

Fusion de données vidéo et télémétriques appliquée à l'étude du geste sportif.

Pascale Canal Ligné, Joseph Alizon, François Collange.

LABoratoire des Sciences et Matériaux pour l'Electronique, et d'Automatique (LASMEA),
UMR 6602 du CNRS, Université Blaise Pascal de Clermont-Ferrand,
24 avenue des Landais, 63177 Aubière Cedex. France.

RÉSUMÉ

Nous présentons une méthode d'identification et de localisation des membres d'un coureur à pied dans des images, par fusion de données multisensorielles. Notre démarche consiste à utiliser des informations délivrées par un capteur qui associe une caméra vidéo avec un télémètre laser à balayage. En relation avec un modèle du corps humain choisi sous forme de bâtonnets, nous recherchons l'attitude du sportif dans l'image. Une segmentation de l'image télémétrique permet d'obtenir une correspondance approximative entre la position du modèle et la position du personnage. Cette attitude est ensuite améliorée en fusionnant les résultats de cette segmentation dans un processus de segmentation de l'image de luminance. Afin de valider la méthode, des essais ont été réalisés sur des images vidéo/télémétriques, prises en environnement extérieur sur un pantin, dont les différentes attitudes statiques décrivent deux foulées d'un coureur à pied.

ABSTRACT

We present a method for identify and locate athlete limbs in images by multisensorial data fusion. These data are composed of 2D intensity images delivered by a video camera, and of 3D depth images delivered by a scanning range finder. With a human body model chosen as a stick figure, we search the athlete pose in images. At first, a telemetric image segmentation allows to obtain an approximative matching between the model and the body. This pose is then improved by injecting telemetric segmentation results in a process of intensity image segmentation. In order to validate the method, experiments have been conducted with video/telemetric images, taken in outdoor environment with an articulated puppet, which static poses describe two strides of a runner.

1 Introduction

L'analyse du mouvement humain est un sujet complexe de part la variété des situations et la nature même du corps de l'individu. En effet, le corps humain est un objet 3D, articulé, non rigide, et capable d'effectuer des mouvements très divers. Les conditions d'observation sont souvent mal contrôlées, en particulier lorsque la scène se déroule à l'extérieur. De plus, au cours du mouvement certains membres peuvent se trouver en partie ou complètement occultés.

Afin de s'affranchir de ces inconvénients, les méthodes traditionnelles utilisent des marqueurs positionnés sur les articulations du sujet [1]. Cependant la non rigidité de l'enveloppe corporelle durant le mouvement induit des déplacements relatifs entre les marqueurs et le corps, et donc des incertitudes sur les résultats.

Fort de ces constatations, l'objectif de notre projet de développer une méthode pour analyser le mouvement des différentes parties du corps d'un athlète dépourvu de marqueurs, dans des conditions d'expérimentation telles que celles rencontrées pour une course en stade extérieur.

Des études s'affranchissent déjà de ces marqueurs. Nous les avons classées en deux groupes. Les travaux appartenant au premier groupe concernent le suivi d'une seule partie du corps humain, comme la jambe pour Lerasle [3], ou encore la main pour Reh et Kanade [4]. Dans ces deux approches, les

scènes se déroulent à l'intérieur, avec une faible profondeur de champs (environ 1 mètre). De plus l'objet et le fond sont discriminés par des textures très différentes ou par un fort contraste. Ces conditions leur permettent d'utiliser des modèles très précis ou très complexes, voire les deux, pour reconnaître la position 3D du membre considéré.

Dans le second groupe, des études, comme celle de Hogg [2], effectuent le suivi du corps humain pris dans sa totalité. La phase initiale de l'approche de Hogg consiste à rechercher une zone d'intérêt dans l'image par un algorithme de détection de mouvement, et ensuite à effectuer un suivi guidé par le modèle. Cependant dans sa séquence d'images, prise en extérieur, seul le personnage est en mouvement, ce qui n'est pas le cas pour nous, où de nombreux objets paraissent en mouvement de part les variations d'illumination et l'influence du vent sur notre environnement (arbres, herbes). De plus, le recalage 3D de son modèle du corps humain, ne permet pas d'obtenir une grande précision sur l'orientation des membres.

