Conception d'un système de reconnaissance de gestes dansés

Design of a dance gesture recognition system

S. Boukir, F. Chenevière
Laboratoire L3i, Université de La Rochelle, Avenue Michel Crépeau,17042 La Rochelle cedex 1
sboukir@univ-lr.fr, fchenevi@univ-lr.fr

Manuscrit reçu le 2 septembre 2003

Résumé et mots clés

Nous présentons ici un prototype complet et opérationnel intégrant la compression et 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. Nous proposons un outil efficace pour le sous-échantillonnage non uniforme de signaux spatio-temporels. Notre approche utilise une approximation polygonale des contours pour construire une représentation compacte et efficace des trajectoires de mouvement. Notre méthode de reconnaissance de gestes dansés repose sur un ensemble de Modèles de Markov Cachés (MMC) chacun étant associé à la trajectoire d'un marqueur. Nous avons validé notre système de reconnaissance sur 12 mouvements de base effectués par 4 danseurs d'un ballet contemporain.

Approximation polygonale, compression du signal, modèles de Markov cachés, reconnaissance de gestes, trajectoires de mouvement.

Abstract and key words

We present here a whole operational prototype for the compression and recognition 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. We propose a suitable tool for nonuniform sub-sampling of spatio-temporal signals. The key of our approach is the use of polygonal approximation to provide a compact and efficient representation of motion trajectories. Our dance gesture recognition method involves a set of Hidden Markov Models (HMMs), each of them being related to a motion trajectory followed by the joints. We have validated our recognition system on 12 fundamental movements from contemporary ballet performed by 4 dancers.

Gesture recognition, hidden Markov model, motion trajectory, polygonal approximation, signal compression.

1. Introduction

L'intérêt grandissant que porte la communauté des chercheurs en vision artificielle à l'analyse du mouvement humain [1] se manifeste dans une très grande variété de contextes : de l'analyse de performance en sport [7] à la surveillance, en passant par les interfaces homme-machine [32], la réalité virtuelle et augmentée [27, 30, 12], l'indexation vidéo et la danse [6, 9]. Rares sont les travaux consacrés au dernier thème en dépit de sa dimension artistique et culturelle. Les différentes contributions de l'informatique à la danse sont décrites dans [13]. Notre contribution concerne l'interprétation et la caractérisation de gestes dansés et implique deux autres partenaires : le BARC[1], un ballet contemporain de renommée internationale et XD-Productions, un studio de capture du mouvement. Nous ne présentons pas ici de nouvel algorithme mais plutôt une association judicieuse de méthodes existantes que nous ajustons pour aboutir à un prototype applicatif complet et opérationnel. Nous proposons néanmoins une version améliorée d'un algorithme de compression de courbes existant.

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, positionnés 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 une approximation polygonale de courbes pour construire une représentation compacte et efficace des trajectoires de mouvement. La reconnaissance de tels ensembles de trajectoires est alors effectuée par un système à base de modèles de Markov cachés [9, 8] largement utilisés pour la représentation de séries temporelles [24, 26, 7, 29, 32, 23].

2. Compression de données

La compression de courbes est un pré-traitement fondamental dans notre système de reconnaissance automatique de gestes dansés. L'objectif est de trouver un bon compromis entre compression et préservation de la forme d'origine. Une bonne entrée pour l'étape de reconnaissance est tout d'abord définie par un taux de compression élevé. La compression permet en effet non seulement d'atténuer les variations intra-gestes mais aussi d'augmenter la discriminance inter-classes. Plus on garde d'informations des gestes, plus le risque de rencontrer des ambiguïtés entre eux est important et inversement. Il est important aussi de préserver les pics des trajectoires associées a chaque geste.

2.1 État de l'art

Il existe de nombreux moyens de réduire la dimension des données [5, 11] mais, nous n'aborderons ici que les techniques applicables aux séries temporelles.

Une méthode de compression très répandue est l'Analyse en Composantes Principales (ACP) [4]. Cette technique transforme un ensemble de variables corrélées en un nouvel ensemble de variables non corrélées, appelées composantes principales, classées par ordre d'importance décroissant. L'ACP peut être utilisée pour la description de courbes en général [11] et les trajectoires de mouvement en particulier [21]. Le principal inconvénient de ce type de techniques est la difficulté d'interprétation des nouvelles variables engendrées, à savoir, les *composantes principales*.

Les descripteurs de Fourier sont aussi souvent utilisés pour lisser les détails d'une forme. Ils permettent une décomposition hiérarchique d'une courbe (contour d'une forme par exemple).

Si on retient uniquement un sous-ensemble de descripteurs basses fréquences, on obtient une courbe approchant juste l'aspect global de la forme. En augmentant le nombre de composantes de la description, les hautes fréquences sont aussi rendues et les détails de la forme peuvent ainsi être générés [22, 5, 11]. Malheureusement, dans certaines applications, trop d'informations sont perdues lors de l'approximation et on ne peut garantir que la reconstitution de la forme sera fiable après la troncature de l'ensemble des descripteurs de Fourier associé [22, 5].

Une technique relativement similaire aux descripteurs de Fourier sachant qu'elle opère dans le domaine fréquentiel et qu'elle permet aussi une représentation hiérarchique des contours est basée sur les ondelettes [20]. Le principal avantage des ondelettes est qu'elles représentent mieux les caractéristiques locales d'un contour. Si un coefficient est modifié, seule une partie du contour en sera affectée, alors que la modification d'un descripteur de Fourier a une influence sur l'aspect global du contour.

Les descripteurs stochastiques, issus généralement des modèles AR (Auto-Régressif), permettent, comme les descripteurs de Fourier, une description globale de la forme. Un modèle AR linéaire estime la valeur d'une fonction à partir de *m* valeurs précédentes en combinaison linéaire. Ce type de modèle est souvent inadéquat pour représenter des formes non convexes. La précision d'approximation de la forme peut néanmoins être améliorée en augmentant l'ordre *m* du modèle ou en considérant des modèles non linéaires. Kaupinnen *et al.* ont effectué une étude comparative entre descripteurs de Fourier et descripteurs stochastiques (AR) qui montre la supériorité des descripteurs de Fourier dans un contexte de classification [15]. Ils expliquent cette supériorité par leur excellente aptitude à décrire l'aspect global d'une forme ce que confirment nos expérimentations [5].

