



HAL
open science

Apprentissage des Mouvements Humain en Situation Informatisée: Avantages, Limites et Perspectives Apportées par l'Approche Automatique

Quentin Couland, Ludovic Hamon, Sébastien George

► **To cite this version:**

Quentin Couland, Ludovic Hamon, Sébastien George. Apprentissage des Mouvements Humain en Situation Informatisée: Avantages, Limites et Perspectives Apportées par l'Approche Automatique. 8ème conférence des Environnements Informatiques pour l'Apprentissage Humain (EIAH 2017), Jun 2017, Strasbourg, France. pp.137-148. hal-01713082

HAL Id: hal-01713082

<https://hal.science/hal-01713082v1>

Submitted on 26 Jan 2022

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Apprentissage des mouvements humains en situation informatisée : avantages, limites et perspectives apportées par l'approche automatique

Quentin Couland, Ludovic Hamon, Sebastien George

UBL, Université du Maine, EA 4023, LIUM, 72085 Le Mans, France
{quentin.couland, ludovic.hamon, sebastien.george}@univ-lemans.fr

Résumé. Les applications utilisant les mouvements des utilisateurs afin de renforcer la construction des connaissances et d'améliorer l'immersion, sont de plus en plus utilisées dans des domaines tels que le sport, la chirurgie ou l'éducation. Actuellement, les informations sur les actions, intentions, situations et comportements, sont rarement extraites à partir de données de mouvement 3D. Cela peut être expliqué par l'hétérogénéité, la complexité et la grande dimensionnalité de ces informations d'une part et leurs corrélations avec les besoins d'observation des enseignants d'autre part. Cependant, les méthodes de traitement automatique (*machine learning*) pourraient être utilisées afin de surmonter ces contraintes. Une meilleure analyse de ces données pourrait être obtenue pour mieux modéliser le comportement de l'apprenant lors de la situation d'apprentissage. Cet article introduit les défis propres à ces problèmes et propose un système d'aide à l'analyse des mouvements humains.

Mots-clés. Mouvement, apprentissage humain, apprentissage automatique

Abstract. In order to improve information retention, as well as immersion during learning activities, motion is more and more used in domains such as sport, surgery or education. Nowadays, relevant feedback and learner models regarding their actions, intentions, situations, behaviors, etc. are rarely extracted from 3D captured motions. This can be explained, on one hand, by the heterogeneity, the complexity and the highly dimensional nature of such information and, on the other hand, by its correlation with the teacher's observation needs. However, automatic methods such as machine learning could be used to overcome these constraints and provide more efficient and complementary pedagogical feedback from motion data. A better analysis of these data could provide a better learner's model from the learning situation. This article introduces the challenges related to these problems, and proposes a system dedicated to human motion analysis.

Keywords. Motion, human learning, machine learning

1 Introduction

De nos jours, la capture de mouvements est de plus en plus utilisée dans de multiples domaines, tels que les jeux-vidéo, les films d'animations, la réalité virtuelle, le sport, la médecine, l'industrie et l'éducation. Grâce aux avancées réalisées dans l'électronique, les Interfaces Homme-Machine (IHM) et le traitement des données, il est raisonnable

de penser que la capture, l'édition et le partage des mouvements humains 3D seront bientôt démocratisés. Cela peut avoir un impact fort en éducation et dans tous les domaines qui impliquent des gestes. En effet, le mouvement humain est porteur d'information : il est possible, par exemple, d'extraire des informations bas-niveau, e.g. liées à la cinétique et la dynamique [1]. Le geste est également porteur d'informations sémantiques, dans le cadre d'une communication verbale [2] ou non-verbale [3]. En plus de cela, il est également possible d'inférer des informations de haut-niveau, e.g. par rapport à l'émotion [4], l'intention [5] et l'action [6]. Tracer les activités de l'apprenant implique la génération d'une grande quantité de données de mouvement. Traiter et analyser ces informations manuellement n'est pas envisageable [7]. Le développement de techniques de traitements automatiques, (e.g. machine learning) peut aider à alléger cette tâche. Ces méthodes peuvent analyser des données de haute dimensionnalité, dans un but de classification, d'extraction de caractéristiques, de résolution de problèmes de régression, etc. [8]. Dans des situations d'apprentissage, ces algorithmes ont déjà été largement utilisés, afin d'étudier les actions et les comportements d'apprenants [9] [10].

Il y a cependant un manque de travaux sur l'analyse automatique de mouvements de l'apprenant afin d'en extraire des informations pédagogiquement utiles. Cela peut être expliqué par un certain nombre de verrous techniques et scientifiques. L'objectif de ce travail est d'étudier les besoins, défis et perspectives relatifs à la démocratisation de cette approche. La partie suivante aborde l'utilisation du mouvement, ainsi que son apprentissage, qu'il soit la cible de l'apprentissage, ou une étape menant à l'objectif voulu. L'apprentissage automatique (*machine learning*), ainsi que ses utilisations dans des contextes d'apprentissage et d'analyse de l'apprenant est le sujet de la partie 3. La partie 4 est consacrée à la proposition d'un système dédié à l'analyse du mouvement d'un apprenant.

