

Compression et caractérisation de systèmes de trajectoires

Frédéric CHENEVIÈRE, Samia BOUKIR, Bertrand VACHON

Laboratoire Informatique, Image, Interaction
Pôle Sciences et Technologie 17042 La Rochelle Cedex 1 - FRANCE
fchenevi@univ-lr.fr, sboukir@univ-lr.fr
bvachon@univ-lr.fr

Résumé – Le but de ce travail est la reconnaissance de gestes dansés issus d’un ballet contemporain. Les données traitées sont des trajectoires de mouvement suivies par les articulations d’un corps dansant. Ces courbes spatio-temporelles sont fournies par un système de capture du mouvement. Il est évident que l’utilisation des signaux bruts sans pré-traitement préalable est inappropriée et coûteuse. Nous proposons donc un outil efficace pour le sous-échantillonnage non uniforme de signaux spatio-temporels. Notre approche utilise une adaptation des modèles de contours actifs pour construire une représentation compacte et efficace des trajectoires de mouvement. Notre méthode de reconnaissance de gestes dansés repose sur un ensemble automatiquement pondérés de modèles de Markov cachés (MMC) continus, chacun étant associé à la trajectoire d’un marqueur.

Abstract – We aim at recognizing a set of dance gestures from contemporary ballet. Our input data are motion trajectories followed by the joints of a dancing body provided by a motion-capture system. It is obvious that direct use of the original signals is unreliable and expensive. Therefore, we propose a suitable tool for nonuniform sub-sampling of spatio-temporal signals. The key of our approach is the adaptation of active models to provide a compact and efficient representation of motion trajectories. Our dance gesture recognition method involves a automatic weighted set of continuous Hidden Markov Models (HMMs), each of them being related to a motion trajectory followed by the joints.

1 Introduction

L’intérêt grandissant des chercheurs en vision artificielle pour l’analyse du mouvement se manifeste dans une très grande variété de contextes : de la robotique à l’indexation vidéo en passant par les interfaces homme-machine, la réalité virtuelle ou augmentée et la danse. Les travaux consacrés à cette dernière application ne sont pas légion et se focalisent généralement sur l’assistance à la notation du mouvement [6]. Excessivement lourde si elle est exécutée manuellement, cette transcription pourrait être automatisée par une conversion automatique de séquences spatio-temporelles. C’est l’une des finalités de notre projet qui implique deux autres partenaires : le BARC¹, un ballet contemporain de renommée internationale et XD-Productions, un studio de capture du mouvement.

Dans le cadre de l’analyse du mouvement humain, nous considérons tout mouvement comme un ensemble de courbes spatio-temporelles qui représentent les trajectoires suivies par les différentes articulations du corps. Les systèmes de capture du mouvement transforment ces courbes en signaux discrets en sous-échantillonnant la position des marqueurs, placés judicieusement sur le corps en mouvement, à des intervalles réguliers. L’énorme masse de données induite impose le recours à des techniques de compression d’information. Notre approche utilise les contours actifs pour construire une représentation compacte et efficace des trajectoires de mouvement.

Nous proposons ensuite un système de reconnaissance de gestes dansés à partir de données 3D basé sur des

modèles de Markov cachés, largement utilisés pour la représentation de séries temporelles [3].

2 Compression de courbes spatio-temporelles

2.1 Techniques existantes

L’utilisation directe des signaux spatio-temporels générés par les systèmes de capture du mouvement est peu fiable et coûteuse en temps de calcul. L’exploitation convenable de ces données passe donc par une représentation efficace et compacte. La littérature fournit différents types de solutions.

Si on met de côté le sous-échantillonnage uniforme et ”aveugle”, on trouve les méthodes fonctionnelles (B-splines non uniformes) [11]. Cependant, leurs bons résultats ne doivent pas faire oublier la problématique du choix des points de contrôle. Les procédés qui se basent sur l’Analyse en Composantes Principales (ACP) [9] sont sources de difficultés d’interprétation des nouvelles variables engendrées. Les descriptions de Fourier [5] ne constituent pas une représentation efficace dans le cadre de notre application (trop de pertes d’informations [1]). Enfin, l’approximation polygonale de contours est très répandue [4], mais il paraît difficile de les faire évoluer vers des modèles plus précis tout en gardant le caractère automatique des algorithmes efficaces employés.

Nous proposons ici d’adapter les modèles de contours actifs au sous-échantillonnage non uniforme de trajectoires spatio-temporelles.

1. Ballet Atlantique Régine Chopinot

2.2 Sous-échantillonnage non uniforme de signaux spatio-temporels par contours actifs

2.2.1 Principe

Les contours actifs sont désormais largement utilisés en analyse d'images pour la segmentation. Cette approche par modèle combine géométrie, physique et théorie de l'approximation pour offrir une description compacte et analytique des contours grâce à une représentation paramétrique [7].

