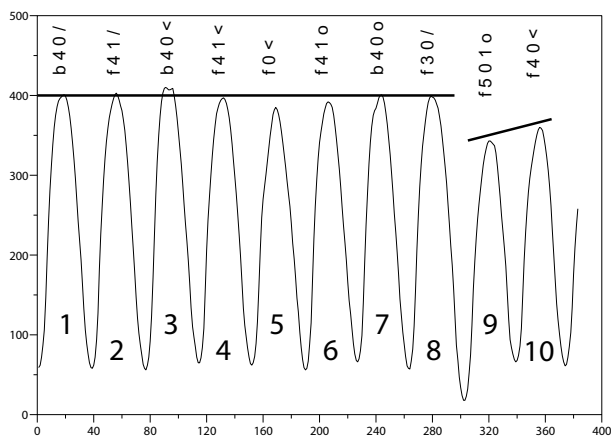


FIG. 8 – Séquence corrigée (adaptée).

Dans ces deux exemples, le système évalue les vitesses angulaires sur le premier exemple et les variations d'amplitudes sur le deuxième. Ces informations sont difficilement accessibles par l'œil humain. Dans les prochaines améliorations, le système devrait être capable de reconnaître et évaluer ces éléments.

## 6 Perspectives

La première perspective du système est de pouvoir évaluer quantitativement les vrilles. L'utilisation du flot optique pourrait résoudre ce problème. Les premières mesures du flot optique [8] nous poussent à continuer nos investigations. Le flot optique reste une technique extrêmement coûteuse en temps de calcul et pourrait casser l'atout du temps réel. Nous nous proposons alors de calculer le flot optique sur des parties de la zone de traitement après avoir compensé les composantes translationnelles et rotationnelles déjà détectées. Ce calcul ne devrait faire apparaître que les composantes rotationnelles longitudinales. La deuxième perspective est de finaliser le système pour en faire un système de reconnaissance robuste des figures acrobatiques. A terme le système se veut être un système de reconnaissance des figures acrobatiques par mesure globale. Il pourra être d'une grande aide pour les entraîneurs mais aussi pour les juges en compétition.

## 7 Conclusion

L'article présente un système d'analyse du geste sportif par des mesures globales. On entend par mesure globale le fait que le système n'identifie pas les parties du corps humain mais se base sur des mesures de celui-ci. Le système est

non intrusif. Il n'y a pas de contraintes d'utilisation pour le sportif (aucun capteur gênant le sportif), il conserve la naturalité du geste analysé. La simplicité de mise en œuvre et le faible coût du matériel en font un système accessible au grand public. La mesure globale peut conduire à une reconnaissance robuste grâce à une caractérisation adéquate. La faible complexité des algorithmes utilisés permet une utilisation en temps réel. Le système montre ses limites quand la caméra n'est pas fixe ou quand une foule dense de personnes est en mouvement dans l'arrière plan. De même le système ne permet pas de traiter deux personnes dans le champ de la caméra. A terme le système se veut être un système de reconnaissance des figures acrobatiques par mesure globale. La caractérisation du mouvement utilisée découle d'une terminologie employée et reconnue dans la communauté internationale de gymnastique. Elle est certes peut-être agréable mais chaque élément a sa traduction dans toutes les langues. Les systèmes de ce type sont d'une grande aide pour les entraîneurs mais aussi pour les juges en compétition.

## Références

- [1] H. Lakany and G. Hayes, "An algorithm for recognising walkers," *Second IAPR Workshop on Graphics Recognition*, Nancy, France, pp. 112-118, August 22-23, 1997.
- [2] Ryan Cassel and Christophe Collet, "Tracking of Real Time Acrobatic Movements by Image Processing," *5th International Gesture Workshop*, Genova, Italy, pp. 164 - 171, April 15-17, 2003.
- [3] A. Braffort, A. Choisier, C. Collet, et al., "Toward an annotation software for video of Sign Language, including image processing tools and signing space modelling," *4th International Conference on Language Resources and Evaluation*, Lisbon, Portugal, 2004.
- [4] B. Bossard, T. Convard, A. Braffort, et al., "Un système de reconnaissance de gestes générique pour la Réalité Virtuelle," *14ème Congrès Francophone AFRIF-AFIA de Reconnaissance des Formes et Intelligence Artificielle*, Toulouse, France, 2004.
- [5] Masanobu Yamamoto, Takuya Kondo, Takashi Yamagiwa and Kouji Yamanaka, "Skill Recognition," *3rd IEEE International Conference on Automatic Face and Gesture Recognition*, pp.604-609, 1998.
- [6] Greg Welch, Gary Bishop, "An Introduction to the Kalman Filter," *TR 95-041*, Technical Report, Department of Computer Science, University of North Carolina, NC, USA, 2002
- [7] Ryan Cassel and Christophe Collet, "Characterization and Tracking of Acrobatic Movements for Recognition," *Workshop in 14ème Congrès Francophone AFRIF-AFIA de Reconnaissance des Formes et Intelligence Artificielle*, Toulouse, France, 2004.
- [8] Changming Sun, "Fast Optical Flow Using 3D Shortest Path Techniques," *Image and Vision Computing*, Vol. 20, no.13/14, pp.981-991, December, 2002.
- [9] Gopal Pingali, Agata Opalach, Yves Jean, "Ball Tracking and Virtual Replays for Innovative Tennis Broadcasts," *International Conference on Pattern Recognition (ICPR'00)*, p. 4152, Volume 4, September 03 - 08, Barcelona, Spain, 2000.

- [10] Fanch Lejeune, Annelies Braffort and Jean-Pierre Desclés , “Study on Semantic Representations of French Sign Language Sentences,” *4th International Gesture Workshop on Gesture and Sign Languages in Human-Computer Interaction*, P. 197-201, London, UK, 2001.
- [11] Fédération Internationale de Gymnastique. “Trampoline Code of Points 2005,” [www.fig-gymnastics.com](http://www.fig-gymnastics.com).