[1] Ballet Atlantique Régine Chopinot

Une autre méthode de réduction de données est l'interpolation polynomiale qui est largement utilisée pour une représentation lisse de courbes [19] ou de signaux basses fréquences. Il s'agit en général d'une représentation en termes de splines ou de B-splines, modèles de base pour l'approximation de courbes [25]. Dans [27], une représentation hiérarchique en termes de B-splines non uniformes fournit une description compacte et efficace des trajectoires de mouvement. L'utilisation de B-splines non uniformes [27] (plutôt qu'uniformes [25]) assure plus de flexibilité durant le processus d'interpolation. Les techniques de réduction de données, basées sur des splines, les plus répandues, sont les stratégies de suppression de nœuds [19]. Ces méthodes éliminent les nœuds les moins significatifs de la spline d'origine en accord avec une tolérance donnée qui autorise une déformation plus ou moins importante de la spline d'origine. Cependant, le succès de ces techniques dépend du choix problématique du nombre de points de contrôle ainsi que de leur position initiale.

Une alternative à l'interpolation polynomiale est l'approximation polygonale de courbes qui s'avère de loin moins coûteuse en temps de calcul [25]. C'est une technique bien connue et très utilisée en pratique [18, 11]. Elle permet de représenter un contour par un certain nombre de segments dont les extrémités correspondent à des points caractéristiques de la forme. Dans [25], les auteurs ont effectué un comparatif approfondi des méthodes d'interpolation par B-splines avec les méthodes d'approximation polygonale de courbes. Leur approche, naturelle en compression de courbes, consiste à déterminer le nombre minimum de points représentatifs tel que la courbe approximative résultante induise une erreur maximum au plus égale à une tolérance fixée ϵ. Les auteurs utilisent la métrique de Hausdorff, qui permet de calculer dans leur cas la distance entre deux courbes polygonales, pour déterminer la déviation maximum par rapport à la courbe d'origine. Ce critère de précision local permet de garantir que chaque point de la courbe satisfait la tolérance ϵ. Les taux de compression qu'ils ont obtenus, avec la même tolérance, sont de loin plus avantageux en utilisant les méthodes d'approximation polygonale. Les différences dans les taux de compression entre l'interpolation par B-splines et l'approximation polygonale varient de 12 à 56 % pour les courbes complexes et de 6 à 50 % pour les courbes lisses selon les algorithmes utilisés. Cependant, l'interpolation polynomiale préserve mieux l'apparence visuelle de la courbe d'origine.

Nous avons donc opté pour une technique d'approximation polygonale. De plus, comme dans notre cas il s'agit de sous-échantillonner des signaux, l'utilisation de segments de droite plutôt que des segments de courbe n'a aucun impact sur notre tâche ultérieure de reconnaissance. Nos primitives de base sont des points 3D [5, 9].

2.2 Approximation polygonale de courbes

Une technique d'approximation polygonale des contours puissante et originale a été récemment proposée par Latecki et Lakämper [17]. Contrairement aux méthodes conventionnelles basées sur une stratégie descendante[2] [11], l'algorithme de Latecki et Lakämper repose sur une décomposition ascendante. Le principal avantage de cet algorithme par rapport aux méthodes plus classiques est qu'il ne requiert le réglage d'aucun paramètre. L'idée de base du processus discret d'évolution de la forme est très simple. Soit $\mathcal{P}_0 = \{v_0, \ldots, v_{n-1}\}$ le polygone constitué par la liste des points consécutifs de la courbe d'origine. Pour chaque étape d'évolution k :

1. $\forall\, v_i \in \mathcal{P}_k, v_i \mapsto W(v_i, \mathcal{P}_k), 0 < i < n-1$

2. $\mathcal{P}_{k+1} = \left\{ s : s \in \mathcal{P}_k \text{ et } s \notin \arg\min_{v \in \mathcal{P}_k} W(v, \mathcal{P}_k) \right\}$

Le critère utilisé $W(v, \mathcal{P}_k) = W(u, v, w)$, u et w étant les voisins directs du sommet v, est défini comme suit :

$$W(v, \mathcal{P}_k) = \frac{\theta \cdot |\widehat{uv}| \cdot |\widehat{vw}|}{L \cdot (|\widehat{uv}| + |\widehat{vw}|)} \qquad (1)$$

avec $\theta = (\vec{uv}, \vec{vw})$ l'angle entre les segments consécutifs uv et vw dans $\mathcal{P}_k$, $|\widehat{uw}|, |\widehat{vw}|$ les longueurs des arcs uv et vw, et L la longueur totale de la courbe.

2.3 Application au sous-échantillonnage de signaux

Pour sous-échantillonner des signaux spatio-temporels, nous utilisons une version modifiée de l'algorithme de Latecki et Lakämper [17]. Nous proposons en effet une nouvelle énergie E qui s'avère plus efficace que W pour la compression et la reconnaissance de trajectoires et permet un réglage plus naturel de la précision souhaitée (cm).

$$E(v_i, \mathcal{P}_k) = \max_{u_j \in \mathcal{P}_0,\, l<j<m} h_j \quad \backslash \qquad (2)$$

$$u_l = v_{i-1}, u_m = v_{i+1}; \quad u_l, u_m \in \mathcal{P}_0$$

avec h_j la distance euclidienne du sommet u_j du polygone d'origine $\mathcal{P}_0$ au segment $v_{i-1}v_{i+1}$ du polygone $\mathcal{P}_k$. Cette distance est calculée pour tous les pts u_j de la courbe d'origine qui sont situés entre les sommets v_{i-1} et v_{i+1} du polygone $\mathcal{P}_k$ (voir fig. 1).

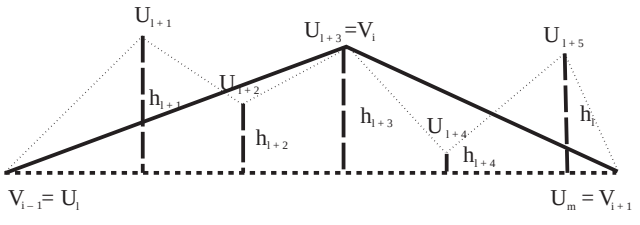


Figure 1. Exemple d'évaluation du critère E : $E(V_i, \mathcal{P}_k) = h_{l+1}$ avec $\mathcal{P}_k = \{V_{i-1}, V_i, V_{i+1}\} \subset \mathcal{P}_0 = \{U_l, \ldots, U_m\}$

[2] évolution d'une approximation grossière vers une approximation plus fine de la forme.