Alors que le contour actif est traditionnellement défini par un ensemble de points de l'image, nous le définissons dans ce travail par un n-uplet de points de la trajectoire. Classiquement, une énergie propre est associée à ce contour, elle maintient les caractéristiques intrinsèques de tension et de rigidité du contour (énergie interne) et le lie aux propriétés de la trajectoire (énergie externe). L'expression de l'énergie interne reste classique, en revanche, nous proposons d'adapter l'énergie externe à notre problème. Les points marquants de la trajectoire (dérivée nulle ou inflexion) forment des bassins d'attraction pour les noeuds du contour.

Après l'initialisation du contour actif (peu de noeuds répartis uniformément sur la trajectoire, 2 points ici), un cycle de deux étapes démarre. La première étape, dite d'optimisation, consiste à minimiser l'énergie du contour par un processus itératif déterministe. La seconde étape (multi-résolution) assure que le contour approxime la trajectoire avec la résolution δ choisie. δ correspond au rapport entre la distance entre deux noeuds consécutifs du contour, et la longueur du chemin reliant ces deux noeuds le long de la trajectoire. Quand ce rapport vaut 1, le contour suit exactement la trajectoire; plus le rapport est faible, plus on autorise le contour à s'éloigner de la trajectoire. Pour les intervalles de deux noeuds consécutifs qui n'assurent pas le critère de résolution δ , un nombre de noeuds supplémentaires, fixe et arbitrairement petit (un seul noeud ici), est inséré dans le contour.

Tant que de nouveaux noeuds sont insérés, le cycle se poursuit. Le contour actif obtenu retient donc de manière non uniforme l'ensemble des points pertinents de la trajectoire initiale.

2.2.2 Résultats

La figure 1 montre les résultats de compression obtenus en maintenant un bon compromis entre la résolution δ et la compression du signal d'entrée et ce, même à un taux de compression dépassant les 90 %. Une résolution δ à 0,90 (respectivement à 0,95) conduit à un taux de compression de 92% du signal d'entrée (respectivement de 79%).

On peut bien évidemment arrêter le processus itératif de compression de signal avant terme en se fixant comme condition d'arrêt un taux de compression donné.

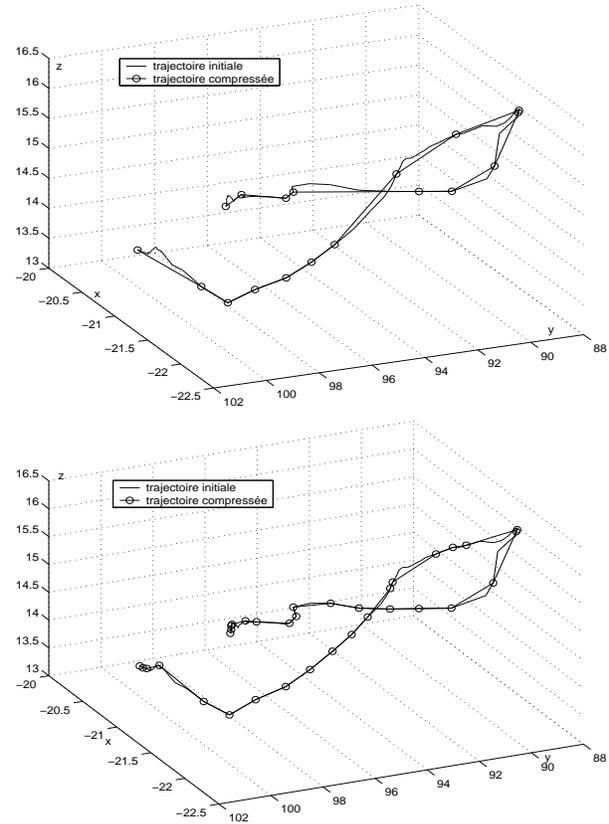


FIG. 1: Approximation d'un signal 3D - $\delta = 0,90$ et $\delta = 0,95$

3 Reconnaissance de gestes dansés par modèles de Markov cachés

3.1 Principe

Des ensembles de ces trajectoires sous-échantillonnées font ensuite l'objet d'une classification, notre but étant de parvenir à un système de reconnaissance supervisé capable de différencier divers mouvements de danse contemporaine indépendamment de la personne qui les réalise. La littérature offre essentiellement trois types de méthodes efficaces [3]. Les deux premières, les techniques de mises en correspondance comme le Dynamic Time Warping [8] et les réseaux de neurones temporels [2] en l'occurrence, n'affichent pas un aussi bon compromis que les modèles de Markov Cachés (MMC) en termes de robustesse, de souplesse et de rapidité [10]. Le système proposé s'appuie donc sur ces modèles probabilistes [3]. Puisqu'il modélise des ensembles cohérents de trajectoires, il regroupe autant de MMC continus que de trajectoires. A chaque MMC est affectée une pondération qui est initialisée arbitrairement (ce qui permet l'ajout de connaissances a priori sur l'importance de certains marqueurs vis à vis d'un mouvement donné), puis affinée par apprentissage.

