

# Indexation et recherche d'actions humaines 3D basées sur l'analyse des courbes surfaciques

R. Slama<sup>1</sup>, H. Wannous<sup>1</sup> et M. Daoudi<sup>1,2</sup>

<sup>1</sup> LIFL Laboratoire fondamentale d'informatique de Lille/ UMR CNRS 8022, Villeneuve d'Ascq, France

<sup>2</sup>Institut Mines-Telecom / Telecom Lille1, France

{rim.slama,hazem.wanous,mohamed.daoudi}@telecom-lille1.eu

## Résumé

*Dans ce travail, nous proposons une nouvelle approche pour la recherche par le contenu de mouvement dans des bases de vidéos 3D. L'action humaine est représentée dans un cadre géométrique par une séquence temporelle de poses, chacune caractérisée par un descripteur extrait sur la surface du corps humain associée. Ce descripteur, nommé EHC (Extremal Human Curves), est représenté par une collection de courbes ouvertes décrivant la forme du corps humain dans une pose donnée. Chacune de ces courbes peut être considérée comme un point de l'espace des formes de courbes ouvertes et, par conséquent, chaque action est modélisée par une trajectoire dans cet espace. Enfin, pour comparer deux séquences représentées par leurs trajectoires nous utilisons la déformation temporelle dynamique (DTW) afin de les aligner en utilisant la métrique élastique sur l'espace des formes. Notre approche est évaluée sur la base i3dPost qui contient différentes actions réalisées par différents acteurs.*

## Mots clefs

Variété Reimannienne, vidéo 3D d'actions humaines, courbes ouvertes, segmentation de l'action, DTW.

## 1 Introduction

Les séquences de vidéo 3D du mouvement humain sont de plus en plus disponibles. En effet, leur acquisition avec des systèmes de reconstruction multivue ou par animation et des approches de synthèse a reçu un intérêt considérable au cours de la dernière décennie. La plupart des recherches sur la vidéo 3D se concentrent principalement sur l'amélioration de la qualité d'acquisition et la performance des méthodes de compression de données. Par conséquent, les vidéos 3D sont principalement utilisées pour la visualisation. Toutefois, l'acquisition de longues séquences produit d'énormes quantités de données qui rendent la navigation et la recherche de l'information dans ces bases de données une tâche difficile. Il est donc nécessaire de développer un système de segmentation et de recherche par contenu efficace pour gérer les bases de vidéo 3D et qui permet d'effectuer une recherche de l'information rapidement et efficacement.

Dans la littérature, certaines approches représentent le maillage du corps humain par un histogramme [1], [2], [3], ensuite proposent d'apparier les mouvements à l'aide de la programmation dynamique, utilisant une distance euclidienne entre histogrammes pour comparer les poses correspondantes. Dans le même contexte [4], propose de représenter la pose non pas par un histogramme mais plutôt par un squelette de Reeb graph et de modéliser l'action par un modèle de Markov caché HMM.

D'autres travaux ont tendance à cumuler des descripteurs de forme statique pour former un modèle de l'action ou encore former des descripteurs temporels pour décrire l'évolution et le changement de la pose au cours de la séquence [5], [6], [7]. Parmi ces représentations, on trouve : l'historique du mouvement de volume (MHV), le flux optique 3D, le modèle du corps ellipsoïde cylindre.

Ce travail a pour objectif de développer une approche de recherche de mouvement similaire dans des vidéo 3D. Notre approche peut être décomposée en trois étapes clés : (1) l'extraction de descripteur du maillage 3D, (2) la segmentation de mouvement (3) la comparaison des actions.