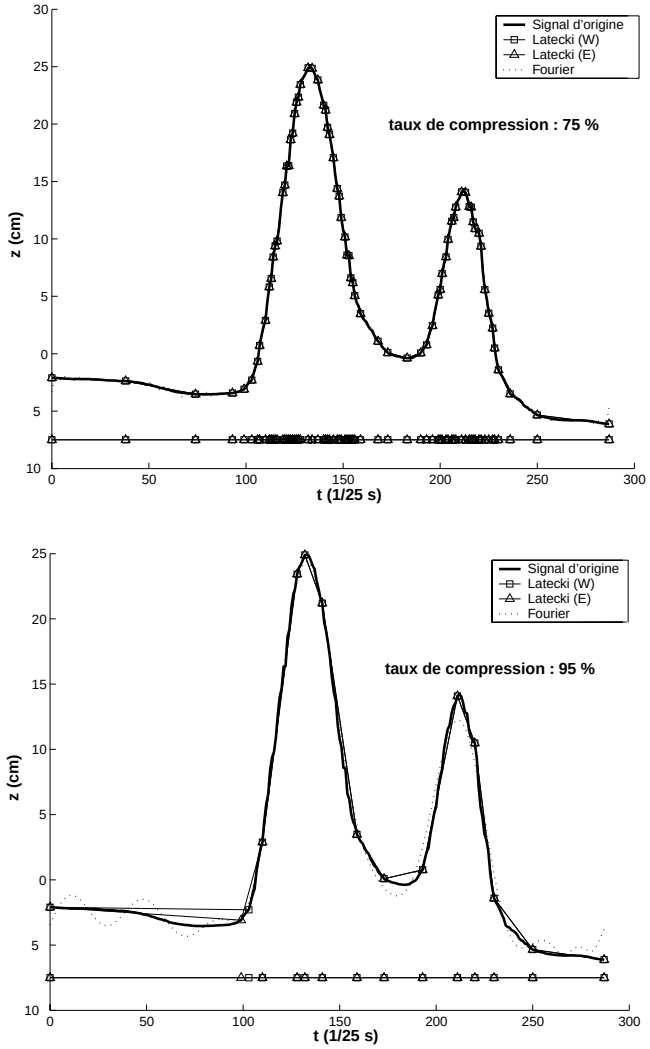


Figure 2. Approximation d'un signal 1D lisse

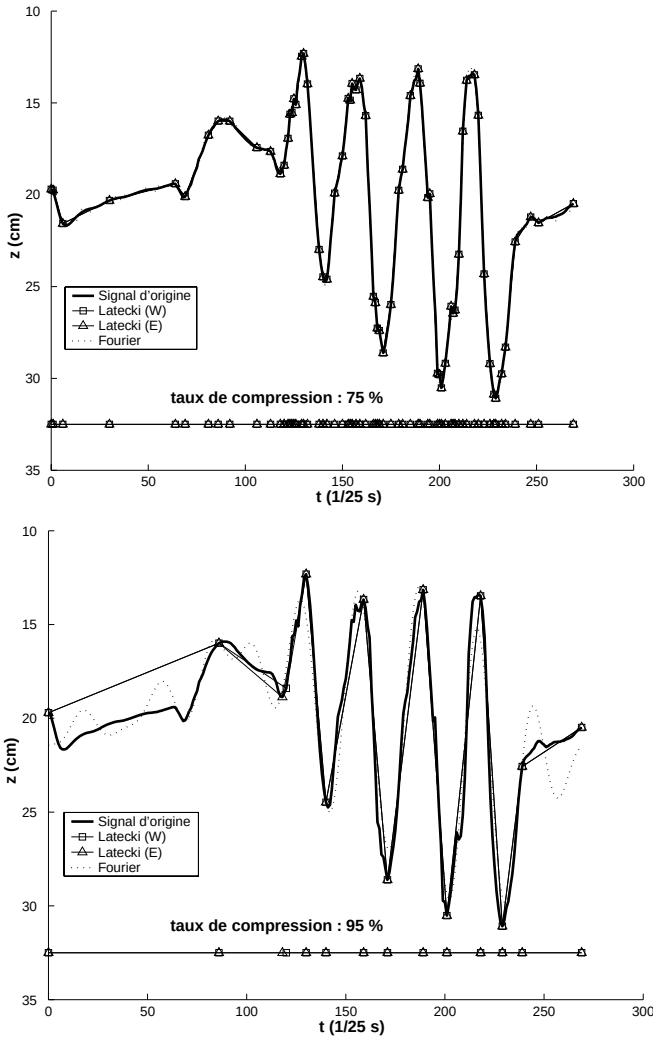


Figure 3. Approximation d'un signal 1D complexe

L'algorithme d'évolution d'un polygone $\mathcal{P}_0$ est le suivant :

ALGORITHME. (Procédure d'approximation de courbes $(\mathcal{P}_0)$)
1. $k = 0$;
2. **Tant que** $\min_{v \in \mathcal{P}_k} E(v, \mathcal{P}_k) \leq \epsilon$
3. // ou tant que $\dfrac{\text{card}(\mathcal{P}_0) - \text{card}(\mathcal{P}_k)}{\text{card}(\mathcal{P}_0)} \leq \tau$
4. **Faire**
5. $\mathcal{V}^*(\mathcal{P}_k) = \arg\min_{v \in \mathcal{P}_k} E(v, \mathcal{P}_k)$;
6. $\mathcal{P}_{k+1} = \mathcal{P}_k \backslash \mathcal{V}^*(\mathcal{P}_k)$;
7. $// \Leftrightarrow \mathcal{P}_{k+1} = \{v : v \in \mathcal{P}_k \text{ et } v \notin \mathcal{V}^*(\mathcal{P}_k)\}$;
8. $k = k + 1$;
9. **Fait**

Selon la tolérance ϵ ou le taux de compression τ désiré, plus ou moins de détails de la courbe d'origine sont préservés dans la forme simplifiée engendrée.

