Annotation sémantique des mouvements humains par une approche ontologique

Sawsan Saad[1] Dominique De Beul[1] Saïd Mahmoudi[1] Pierre Manneback[1]

[1] Service d'informatique

Université de Mons (UMONS) - Faculté Polytechnique
20, Place du Parc
7000 Mons - Belgique

{Sawsan.Saad,Dominique.Debeul,Said.Mahmoudi,Pierre.Manneback}@umons.ac.be

Résumé

Dans cet article, nous proposons une nouvelle approche d'annotation sémantique de mouvements basée sur le langage OWL (Ontology Web Language). Notre modèle emploie la notation de mouvements de Benesh et se compose de concepts d'ontologies de mouvements ainsi que de leurs relations. Afin de réaliser nos annotations, nous avons développé des règles sémantiques (SWRL : Semantic Web Rules Language). Les résultats montrent l'efficacité du système proposé en produisant un ensemble d'annotations cohérentes [1].

Mots clefs

Fossé sémantique, ontologie, descriptions logiques, SWRL, annotation de mouvements, notation de mouvements de Benesh.

1 Introduction

Au cours de ces dernières années, le fossé séparant les concepts sémantiques et les caractéristiques numériques de bas niveau a retenu une grande attention de la part de la communauté scientifique [1, 2], et reste l'un des principaux défis dans le domaine de la vision par ordinateur. Afin de réduire ce fossé sémantique [3], nous proposons un système d'annotation basé sur une ontologie et une approche de segmentation temporelle. Ce système extrait ou segmente dans un premier temps les mouvements du corps humain dans des scènes fixes, et les associent ensuite à des annotations sémantiques.

La segmentation par soustraction d'arrières-plans [4] permet d'extraire les corps en mouvement et d'éliminer les objets immobiles. Pour distinguer les mouvements rapides des mouvements lents (voire nuls), nous utilisons une méthode qui se base sur les flots optiques. Ces mouvements sont ensuite convertis dans un des systèmes de notation de mouvements répandu à l'échelle internationale et connu sous le nom de "notation de Benesh" [5]. Ces mouvements convertis sont enregistrés dans un fichier XML (eXentensible Markup Language).

La notation présentée par Joan et Rudolf Benesh modélise les mouvements humains à l'aide de symboles et s'écrit de gauche à droite sur une portée de cinq lignes (comme les portées musicales) pour illustrer les positions des membres du corps humain dans l'espace. D'autres notations telles que la cinétographie de Laban [6] (utilisée dans l'ontologie de El Rahed [7]), de Conté (`http://conte.notation.free.fr/index.php`), de Sutton (`http://www.dancewriting.org`) et de Eshkol-Wachman (`http://www.movementnotation.com/`) permettent également de modéliser les mouvements humains.

L'ontologie de mouvements dans les vidéos que nous développons traduit la sémantique des symboles du système de notation vers des concepts, des propriétés et leurs relations. Cela fournit une description formelle du domaine de mouvements. Dans cet article, nous avons défini une ontologie qui permet d'annoter les mouvements humains dans les vidéos. Pour illustrer cette ontologie, nous employons le langage sémantique (OWL) et les règles sémantiques (SWRL) proposé par le W3C (World Wide Web Consortium).

L'article est organisé comme suit : la section 2 présente l'état de l'art dans le domaine de l'analyse sémantique de vidéos. Différentes étapes sont illustrées dans la section 3 pour réaliser l'analyse sémantique des mouvements. Les résultats expérimentaux sont présentés dans la section 4. Et pour terminer, nous concluons et donnons quelques perspectives en section 5.

2 État de l'art

Une ontologie est une description formelle des entités, de leurs propriétés, de leurs relations et contraintes, pour un domaine, exprimés par des axiomes [8]. Le domaine ontologique [9] présente des concepts qui sont liés, par

1. Ce travail est supporté par le projet OLIMP, FWB, Belgique, ARC-AUWB-2008-08/12-FPMS11 et le Programme d'Excellence Numediart, Wallonie, Belgique, 716631

exemple, à la médecine, au multimédia, au transport, etc. Les ontologies sont étudiées et développées grâce aux possibilités de représentation des bases de connaissances et aux facilités à partager ces connaissances [10].