Chaque maillage 3D sera représenté par un descripteur, dit "Extremal Human Curve" (EHC), efficace dans sa représentation et discriminant lors des comparaisons des poses [8]. Ce descripteur est un ensemble de courbes surfacique 3D ouvertes dont la comparaison sera effectuée dans un espace de forme Riemannien utilisant une distance élastique. La segmentation et la comparaison des séquences de l'action se fera, par la suite, à l'aide de ce descripteur. D'abord, la segmentation du mouvement joue un rôle important dans la recherche par le contenu en divisant les données de vidéo 3D en segments significatifs qui décrivent les mouvements de base, appelés clips. Ces clips seront utilisés par la suite comme unité de base pour la recherche d'un mouvement donné. Nous avons développé une technique de segmentation en analysant le vecteur vitesse de la séquence extrait à partir des données fournies par le descripteur EHC.

Afin d'apparier deux actions, nous basons notre comparaison de séquence d'actions sur la "DTW" entre chaque paire de clips. Cependant, puisqu'un clip peut être vu comme une séquence de EHC, il sera représenté par une trajectoire dans l'espace de formes. La "DTW" sera donc adapté

à cet usage en utilisant la distance spécifique à la comparaison des EHC, à savoir la distance géodésique sur la variété Riemannienne de l'espace de formes. Prenant avantage de cette représentation en trajectoire de chaque clip, on propose même de calculer des moyenne de clip sur la variété Riemannienne et d'effectuer une recherche hiérarchique dans la base d'actions.

Le plan de cet article est le suivant : L'extraction des courbes et la métrique élastique utilisée pour la comparaison des poses sont présentés dans la section 2. Dans la section 3, notre cadre de travail utilisé pour la segmentation de mouvement et la recherche par le contenu est présentée. Dans la section 4, l'évaluation de notre descripteur et les résultats expérimentaux pour la segmentation de vidéo et la recherche d'information sont effectuées. Enfin, nous concluons dans la section 5 en résumant nos résultats et discuter des futurs travaux.

## 2 Descripteur de pose

Nous proposons de représenter la pose du corps humain par une représentation à base de squelette surfacique 3D. Ce squelette correspond à une collection de courbes ouvertes 3D extraites sur la surface du maillage. Une courbe surfacique résulte de la liaison des points caractéristiques situées sur les extrémités du corps. L'idée principale derrière l'utilisation de cette représentation est d'estimer la variation de la pose en analysant le changement de formes de ces courbes surfaciques sur le corps, tenant en compte l'élasticité du corps.

### 2.1 Courbes surfaciques d'extrémité EHC

**Extraction des courbes d'extrémité** Afin d'extraire nos descripteurs surfacique, notre approche est basée sur la détection des 5 extrémité sur le modèle 3D du corps humain. Pour cette extraction, nous avons choisi l'approche proposée par [9] qui est basée sur la distance géodésique. Les expérimentations montrent que les extrémités trouvées par cette approche sont stables et invariants aux transformations géométriques et changement de poses. Soit  $M$  une surface du corps, et  $E = \{e_1, e_2, e_3, e_4, e_5\}$  l'ensemble de points caractéristiques sur le corps qui représente la sortie de l'approche d'extraction d'extrémités. Soit  $\beta$  une courbe surfacique ouverte qui relie deux points caractéristiques  $\{e_i, e_j\}$ . Pour obtenir  $\beta$ , nous recherchons le chemin géodésique  $P_{ij}$  entre  $e_i$  et  $e_j$ . Nous répétons cette étape dix fois pour extraire toutes les courbes d'extrémité possibles de la surface du corps. La Figure 1 illustre un exemple de résultats d'extraction des extrémités sur le corps et de leurs courbes surfaciques (EHC) correspondantes.

Le corps sera donc représenté par cette collection de courbes extrémales qui donne l'information sur la pose faite par la personne. Nous avons choisi de représenter le corps humain par un ensemble de courbes surfaciques pour deux raisons. Tout d'abord, ces courbes relient les membres du corps et donnent une information locale sur la position spatiale des membres du corps, tout en uti-

lisant une représentation réduite du maillage. Deuxièmement, pour analyser la variation de pose, nous allons analyser la variation de forme de ces courbes surfaciques qui change avec le changement de pose. Toutefois, afin de comparer les courbes correspondantes nous avons besoin d'une distance afin d'évaluer combien leur forme est similaire.