Cet algorithme s'avère efficace pour le sous-échantillonnage non uniforme de signaux qu'ils soient lisses (*cf* fig. 2) ou complexes (*cf* fig. 3). La figure 2 montre une courbe décrivant l'évolution temporelle de la position de la tête le long de l'axe Z durant un mouvement dansé (« enroulé déroulé tête »). La courbe approximative correspondante définie par seulement 25 % et 5 % de points représentatifs, en utilisant respectivement l'approximation polygonale de courbes (avec les énergies W et E) et les descripteurs de Fourier, est affichée sur la même figure. Le sous-échantillonnage par approximation polygonale du signal est mis en évidence sur l'axe des temps. Comme prévu, les nœuds sont plus nombreux dans les régions de forte courbure. La figure 3 montre une courbe décrivant l'évolution temporelle de la position de la main gauche le long de l'axe Z durant une marche avant et après compression.

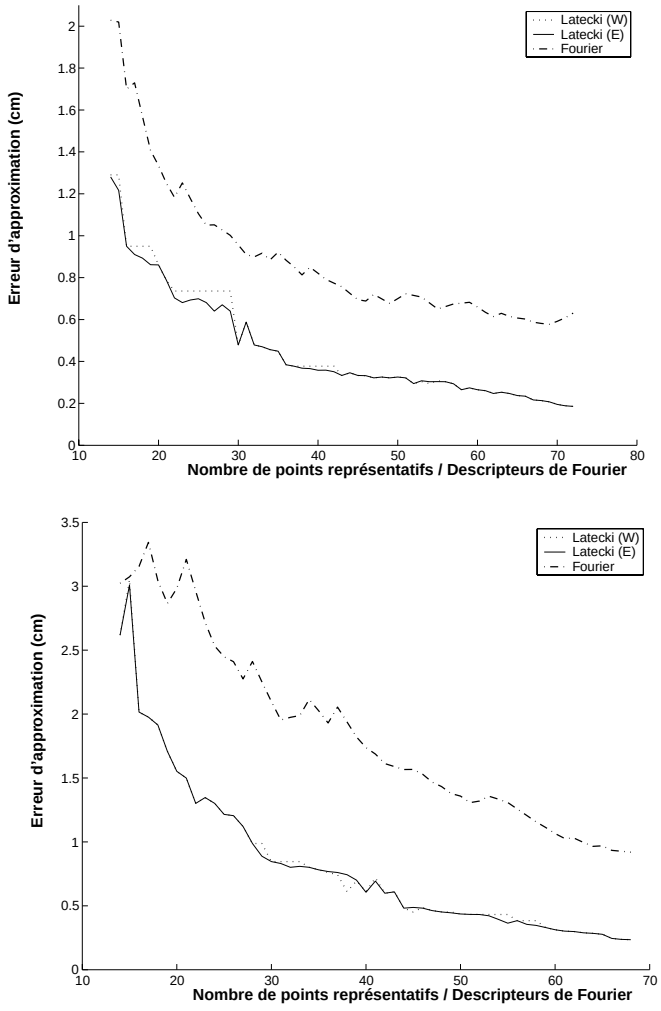


Figure 4. Erreur d'approximation versus la complexité des signaux 1D lisse (en haut) et complexe (en bas)

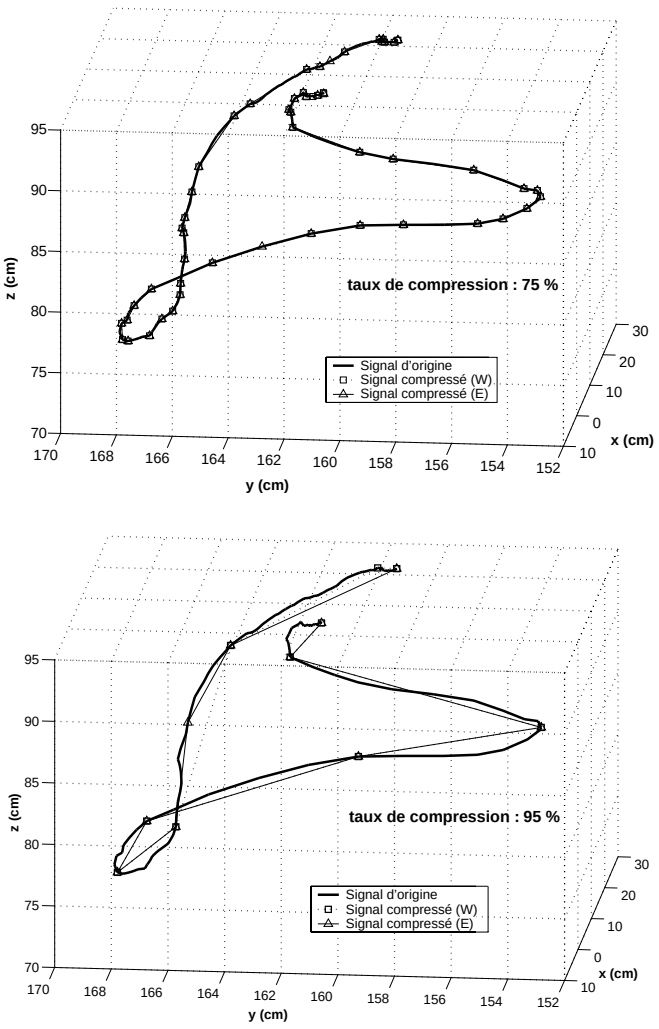


Figure 5. Approximation d'un signal 3D

Nous retenons ici les descripteurs de Fourier de plus haute énergie qui s'avèrent bien plus efficaces pour la compression de signaux que la conservation systématique des fréquences les plus basses [5].

La qualité de compression à 75 % des deux signaux par approximation polygonale est légèrement meilleure que par descripteurs de Fourier mais la différence de précision est significative à un taux plus conséquent (95 %). On peut noter aussi que l'utilisation du critère E que nous proposons permet une meilleure préservation des pics du signal que le critère *W* (voir 2ème point représentatif sur fig. 2 et 3ème point caractéristique sur fig. 3 à 95 % de compression).