Au cours de la dernière décennie, différentes approches ont proposé l'utilisation des ontologies pour la réalisation des systèmes d'annotations vidéos. Plusieurs suggestions sont faites pour réduire le fossé sémantique, par exemple, en développant des modèles de reconnaissances de concepts visuels. Bertini et al. [11] présentent un framework ontologique qui inclut des instances de données en relation avec des concepts de haut-niveau de vidéo de sport et leur modèle spatio-temporel. D'autres travaux utilisent le raisonnement qui se base sur des règles sémantiques afin d'obtenir une annotation plus détaillée. Hollink et al. [12] proposent une méthode semi-automatique d'annotation d'images de cellule du pancréas qui utilise une nouvelle ontologie avec un nouveau vocabulaire. Zhang et al. [13] propose un framework sur base d'ontologie pour analyser le comportement des personnes (reconnaissance d'activité humaines) dans le but de les rappeler à une tâche.

D'un autre côté, l'utilisation des ontologies dans l'annotation des données multimédias nécessite souvent une segmentation préalable des données (images, vidéos et objets 3D). La segmentation d'informations est une étape critique pour la compréhension de vidéos, la reconnaissance de forme et la vision par ordinateur. Pour un être humain, il semble évident de distinguer les divers objets d'une vidéo et de pouvoir ensuite les interpréter. Malheureusement, pour les systèmes informatisés, cela semble moins évident. Pour distinguer les objets animés ou non animés dans une vidéo, ces systèmes doivent segmenter efficacement les informations pertinentes. Plusieurs méthodes de segmentation sont présentées dans la littérature [14]. Certaines proviennent de la théorie du traitement des images où la notion de temps est quasi inexistante. A l'opposé, d'autres méthodes tiennent compte de l'aspect temporel. Nous utilisons la méthode de soustraction d'arrière-plan qui compare les trames contiguës pour détecter les objets en mouvement [15].

Dans notre démarche, nous souhaitons automatiser la détection du mouvement dans les vidéos en utilisant les concepts et les règles sémantiques. Ainsi, nous présentons une méthode qui détecte les concepts de mouvements et permet de réduire le fossé sémantique.

3 Analyse sémantique des mouvements

Nous développons un modèle d'ontologie représenté par la figure 1, à partir des informations résultantes de la segmentation (Section 4). Il existe plusieurs modèles qui tiennent compte du but de l'ontologie, de la conceptualisation, de la formalisation. Le but de l'ontologie de mouvement dans les vidéos est de réaliser une annotation efficace de contenu de vidéos. La définition de l'ontologie vidéo que nous proposons passe par les deux étapes suivantes : la conceptualisation (Section 3.1) et la formalisation (Section 3.2).

3.1 La conceptualisation

La conceptualisation définit les concepts, les relations et les contraintes liées au domaine de mouvement. Pour commencer, nous segmentons la vidéo en plusieurs séquences. Chaque séquence se décompose en plusieurs poses de référence. Nous associons à chaque pose de référence un descripteur qui contient la durée, la position des membres et leur distance. Pour enregistrer une pose de référence (KeyPose), la notation de Benesh indique la position occupée par les mains et les pieds du corps humain dans un plan. Ce Plan est utilisé pour représenter la position anatomique du corps.

La position des extrémités du corps dans la portée, nommé symbole, indique pour chaque pose de référence la position des mains et des pieds en tenant compte des mesures anthropométriques. La portée comporte cinq lignes, la première au niveau du sol (hauteur zéro), la deuxième au niveau des genoux, la troisième au niveau de la taille, la quatrième au niveau des épaules et la dernière au niveau de la tête.

Pour représenter les membres en trois dimensions, trois plans sont employés : Sagittal, Coronal et Transversal. Ils sont représentés par les symboles : devant (|), arrière (.), niveau (-). La figure 2 montre une partie de l'ontologie qui représente le mouvement dans notre système de notation de Benesh.