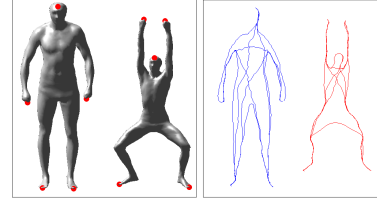


Figure 1 – Points d'extrémité extraits sur le corps humain et les courbes d'extrémité correspondantes.

### Analyse Riemannienne des courbes ouvertes dans $\mathbb{R}^3$

Soit  $\beta : I \rightarrow \mathbb{R}^3$ , pour  $I = [0, 1]$ , une courbe d'extrémité sur le maillage du corps humain. Afin d'étudier la forme de  $\beta$ , nous représentons la courbe par la *square-root velocity function* (SRVF), notée  $q(t) \doteq \dot{\beta}(t) / \sqrt{\|\dot{\beta}(t)\|}$ .  $q(t)$  est une fonction assez particulière introduite par Joshi et al.[10] qui capture la forme de la courbe  $\beta$  tout en offrant des facilités de calcul. En effet, dans l'espace de forme, noté  $\mathcal{C}$ , la norme  $\mathbb{L}^2$  représente la métrique élastique de comparaison de courbes [10].

$$\mathcal{C} = \{q : I \rightarrow \mathbb{R}^3 \mid \|q\| = 1\} \subset \mathbb{L}^2(I, \mathbb{R}^3) \quad (1)$$

Avec la norme  $\mathbb{L}^2$  sur son espace tangent, cette variété  $\mathcal{C}$ , devient alors une variété Riemannienne. On définit ainsi la distance entre  $q_1$  et  $q_2$  par la longueur de ce chemin géodésique dans  $\mathcal{C}$ .  $\mathcal{C}$  étant l'hypersphère de l'espace de Hilbert, le chemin géodésique ainsi que sa longueur peuvent être calculés analytiquement. La distance géodésique entre deux points quelconques  $q_1$  et  $q_2 \in \mathcal{C}$  est donnée par :

$$d_c(q_1, q_2) = \cos^{-1}(\langle q_1, q_2 \rangle), \quad (2)$$

### 2.2 Comparaison des poses

Étant donné deux maillages 3D  $x, y$  et leurs descripteurs  $x' = \{q_1^x, q_2^x, q_3^x, \dots, q_N^x\}$  et  $y' = \{q_1^y, q_2^y, q_3^y, \dots, q_N^y\}$ , la similarité entre deux maillages peut être représentée par la moyenne des distances comme suit :

$$s(x, y) = d(x', y') = \frac{\sum_{i=1}^N d_s(q_i^x, q_i^y)}{N}. \quad (3)$$

où  $N$  est le nombre de courbes utilisées pour décrire la grille. En cas de changement de forme dans une courbe donnée, la distance globale sera affectée et augmentera indiquant que les poses sont différentes. Dans la Figure 2, un chemin géodésique entre chacune des deux courbes extrémales correspondantes est représenté. L'évolution de la

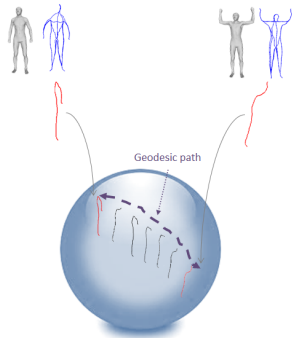


Figure 2 – Chemin géodésique entre les courbes d'EHC, avec une pose neutre contre une pose avec un sujet qui lève les deux mains.

courbe, qui relie la main au pied, entre une pose avec la main baissée et l'autre levée paraît naturelle. En effet, la géodésique illustre la déformation que subit la courbe pour passer d'une forme à l'autre.

