CLUSTERING AND ANALYSIS OF USER MOTIONS TO ENHANCE HUMAN LEARNING: A FIRST STUDY CASE WITH THE BOTTLE FLIP CHALLENGE

Quentin Couland, Ludovic Hamon and Sébastien George
Laboratoire d'Informatique de l'Université du Mans, LIUM - EA 4023, Le Mans Université, Avenue Olivier Messiaen, 72085 Le Mans, Cedex 9, France

ABSTRACT

More and more domains such as industry, sport, medicine, Human Computer Interaction (HCI) and education analyze user motions to observe human behavior, follow and predict its action, intention and emotion, to interact with computer systems and enhance user experience in Virtual (VR) and Augmented Reality (AR). In the context of human learning of movements, existing software applications and methods rarely use 3D captured motions for pedagogical feedback. This comes from several issues related to the highly complex and dimensional nature of these data, and by the need to correlate this information with the observation needs of the teacher. Such issues could be solved by the use of machine learning techniques, which could provide efficient and complementary feedback in addition to the expert advice, from motion data. The context of the presented work is the improvement of the human learning process of a motion, based on clustering techniques. The main goal is to give advice according to the analysis of clusters representing user profiles during a learning situation. To achieve this purpose, a first step is to work on the separation of the motions into different categories according to a set of well-chosen features. In this way, allowing a better and more accurate analysis of the motion characteristics is expected. An experimentation was conducted with the Bottle Flip Challenge. Human motions were first captured and filtered, in order to compensate for hardware related errors. Descriptors related to speed and acceleration are then computed, and used in two different automatic approaches. The first one tries to separate the motions, using the computed descriptors, and the second one, compares the obtained separation with the ground truth. The results show that, while the obtained partitioning is not relevant to the degree of success of the task, the data are separable using the descriptors.

KEYWORDS

Human Motion, Human Learning, Machine Learning, Clustering

1. INTRODUCTION

Motion capture is increasingly used in multiple domains such as video-game, animation movies, Virtual Reality (VR), sport, medicine, industry and education. Thanks to breakthroughs made in electronics, Human-Computer Interface (HCI) and data processing, it is reasonable to assume that capturing, editing and sharing human gestures will be soon generalized. This assumption has a strong impact on education and on every domain implying human movements. Indeed, different kinds of information can be extracted from human motion analysis. One can easily generate low-level descriptors such as kinematic and dynamic data (Nunes & Moreira, 2016)(Larboulette & Gibet, 2015). Gestures may have a meaning in verbal (Huang, *et al.*, 2015) or non-verbal communication (Chang, *et al.*, 2013). In addition, high-level data linked to human emotion (Kobayashi, 2007), intention (Yu & Lee, 2015) and action (Kapsouras & Nikolaidis, 2014) can be reified and built. Monitoring learner activities can imply the generation of a large amount of motion data that cannot be manually analyzed (Gu & Sosnovsky, 2014). Automatic methods, such as machine learning techniques, can ease such a task. This set of techniques can process high-dimensional data for classification purposes, features extraction, regression problems, etc. (Ng, 2016). In an educational context, these algorithms are used for learning analytics for instance, to study learner actions (Lokaiczyk, *et al.*, 2007) and/or behavior (Markowska-Kaczmar, *et al.*, 2010). Supervised learning can be used in order to classify motions. However, this kind of algorithms implies (i) that a large database of specific motions exists, (ii) that the different classes are known in advance. Furthermore, these works are rarely focused on motions that requires a learning effort from the user. There is a lack of work regarding the automatic extraction of relevant information in pedagogical situations from learner motions. This can be explained by several technical and scientific issues. Some of these issues could be overcome by the use of clustering algorithm, in order to avoid the constraints specific to supervised learning (database size, labeling), and by using morphology-invariant descriptors relevant to the given context. The goal of this work is to use kinematic descriptors along with clustering techniques in order to have a relevant data separation.

The remainder of the paper is structured as follows: section two presents a review of related work, our new approach are shown in section 3, the experimentation and its related protocol, results and discussion are detailed in section 4. Finally, perspectives and future work ends this study.