Un mouvement est une manière de présenter les parties du corps qui peuvent se déplacer afin de définir les gestes de mobilité du corps humain. Les mouvements du corps sont des blocs d'une partition qui spécifie la position des parties du corps pour un certain temps. L'élément partie du corps (BodyPart) est la classe de base pour toutes les classes de mouvements. Nos mouvements sont hiérarchisés (Fig.3) en mouvement simple (pose de référence) et en mouvement complexe. Un mouvement est considéré comme simple quand il est constitué de deux poses de référence, il devient complexe lorsque plus de deux poses de référence sont utilisés. La marche est un mouvement complexe, car de manière répétitive plus de deux poses de référence sont employées. Nous utilisons également des descripteurs pour annoter les distances entre les parties du corps et le temps. Nous employons les concepts sémantiques pour décrire les actions (déplacer, lever, etc.).

3.2 Formalisation

Nous utilisons le langage OWL et le logiciel Protégé 4.1 pour formaliser notre proposition d'ontologie. Les expressions formelles ou les règles sémantiques en SWRL perme-

ttent de classer chaque mouvement par déduction. La figure 4 (en bas) présente une règle d'un mouvement nommé (Mouvement (?a)) classé comme un mouvement simple (Hands-Moved-Horizontally-FrontPlane (?a)). Cette règle comporte les concepts (Movement, Hand, Actions and Sagittal) et les relations entre ces concepts (hasPart, hasSide and hasPlan).

4 Résultats expérimentaux

Nous avons créé plusieurs instances de membres du corps humain en mouvements à l'aide de vidéos utilisés pour l'apprentissage des poses de référence. Pour tester et évaluer ces instances, nous employons 10 autres vidéos d'une durée de 2 minutes par vidéo. La segmentation de ces membres en mouvement s'effectue suivant la technique de soustraction d'arrière plan. Cette technique consiste à soustraire chaque trame contiguë, le résultat nous donne une trame qui comporte deux silhouettes, ce qui n'est pas souhaité. Pour éviter ce problème, nous employons une trame de référence qui est déterminée statistiquement en prenant la moyenne des dix premières trames. Nous utilisons une approche basée sur les flots optiques pour détecter la présence de mouvements dans la vidéo. Lorsqu'il n'y a plus de mouvement, nous considérons que nous sommes en présence d'une pose de référence (Fig. 5). Pour le traitement des vidéos, nous utilisons la librairie de traitement d'images OpenCV (2.3.1).

Les poses de référence détectées sont comparées avec des formes pré-définies, ce qui permet de déterminer les caractéristiques de chaque pose. Les caractéristiques sont ensuite enregistrés dans un fichier XML et vont nous servir à annoter les vidéos. Ce fichier contient des informations telles que le nombre de poses de référence, le temps, les parties du corps et la notation de Benesh (plan, position et distance).

Pour vérifier que tout les concepts et les informations nécessaires soient utilisés, différentes règles sémantiques sont créées (Fig. 6). Chaque règle est formulée en utilisant le plug-in de Protégé, SWRLtab. Nous utilisons le moteur d'inférence JESS (Java Expert System Shell) intégré dans Protégé pour exécuter les règles sémantiques et en déduire les résultats. Les règles sémantiques d'inférences se basent sur une cartographie des caractéristiques de mouvements et des concepts de l'ontologie. La figure 4 reprend le code XML (à gauche) d'une pose de référence (à droite). Les résultats d'annotation obtenus par le système développé, montre la cohérence entre les concepts, les règles sémantiques et les annotations des poses d'image.

5 Conclusion et perspective

Dans cet article, nous avons proposé et défini une ontologie pour automatiser l'annotation de mouvements dans les vidéos. Notre contribution est la description détaillée d'une ontologie de mouvements de vidéos, la définition de concepts, la spécification de concepts et la formalisation d'axiomes liés au domaine du mouvement. Nous avons décrit également la manière dont cette ontologie peut être utilisée pour annoter les vidéos et récupérer par inférence des axiomes définis dans l'ontologie. Nous avons rassemblé des concepts de données de vidéos de bas niveau, qui ont été enregistrées dans un fichier XML basé sur la notation de mouvement Benesh. Le sytème que nous proposons permet de produire des annotations efficaces et cohérentes.

Nous avons comme objectif de développer un système qui détecte automatiquement les instances dans différents fichiers XML. Nous souhaitons analyser et intégrer les différentes possibilités d'inférer des mouvements complexes dans l'ontologie. Nous voulons également faire évoluer l'ontologie vers d'autres types de mouvement, comme les mouvements de danses, de gymnastiques ou en kinésithérapie.