### 3 Segmentation du mouvement et comparaison d'action 3D

En se basant sur notre représentation en EHC du maillage du corps humain dans une pose donnée, il est possible de comparer deux séquences de vidéo 3D en comparant chaque EHC correspondant. Cependant, une séquence d'action humaine peut être composée par plusieurs actions distinctes, et chacune peut être répétée plusieurs fois. Par conséquent, la segmentation de mouvement peut jouer un rôle important dans le processus de recherche d'information dynamique en divisant l'ensemble des données de vidéos 3D en petites actions élémentaires significatives appelées clips.

#### 3.1 Segmentation du mouvement

L'objectif de la segmentation de mouvement est de diviser la séquence continue en sous séquences de mouvements de base, appelés clips. Pour une segmentation significative, la vitesse du mouvement est le facteur qui doit être pris en compte. En effet, lorsque l'humain change le type ou la direction du mouvement, sa vitesse de déplacement diminue et cela se traduit par une chute dans le vecteur de vitesse. Afin de détecter cette baisse, on cherche les minima locaux pour le changement de type de mouvement et des maxima locaux pour le changement de direction. Les extrema détectés sur la courbe de vitesse doivent être sélectionnés en tant que points saillants délimitant le début et la fin de chaque clip. La Figure 3 illustre un vecteur de degré de mouvement pour l'action " marcher ". On peut constater que les trames détectées comme maxima correspondent au moment où l'acteur change l'orientation du mouvement du pied (du bas vers le haut), et les trames détectées comme minima correspondent au moment où l'acteur répète le même mouvement à savoir le changement de

pas. Afin d'éviter les clips avec de légères variations et un petit nombre de trames, nous ne prenons en compte que le changement dans le type de mouvement comme clips significatifs.

Pour calculer la variation de vitesse, le degré de mouvement sera calculé grâce à la distance entre chaque deux EHC successifs dans la séquence. Les variations d'une séquence sont représentées dans le vecteur de vitesse et ensuite un lissage est appliqué afin d'obtenir le vecteur résultant de degrés de mouvement.

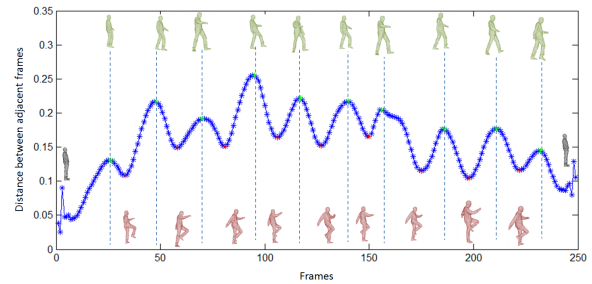


Figure 3 – Extrema détectés sur la courbe de degré de vitesse pour l'action "marcher".

#### 3.2 Appariement de clips

Dans cet article, les requêtes à base d'exemple sont employées. Un clip à partir d'une certaine vidéo 3D est donné comme requête et le but est de retrouver les clips similaires dans la base de données.

Deux clips sont similaires s'ils contiennent la même action élémentaire et ce même si : la forme de l'acteur, la vitesse, la taille de la séquence, ou la position spatiale de l'acteur change. Pour ce fait, chaque clip est représenté par une séquence temporelle de poses, caractérisé par leur représentation d'EHC associée à la forme du modèle. Chaque courbe d'EHC est représentée par un point sur la variété Riemannienne de courbes ouvertes, et donc une séquence de la même courbe d'un clip forme une trajectoire sur la variété Riemannienne. La Figure 4 illustre un exemple de trajectoire d'une séquence où une personne est en train de courir.

Enfin, les trajectoires de chaque courbe sont appariées pour donner une mesure de similarité. Toutefois, chaque personne met un temps différent pour faire la même action et donc deux trajectoires n'ont pas nécessairement la même longueur. Par conséquent, un alignement temporel de ces trajectoires est crucial avant le calcul de la mesure de similarité globale. Afin de résoudre ce problème de variation temporelle, nous utilisons l'algorithme DTW [11]. Cet algorithme utilise une approche de programmation dynamique, permettant de trouver le chemin optimal qui permet d'apparier au mieux les deux séries temporelles qui varient dans le temps comme illustré à la droite de la Figure 4.