2 Usage du mouvement

2.1 Capture et modélisation du mouvement

La capture du mouvement intervient dans de nombreux domaines, comme le film, les jeux-vidéo, etc. Dans ces contextes, le mouvement est principalement considéré comme une évolution séquentielle de postures au cours du temps [11]. Chaque posture, appelée, « *animation frame* », ou « *frame* », est séparée de la suivante par un intervalle de temps, généralement régulier. Une manière de modéliser informatiquement ces postures est d'utiliser une structure de graphe afin de décrire la hiérarchie entre les différentes articulations (*i.e.* les nœuds du graphe), selon une représentation squelettique sous la forme d'un arbre, dont la racine est usuellement la partie basse du torse (*i.e.* les hanches), et dont les fils sont les différentes articulations du corps. Chaque nœud contient des informations, comme la position et/ou l'orientation relative au parent de ce nœud. Des informations additionnelles peuvent être incorporées, telles que les *offsets* des capteurs, le nombre de *frames*, le temps *inter-frame*, la longueur et les dimensions de chaque membre du corps, etc. Les méthodes de capture, de traitement et

de rendu impliquent généralement des compétences techniques en image de synthèse et en animation, ainsi que l'usage de logiciels dédiés [1] [12].

L'acquisition d'un mouvement peut avoir différents objectifs : un aspect purement ludique (*e.g.* jeux-vidéo), une étude médicale (*e.g.* analyse de la démarche d'un patient en rééducation, analyse *a posteriori* d'une intervention chirurgicale) [13], l'inférence d'informations comportementales [14], etc. Le besoin en terme d'observation et d'analyse de chaque cas est différent, et donc le matériel utilisé n'est pas le même. Une des familles de capteurs qui tend à se développer de plus en plus est la famille des capteurs inertiels, *i.e.* les accéléromètres et les gyroscopes¹. Ces capteurs renvoient notamment des informations liées à la vitesse et à l'accélération du dispositif en mouvement, à partir desquelles l'orientation et la position relatives peuvent être approximées. À partir de ces données, il est possible de reconstituer un mouvement, en affichant les données successives reçues. L'avantage de ce type de capteurs est qu'il est possible d'en utiliser plusieurs en parallèle, afin de capter l'ensemble d'un corps en mouvement, malgré un phénomène de dérive nécessitant une recalibration régulière. Les caméras RGB-D (*Red Green Blue Depth*) ont été popularisées grâce à l'apparition de systèmes grands publics (*e.g.* Kinect, de Microsoft) et sont beaucoup utilisés dans des contextes de recherche, afin notamment de suivre le squelette d'un ou plusieurs utilisateurs (à raison de 25 articulations par utilisateur, 6 utilisateurs maximum pour Kinect v2 par exemple). Les systèmes les plus précis et les plus coûteux sont ceux usuellement utilisés dans la capture de mouvement dite « professionnelle », *e.g.* pour le cinéma, les jeux-vidéo. Ces systèmes sont composés de multiples caméras infrarouges, disposées tout autour de la scène, et permettent de capturer une information haute-fidélité, à l'aide d'une combinaison réfléchissante, portée par l'acteur².

2.2 Apprentissage et analyse du mouvement

Le mouvement humain devient un moyen privilégié d'interaction avec les environnements informatiques, y compris dans le domaine de l'éducation [15]. L'apprentissage des mouvements humains peut être considéré selon deux points de vue en fonction de la finalité souhaitée : (i) la finalité est de reproduire/copier une succession de postures spécifiques, et (ii) la finalité est d'atteindre l'état cible d'un objet et/ou d'un corps à l'issue du mouvement, sans se focaliser sur l'apprentissage de tout ou partie de la succession de postures générant ce mouvement. Ces deux points de vue ne sont bien évidemment pas exclusifs. Cette section explore l'utilisation du mouvement dans un contexte d'apprentissage.

Reproduire le mouvement comme une suite de postures dans le temps

Dans le domaine médical, le mouvement est utilisé dans plusieurs buts. Aminian et Najafi ont analysé la démarche du patient [16], alors que Zhou et Hu [17] ont travaillé sur l'apprentissage de mouvements pour la rééducation. L'approche de représentation par un squelette n'a pas été systématiquement considérée, car les données de mouvement ont été recueillies sur des capteurs différents, selon l'objet de l'étude. Dans

¹ <https://neuronmocap.com/>

² <https://www.vicon.com/>

ces travaux, il n'y a aucune analyse automatique de ces mouvements, l'observation et la déduction d'information se faisant par un expert humain.