2. RELATED WORK

Human learning motion can use captured motion, in order to assist the student in his learning task. In this context, the motion is mainly represented as a sequential evolution of human postures through time. Usually, a fixed time-step separates each posture (called "frame"). One way to represent a posture is to build a set of joints, hierarchically structured thanks to a graph, each node describing a joint. This set of joints is organized according to a skeleton model, *i.e.* a tree data structure, in which the root represents the low body part of the torso (i.e. the hip bone) and the nodes represent the body joints. Each node contains the position and the orientation, related to its parent node. It is possible to extract kinematic and dynamic descriptors from this structure such as the speed of the joints, the acceleration, the displacement through time (Nunes & Moreira, 2016) (Larboulette & Gibet, 2015). Zhu and Hu worked on the learning of specific motions for reeducation (Zhou & Hu, 2008). The skeleton model was not systematically considered, because different kinds of sensors were used to gather motion data, depending on the observed movement ; thus, it wasn't systematically possible to construct a skeleton from these data. The data were used in order to analyze the patient's gait. No automatic analyses of the recorded movements were made, the observations and deductions of information were always made by a human expert. For Japanese archery learning, Yoshinaga and Soga developed a system based on a Kinect sensor to capture learner skeletons and its variations through time (Yoshinaga & Soga, 2015). Expert movements were also recorded and learners could compare their motions with the expert ones. The analysis was empirically made by humans.

Works using supervised and unsupervised algorithms to analyze facial expressions, gestures and actions exist. Among them, some were based on 3D captured data. Patrona *et al.* presented a framework for action recognition and evaluation based on extreme learning machine (Patrona, *et al.*, 2018). Using fuzzy-logic, a semantic feedback (depending on the activity context) is given to the learner, such as information about the velocity at specific frames, in order to improve the motion realized. This feedback requires a reference motion and a large corpus of existing motions, as the goal here is to classify the motion into predefined categories in different datasets (CVD exercise, MSRC-12 and MSR-Action3D). Hachaj and Marek used a set of expert rules relating to the learner displacements, *e.g.* the distance covered by the learning in a time step, in order to classify motions (Hachaj & Marek R., 2015). Although these approaches are efficient, the motion does not require a cognitive effort in terms of human learning. Furthermore, the goal is not to evaluate the success degree of the motion and the descriptors cannot be used to give pedagogical feedback. Lui *et al.* worked on video databases from which two sets of descriptors were extracted (Lui, *et al.*, 2011). These descriptors are, on the one hand, localized space-time features that are used with a Bag Of Features approach, and a manifold product on the other hand. The results showed a good data partitioning, especially with the manifold product set of descriptors. The performed motions are trivial in terms of cognitive effort, such as walking, jogging, running, and the descriptors cannot be used to give feedback to the learner. Due to the nature of the motions, the degree of success of the task is not evaluated.

With a sufficient amount of data for the training phase, supervised machine learning algorithms are efficient when the searched and estimated hypothesis is well designed for the problem complexity. However, these kinds of algorithms need a large amount of labeled data related to the given context. The data labeling is usually a costly task in terms of time and resources. Furthermore, some pre-processing steps can change the nature of the data (e.g. PCA), and some decision/separation frontier cannot be easily interpreted by humans (e.g. SVN, Neural Networks). Consequently, analyzing and giving feedback to the learner can be a hard task or impossible to perform. Unsupervised learning approaches, by nature, do not need labeling data to group them into different clusters. It seems that there's a lack of works using unsupervised machine learning algorithms to automatically extract useful pedagogical information from 3D motion data. This could allow to automatically detect the most distinguishing features of a set of motions, group them into learner profiles according to the observation needs of the teachers (i.e. high level descriptors) and help the expert in giving a better feedback to the learner. The presented work is based on the two following hypothesis: (i) for one identified task, it is possible to group motions in separable clusters, with each cluster made of motions with common features, and that (ii) it is possible to automatically group gestures according to the degree of success of the motion-based task. This approach, as well as an experiment are detailed in the next sections.