Les coûts globaux cumulés le long du chemin optimal fournissent une distance globale entre les deux clips à com-

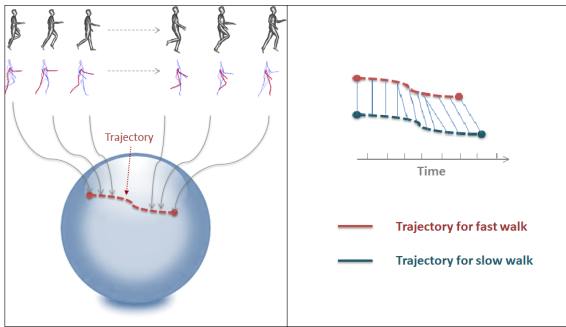


Figure 4 – Processus d’alignement des deux trajectoires qui modèlent deux actions.

parer. Puisque le DTW peut fonctionner avec n’importe quelle mesure de similarité, nous l’adaptions aux caractéristiques qui résident sur la variété riemanniennes. Par conséquent, nous utilisons la distance géodésique entre les différents points de l’espace de forme  $d_s(q_i, q_j)$  comme fonction de distance à différents instants.

## 4 Résultats expérimentaux

Pour valider notre approche, nous avons testé la robustesse de notre descripteur en utilisant deux bases. Une partie des expérimentation sur des données synthétiques issues de la base de donnée i3DPost [12] et l’autre sur des données réels de la base de données [13].

### 4.1 Segmentation du mouvement

La courbe de vitesse, représentée par la distance entre EHC de trames successives, donne une courbe très bruitée. Les minimas sur cette courbes ne définissent pas les limites séparant les clips sémantiques d’où une sur-segmentation de la séquence (voir Figure 5 (Top)). Pour obtenir des minima locaux plus significatif, la courbe est convoluée avec un filtre temporel qui permet de prendre en compte la variation de mouvement, non seulement entre les deux trames successives, mais également dans une fenêtre de temps. Le vecteur de vitesse de mouvement après convolution est illustré par la Figure 5 (en haut). On remarque que les points d’arrêt sont plus précis après convolution et correspondent à l’étape du changement de mouvement dans la séquence de vidéo. La taille de la fenêtre est définie de manière empirique et fixée à 6 pour tous les types d’actions.

Dans la Figure 6, nous montrons des résultats de segmentation de mouvement pour une action "marcher" lente et une autre rapide. Bien que la vitesse de marche varie, la segmentation de l’action reste stable et correspond aux changement de pas fait par l’acteur. Un résultat de segmentation sur une action plus complexe est illustrée par la Figure 6 (en bas) où l’acteur danse du "rockn’roll". Grâce à la sélection des minima locaux dans un voisinage précis, seuls les points de rupture significatifs sont sélectionnés.

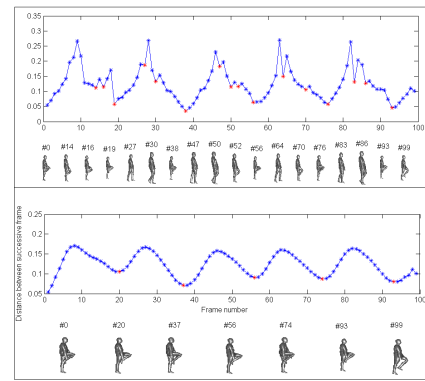


Figure 5 – Vecteur de degré de mouvement, sans et avec lissage.



Figure 6 – Résultats de segmentation sur différents mouvements : (en haut) marche lente (au milieu) marche rapide (en bas) danse : "rockn’roll".