Dans le cadre de l'apprentissage de l'archerie japonaise, Yoshinaga et Soga ont développé un système basé sur un capteur Kinect, afin de capturer le squelette de l'apprenant et sa variation dans le temps. Les gestes d'experts sont également modélisés [18]. L'apprenant peut ainsi se comparer avec une multitude de modèles d'experts stockés. L'analyse se fait empiriquement, soit par l'apprenant lui-même, soit par un expert qui observe le geste. Kora *et al.* [19] ont utilisés de nombreux dispositifs (casque de réalité virtuelle avec une caméra, une Kinect, *etc.*) afin de construire un système d'apprentissage du golf. Grâce à la superposition du squelette de l'apprenant et du squelette de l'expert, l'utilisateur est capable de voir son mouvement en temps réel, et ainsi essaye de coller au mieux au mouvement expert. Cependant, cet affichage en temps-réel peut être perturbant et aucune information pédagogique n'est extraite de ce processus.

L'entraînement sportif est un domaine qui peut beaucoup bénéficier de la capture de mouvement. Dong *et al.* [20] ont passé en revue les systèmes informatisés utilisés dans des contextes d'entraînement physique. Il ressort que les environnements de réalité virtuelle, ainsi que l'analyse de vidéos, peuvent améliorer l'apprentissage de gestes, à l'aide d'un expert ou non. Il est également possible d'inférer des informations quant à la forme physique et les tactiques d'un joueur à partir d'informations plus générales (telles que les profils d'activités, les tactiques d'équipes, *etc.*), comme montré par Beetz *et al.* [21]. Cette analyse permet au coach d'avoir une meilleure lecture du jeu, ainsi que d'apporter de meilleurs conseils aux joueurs.

Atteindre l'état cible d'un objet ou d'un corps à l'issue du mouvement

La capture de mouvements peut être utile dans de nombreuses tâches où une action manuelle est nécessaire, dans des domaines tels que la médecine, la chirurgie, l'industrie, *etc.* Dans ce cas, l'imitation d'un geste n'est pas toujours la cible de l'apprentissage ; l'apprenant doit atteindre certains états spécifiques de l'objet manipulé au cours du temps. Dans cette optique, plusieurs systèmes informatisés ont été développés pour enregistrer et analyser les gestes humains, afin de donner un retour utile à l'apprenant.

Dans le domaine de la chirurgie, Toussaint et Luengo [22] ont utilisé des paradigmes d'interaction avancés, en combinaison d'un EIAH. L'objectif était de reconnaître et d'évaluer automatiquement les différentes phases d'une intervention chirurgicale. Des techniques de *datamining* ont été employées sur des données fournies par un bras haptique, ainsi qu'un système d'*eye-tracking*, afin d'en extraire des motifs et des procédures d'intervention chirurgicale. Ces données sont ensuite analysées par un expert pour analyser le lien entre le comportement et les performances de l'apprenant.

Du point de vue de l'apprenant, le mouvement capturé peut apporter des retours pertinents et exhaustifs. Dans le domaine médical, Choi *et al.* [23] utilisent des bras haptiques afin de simuler l'insertion d'une sonde nasogastrique. Une simulation informatique permet de simuler les collisions avec les parties du corps concernées, et ainsi donne un retour de force à travers le bras haptique à l'apprenant. Le résultat de l'entraînement est analysé à l'aide de données telles que le temps nécessaire, l'insertion moyenne et maximale, le nombre d'insertions, le nombre de fois où le tube a plié, *etc.*

Ces indicateurs sont ensuite utilisés pour compléter le jugement d'un expert, afin de décider du degré de réussite de la procédure.

Dans le domaine industriel, Chamaret *et al.* [24] ont capturé des mouvements à l'aide d'un dispositif appelé « spidar ». Le but était d'étudier et d'évaluer différentes procédures d'assemblage et de démontage d'un phare de voiture, dans un environnement virtuel. Les modèles 3D, ainsi que le retour de force, étaient adaptés à la morphologie de l'utilisateur. L'objectif était d'évaluer empiriquement la capacité de l'utilisateur à accomplir la tâche demandée, en accord avec le nombre de collisions. Cependant, aucune donnée relative à la manipulation n'était relevée, et aucun indicateur pédagogique n'était calculé.

Les systèmes de capture de mouvement ne sont plus seulement réservés aux jeux vidéo, films d'animation et à la réalité virtuelle. De nombreux domaines professionnels utilisent désormais ces systèmes et méthodes afin de développer leurs propres applications. Cette section a présentée deux types de situation où le mouvement est utilisé à des fins d'apprentissage humain : (i) apprendre une séquence de postures spécifiques, et (ii) atteindre l'état cible d'un objet et/ou d'un corps à l'issue du mouvement. Ces deux catégories ne sont pas exclusives. De plus, les données de mouvement peuvent être combinées avec d'autres appareils, tels que les casques de réalité virtuelle, les bras haptiques, *etc.* afin d'améliorer l'immersion et d'obtenir des informations supplémentaires sur le déroulement de l'activité de l'apprenant. Grâce à ce type de données, il est possible d'observer, de décomposer et d'analyser les mouvements d'un expert et/ou d'un apprenant selon les besoins.