En vue de faire l'analyse du mouvement des différentes parties du corps humain, notre but consiste à effectuer une étape d'initialisation automatique de la localisation des membres. Dans notre approche, nous utilisons un modèle 3D simple et facilement adaptable à chaque individu. La profondeur de champs utilisée est de 10 mètres minimum. Pour localiser grossièrement l'individu, notre zone d'intérêt est construite non pas à partir d'une détection de mouvement, mais à partir

de l'information de profondeur de la scène. De plus cette information, associée à l'information de luminance, permet d'obtenir directement une mesure 3D sur la position des membres, fonction de la précision du capteur télémétrique.

2 L'approche

La démarche novatrice que nous avons adoptée, consiste à associer deux types d'informations fournies par un capteur multisensoriel. D'une part des images 2D délivrées par une caméra vidéo, et d'autre part, des images de profondeur de la scène délivrées par un imageur télémétrique laser 3D. L'intérêt d'utiliser un télémètre est de disposer directement de l'information de profondeur sans avoir à effectuer une reconstruction 3D dans l'image. Si l'information 3D est suffisamment précise, il permet de s'affranchir des problèmes causés par des variations d'illumination et de dissocier deux membres se recouvrant dans le cas où l'on a des impacts télémétriques sur les deux.

Le télémètre laser dont nous disposons actuellement a été développé dans le cadre du projet PROMETHEUS [5] pour la détection d'obstacles sur route. En ce qui concerne ses performances actuelles, il est bien adapté en vitesse et en précision pour détecter des véhicules automobiles mais beaucoup moins pour localiser les membres d'un individu qui nécessitent une détection plus précise. En effet, l'angle solide du faisceau laser (3 mstr) ne permet pas d'obtenir une mesure suffisamment ponctuelle de la profondeur pour notre application. De plus, la vitesse d'acquisition du télémètre est beaucoup trop lente (200 ms pour l'acquisition d'une image télémétrique de 800 points) par rapport à la cadence vidéo (40 ms). C'est pour cette raison que nous avons fait l'acquisition de différentes positions statiques du personnage. Dans notre approche, nous associons les données télémétriques avec les données de luminance pour obtenir une localisation des membres plus précise (voir Figure 1).

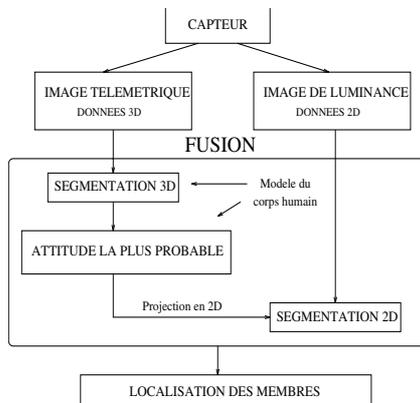


FIG. 1 — La démarche.

Notre approche consiste tout d'abord à effectuer une segmentation des données télémétriques permettant d'obtenir une localisation grossière des différents membres du personnage dans l'image. Ensuite, une segmentation de l'image de luminance, fusionnée avec les résultats de la segmentation de

l'image télémétrique, permet d'apporter de la précision sur la localisation des membres.

2.1 Le modèle du corps humain

Pour initialiser le processus d'identification et de localisation des membres dans l'image, nous utilisons un modèle du corps humain. Compte tenu de nos données télémétriques peu précises, nous avons choisi d'utiliser une représentation simple par bâtonnets qui évoluera en complexité au cours du processus. Notre modèle a pour référence le modèle 3D de Winter [7] utilisé en biomécanique. Il est simple, défini avec peu de paramètres, et adaptable à chaque individu étudié. C'est un modèle où chaque membre est représenté par un segment de longueur paramétré selon la taille de l'individu. Notre modèle est constitué de 10 segments caractérisés par leur longueur, leur position et leur orientation (voir Figure 2). La position de chaque segment est définie comme étant la hauteur séparant le sol du centre du segment considéré. De part l'information 3D, les dimensions du modèle sont ajustées en fonction de la distance à laquelle se situe l'individu.