3. A CLUSTERING APPROACH FOR MOTION ANALYSIS

A motion is not usually described by a perfect example. Instead, in most of the cases, a targeted gesture is defined by one or several experts. Establishing the relevant features allowing to tell if the motion is successful or not depends on the context, the expectations of the experts, which can vary from one to another (i.e. given a learning situation, the set of discriminant features is not the same for every expert). Using supervised learning algorithms implies that a database containing non-trivial and labeled motions in terms of cognitive effort exists. The degree of success of the task of each sample must be stored within the labels. In practice, most of the databases focus on trivial motions, such as sitting, running, walking, etc. The chosen approach relies on the automatic analysis of motions through clustering techniques, in order to avoid most of the drawbacks of the supervised approach. The global context can be seen in Figure 1. From a motion corpus, a first pre-processing step applies several filters, in order to clean the data if needed (frames loss or corrupted, framerate variation, etc.). The next step extracts descriptors from the cleaned motions and the extraction of a wide range of descriptors is possible (Larboulette & Gibet, 2015). One should be careful about them, as some descriptors are morphology-invariant (*e.g.* the ones related to the joints distance), and some are not (*e.g.* the rotation of joints). From here, according to the observation needs of the teacher, the data are analyzed through their descriptors. These descriptors are then used in a clustering process, using the k-means algorithm, from which several metrics are computed to assess its quality. The use of an IT environment and especially a 3D virtual environment allows observing the motion and offering interactions that are hard, or not possible to perform in real environment, e.g. replay motion from several viewpoints, slow down, speed up, pause, etc. From these observations, the expert can then give feedback to the learner, while refining his observation needs.

This paper focuses on the clustering part of Figure 1., implying that clean data are available. An example of such data can be seen in Figure 2c. The goal is to find a set of descriptors, algorithms and metrics, such as (i) the motion corpus can be separated in different groups and (ii) the obtained separation can give an indication of the degree of success of the motion. Such separation would allow analyzing the properties of the clusters, giving information about what the characteristics of each motion type are, and thus giving a more accurate feedback about the needed advice to give for the improvement of the learner motion. The next section presents the experimentation conducted, in order to validate the presented hypotheses.

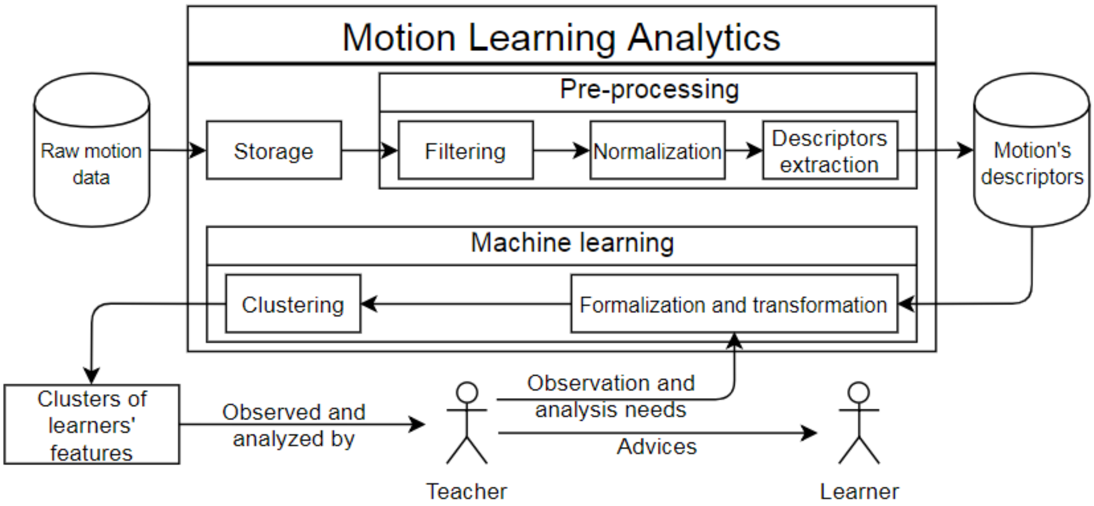

Figure 1. The main principles of a system for assisting the human motion analysis

4. EXPERIMENTATION ON CLUSTERING WITH KINEMATIC DESCRIPTORS

This section is dedicated to an experimentation for the validation of the two previous hypotheses. As a reminder, these assumptions are: (i) it is possible to separate the data into well-defined clusters, and (ii) it is possible to obtain a separation corresponding to the degree of success of the motion. The next paragraphs focus on the protocol used to test the hypotheses, present and discuss the results.