Malgré les intérêts d'un environnement informatique support à l'étude du mouvement humain, le retour de l'expert sur une tâche d'apprentissage est principalement empirique, et non-formel. Cela peut être expliqué par la taille des données récoltées, qui sont difficilement réifiables et/ou traitables manuellement. Il existe cependant un ensemble de techniques et de méthodes qui permettent d'inférer des informations à partir de données de grande dimensionnalité [25]. Ce point sera abordé dans la partie suivante.

3 Apport des méthodes d'analyse automatique

Un objectif récurrent des travaux de recherche en informatique est d'analyser des activités humaines pour en extraire automatiquement des informations de plus haut-niveau *i.e.* des caractéristiques humaines difficilement calculables à l'aide de méthodes formelles, telles que l'émotion, l'intention et l'action. Après une introduction sur les principes et intérêts de l'extraction automatique et la classification, à l'aide de techniques de *machine learning*, cette partie présente les différentes modalités et conditions nécessaires pour inférer ces caractéristiques, au travers de cas d'études centrés sur l'apprentissage humain.

3.1 Principes et intérêts des méthodes d'apprentissage automatique

Extraire des caractéristiques de haut-niveau à partir de données brutes issues de l'activité humaine est une tâche non triviale. Plusieurs phases de prétraitement et

traitement sont nécessaires, telles que la réduction du bruit et le filtrage, afin notamment de ne garder que les données pertinentes. À partir de ces traces nettoyées, de nombreuses méthodes existent afin d'inférer des informations utiles. Ces méthodes, regroupées sous le nom de *machine learning*, permettent de traiter automatiquement de grands volumes de données. Il est possible de les regrouper dans deux catégories : (i) l'apprentissage supervisé, où l'algorithme utilise des données étiquetées avec des catégories prédéfinies, afin d'apprendre une fonction qui lui permettra de classer les futures données et (ii), l'apprentissage non-supervisé, ou *clustering*, où l'algorithme regroupe les données dans des clusters qui sont déterminées automatiquement à partir de caractéristiques communes préalablement définies. Ces techniques peuvent servir, entre autres, à inférer des informations sur l'émotion, l'intention et le comportement humain.

La reconnaissance d'actions a été explorée par Kapsouras et Nikolaidis [6], sur plusieurs bases de données. Chaque base était composée de plusieurs mouvements, tels que sauter, lancer, taper dans les mains, etc. À partir de la combinaison de schémas récurrents dans les mouvements, l'utilisation des algorithmes des k plus proches voisins, ainsi que d'un SVM (*Support Vector Machine*), une bonne reconnaissance du mouvement a été obtenue. La reconnaissance d'action humaine est un domaine très actif, dans le cadre de la vision par ordinateur. Sanchez-Mendoza *et al.* [26] utilisent des caractéristiques faciales, afin de déterminer l'émotion d'un sujet. À partir de vidéos, des mouvements faciaux atomiques ont été isolés, dans le but de classer l'émotion du sujet, à l'aide d'un SVM. Le mouvement humain a déjà été utilisé pour inférer des émotions, comme c'est le cas dans le domaine de l'*affective computing*. Kapur *et al.* utilisent des données de mouvement capturé à l'aide d'un système de marqueurs et de caméras, afin de déterminer l'émotion du sujet, répartie entre quatre catégories (*sad, joy, anger, fear*) [14]. La vitesse, ainsi que l'accélération, sont extraites de ces données. Plusieurs algorithmes ont été utilisés : régression logistique, classification naïve bayésienne, arbre de décision, réseau de neurones et SVM. Les résultats montrent un taux de reconnaissance de 92%, juste en dessous du taux de reconnaissance humain (93%).

3.2 Analyse de l'activité de l'apprenant

La détection automatique des actions, de l'intention et des émotions est fondamentale pour la recherche interdisciplinaire. Dans le contexte de l'apprentissage humain, ces caractéristiques ont été prises en compte en psychologie et en neuroscience par exemple, afin de mieux comprendre les mécanismes d'apprentissage, développer de nouvelles méthodes d'enseignement et proposer des méthodes d'apprentissage adaptées. Un des objectifs est de proposer des scénarios d'apprentissage personnalisés, à l'aide d'EIAH [22]. Construire une représentation de l'apprenant à l'aide de ces données d'interaction peut aider à cette adaptation.

Les techniques de *machine learning* jouent un rôle important dans la prédiction des actions des apprenants, la détection de la situation de l'utilisateur et la construction d'un modèle d'apprenant. Markowska-Kaczmar *et al.* [10] se sont intéressés à la génération automatique de modèles d'apprenants, basés sur l'analyse de styles d'apprentissages rapportés à des modèles cognitifs. Le système ASPOLE, développé par Lokaiczuk *et*

al. permet de prédire la tâche courante de l'utilisateur grâce à l'analyse des actions de ce dernier [9].