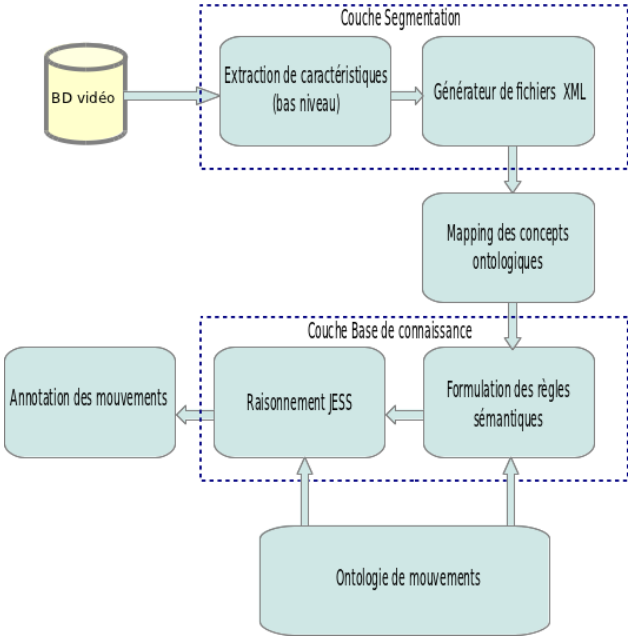

Figure 1 – *Schéma fonctionnel du modèle d'ontologie*

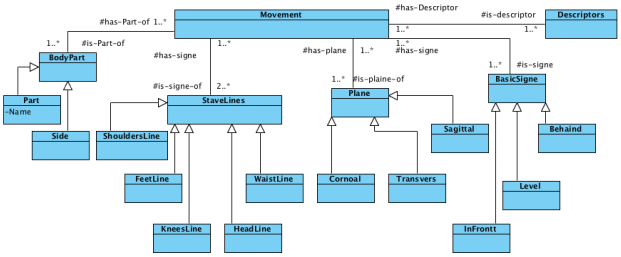

Figure 2 – *La partie de l'ontologie développée représentant les mouvements*

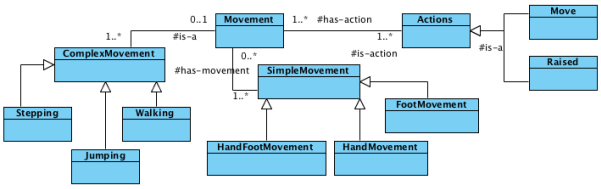

Figure 3 – *Partie de l'ontologie liée à différents types de mouvement*

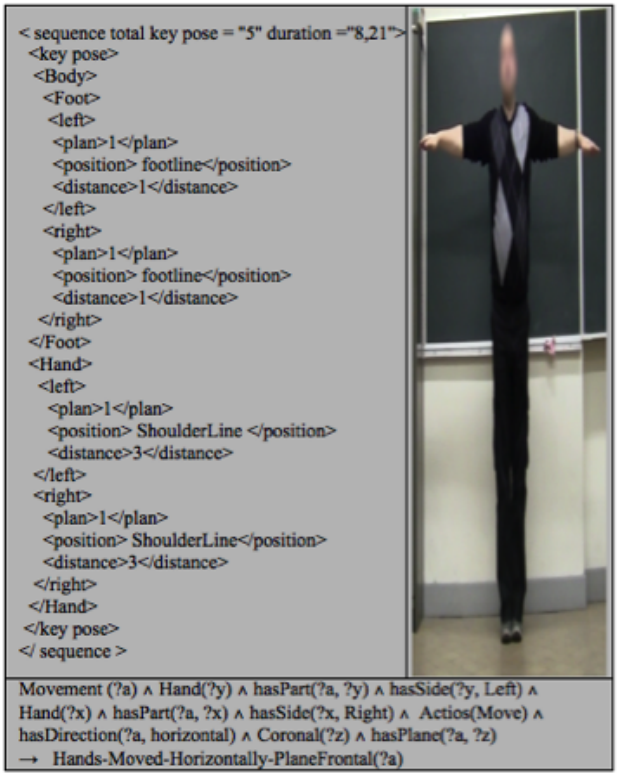

Figure 4 – *Représentation XML et SWRL d'une pose de référence*

Figure 5 – *Segmentation de vidéo : poses de références*

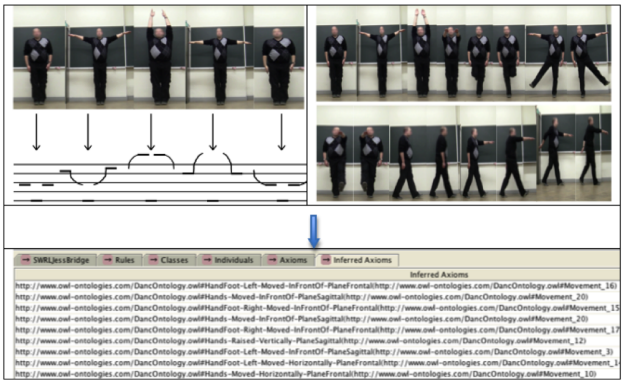

Figure 6 – *Représentation SWRL et inférence des mouvements par l'utilisation du moteur Jess*

Références

[1] Anil Kokaram, Niall Rea, Rozenn Dahyot, Murat Tekalp, Patrick Bouthemy, Patrick Gros, et Ibrahim Sezan. Browsing sports video : trends in sports-related indexing and retrieval work. *IEEE Signal Processing Magazine*, 23(2) :47–58, Mars 2006.

[2] Cees G.M. Snoek et Marcel Worring. Multimodal video indexing : A review of the state-of-the-art. *Multimedia Tools and Applications*, 25 :5–35, 2003.

[3] Arnold W. M. Smeulders, Marcel Worring, Simone Santini, Amarnath Gupta, et Ramesh Jain. Content-based image retrieval at the end of the early years. *IEEE Trans. Pattern Anal. Mach. Intell.*, 22 :1349–1380, December 2000.

[4] Y J Zhang. An overview of image and video segmentation in the last 40 years. *Advances in Image and Video Segmentation*, pages 1–15, 2006.

[5] Benesh J. Benesh R.. *Reading dance : The birth of Choreography*. Souvenir Press LTD, first édition, 1983.

[6] Minako Nakamura et Kozaburo Hachimura. An xml representation of labanotation, labanxml, and its implementation on the notation editor labaneditor2. Dans *In first South-Eastern European Digitization Initiative (SEEDI) Conference*, 2005.