4.1 Protocol

For this experimentation, a database made of short motions requiring some dexterity was created. The Bottle Flip Challenge was the chosen task. The goal is to throw a bottle, such as it completely rotates once on the horizontal axis, and then lands correctly on a table. The distance from the person performing the gesture to the table was empirically set to 70cm (27.5 inches), indicated by a mark on the floor. A MOCAP suit named Perception Neuron and based on Inertial Measurement Units (IMU) was used to capture the motions (https://neuronmocap.com/). It allows capturing 72 joints (some of which are interpolated) at the rate of 60 frames per second. The skeleton of the subject was measured according to the measuring guide, in order to have data skeletons made in accordance with the user morphology. Due to the nature of the sensors, the experimental protocol ensures that (i) no device generating electromagnetic perturbation was close to the user, and (ii) all metallic accessories were removed (including rings, bracelets, watches, belt with metallic buckle, etc.). During the experiment, the MOCAP suit had to be regularly recalibrated, due to the inherent drift of the sensors. Each subject had to perform the motion a hundred times and for every throw, the success (or not) of the task was recorded.

Figure 1.a shows the artifacts of the suit sensors, on the hand's data. Such data are not usable, as the original signal is distorted by the noise. In order to compensate these errors, a Savitsky-Golay filter was applied on each motion (Figure 1.b). Then, the throwing part of the motion was automatically segmented to extract the motion part of interest. (Figure 1.c). From those cleaned data, descriptors were computed. Since the subjects have different body types, morphology-invariants descriptors were chosen: speed and acceleration (vector norm and direction, components along each axis in both cases). The descriptors were computed from three moments of each cleaned motion: beginning of the throw, maximum value of the speed norm for the dominant hand (corresponding to the release of the bottle), and end of the throw. The chosen clustering algorithm is the k-means, as it can gives an insight of the data possible separations, is faster to run than other clustering algorithms (execution time scales linearly with data size), and has easily explainable results. The *k* values ranged from 2 to 10 for this experimentation.

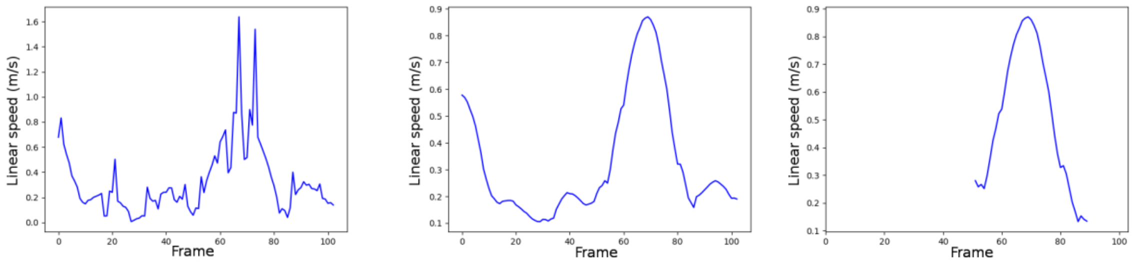

Figure 2. (a): speed of the captured motion through time of the right-hand of a user (b): initial speed filtered (c): extracted throwing part (Couland, et al., 2018)

In order to analyze the clustering results, a few metrics suited to our approaches were chosen.