Cependant, les données de haute dimensionnalité, fournies par des systèmes d'interactions avancées, semblent être peu étudiées dans ce contexte. Gu et Sosnovsky ont pointés les principaux défis concernant le traitement des traces sous-jacentes [7] *i.e.* quand effectuer la reconnaissance d'intention, comment distinguer les données utiles du bruit, comment formaliser et représenter la dynamique de l'intention de l'apprenant, comment inférer les processus perceptuels cachés à partir du mouvement, et comment prendre en compte le comportement et les intentions erronés. Toutefois, l'origine de ces données, ainsi que leur hétérogénéité, au vu des différentes interfaces possibles, n'ont pas été prises en compte. Toussaint *et al.* ont également travaillé sur ce sujet, dans le cadre de la chirurgie [22]. Des données hétérogènes ont été structurées, provenant de sources diverses (*i.e.* bras haptique, position d'un trocart, suivi des yeux, *etc.*). Les données sont ensuite classées selon différentes catégories, par rapport à la phase de la procédure chirurgicale, dans laquelle elles ont été récoltées. Après une formalisation de ces données dans une représentation d'ensembles/sous-ensembles, un traitement par fouille de données est utilisé, afin d'extraire automatiquement des règles fréquentes des actions séquentielles réalisées par les élèves, pour ensuite les comparer aux procédures établies par les experts.

La génération automatique d'informations de haut-niveau sur l'activité et le comportement humain est devenue cruciale. Elle peut en effet amener à une meilleure compréhension des mécanismes d'apprentissage, l'anticipation des besoins de l'utilisateur et l'adaptation de l'EIAH pour ce dernier. L'efficacité des méthodes d'apprentissage automatique a été prouvée pour la détection d'actions, d'intentions, d'émotions et, conséquemment, pour l'évaluation de la situation de l'apprenant. Ces processus reposent majoritairement sur l'analyse d'images, de vidéos ou d'autres types de traces (*e.g.* provenant d'IHM de type WIMP, de fichiers log, *etc.*). Cependant, les avancées technologiques ont amenées de nouveaux paradigmes d'interactions qui tendent à considérer le mouvement humain comme l'objet principal d'étude. Considérer les mouvements capturés comme un nouveau type de données brutes afin d'inférer automatiquement des informations de plus haut-niveau soulève plusieurs défis et verrous : (i) l'acquisition du mouvement, et conséquemment la constitution d'un corpus de mouvements, (ii) l'hétérogénéité des données de mouvements capturés, qui nécessitent un traitement afin de les normaliser entre elles, (iii) la grande dimensionnalité de ces données, qui font qu'un traitement en temps réel est impossible, (iv) la formalisation des besoins d'observation de l'enseignant, ainsi que la transformation de ceux-ci en caractéristiques pertinentes, (v) le choix, la convergence, ainsi que les résultats d'un algorithme de *machine learning* adapté. La section suivante propose un système qui vise à lever certains de ces verrous.

4 Proposition d'un système d'aide à l'analyse de mouvements humains

À partir des verrous identifiés, l'objectif est de proposer et d'implémenter des solutions à certains d'entre eux. Le système, présenté sur la Figure 1, a pour objectif d'aider les

enseignants et les apprenants lors de l'apprentissage d'un geste. L'utilisation d'un système informatisé, basé sur un environnement virtuel 3D d'observation du geste, permet de réaliser des opérations autrement difficiles : revoir un geste sous différents points de vue, regarder de nouveau un geste à différentes vitesses, faire un arrêt sur image, etc. Il est également possible d'aider l'enseignant dans sa tâche d'évaluation, en lui proposant de formaliser ses besoins d'observation, ce qui permet ensuite d'utiliser des algorithmes adaptés pour traiter automatiquement les données de mouvement. Cette section propose d'illustrer ce système au travers du cas de l'enseignement de la guitare.

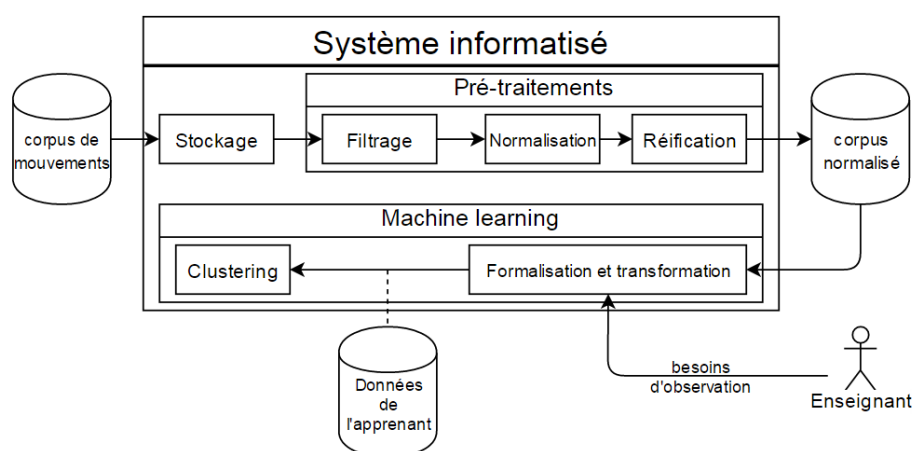


Figure 1 : proposition d'un système d'aide à l'analyse de mouvements humains.