Notre démarche consiste à recalculer au mieux les segments 3D de ce modèle avec des segments représentant le personnage, qui sont obtenus par une séquence algorithmique de segmentation des données télémétriques.

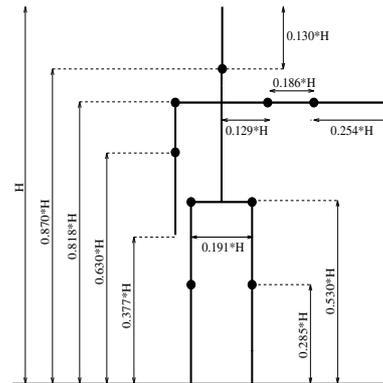


FIG. 2 — Le modèle du corps humain.

2.2 Segmentation de l'image télémétrique

Tout d'abord, pour localiser grossièrement l'individu, nous seuillons les images télémétriques selon un intervalle de distance connu où se situe l'individu. Seuls les impacts télémétriques ayant atteints le sportif sont ainsi extraits de l'image.

Afin de représenter le corps humain sous la forme de segments, nous appliquons sur l'image télémétrique seuillée une opération d'extraction du squelette, suivi d'un chaînage, et d'une approximation par les segments.

Les segments sont ensuite classés suivant leur appartenance aux différentes parties du modèle par une méthode Bayésienne. Cette méthode consiste, dans un premier temps, à définir des lois de probabilités d'appartenance aux membres selon les caractéristiques des segments du modèle (longueur, position, et orientation) en fonction de la taille de l'individu.

Pour chaque membre, ces lois sont établies selon des intervalles de confiance dans lesquelles les caractéristiques considérées sont les plus probables. Ne connaissant pas la forme de la distribution liée aux longueurs des segments du modèle de Winter, nous l'avons modélisée comme étant une équiprobabilité sur un intervalle borné, avec de part et d'autre une décroissance linéaire. Les bornes minimale et maximale ont été estimées de manière empirique à 10% de la valeur moyenne de la longueur du modèle.

En ce qui concerne la position et l'orientation des segments, les lois associées sont du même type que celles relatives à la longueur. Cependant les valeurs minimales et maximales sont issues d'observations effectuées sur plusieurs kinogrammes décrivant des courses différentes.

En exemple, la Figure 3 présente une loi de probabilité d'appartenance à un membre en fonction d'un paramètre x du segment considéré.

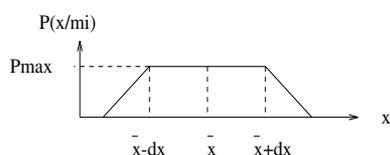


FIG. 3 — Exemple d'une loi de probabilité.

Dans un deuxième temps, par application de la règle de Bayes (voir équation (1)), nous en déduisons les probabilités qu'un segment du squelette, extrait de l'image, de longueur l , de position p , et d'orientation o , appartienne à chacun des membres.

$$P(mi/l.p.o) = \frac{P(l/mi) * P(p/mi) * P(o/mi) * P(mi)}{\sum_{mj=1}^M (P(l/mj) * P(p/mj) * P(o/mj) * P(mj))} \quad (1)$$

avec :

$M = 10$ (nombre de membres).

$P(mi/l.p.o)$ = Probabilité qu'un segment de longueur l , de position p et d'orientation o appartienne au membre mi .

$P(mi/l)$ = Probabilité qu'un segment de longueur l appartienne au membre mi .

$P(mi/p)$ = Probabilité qu'un segment de longueur p appartienne au membre mi .

$P(mi/o)$ = Probabilité qu'un segment de longueur o appartienne au membre mi .

$P(mi)$ = Probabilité que le segment appartienne au membre mi .