La figure 4 montre la précision de l'algorithme d'approximation de courbes en fonction de sa complexité (nombre de points représentatifs/descripteurs de Fourier) et ce, en gardant 5 à 25 % de points du signal. Le critère utilisé est la déviation maximale par rapport à la courbe d'origine. Ces résultats montrent que l'algorithme d'approximation polygonale de courbes permet une bien meilleure qualité d'approximation, pour un taux de compression donné, que les descripteurs de Fourier.

La figure 5 montre des données spatio-temporelles décrivant la position 3D de la tête durant un enveloppé. Les projections de ce signal sur les plans XY et YZ sont montrées sur la figure 6. On peut noter encore une fois l'amélioration apportée par l'utilisation de la nouvelle énergie (E) par rapport à l'originale (W) (voir 2ème segment de courbe gauche à 95 % de compression). L'erreur d'approximation atteinte en retenant 5 à 25 % de points représentatifs du signal est plus faible en utilisant E que W (voir fig. 7). La différence de précision entre les deux critères est significative pour un sous-échantillonnage en dessous de 15 % comme le montre la figure 7. Bien sûr, le taux de sous-échantillonnage adéquat dépend de la géométrie intrinsèque des données. Une courbe lisse autorise évidemment un taux plus bas qu'une courbe complexe.

3. Reconnaissance de mouvements de danse

Notre objectif est d'aboutir à un système de reconnaissance supervisé capable de différencier différents gestes de danse contemporaine définis par un ensemble de trajectoires de mouvement. Parmi toutes les activités humaines, la danse contemporaine est certainement l'une des plus complexes tant les mouvements sous-jacents sont libres et seulement contraints par les limites physiques du corps [13]. Tous les degrés de liberté du corps sont autorisés.

3.1 Travaux relatifs

L'analyse de signaux spatio-temporels passe par des techniques basées sur les principales approches suivantes : la mise en correspondance dynamique (DTW - Dynamic Time Warping) [3, 16], les approches neuronales [31, 30] et les Modèles de Markov Cachés (MMC - Hidden Markov Models, HMM) [24, 26, 7, 29, 32, 23, 9, 8].

Le DTW a été appliqué à la reconnaissance de gestes après avoir fait ses preuves en matière de reconnaissance vocale dynamique. C'est une technique de mise en correspondance dynamique basée sur des modèles connus. Le but est d'aligner le modèle et le signal à reconnaître par déformation non-linéaire de telle sorte que la différence entre les deux signaux devienne minimale [3, 16]. Cette méthode est efficace lorsque le nombre de phénomènes à reconnaître est faible (les variations temporelles en rythme et durée sont prises en compte). Cependant, elle nécessite un grand nombre de modèles lorsque l'on souhaite élargir le champ de reconnaissance. De plus, elle ne reconnaît pas les modèles non définis. Enfin, cette technique pêche par sa lourdeur, même si des améliorations ont été apportées pour accélérer les calculs [16].

Les approches neuronales sont particulièrement adaptées à la classification de modèles dans les ensembles de données [31, 30]. Le principal problème de ces techniques reste la prise en compte du temps. Dans une approche, le temps est vu comme un mécanisme externe. Dans ce cas, le TDNN (Time Delay Neural Network – réseaux de neurones à lignes de retard) est l'un des modèles les plus efficaces. Inspiré du réseau de neurones multi-couches, le TDNN transforme le problème temporel en un problème spatial. Il a montré son efficacité en reconnaissance de gestes en langage des signes [31], mais les systèmes basés sur ce principe ou leurs variantes restent d'une trop grande complexité et demandent une étape de pré-traitement des données trop lourde. Dans une autre approche, le temps est intégré en tant que mécanisme interne au système [28]. Dans tous les cas, le principal inconvénient reste le coût de calcul prohibitif et la complexité de l'apprentissage ainsi que la difficile interprétation des résultats.

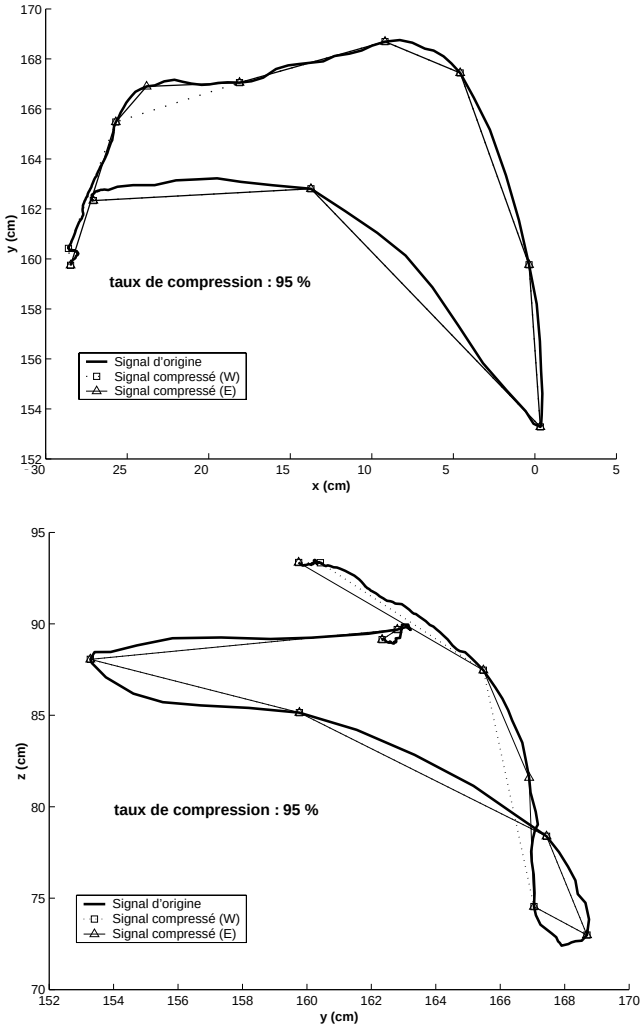


Figure 6. Approximation du signal 3D : projection sur le plan XY (en haut) et YZ (en bas)