Le mouvement étant au cœur du système, la première étape est la constitution d'un corpus de mouvements. Les études portant sur le mouvement utilisent majoritairement des corpus existants ou des méthodes d'acquisition de gestes normalisés. Le système étant pensé dans l'optique d'être indépendant du domaine d'application, l'homogénéité des données de ce corpus initial n'est pas garantie. Il convient donc de choisir une méthode de capture adaptée au domaine ciblé, selon des critères de précision, d'encombrement et de durée d'enregistrement. Dû à la nature des algorithmes utilisés, il est cependant souhaitable de disposer de suffisamment de données en entrée, afin que l'algorithme puisse fournir des résultats cohérents [27]. Dans un cours de guitare, il serait nécessaire de procéder à plusieurs enregistrements des mouvements, autant de la part du professeur (geste cible) que de l'étudiant. De plus, capturer plusieurs enregistrements d'un même mouvement, réalisé par différents experts, permet de créer un modèle de geste cible. Il en va de même pour le geste de l'étudiant.

Lors de la capture, des erreurs peuvent émaner des dispositifs de captures eux-mêmes, mais également des supports de transmission. Des données peuvent être parasitées ou corrompues. Une étape de filtrage permet de vérifier l'intégrité des données et de les nettoyer si nécessaire. Dans le cas où la forme des traces est connue d'avance, il est possible, à partir d'une liste établie d'erreurs, de les corriger en parcourant les données à l'aide d'expressions régulières, puis en les remplaçant par

l'expression attendue [13]. La détection des données aberrantes (*outliers*) permet également d'écarter rapidement les données corrompues. La méthode des moindres carrés permet de comparer les données expérimentales à un modèle mathématique [28]. La détection d'une rupture dans la continuité du mouvement peut se faire en comparant les points successifs de ce mouvement, et permet de corriger l'erreur en extrapolant à partir des points voisins existants. Cette étape peut donc induire une modification, voire une suppression des données.

Les algorithmes de *machine learning* ont besoin que les données soient homogènes au regard de leurs dimensions. Il existe, pour ce cas, plusieurs méthodes de *data mining* pour combler les vides au sein du corpus, comme la suppression des données incomplètes, le calcul du maximum de vraisemblance, ou encore l'estimation des points de données manquants [29] à partir du modèle de référence, constitué des points existants. Afin de prendre en compte les besoins d'observation des enseignants, une réification des données peut s'avérer nécessaire (*i.e.* passer d'une information bas niveau, telle que la vitesse ou l'accélération, à des caractéristiques plus haut niveau telles que l'émotion de l'apprenant). De telles transformations ont été régulièrement utilisées dans l'étude du geste [14] [30], permettant d'une part, d'avoir des données plus significatives et, d'autre part, de réduire leur dimensionnalité.

Il est également possible de segmenter le mouvement en sous-ensembles qui se répètent au sein du corpus [31]. Cette représentation permet de repérer les éléments récurrents d'un geste, par exemple les accords d'un morceau de guitare. La présence (ou l'absence) d'un de ces mouvements peut être décisif au cours de l'apprentissage. Une fois ces prétraitements appliqués, un corpus normalisé peut être obtenu.

Une étape importante est la formalisation, ainsi que la transformation des besoins de l'enseignant en propriétés de *clustering* ou en classes pour l'apprentissage supervisé. Les besoins en termes d'observation peuvent être variés. Pour un professeur de guitare, par exemple, l'objectif peut être d'analyser une progression d'accord qui sonne faux, lorsque l'élève la réalise. Dès lors, l'expertise du professeur lui permet d'orienter son observation sur les transitions entre ces accords. L'observation du geste en temps-réel peut être compliquée, dû à la position des mains et des doigts, bloquant ainsi la vue. À partir du besoin (*e.g.* observer les transitions entre les accords), on peut sélectionner les parties importantes des données, afin de focaliser les algorithmes sur l'étude de cette partie du mouvement.

La dernière étape est l'utilisation d'algorithmes de *machine learning*, afin de satisfaire les besoins d'observation de l'enseignant. Dans cette étape, l'apprentissage non-supervisé semble plus approprié afin de pouvoir offrir aux enseignants un moyen de regrouper les apprenants en *clusters* identifiés selon leurs besoins d'observation. Cette direction permettra : (i) de voir si des types de comportements récurrents (*i.e.* identifiés par les propriétés des clusters) apparaissent selon ces besoins et (ii), d'adapter l'enseignement à chaque cluster. Il est possible de déterminer automatiquement les paramètres de certains de ces algorithmes (*e.g.* le *k* pour les *k-means*, les niveaux de coupe pour le clustering hiérarchique), à travers une validation croisée par exemple. Cependant, l'automatisation devra être utilisée avec parcimonie, pour ne pas alourdir notablement le temps d'apprentissage. En outre, ces paramètres devront aussi être inférés en fonction du contexte et des besoins de l'enseignant. En sortie de l'algorithme, l'enseignant pourra constater que le mouvement de l'apprenant n'est pas en phase (*i.e.* dans le même *cluster*) avec ceux réalisés par les étudiants ayant réussi l'exercice.