The first approach was based on the hypothesis that there are various types of motions that can be gathered in separable clusters. In this context, the computed metric is the Average Silhouette Score (*ASS*) (Rousseeuw, 1987). The Silhouette Score (SS) is a metric which compute if a sample belongs well to the cluster it has been assigned (compared to other clusters). The Average Silhouette Score (ASS) is the mean of every sample SS . It gives an indication about the clusters homogeneity: the highest this value is, the better the clusters are separated. This value ranges from -1 to 1, with 1 meaning that every sample is close to the others in the same clusters (the clusters are well separated), and 0 indicating that the clusters are overlapping. This last case, a possible explanation is that the number of clusters is either too low or too high. An ASS between 0 and 0.25 means that no structure is found in the data, a value between 0.25 and 0.5 indicates that a weak structure is found (potentially artificial), an ASS above 0.5 suggests that a reasonable structure is found, while an ASS value above 0.7 means that a strong structure is found (Struyf, et al., 1997). In this context, the metric allows verifying the separation of the clusters, thus giving an indication about the degree of separation with the computed descriptors.

The second approach was based on the hypothesis that it is possible to obtain clusters corresponding to the degree of success of the motion. In our case, our degrees of success are either a successful, or failed throw. A metric such as the accuracy of the clustering seems to not be a relevant indicator. For example, if the k-means algorithm is considered, this metric, based on the computation of a Euclidian distance, is relative to the measured data, the required accuracy of the measuring system and the learning situation. This accuracy is often ascertained by an advanced expert both in the application domain and in computer sciences. In order to verify the difference between the ground truth and the obtained labeling (*i.e.* failed/success motion), the precision, the recall, the F1-score and the Adjusted Rand Index (*ARI*) were chosen. These metrics were only computed for $k=2$, as the ground truth is defined for $k=2$ (successful/failed). As a reminder, the F1-score is a combination of two metrics (recall and precision) representing the labeling accuracy. This value ranges from 0 to 1, with 1 indicating a perfect matching. The ARI is a measure of the similarity between two data partitioning. This index's maximum value is 1, corresponding to a perfect matching between the two labeled clusters and their labeled data. 0 corresponds to a random cluster assignment, and negative values are obtained if the clustering is orthogonal, to an extent.

4.2 Results

The recorded data consisted in 1300 motions, performed by 13 different subjects. 11 subjects were right-handed, and 2 were left-handed. For the clustering, different sets of joints have been considered: hand (H), forearm (FA), arm (A), these body parts being the most solicited during the movement. The computed descriptors were: Speed Norm (SN), Speed value in x, y and z (Sxyz), Speed directions in x, y, and z (SDxyz), and Speed Norm and directions in x, y and z (SNDxyz). The precision (*P*), recall (*R*), F1-score (*F1*) and Adjusted Rand Index (*ARI)* are given for $k=2$, as it corresponds to the ground truth. The Average Silhouette Score (*ASS*) results are also given for $k=2$, as it is the k value that yields the best value in most of the case (the *ASS* values show non-significant variations for other k values when k=2 doesn't yield the best *ASS* values). The clustering was performed on: (i) the mixed data (left and right-handed together) (ii) left-handed data only and (iii) right-handed data only. Table 1. shows the results obtained on this experimentation. F1-score, *ASS* and *ARI* values slightly decreased when joints were added to the dominant hand, meaning that the dominant hand was the most important joint for this case. The highest *ASS* scores were obtained for speed values along the three axis, in the right-handed (0.73) and mixed data (0.54). Left-handed best *ASS* values are for the speed norm values (0.41), yet they are lower than the right-handed and mixed data *ASS* values for the same data (0.42 and 0.48). The *ARI* stayed close to 0, regardless of the joints and descriptors combination (ranging between 0.05 and 0).