### 4.2 Recherche d’action

Comme dans une procédure de recherche classique, en réponse à une requête donnée par un clip, notre approche cherche dans la base de données les clips similaires et renvoie une liste ordonnée de clips du plus similaire au moins similaire. L’évaluation de l’algorithme est ensuite basée sur l’évaluation de la qualité de cette liste de clips proposées en premier. Pour nos expérimentations, nous utilisons 13 actions différentes de la base de donnée i3DPost [12], interprétées par deux acteurs (26 actions au total). La segmentation de mouvement effectuée sur ces séquences donne un total de 144 clips classés en 14 catégories. Les séquences d’action sont principalement constituées de différents styles de marche, course et quelques séquences de danse. Les classes regroupées représentent différents styles de marche, courir et de pas de danse. La mesure de similarité représentée par la métrique élastiques entre chaque paire de clips permet de générer une matrice de confusion pour toutes les classes de clips. Afin d’évaluer la performance de la reconnaissance, nous comparons ces résultats avec une vérité terrain annotée manuellement. La Figure 7, illustre la matrice représentant la similarité entre les clips dans les séquences d’actions exécutées par une actrice fé-

minine contre les clips de séquences d'actions exécutées par un acteur masculin. La similarité entre clips diminue avec la diminution de la chaleur de la couleur. Grâce à

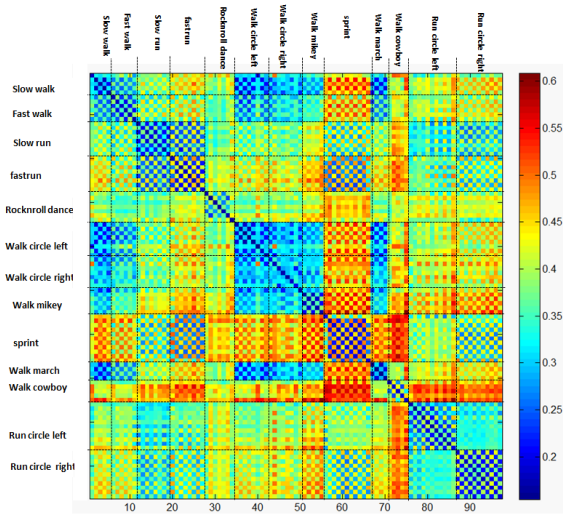


Figure 7 – Matrice de similarité entre clips. La distance entre les clips diminue avec la diminution de l'intensité de chaleur de la couleur.(de préférence à voir en couleur).

l'utilisation de DTW, il est remarqué que le score de similarité entre les clips de même actions réalisés en vitesse différente est faible (voir la Figure 7). En outre, notre approche réussit à trouver les clips faisant la même action de différentes façons. Par exemple, les clips de "marcher en cercle" sont identifiés comme similaires aux clips de "marcher en ligne droite" (voir la Figure 7). Cela peut s'expliquer par l'utilisation de la métrique élastique dans l'espace de formes pour comparer les trajectoires de courbe, ce qui leur rend indépendantes à la rotation.

Dans l'expérimentation, chaque clip de séquences est utilisé comme requête. Les clips à partir des séquences segmentées présents dans la base de données sont utilisés comme candidats. On note que la requête elle-même n'est pas inclus dans les candidats. Pour évaluer notre approche nous utilisons les mesures d'évaluation de l'indexation classiques à savoir le "First-tier", le "Second-tier" et le "PPV". Le "First-tier", permet de trouver le pourcentage des clips correctement trouvés dans les meilleurs k plus proches clips retrouvés, où k représente la taille de la classe des actions semblables à la requête. Idéalement, on ne trouve pas de faux positifs et donc on aura un score de 100 % de clips appartenant à la même classe du clip requête. Le "Second-tier" donne le même type de résultat que le "First-tier", mais pour les meilleurs  $2 * k$  plus proches clips. Le "PPV", plus proche voisin, indique le pourcentage avec lequel on trouve le premier clip classé plus proche est correcte. Dans nos expérimentations, 79.26 % de clips sont trouvés au "First-tier" et plus de 90 % (93 %) des clips sont correctement trouvés dans le deuxième niveau. En outre, le taux du "PPV" est de 99.1 %. L'exemple de la recherche

d'une action dans une base des données d'actions [13] est illustrée dans la Figure 8. Le clip requête, représentant un pas de marche droite-gauche, est correctement apparié avec les clips représentant un pas de marche similaire quoique l'acteur est entrain de marcher en tournant et en variant la vitesse de sa marche.