A l'issue de ces calculs, chaque membre mi est représenté par une classe ci de segments pour lesquels $P(mi/l.p.o)$ est non nulle. Pour chaque classe ci , nous avons donc plusieurs représentants. A terme, chaque classe ci doit comporter des segments dont l'union soit représentative du segment du modèle correspondant au membre associé à cette classe. Cependant l'algorithme d'approximation polygonale appliqué précédemment est non supervisé, c'est à dire qu'il ne prend pas en compte le modèle. Aussi la sélection des segments de chaque classe nécessite d'être améliorée. Nous avons donc développé un algorithme de segmentation supervisé par notre modèle à partir d'un critère sur les longueurs des membres. Pour cela, nous définissons la mesure de "longueur de la classe" li comme étant la longueur de la chaîne de pixels dont sont issus les segments de la classe. Ensuite, partant d'une

extrémité d'un membre, nous limitons cette longueur li à la longueur du membre du modèle correspondant. Ce critère est spécifique à chaque classe. Puis, nous appliquons, de façon récursive sur toutes les classes, la séquence algorithmique précédente : approximation des chaînes par les segments, et classement des segments par la méthode Bayésienne.

Les premières classes traitées sont celles contenant un point extrémité du squelette (pieds, avant-bras, tête). A chaque itération nous affectons les nouveaux segments dans les différentes classes. Les nouveaux points, correspondant à la position des articulations, sont alors utilisés comme point de départ pour les autres classes voisines.

La Figure 5a montre les résultats obtenus sur la localisation des avant-bras, où les parties correspondantes au squelette sont projetées dans l'image de luminance.

De part l'information télémétrique grossière dont nous disposons, la localisation des différentes parties du corps humain manque quelque peu de précision. Les résultats de cette segmentation de l'image télémétrique sont alors injectés dans un algorithme de segmentation de l'image de luminance pour apporter plus de précision sur l'attitude recherchée.

2.3 Segmentation de l'image de luminance

La segmentation de l'image de luminance a pour objectif de localiser plus précisément la position de chacun des membres. Comme vu précédemment une segmentation au sens du mouvement n'est pas exploitable sur nos données expérimentales. Il en est de même pour une segmentation au sens de la texture compte tenu des variations d'illumination. Aussi, nous avons choisi d'effectuer une segmentation au sens des contours, comme Hogg l'a également appliquée après sa détection du mouvement. Cette segmentation consiste à extraire de l'image de luminance les contours 2D de chacun des membres.

Disposant de la localisation approximative des membres, nous avons choisi d'introduire cette connaissance dans la segmentation afin d'éliminer certains contours non pertinents. Pour cela, nous utilisons un algorithme de Ligne de Partage des Eaux (LPE) [6] appliqué à l'image du gradient sur laquelle nous avons imposé les minima locaux comme étant des marqueurs issus de la localisation télémétrique. Nous imposons pour chaque membre, un marqueur correspondant à la projection de la partie du squelette appartenant à ce membre issue de la segmentation de l'image télémétrique, et d'un autre marqueur qui entoure ce membre (voir Figure 4 et Figure 5a). Ce dernier est créé à partir de la longueur, de la largeur, et de l'orientation approximative du membre.

Nous appliquons la LPE dans une région de recherche préalablement définie autour de chaque membre (RImi : Région d'Intérêt du membre mi).

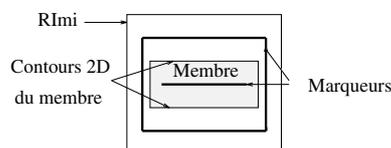


FIG. 4 — Marqueurs d'un membre.

En exemple, la Figure 5b montre les contours 2D des avant-bras extraits de l'image vidéo.

Les axes de symétrie, issus des contours des membres extraits de l'image vidéo, permettent alors d'affiner la précision de l'orientation de chacun des membres. Ces axes de symétrie sont obtenus par la méthode des moindres carrés appliquée sur les points de contours extraits.