3.2 Expérimentations

Pour expérimenter notre système de reconnaissance, une base de 170 gestes dansés a été construite au studio de capture du mouvement d'XD-Productions avec l'aide de

quatre danseurs professionnels du BARC. Elle correspond à une liste de 12 mouvements fondamentaux du ballet contemporain. Nous modélisons un mouvement par un ensemble de 15 MMC de structure identique, décrivant chacun le comportement spatio-temporel de chacun des 15 marqueurs placés judicieusement sur le corps dansant [1].

La table 1 montre les résultats de reconnaissance obtenus avec différents taux δ de résolution de sous-échantillonnage des trajectoires de mouvement.

D'une part, ces résultats démontrent que l'utilisation directe des signaux d'origine, sans pré-traitement, est inappropriée. Des améliorations remarquables sont obtenues en utilisant un algorithme de sous-échantillonnage de signaux adéquat. D'autre part, une résolution trop grande ou trop faible dégrade nettement la tâche de reconnaissance.

TAB. 1: Résultats de reconnaissance de gestes dansés

Résolution δ	Résultats de reconnaissance de gestes dansés	
	Base d'apprentissage	Base de test
1	91.3 %	82.1 %
0,95	91.9 %	83.1 %
0,90	95.2 %	91.7 %
0,80	94.5 %	90.8 %
0,50	86.2 %	84.5 %

Les résultats de l'apprentissage sur cette base montrent l'existence d'une valeur critique de résolution du sous-échantillonnage amenant à un taux de reconnaissance optimal (92 % sur la base de test contre 82 % avec les données brutes). Les variations naturelles des gestes d'une même classe sont atténuées par le processus de réduction des données. En outre, les différences de gestes d'une classe à une autre sont amplifiées par le système de pondération.

4 Conclusion

Nous avons présenté un outil puissant pour le sous-échantillonnage non uniforme de signaux spatio-temporels. L'adaptation des contours actifs à ce problème engendre, pour une résolution donnée, une description compacte et efficace d'un ensemble de trajectoires de mouvement compressées à des taux différents. La tâche délicate de reconnaissance s'en trouve facilitée. Notre système de MMC continus et pondérés assure à 92 % la reconnaissance de 12 classes de mouvements de danse contemporaine. Notre perspective à court terme est d'étoffer notre système de reconnaissance pour l'amener à distinguer des critères qualitatifs (excentricité, concentricité, symétries du mouvement).

Références

[1] Boukir S., Beets E., Chenevière F. Représentation et compression de signaux spatio-temporels. Technical Report RT-2002-09-001, Laboratoire L3i, Université de La Rochelle, September 2002.

[2] Chappelier J.C. *RST: une architecture connexionniste pour la prise en compte de relations spatiales et temporelles*. PhD thesis, École Nationale Supérieure des Télécommunications, January 1996.

[3] Chenevière F. Prétraitements et apprentissage de données spatio-temporelles dans le cadre du mouvement dansé. Technical Report RT-2002-01-001, Laboratoire L3i, Université de La Rochelle, January 2002.

[4] Chenevière F., Boukir S., Vachon B. A HMM-based dance gesture recognition system. *IWSSIP 02, 9th Int. Workshop on Systems, Signals and Image Processing, Manchester, UK*, pages 322–326, November 2002.

[5] Gonzalez R.C., Woods R.E. *Digital image processing*. Prentice Hall, 2002.

[6] Herbison-Evans D. Dance and the computer: A potential for graphic synergy. Technical Report 422, Basser Department of Computer Science, University of Sydney, January 1991.

[7] Kass M., Witkin A. and Terzopoulos D. Snakes: Active contour models. *Proceedings of the First International Conference of Computer Vision*, pages 259–268, 1987.

[8] Keogh E.J. and Pazzani M. Scaling up dynamic time warping to massive datasets. *Proceedings of European Conference on Principles of Data Mining and Knowledge Discovery (PKDD)*, 1999.

[9] Ormoneit D. et al. Learning and tracking cyclic human motion. *Advances in Neural Information Processing Systems 13*, Leen, Todd K. and Dietterich, Thomas G. and Tresp Volker Eds., The MIT Press, pages 894–900, 2001.

[10] Rabiner L.R. A tutorial on hidden Markov models and selected applications in speech recognition. *Proceedings of the IEEE*, 77(2):257–286, 1989.

[11] Sudarsky S., House D. Motion capture data manipulation and reuse via B-splines. *CAPTECH 98, Int. Workshop on Modelling and Motion Capture Techniques for Virtual Environments*, pages 55–69, 1998.