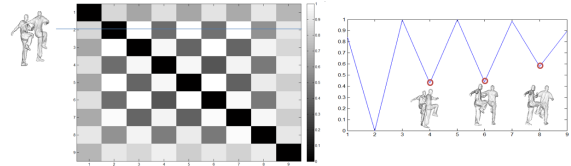


Figure 8 – Exemple de recherche d'un clip dans une longue séquence de vidéo 3D pour l'action "marcher".

### 4.3 Recherche Hiérarchique et classification

Pour un modèle de maillage d'une taille de 1 Mo, la taille de la séquence vidéo 3D augmente linéairement de 1 Mo par trame. Par conséquent, la recherche de l'action devient très difficile dans ces longues séquences.

Notre approche de recherche hiérarchique peut être décomposée en trois étapes : Premièrement, la séquence de vidéo est segmentée et regroupés en clips similaires. Deuxièmement, chaque groupe de clips similaires est représenté par un clip "modèle". Ce dernier est le clip moyen le plus proche de tous les éléments dans le groupe. Troisièmement, nous constituons une nouvelle séquence représentative de la séquence d'origine à partir de l'ensemble de tous clips "modèles". Pour trouver donc les actions similaires à une requête, nous comparons ce clip requête avec les clips "modèles", ce qui permet de réduire le temps de calcul. En effet, la complexité passe de  $n$  à  $\log n$ . Nous procédons à une expérimentation sur les mêmes données de la section 4.2 sur les 26 actions de la base de données i3DPost [12]. La matrice de comparaison est celle de tous les clips avec les clips "modèles" des classes (clusters). Cette matrice est illustrée par dans Figure 9. Les performances de recherche obtenus à partir de cette matrice pour le First-tier et le Second-tier sont respectivement 84.09%, 95.83%.

En terme de classification de clips, 93.75% sont correctement classifiés. L'analyse des résultats donnée par la matrice binarisée montre que les clips les plus mal classés sont ceux des classes : "course" et "course rapide". En effet, ces clips sont dans la vérité terrain mis dans deux classes différentes et la seule différence est que la personne lève le pied un peu plus haut pour courir encore plus rapidement.

## 5 Conclusion

Dans cet article, des technologies clé ont été développées pour la recherche par l'exemple d'actions dans des vidéos 3D. Tout d'abord, la pose est représentée par un ensemble de courbes géodésiques extraites sur la surface du corps humain à l'aide des points caractéristiques d'extrémités. Ensuite, une métrique élastique est calculée entre

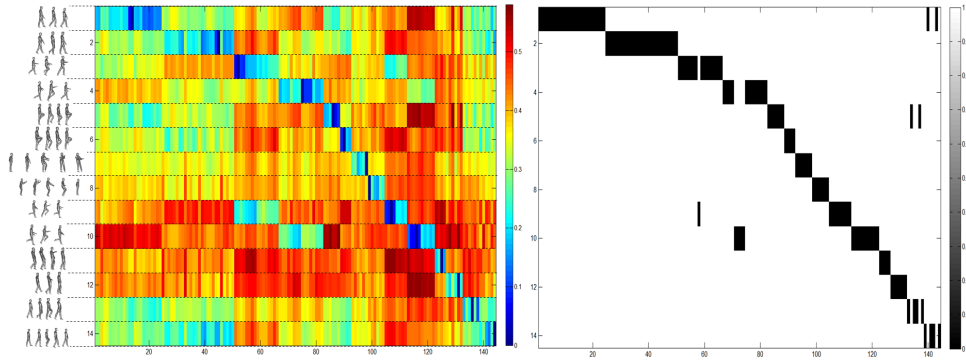


Figure 9 – Matrice de similarité et sa binarisation pour la comparaison du clip "modèle" contre tous les autres clips de la base.