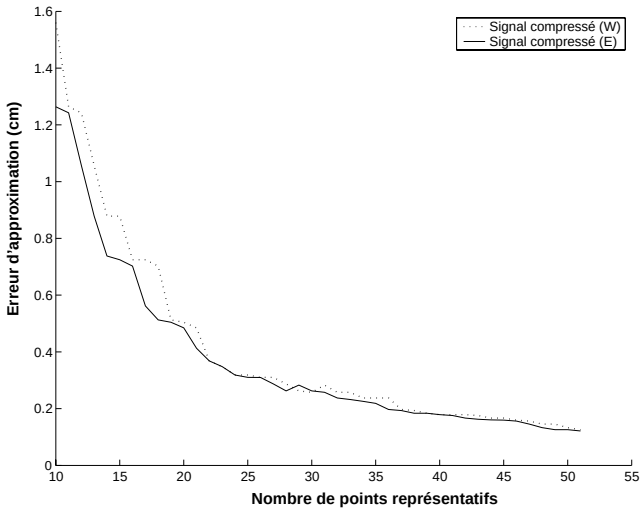


Figure 7. Erreur d'approximation versus la complexité du signal 3D

Les modèles de Markov cachés sont aussi très appliqués dans l'analyse de séries temporelles contenant des variations spatiales et temporelles [24]. Ils ont rencontré beaucoup de succès face aux problèmes de reconnaissance de l'écriture, de la voix et, plus récemment, des gestes [6, 7, 32, 23, 9, 8] en particulier les gestes de la langue des signes [26, 29]. Particulièrement adaptés aux signaux spatio-temporels [24], les modèles de Markov cachés semblent fournir le meilleur compromis en termes de flexibilité, de rapidité et de robustesse aussi bien vis à vis du bruit que des variations temporelles des mouvements. De plus, ils sont moins opaques et moins difficilement interprétables que les réseaux de neurones.

3.2 Reconnaissance de gestes de danse par MMC

Notre système de reconnaissance de gestes de danse regroupe un ensemble de 15 MMC discrets, chacun d'entre eux étant associé à une trajectoire de mouvement suivie par un point d'articulation du corps du danseur et mesurée par un système de capture de mouvement. Nous ne décrirons pas ici les concepts de base des MMC désormais bien connus. Une explication détaillée de la théorie des MMC est disponible dans [24].

Modélisation de trajectoires de mouvement par MMC

Il existe plusieurs travaux dans la littérature qui utilisent les MMC pour la reconnaissance de trajectoires de mouvement 2D [32, 23] ou 3D [6, 7, 29, 8] ; brutes [23] ou définies dans un espace de représentation plus élaboré [7]. Les trajectoires de mouvement sont en général représentées par trois caractéristiques principales : position, orientation et vitesse. L'orientation est le paramètre de loin le plus efficace pour la reconnaissance de trajectoires comme le montrent les expérimentations conduites dans [32]. C'est celui que nous utilisons dans notre cas (orientation locale entre deux points successifs plus précisément) alors que seule la vitesse (polaire et angulaire) est utilisée dans [7] tandis que Vogler et Metaxas utilisent la position et la vitesse [29].

Topologie du modèle

Nous modélisons un mouvement par un ensemble de 15 MMC d'ordre 1, autrement dit, toutes les transitions vers un autre état ne dépendent que de l'état courant. Ces MMC, de structure identique, sont construits selon un modèle gauche-droite complet avec des transitions réflexives ou vers chacun de tous les états suivants [24]. Il existe d'autres topologies de MMC tel que le modèle ergodique ou complètement connecté, chaque état étant atteignable depuis tous les autres états en un nombre fini de transitions. Cependant, la topologie gauche-droite semble la plus adéquate pour modéliser des séries temporelles de vitesse et de durée variables [24]. Nos choix de modèle, aussi bien dans l'ordre que dans la structure, rejoignent les choix de la majorité des travaux utilisant les MMC pour la reconnaissance de gestes [7, 32, 29]. Soulignons cependant que nous utilisons une structure plus générale puisque tous ces travaux utilisent le modèle de Bakis (transitions réflexives ou vers chacun des deux états suivants) alors que nous utilisons un modèle complet.

Paramètres du modèle

Notre modèle MMC, désigné par $\lambda = (A, B, \pi)$, est décrit par l'ensemble des paramètres statistiques suivants :

– N, le nombre d'états, est fixé expérimentalement à 20. En fait, nous avons expérimenté l'algorithme de détermination automatique du nombre d'états de Biernacki, Celeux et Govaert [10]. Cependant, le gain en taux de reconnaissance obtenu par rapport au cas d'un nombre d'états homogène égal à 20 est faible : gain de 1 % à 95 % de compression (7 % à 0 % de compression mais ce gain est sans intérêt car il reste bien en dessous du taux obtenu à 95 % de compression, voir fig. 8), et ne justifie pas le coût de calcul induit qui s'avère prohibitif même pour une dynamique de 10 valeurs.

- $S = \{(0,0),...,(0,3),...,(7,0),...,(7,3)\}$ est l'ensemble des symboles observables de notre représentation. En effet, nous représentons l'orientation locale des trajectoires 3D par 2 angles $(\Psi \bmod 2\pi, \beta \bmod \pi)$ dans l'espace 3D (coordonnées sphériques). En utilisant le codage de Freeman pour discrétiser l'espace de variation de chacun de ces angles, nous obtenons un ensemble de 32 symboles (8×4 directions).

- M, le nombre de symboles, vaut 32 selon notre choix de représentation.

- Les matrices des probabilités de transition A et d'observation B sont initialisées de manière aléatoire par un processus itératif.

- Le vecteur des probabilités des états initiaux π est arbitrairement initialisé pour que le premier état soit l'état s_1 en accord avec le modèle « gauche-droite » que nous utilisons.

Après initialisation, les paramètres des MMC sont affinés par l'algorithme de Baum-Welch, un algorithme bien connu d'apprentissage supervisé [2]. Ainsi, chaque MMC résultant représente un modèle de séquence. L'algorithme de recherche de Viterbi [14] nous permet ensuite de comparer toute nouvelle séquence aux séquences apprises et de reconnaître la plus proche.