3 Résultats expérimentaux

Des essais ont été réalisés sur des images vidéo/téléométrique décrivant les deux foulées d'un coureur à pied. Compte tenu de la trop faible vitesse d'acquisition du télémètre pour notre application, nous utilisons un pantin 3D, en vrai grandeur, articulé, et qui permet de tenir les positions de course figées pendant le temps nécessaire à l'acquisition des données 3D. Les données expérimentales ont été obtenues, en environnement extérieur, avec des variations d'éclairage notables, et avec un fond de la scène non uniforme dans le temps (influence du vent sur l'herbe et les arbres). Sur ces images, des méthodes de soustraction du fond et/ou de détection du mouvement, en vue de sélectionner une zone d'intérêt comprenant le sportif, ont été mises en échec. A partir d'une série de 10 images vidéo/téléométriques décrivant deux foulées, et prise dans les conditions décrites précédemment, la segmentation de l'image téléométrique apporte des résultats satisfaisants dans la localisation approximative des membres de l'individu. Cette localisation est ensuite améliorée par une fusion de ces résultats avec la segmentation de l'image de luminance. Les contours 2D d'un membre nous permettent ensuite de sélectionner les impacts téléométriques relatifs à ce membre, et ainsi de mesurer une position 3D du membre fonction de la précision de notre capteur. L'analyse des différents résultats montre cependant que la précision de la localisation des membres issue de la segmentation de l'image de luminance peut être encore améliorée en respectant deux contraintes simples à mettre en oeuvre pour notre application. D'une part en utilisant un habit non texturé pour le coureur, et d'autre part, un contraste plus significatif entre le coureur et le fond. Dans le but de quantifier la précision de notre détection, il est nécessaire d'effectuer une nouvelle acquisition d'images en respectant les contraintes précédentes.

4 Conclusion et perspectives

Ces premiers résultats valident notre première approche de fusion des données multisensorielles caméra vidéo/capteur 3D. Cependant, compte tenu de la précision de notre télémètre, certains membres peuvent être non détectés à partir des données 3D. Afin de palier à ce problème, nous envisageons actuellement une deuxième approche parallèlement à la première, basée sur la prise en compte de données de chrominance. Sur l'habit du coureur, chaque membre sera caractérisé par une couleur particulière. Les parties du corps humain non détectées par les données 3D pourront ainsi l'être grâce à une segmentation des données de chrominance. Nous avons

orienté la suite de nos travaux dans cette voie. Le cas échéant, en fonction des résultats que nous obtiendrons, les deux approches pourront être fusionnées.

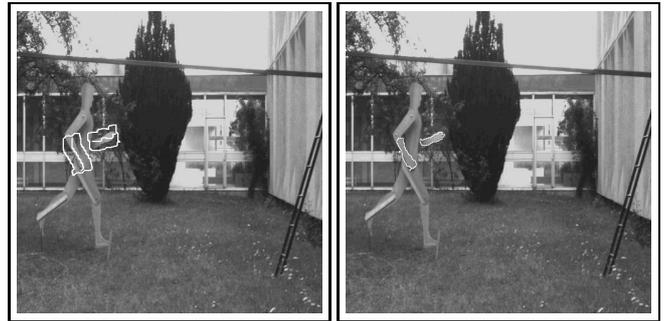


FIG. 5 — Les avant-bras : a - résultats de la segmentation téléométrique reprojétés dans l'image de luminance, et marqueurs associés. b - les contours 2D, résultats de la segmentation de l'image de luminance.

Références

- [1] F. Elsner and K. Hassine and J.C. Angué. The Use of Image Sequences to Track Human Motion by Spatio-temporal Segmentation. *Proceedings of the Workshop on Computer Vision and Space Applications*, pp 208-219, Antibes - Juan les Pins, 22-24 septembre 1993.
- [2] D. Hogg. Model-based Vision : a Program to See a Walking Person. *Image and Vision Comput.*, 1(1), pp 5-20, 1983.
- [3] F. Lerasle. *Vers le suivi du geste sportif par vision artificielle*, Thèse de Doctorat, Université Blaise Pascal, Clermont-Ferrand, 1997.
- [4] J.M. Rehg and T. Kanade. Visual Tracking of High dot Articulated Structures : an Application to Human Hand Tracking, *In European Conf. Comp. Vision*, pp 35-46, 1994.
- [5] L. Trassoudaine. *Solution multisensorielle temps réel pour la détection d'obstacles sur route*, Thèse de Doctorat, Université Blaise Pascal, Clermont-Ferrand, 1993.
- [6] L. Vincent. *Algorithmes morphologiques à base de files d'attente et de lacets. Extension aux graphes*, Thèse de Doctorat, Ecole Nationale Supérieure des Mines de Paris, mai 1990.
- [7] D.A. Winter. *Biomechanics and Motor Control of Human Movement*. John Wiley et Sons, 1979.