L'enseignant peut, à partir de là, proposer des exercices spécifiques sur la transition entre accords.

À travers ce système, nous souhaitons proposer un système aidant à l'analyse du mouvement humain dans des situations d'apprentissage. Les différents verrous proposés (techniques et scientifiques) sont visibles à travers les différentes étapes et modules de l'application. Le choix et la manipulation des algorithmes, l'identification des clusters ainsi que la présentation des résultats, doit se faire en conjonction avec l'enseignant et ses besoins. Ce lien est crucial, car il déterminera l'utilisation faite des données, ainsi que les résultats obtenus et proposés à l'enseignant.

5 Conclusion

L'apprentissage de gestes est devenu de nos jours un sujet d'étude crucial. Pour un même geste, les méthodes d'enseignements peuvent différer d'une personne à une autre. De plus, bien qu'une évaluation subjective soit suffisante pour certains domaines, il est essentiel de pouvoir fournir une évaluation exhaustive de la procédure complète, afin d'éviter des erreurs qui peuvent être critiques (*e.g.* chirurgie) [22]. La création d'outils dédiés à l'analyse automatique de mouvement permet d'assister l'utilisateur dans sa tâche d'apprentissage, d'observation et d'évaluation. Actuellement, il semble que les mouvements capturés 3D soient rarement utilisés d'une manière efficace, afin de générer des retours de haut-niveau sur l'apprentissage, autant pour l'apprenant que l'enseignant. Le coût des systèmes d'acquisition, ainsi que la grande dimensionnalité et hétérogénéité des données produites, sont deux obstacles majeurs à ce développement. Cependant, l'utilisation d'algorithmes de *machine learning*, ayant fait leur preuve dans l'analyse des actions et des comportements humains, pourrait aider à cette tâche.

Plusieurs défis techniques et scientifiques existent pour la réalisation d'un tel système : la constitution d'un corpus de mouvements propre au problème, l'hétérogénéité des données, leur dimensionnalité, la formalisation et la transformation des besoins d'observation de l'enseignant, ainsi que le choix d'un algorithme adapté au problème. De plus, un tel système se doit d'être facilement compréhensible et utilisable par des personnes non-expertes. Cet outil doit également être testé et expérimenté auprès d'utilisateurs finaux (*i.e.* enseignants et apprenants), afin de prouver l'utilisabilité et l'adaptation d'un tel système. Bien que les verrous soient identifiés, il est prématuré de dire dans quelle mesure le système sera à même de les lever. La phase d'expérimentation et d'évaluation permettra de déterminer si les pistes envisagées par la littérature seront couronnées de succès ou non. Si les expérimentations fournissent les résultats espérés, les travaux autour de ce système pourront s'étendre à d'autres domaines métier.

Références

1. J. F. Nunes et P. M. Moreira, *Handbook of Research on Computational Simulation and Modeling in Engineering*, 2016, pp. 359-388.