Résultats expérimentaux

Tout d'abord, nous avons construit une base de gestes dansés en utilisant le studio de capture du mouvement d'XD-Productions avec l'aide de quatre danseurs professionnels du BARC. Nous avons choisi une liste de 12 mouvements fondamentaux du ballet contemporain (voir fig. 8), mouvements variés dans la forme, dans la durée et dans le rythme, interprétés par des danseurs de corpulence différente et enregistrés à une fréquence de 25 Hz depuis des points de vue différents.Ces gestes font partie d'un langage général de gestes de danse caractérisé par plusieurs familles fondamentales telles que les marches, les sauts, les chutes, les tours, etc. Le nombre total de gestes possibles est,

contrairement au ballet classique, indéterminé puisque les chorégraphes contemporains sont en perpétuelle recherche de nouvelles expressions corporelles.
Il en résulte une base de données de 360 observations dont 60 % ont été utilisées pour la base d'apprentissage et le reste pour la base de test.
La figure 8 montre les résultats de reconnaissance obtenus sur la base de test avec différents taux de compression des trajectoires de mouvement. 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. Cet algorithme est une version améliorée de l'algorithme de Latecki et Lakämper [17]. De plus, la réduction de caractéristiques qui en découle allège le processus d'apprentissage et permet une modélisation plus fiable du comportement spatio-temporel des marqueurs. La figure 9 montre l'évolution du taux de reconnaissance sur la base de test en fonction de la qualité d'approximation des trajectoires de mouvement définie ici par la déviation maximale par rapport à la courbe d'origine. On peut

Mouvement	Reconnaissance sur la base de test									
Taux de compression	0%		50%		75%		90%		95%	
	oui	non	oui	non	oui	non	oui	non	oui	non
marche	11	1	11	1	12	0	12	0	12	0
enveloppé	10	2	10	2	11	1	12	0	12	0
chute pliée	7	5	9	3	11	1	12	0	12	0
enroulé tête	8	4	9	3	11	1	11	1	11	1
enroulé bassin	7	5	8	4	10	2	10	2	10	2
grand plié	9	3	10	2	10	2	11	1	11	1
marche glissée	8	4	9	3	10	2	11	1	11	1
saut attitude	7	5	10	2	10	2	10	2	10	2
chute repliée	7	5	8	4	9	3	10	2	11	1
chute allongée	8	4	9	3	9	3	9	3	9	3
jeté	9	3	10	2	10	2	11	1	11	1
tour	10	2	11	1	11	1	12	0	12	0
TOTAL	**101**	**43**	**114**	**30**	**124**	**20**	**131**	**13**	**132**	**12**
TAUX	**70,1**	**29,9**	**79,2**	**20,8**	**86,1**	**13,9**	**91,0**	**9,0**	**91,7**	**8,3**

Figure 8. Résultats de reconnaissance de gestes dansés

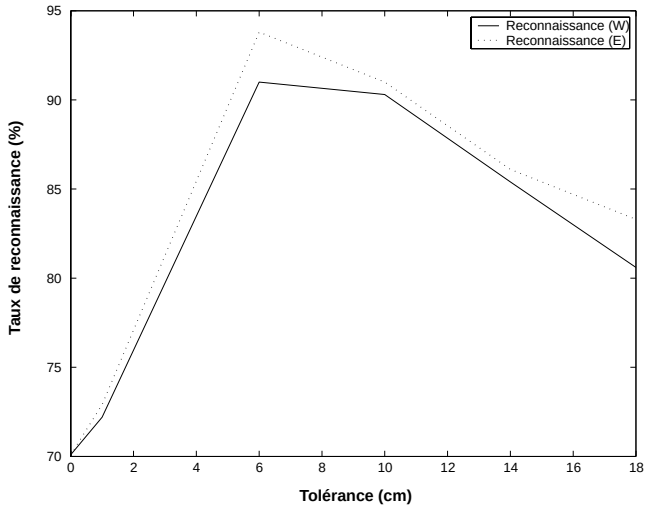


Figure 9. Évolution du taux de reconnaissance en fonction de la qualité d'approximation des trajectoires de mouvement

souligner l'apport significatif de notre version modifiée de l'algorithme de Latecki et Lakämper (nouvelle énergie E *versus* ancienne W) sur le taux de reconnaissance. Le taux de reconnaissance maximal obtenu est à 94 % en utilisant E et à 91 % avec W pour une tolérance de 6 cm. Bien évidemment, plus la tolérance est importante, plus le taux de compression est élevé et inversement.
Enfin, il est évident, comme le montre la figure 9, qu'une perte importante d'informations (sur-compression), tout comme le surplus d'informations, dégrade aussi les performances de la tâche de reconnaissance.

4. Conclusion

Nous avons présenté un prototype complet et opérationnel intégrant le sous-échantillonnage non-uniforme et la reconnaissance automatique des trajectoires de mouvement. Nous avons montré que des trajectoires polygonalisées engendrent une description compacte et efficace des gestes dansés. Cela facilite grandement la tâche délicate de reconnaissance. L'amélioration du taux de reconnaissance par rapport à une utilisation brute des signaux spatio-temporels décrivant les gestes dansés est supérieure à 20 %.
L'une des perspectives à court terme de ce travail est de regrouper le mouvement de certains marqueurs dans des macro-chaînes. Nous souhaitons aussi caractériser le taux de compression approprié d'un signal en fonction de sa complexité.
À plus long terme, le problème majeur à résoudre reste l'extraction robuste des trajectoires 3D directement des données vidéo.

Références

[1] J.K. Aggarwal et Q. Cai, « Human motion analysis: a review », *Computer Vision and Image Understanding*, 73(3):428-440, mars 1999.

[2] L.E. Baum et J. Eagon, « An inequality with applications to statistical prediction for functions of Markov processes and to a model of ecology », *Bull. Amer. Math. Soc.*, 73 : pp. 360-363, 1967.

[3] D. Berndt et J. Clifford, « Using dynamic time warping to find patterns in time series », *Workshop on Knowledge Discovery in Databases*, pp. 359-370, juillet 1994.

[4] M. Berthold, D.J. Hand, « Intelligent data analysis, an introduction », 1st edition, Springer-Verlag, New York, 1999.

[5] S. Boukir, E. Beets et F. Chenevière, « Représentation et compression de signaux spatio-temporels », *Rapport de recherche RT-2002-09-001, Laboratoire L3i, Université de La Rochelle*, Septembre 2002.

[6] L. Campbell et A. Bobick, « Recognition of human body motion using phase space constraints », *ICCV 95, 5th Int. Conf. on Computer Vision, Cambridge MA*, pp. 624-630, 1995.