Table 1. Clustering metrics for various joints combinations

Joints	H					H, FA					H, FA, A				
Metric	ASS	P	R	F1	ARI	ASS	P	R	F1	ARI	ASS	P	R	F1	ARI
	Left and Right-Handed														
SN	0.48	0.25	0.33	0.29	0.04	0.44	0.25	0.33	0.29	0.04	0.43	0.25	0.33	0.29	0.04
Sxyz	**0.54**	0.18	0.67	0.30	0.05	**0.52**	0.27	0.32	0.29	0.05	**0.51**	0.18	0.68	0.29	0.05
Sdxyz	0.24	0.21	0.53	0.30	0.00	0.27	0.25	0.25	0.25	0.04	0.22	0.18	0.72	0.27	0.04
SNDxyz	0.21	0.18	0.47	0.30	0.00	0.27	0.25	0.26	0.26	0.04	0.22	0.26	0.28	0.27	0.04
	Left Handed														
SN	**0.41**	0.39	0.39	0.39	0.02	**0.42**	0.38	0.39	0.39	0.01	**0.41**	0.31	0.61	0.39	0.01
Sxyz	0.35	0.32	0.57	0.39	0.00	0.34	0.32	0.57	0.39	0.00	0.33	0.35	0.43	0.39	0.00
Sdxyz	0.31	0.34	0.48	0.40	0.00	0.27	0.34	0.54	0.39	0.00	0.23	0.34	0.48	0.40	0.00
SNDxyz	0.27	0.34	0.49	0.40	0.00	0.25	0.33	0.48	0.39	0.00	0.22	0.34	0.52	0.41	0.00
	Right Handed														
SN	0.42	0.18	0.29	0.22	0.00	0.36	0.17	0.28	0.21	0.00	0.34	0.17	0.28	0.21	0.00
Sxyz	**0.73**	0.19	0.12	0.15	0.01	**0.71**	0.19	0.12	0.15	0.01	**0.71**	0.19	0.12	0.15	0.01
Sdxyz	0.28	0.15	0.45	0.28	0.00	0.20	0.16	0.49	0.27	0.00	0.26	0.19	0.13	0.15	0.01
SNDxyz	0.26	0.16	0.45	0.28	0.00	0.19	0.19	0.52	0.27	0.00	0.26	0.17	0.87	0.15	0.01

4.3 Discussion

The combination of the speed vectors in each axis is a good separation criterion, as suggested by results shown in section 4.2. The best ASS values were obtained for the descriptors extracted from the dominant hand, suggesting that other body parts only add noise. This can be partially explained by the fact that every joint motion is related to the other, and that the hand movement is the one with the widest range of values (in terms of speed).

While the *ASS* stayed at an acceptable value (*ASS* ≈ 0.5) for the mixed data, better results were obtained when right-handed and left-handed people are separated (*ASS* ≈ 0.75). The acquisition problems of the suite can explain this phenomenon (and are discussed below in this section). In terms of relative distance, the most discriminant features were the maximum speed value, in both Z (forward) and Y (upward) directions (regarding to the subject), as seen in Table 2.

Table 2. Relative distance of the clusters centroids, for the right hand, with the speed directions in x, y, and z, for k=2

	Beginning	Maximum	End
X (Side)	0.0398	0.5071	0.0110
Y (Upward)	0.0415	1.7497	0.0998
Z (Forward)	0.0847	2.0477	0.0536

The clusters were indeed separable, but the ARI stayed close to 0 for every case (max(ARI) ≈ 0.05), indicating a random cluster assignment. That means that the obtained clusters cannot be related to the outcome of the throw. The current descriptors (speed, acceleration and direction) with the proposed separation model are uncorrelated from the degree of success of the task. One can argue that, the considered task itself does not present a significant variation from one throw to another, in terms of speed and acceleration. Furthermore, the computed descriptors all relies on speed or acceleration, and that can possibly limit the variability of the results. Other higher level descriptors exist (Larboulette & Gibet, 2015), and could be used to analyze the motions. For example, the jerk (rate of change of the acceleration during the motion) can give an indication on how smooth the motion is, and the curvature, which is a measure of how fast a curve is changing through time, can give more accurate data about the wrist rotation. The geometric descriptors, such as the rotation of joints through time, and the center of mass displacement are also interesting values to consider.

In this experimentation, several problems arose. First, the distance between the subject and the table was not constant, as some people took a small step back before throwing. The table was also slippery, and the bottle slid on the table, thus the distance between the subject and the impact point of the bottle cannot be measured. Despite the fact that this measure can be an interesting feature to analyze.