2. J. Huang, W. Zhou, H. Li et W. Li, «Sign Language Recognition using 3D convolutional neural networks,» *Multimedia and Expo (ICME), 2015 IEEE International Conference on*, pp. 1-6, 2015.
3. C.-Y. Chang, C.-W. Chang, J.-Y. Zheng et P.-C. Chung, «Physiological emotion analysis using support vector regression,» *Neurocomputing*, n°1122, pp. 79-87, 2013.
4. Y. Kobayashi, «The EMOSIGN - analyzing the emotion signature in human motion,» *Systems, Man and Cybernetics, 2007. ISIC. IEEE International Conference on*, pp. 1171-1176, 2007.
5. Z. Yu et M. Lee, «Human motion based intent recognition using a deep dynamic neural model,» *Emerging Spatial Competences: From Machine Perception to Sensorimotor Intelligence*, pp. 134-149, Septembre 2015.
6. I. Kapsouras et N. Nikolaidis, «Action recognition on motion capture data using a dynemes and forward differences representation,» *Journal of Visual Communication and Image Representation*, vol. 25, n°16, pp. 1432-1445, 2014.
7. Y. Gu et S. Sosnovsky, «Recognition of Student Intentions in a Virtual Reality Training Environment,» *Proceedings of the Companion Publication of the 19th International Conference on Intelligent User Interfaces*, pp. 69-72, 2014.
8. A. Ng, «CS229 - Machine Learning course, Lecture N 19: Stanford Engineering Everywhere, Stanford University,» 2016. [En ligne]. Available: <https://see.stanford.edu/Course/CS229>. [Accès le 2016].
9. R. Lokaiczuk, A. Faatz, A. Beckhaus et M. Goertz, «Enhancing Just-in-Time E-Learning Through Machine Learning on Desktop Context Sensors,» *Modeling and Using Context: 6th International and Interdisciplinary Conference, CONTEXT 2007, Roskilde, Denmark, August 20-24, 2007. Proceedings*, pp. 330-341, 2007.
10. U. Markowska-Kaczmar, H. Kwasnicka et M. Paradowski, «Computational Intelligence for Technology Enhanced Learning,» *Intelligent Techniques in Personalization of Learning in e-Learning Systems*, pp. 1-23, 2010.
11. A. Menache, *Understanding Motion Capture for Computer Animation*, Second Edition, San Francisco, CA, USA: Morgan Kaufmann Publishers Inc., 2010.
12. «Motion Builder overview,» [En ligne]. Available: <http://www.autodesk.com/products/motionbuilder/overview>. [Accès le 2016].
13. B.-M. Toussaint, *Apprentissage automatique a partir de traces multi-sources heterogenes pour la modelisation de connaissances perceptivo-gestuelles*, 2015.
14. A. Kapur, A. Kapur, N. Virji-Babul, G. Tzanetakis et P. F. Driessen, «Gesture-Based Affective Computing on Motion Capture Data,» *ACII'05 Proceedings of the First international conference ACII'05*, pp. 1-7, 2005.
15. T. A. Mikropoulos et A. Natsis, «Educational virtual environments: A ten-year review of empirical research (1999–2009),» *Computers & Education*, vol. 56, n°13, pp. 769-780, 2011.
16. K. Aminian et B. Najafi, «Capturing human motion using body-fixed sensors: outdoor measurement and clinical applications,» *Computer Animation and Virtual Worlds*, pp. 79-94, 2004.
17. H. Zhou et H. Hu, «Human motion tracking for rehabilitation - A survey,» *Biomedical Signal Processing and Control*, pp. 1-18, 2008.
18. T. Yoshinaga et M. Soga, «Development of a Motion Learning Support System Arranging and Showing Several Coaches' Motion Data,» *Procedia Computer Science*, vol. 60, pp. 1497-1505, 2015.
19. T. Kora, M. Soga et H. Taki, «Golf Learning Environment Enabling Overlaid Display of Expert's Model Motion and Learner's Motion Using Kinect,» *Procedia Computer Science*, vol. 60, pp. 1559-1565, 2015.

20. H.-j. Dong, F. Chong et F.-y. Zhang, «Research on Information Technology in Sport Educational Training System,» *Proceedings of the 2nd International Conference on Green Communications and Networks 2012 (GCN 2012)*, vol. 5, pp. 403-409, 2013.
21. M. Beetz, B. Kirchlechner et M. Lames, «Computerized real-time analysis of football games,» *IEEE Pervasive Computing*, vol. 4, n°13, pp. 33-39, 2005.
22. B.-M. Toussaint et V. Luengo, «Representing and Mining Association Rules from Multisource Heterogeneous Simulation Traces,» *RJCIA 2015*, 2015.
23. K.-S. Choi, X. He, V. Chung-Lim et Z. Deng, «A virtual reality based simulator for learning nasogastric tube placement,» *Computers in Biology and Medicine*, vol. 57, pp. 103-115, 2015.
24. D. Chamaret, S. Ullah, P. Richard et M. Naud, «Integration and evaluation of haptic feedbacks: from CAD models to virtual prototyping,» *International Journal on Interactive Design and Manufacturing (IJIDeM)*, vol. 4, n°12, pp. 87-94, 2010.
25. S. Singh et S. P. Lal, «Educational courseware evaluation using Machine Learning techniques,» *e-Learning, e-Management and e-Services (IC3e), 2013 IEEE Conference on*, pp. 73-78, 2013.
26. D. Sanchez-Mendoza, D. Masip et A. Laepdriza, «Emotion Recognition from Mid-level Features,» *Pattern Recogn. Lett.*, vol. 67, n°1P1, pp. 66-74, 2015.
27. C. M. Bishop, *Pattern Recognition and Machine Learning (Information Science and Statistics)*, Secaucus, NJ, USA: Springer-Verlag New York, Inc., 2006.
28. Y. Wang, B. Zhang et Q. Zhao, «Dynamic data outliers eliminating and filtering Method for motion information of hovercraft,» *2016 IEEE International Conference on Mechatronics and Automation*, pp. 1931-1936, 2016.
29. J. Luengo, S. Garcia et F. Herrera, «On the choice of the best imputation methods for missing values considering three groups of classification methods,» *Knowledge and Information Systems*, vol. 32, n°11, pp. 77-108, 2012.
30. K. C. Ng, «Music via motion: transdomain mapping of motion and sound for interactive performances,» *Proceedings of the IEEE*, vol. 92, n°14, pp. 645-655, 2004.
31. B. Kruger, A. Vogegele, T. Willig, A. Yao, K. Reinhard et A. Weber, «Efficient Unsupervised Temporal Segmentation of Motion Data,» *IEEE Transactions on Multimedia*, vol. PP, n°199, pp. 1-1, 2016.