[7] L.W. Campbell *et al.*, « Invariant features for 3-D gesture recognition », *Int. Conf. on Automatic Face and Gesture Recognition*, pp. 157-162, 1996.

[8] F. Chenevière, S. Boukir et E. Beets, « Compression et reconnaissance de séquences spatio-temporelles issues d'un ballet contemporain », *ORASIS'2003, 9ème Congrès Francophone en Vision par Ordinateur, Gérardmer*, pp. 353-361, mai 2003.

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

[10] J.B. Durand, « Modèles à structure cachée : inférence, sélection de modèles et applications », *thèse de doctorat, Université Joseph Fourier, Grenoble*, janvier 2003.

[11] R.C. Gonzalez et R.E. Woods, « Digital image processing », *2nd edition, Prentice Hall, Upper Saddle River*, 2002.

[12] D. Hall, C. Le Gal, J. Martin, O. Chomat et J.L. Crowley, « Magicboard: A contribution to an intelligent office environment », *Robotics and Autonomous Systems*, 35(3-4): 211-220, juin 2000.

[13] D. Herbisson-Evans, « Dance and the computer : A potential for graphic synergy », *Technical Report 422, Basser Department of Computer Science*, University of Sydney, janvier 1991.

[14] X.D. Huang, Y. Ariki et M.A. Jack, « Hidden Markov Models for speech recognition », *1st edition, Edinburgh University Press, Edinburgh*, 1990.

[15] H. Kauppinen, T. Seppanen et M. Pietikainen, « An experimental comparison of autoregressive and Fourier-based descriptors in 2D shape classification », *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 17(2): pp.201-207, 1995.

[16] E.J. Keogh et M. Pazzani, « Scaling up dynamic time warping to massive datasets », *Proceedings of the 3rd European Conference on Principles of Data Mining and Knowledge Discovery (PKDD), Prague, Czech Republic*, pp. 1-11, septembre 1999.

[17] L.J. Latecki et R. Lakamper, « Convexity rule for shape decomposition based on discrete contour evolution », *Computer Vision and Image Understanding (CVIU)*, 73(3): pp. 441-454, mars 1999.

[18] T. C. Le Buhan Jordan et T.C. Ebrahimi, « Progressive polygon encoding of shape contours », *Proc. of the 6th Int. Conference on Image Processing and its Applications (IPA'97), Dublin, Ireland*, 1: pp.17-21, juillet 1997.

[19] M. Daehlen, T. Lyche et L. Schumaker, « Mathematical methods for curves and surfaces », *1st edition, Vanderbilt University Press, London*, 1995.

[20] K. Muller et J.R. Ohm, « Wavelet-based contour descriptor », *Technical report*, MPEG-7 proposal nb. P567, février 1999.

[21] D. Ormoneit *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*, pp. 894-900, 2001.

[22] B. Pinkowski, « Fourier descriptors for characterizing object contour », *ICSPAT 1996, Int. Conf. on Signal Processing Applications and Technology, Boston*, pp. 1007-1011, 1996.

[23] A. Psarrou, S. Gong et M. Walter, « Recognition of human gestures and behaviour based on motion trajectories », *Image and Vision Computing*, 20: pp. 349-358, 2002.

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

[25] E. Saux et M. Daniel, « Data reduction of polygonal curves using B-splines », *Computer-aided design*, (31): pp.507-515, 1999.

[26] T. Starner et A. Pentland, « Visual recognition of american sign language using hidden markov models », *Int. Workshop on Automatic Face and Gesture Recognition, Zurich, Switzerland*, pp. 189-194, 1995.

[27] S. Sudarsky et D. House, « Motion capture data manipulation and reuse via B-splines », *CAPTECH 98, Int. Workshop on Modeling and Motion Capture Techniques for Virtual Environments*, pp. 55-69, 1998.

[28] N. Szilas, « Les réseaux récurrents supervisés : une revue critique », *Rapport de recherche 972-I, Institut National Polytechnique de Grenoble*, mars 1997.

[29] C. Vogler et D. Metaxas, « A framework for recognizing the simultaneous aspects of american sign language », *Computer Vision and Image Understanding*, 81(3): pp. 358-384, mars 2001.

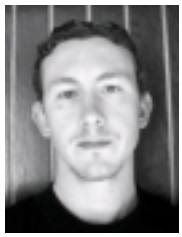
[30] J. Weismann et R. Saloman, « Gesture recognition for virtual reality applications using data glove and neural networks », *IEEE Int. Joint Conf. on Neural Networks, Washington DC*, (3): pp. 2043-2046, 1999.

[31] M.H. Yang et N. Ahuja, « Recognizing hand gesture using motion trajectories », *CVPR 99, IEEE Conf. on Computer Vision and Pattern Recognition, Ft. Collins*, CO, pp. 466-472, Juin 1999.

[32] H.S Yoon, J. Soh, Y.J. Bae et H.S. Yang, « Hand gesture recognition using combined features of location, angle and velocity », *Pattern Recognition*, (34): pp.1491-1501, 2001.

Samia **Boukir**

Samia Boukir a obtenu le DEA de Robotique de Paris 6 en 1990 et le Doctorat en Vision-Robotique de l'INRIA de Rennes en 1993. Elle est membre du laboratoire L3i depuis 1993 et maître de conférences en Informatique à l'université de La Rochelle depuis 1994. Ses activités de recherche concernent principalement le traitement et l'analyse de séquences spatio-temporelles (images et signaux). Actuellement, elle travaille sur la représentation, la compression et la classification de signaux spatio-temporels dans le cadre du projet *Corps dansant* mené en collaboration avec des artistes d'un ballet contemporain.



Frédéric **Chenevière**

Frédéric Chenevière est doctorant au L3i (Laboratoire Informatique - Image - Interaction) à l'Université de La Rochelle depuis octobre 2000. En collaboration avec Samia Boukir et Bertrand Vachon, il a participé à la création et à l'animation du projet *Corps Dansant* visant l'étude des mouvements de danse contemporaine. Ses centres d'intérêts vont de l'acquisition de mouvements humains à leur reconnaissance automatique, en passant par les pré-traitements nécessaires, notamment la compression de données spatio-temporelles.