deux formes de courbes afin d'estimer leur similitude. Nous avons étendu ce descripteur pour la recherche par le contenu de la vidéo 3D. Une segmentation de mouvement est effectuée sur une séquence continue pour la diviser en segments d'actions élémentaires appelés clips. Ceux-ci sont ensuite représentés par des trajectoires temporelles dans l'espace de forme de courbe ouverte. La recherche de l'action est ensuite effectuée en faisant correspondre les trajectoires en utilisant l'algorithme "DTW" adapté avec la métrique élastique définie dans l'espace en forme. Notre approche permet d'obtenir une précision de la performance de 93.44% pour la recherche de vidéo en "Second-tier", ce qui est encourageant et montre le potentiel de cette approche.

## Références

- [1] Toshihiko Yamasaki et Kiyoharu Aizawa. Motion segmentation and retrieval for 3d video based on modified shape distribution. volume 2007, pages 211–211, New York, NY, United States, Janvier 2007. Hindawi Publishing Corp.
- [2] Peng Huang, Adrian Hilton, et Jonathan Starck. Shape similarity for 3d video sequences of people. volume 89, pages 362–381, Hingham, MA, USA, Septembre 2010.
- [3] J. Kilner, Guillemaut, et A. Hilton. 3D Action Matching with Key-Pose Detection. Dans *Search in 3D and Video (S3DV)*, 2009.
- [4] T. Tung et T. Matsuyama. Topology dictionary for 3d video understanding. volume 34, pages 1645–1657, Los Alamitos, CA, USA, 2012. IEEE Computer Society.
- [5] Michael B. Holte, Cuong Tran, Mohan M. Trivedi, et Thomas B. Moeslund. Human action recognition using multiple views : a comparative perspective on recent developments. Dans *Proceedings of the 2011 joint ACM workshop on Human gesture and behavior understanding*, J-HGBU '11, pages 47–52, 2011.
- [6] Daniel Weinland, Edmond Boyer, et Remi Ronfard. Action Recognition from Arbitrary Views using 3D Exemplars. pages 1–7, Rio de Janeiro, Brésil, 2007. IEEE.
- [7] Pingkun Yan, Saad M. Khan, et Mubarak Shah. Learning 4d action feature models for arbitrary view action recognition. Dans *2008 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR 2008)*, 24-26 June 2008, Anchorage, Alaska, USA. IEEE Computer Society, 2008.
- [8] Rim Slama, Hazem Wannous, et Mohamed Daoudi. Extremal human curves : a new human body shape and pose descriptor. 10th IEEE international conference on Automatic Face and Gesture Recognition 2013, April 22-26, Shanghai, China.
- [9] Julien Tierny, Jean-Philippe Vandebois, et Mohamed Daoudi. Invariant high level reeb graphs of 3d polygonal meshes. pages 105–112, Los Alamitos, CA, USA, 2006. IEEE Computer Society.
- [10] S.H. Joshi, E. Klassen, A. Srivastava, et I. Jermyn. A novel representation for riemannian analysis of elastic curves in rn. Dans *Computer Vision and Pattern Recognition, 2007. CVPR '07. IEEE Conference on*, pages 1–7, june 2007.
- [11] Toni Giorgino. Computing and visualizing dynamic time warping alignments in r : The dtw package. volume 31, page 1–24, 06/2009 2009.
- [12] Nikolaos Gkalelis, Hansung Kim, Adrian Hilton, Nikos Nikolaidis, et Ioannis Pitas. The i3dpost multi-view and 3d human action/interaction database. Dans *Proceedings of the 2009 Conference for Visual Media Production, CVMP '09*, pages 159–168, Washington, DC, USA, 2009. IEEE Computer Society.
- [13] Daniel Vlasic, Ilya Baran, Wojciech Matusik, et Jovan Popović. Articulated mesh animation from multi-view silhouettes. Dans *ACM SIGGRAPH 2008 papers, SIGGRAPH '08*, pages 97 :1–97 :9, 2008.