The use of a MOCAP suit limits the experiment to its sensors accuracy and their constraints for a good use, opposed to, for example, an infrared camera system. Having accurate data for the wrist could have been interesting, as its movement is a crucial part of the motion. Furthermore, frame by frame data analysis showed that the data flow was not constant, and that the mandatory software used to gather the data used some undocumented method to counterbalance the data loss, that creates the artifacts seen in Figure 1.a. While the pre-processing steps took care of these problems, nothing can ensure that, the used method did not alter the initial data. Furthermore, the left side of the suit (from the shoulder to the hand) outputted noisy data. When the clustering was performed, mixing left-handed and right-handed data yielded worse results than keeping only the right-handed subjects, due to noisy nature of the left-handed data (Figure 3). This noise was visible on the captured data, and it is due to the fact that the suit has difficulties to handle a capture of the full body.

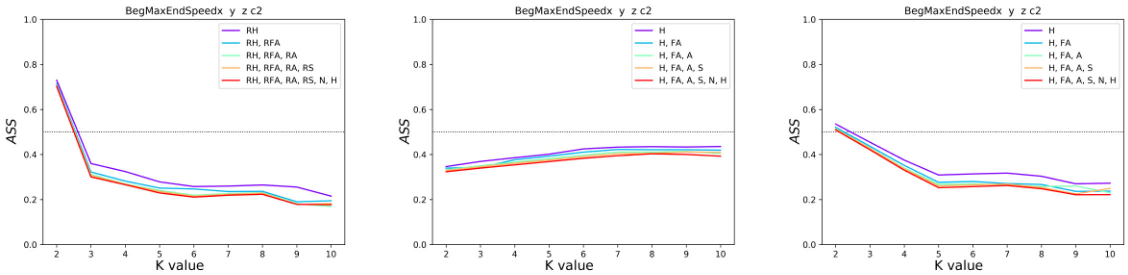

Figure 3. *ASS* score for various joints combinations and k ranging from 2 to 10 of (a) the right-handed subjects (b) the left-handed subjects (c) left and right-handed subjects together

As the motion variability of the chosen task can be discussed, another experiment was conducted to verify if the computed descriptors, combined with the k-means algorithm, can separate the motions according to the ground truth. In this experiment, a subject must throw a ball in one of two bins, placed in a line front on him (one placed 2m (6,56 ft) from them, another one placed 3.5m (11,48 ft) from them). The subject has to perform 100 throws, without any constraints about the throwing motion. For each throw, (i) the degree of success of the throw, (ii) the bin aimed at, and (iii) the type of throw (*i.e.* basket type launch, bowling type launch), were recorded . Having multiple labels for each motion allow for a wider range of tests, and allows to work on the degree of success, as well as the descriptors' ability to discriminate in various cases. Early results have shown that while the *ASS* and *ARS* values stay the same as the first experimentation for the successful/failed labeling, the clustering gives a good *ARS* for the throwing type, with the norm, and "norm + directions" descriptors. Further work is needed in order to validate these results on a larger scale.

5. CONCLUSION AND PERSPECTIVES

A new approach regarding the analysis of 3D motions was presented in this paper. The goal is to give a method to analyze the motion, through explainable descriptors extracted from it, leading to personalized feedback given to the learner in order to improve his motion. After acquiring and processing the motion data, some descriptors based on speed, acceleration and direction were extracted from it. These descriptors were then used in a clustering process, in order to find different explainable types of motions. This approach relied on two hypotheses: (i) it is possible to separate the motions into explainable clusters (ii) it is possible to obtain partitions corresponding to the degree of success of the task. While the second objective did not reach the expectations, the results of the first objective showed that the separation of clusters is indeed possible, validating the hypothesis, and the used descriptors (with the proposed method) in terms of discriminant features. The computation of more descriptors is planned, as the current ones may be limited, regardless of the application context. As the data are time series, the use of Dynamic Time Warping (DTW), computing a distance between motions (Morel, 2017), would provide another similarity measure between them, giving inter and intra-clusters information about the motions. Future work will also focus on performing recursive clustering on obtained clusters, in order to find if the motions, in each cluster, are separable according to the degree of success of the task or other features. The ongoing second experimentation will allow testing the new considered descriptors, as well as generalizing the context in which each descriptor is the best suited.