[7] Y. Ioannidis K. El Raheb. A labanotation based ontology for representing dance movement. *Dans Proceedings of Gesture Workshop 2011*.

[8] M. Grüninger et M. S. Fox. Methodology for the design and evaluation of ontologies. Dans *International Joint Conference on Artificial Inteligence (IJCAI95), Workshop on Basic Ontological Issues in Knowledge Sharing*, 1995.

[9] Natalya F. Noy et Deborah L. McGuinness. Ontology development 101 : A guide to creating your first ontology. Rapport technique, Stanford Knowledge Systems Laboratory and Stanford Medical Informatics, 2001.

[10] Steffen Staab et Rudi Studer. *Handbook on Ontologies*. Springer Publishing Company, 2nd édition, 2009.

[11] Marco Bertini, Alberto Del Bimbo, Giuseppe Serra, Carlo Torniai, Rita Cucchiara, Costantino Grana, et Roberto Vezzani. Dynamic pictorially enriched ontologies for digital video libraries. *IEEE MultiMedia*, 16 :42–51, April 2009.

[12] L. Hollink, S. Little, et J. Hunter. Evaluating the application of semantic inferencing rules to image annotation. Dans *Proceedings of the 3rd international conference on Knowledge capture*, K-CAP '05, pages 91–98, New York, NY, USA, 2005. ACM.

[13] Zhang et al. An ontology-based context-aware approach for behaviour analysis. Dans Activity Recognition in Pervasive Intelligent Environments, éditeur, *Liming Chen, Chris D. Nugent, Jit Biswas et Jesse Hoey*, volume 4, pages 127–148. Atlantis Press, Paris, France, August 2011.

[14] Massimo Piccardi. Background subtraction techniques : a review. Dans *SMC (4)*, pages 3099–3104, 2004.

[15] Sen-Ching S. Cheung et Chandrika Kamath. Robust background subtraction with foreground validation for urban traffic video. *EURASIP J. Appl. Signal Process.*, 2005 :2330–2340, January 2005.