REFERENCES

Chang, C.-Y., Chang, C.-W., Zheng, J.-Y. & Chung, P.-C., 2013. Physiological emotion analysis using support vector regression. *Neurocomputing,* Issue 122, pp. 79-87.

Couland, Q., Hamon, L. & George, S., 2018. Enhancing Human Learning of Motions: An Approach Through Clustering. *European Conference on Technology Enhanced Learning.*

Gu, Y. & Sosnovsky, S., 2014. 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.

Hachaj, T. & Marek R., O., 2015. Full Body Movements Recognition - Unsupervised Learning Approach with Heuristic R-GDL Method. *Digital Signal Processing,* Volume 46, pp. 239-252.

Huang, J., Zhou, W., Li, H. & Li, W., 2015. Sign Language Recognition using 3D convolutional neural networks. *Multimedia and Expo (ICME), 2015 IEEE International Conference on,* pp. 1-6.

Kapsouras, I. & Nikolaidis, N., 2014. Action recognition on motion capture data using a dynemes and forward differences representation. *Journal of Visual Communication and Image Representation,* 25(6), pp. 1432-1445.

Kobayashi, Y., 2007. The EMOSIGN - analyzing the emotion signature in human motion. *Systems, Man and Cybernetics, 2007. ISIC. IEEE International Conference on,* pp. 1171-1176.

Larboulette, C. & Gibet, S., 2015. A Review of Computable Expressive Descriptors of Human Motion. *Proceedings of the 2Nd International Workshop on Movement and Computing,* pp. 21--28.

Lokaiczyk, R., Faatz, A., Beckhaus, A. & Goertz, M., 2007. 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.

Lui, Y. M., O'Hara, S. & Draper, B. A., 2011. Unsupervised Learning of Humain Expressions, Gestures, and Actions. *Face and Gesture 2011,* pp. 1-8.

Markowska-Kaczmar, U., Kwasnicka, H. & Paradowski, M., 2010. Computational Intelligence for Technology Enhanced Learning. *Intelligent Techniques in Personalization of Learning in e-Learning Systems,* pp. 1-23.

Morel, M., 2017. *Multidimensional time-series averaging: application to automatic and generic evaluation of sport gestures,* s.l.: s.n.

Ng, A., 2016. *CS229 - Machine Learning course, Lecture N 19: Stanford Engineering Everywhere, Stanford University.* [Online] Available at: https://see.stanford.edu/Course/CS229.

Nunes, J. F. & Moreira, P. M., 2016. *Handbook of Research on Computational Simulation and Modeling in Engineering.* s.l.:s.n.

Patrona, F., Chatzitofis, A., Zarpalas, D. & Daras, P., 2018. Motion analysis: Action detection, recognition and evaluation based on motion capture data. *Pattern Recognition,* Volume 76, pp. 612-622.

Rousseeuw, P. J., 1987. Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. *Journal of Computational and Applied Mathematics,* Volume 20, pp. 53-65.

Sanchez-Mendosa, D., Masip, D. & Laepdriza, A., 2015. Emotion Recognition from Mid-level Features. *pattern Recognition Lett.,* 4(12), pp. 87-94.

Struyf, A., Hubert, M. & Rousseeuw, P., 1997. Clustering in an Object-Oriented Environment. *Journal of Statistical Software, Articles,* 1(4), pp. 1--30.

Yoshinaga, T. & Soga, M., 2015. Development of a Motion Learning Support System Arranging and Showing Several Coaches' Motion Data. *Procedia Computer Science,* Volume 60, pp. 1497-1505.

Yu, Z. & Lee, M., 2015. Human motion based intent recognition using a deep dynamic neural model. *Emerging Spatial Competences: From Machine Perception to Sensorimotor Intelligence,* Septembre, pp. 134-149.

Zhou, H. & Hu, H., 2008. Human motion tracking for rehabilitation - A survey. *Biomedical Signal Processing and Control,* pp. 1-18.