



Enhancing Human Learning of Motions: An Approach Through Clustering

Quentin Couland, Ludovic Hamon, Sébastien George

► To cite this version:

Quentin Couland, Ludovic Hamon, Sébastien George. Enhancing Human Learning of Motions: An Approach Through Clustering. 13th European Conference on Technology Enhanced Learning, EC-TEL 2018, Sep 2018, Leeds, United Kingdom. pp.591-595, 10.1007/978-3-319-98572-5_52 . hal-02001863

HAL Id: hal-02001863

<https://hal.science/hal-02001863>

Submitted on 27 Feb 2024

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.

Enhancing Human Learning of Motions: An Approach Through Clustering

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
{quentin.couland,ludovic.hamon,sebastien.george}@univ-lemans.fr

Abstract. More and more software applications use human motions to improve the information retention. Some virtual environments are especially built to support the learning of human motions. However, these kinds of applications and their pedagogical feedback are rarely made from the analysis of 3D captured motions. This can be explained by the heterogeneity, the complexity and the high-dimensional nature of such data. However, machine learning techniques could be used to overcome these issues. This paper presents a first step towards the improvement of the human learning process of a motion, thanks to the analysis of clusters representing user profiles. In the context of the Bottle Flip Challenge and using raw captured motions, descriptors based on speed and acceleration are extracted. The motions are then automatically analyzed, according to two different approaches: one with the ground truth, and one without constraints on the number of clusters. The results suggest that the data are separable using the computed descriptors.

Keywords: Human motion · Human learning · Machine learning
Clustering

1 Related Work

Captured motions have been used in various fields related to human learning of motions. Indeed, it is possible to extract cinematic and dynamic data from low level raw data (*i.e.* the evolution of joints position through time) [2,4]. These methods require an expert to analyze the data and give a feedback to the learner. Some works used supervised and unsupervised algorithms to analyze facial expressions, movements and actions. Among them 3D captured data were studied with a set of expert rules relating to the learner displacement [1]. Although these approaches are efficient, the motions do not require a cognitive effort in terms of human learning. Furthermore, the goal was not to evaluate the degree of learning and/or success of the motion, and the descriptors could not be used to give a pedagogical feedback. The use of supervised machine learning algorithms assumes that: (a) labeled data exist for the specific problem, and (b) there is a sufficient amount of data to

allow the algorithm to converge. In practical cases, (a) is often in contradiction with (b). Furthermore, the results are not always explainable as the descriptors can be modified through the algorithm (*e.g.* using PCA), and the separation model is not humanly interpretable, or with difficulty (*e.g.* SVN or Neural Networks). Unsupervised learning approaches have other requirements and, in a human learning context, could help the teachers to regroup the learners in different clusters identified by their observation needs. This would allow (i) seeing if recurring behaviors appear regarding to these needs, and (ii) adapting the learning process to each cluster *i.e.* learner profile.

2 Motion Analysis: A Clustering Based Approach

Establishing the relevant features that allow telling if a motion is successful (or not) is a non-trivial task. For a given task, there is not one or several perfect motions. The use of supervised machine learning algorithms assumes that large databases of labeled motions exist; yet, there are very few specialized database containing the same kind of motions with several degrees of success and that requires a cognitive effort in terms of human learning. Taking into account all of these elements, the automatic analysis of motions through clustering techniques is the chosen approach.

2.1 Protocol

A database made of non-trivial motions to learn (in terms of human learning) was created. For this study, the Bottle Flip Challenge¹ was chosen as the learning task. This task requires some dexterity, and the execution time is short. To capture the motions, a MOCAP suit based on Inertial Measurement Units (IMU) was chosen². Figure 1a shows the hand's captured motion. One can see that there are a lot of artifacts, due to the capture system. These data were filtered using the Savitzky-Golay filter, in order to eliminate the errors and noises related to the MOCAP system (Fig. 1b). For each motion, the throwing part of the motion was automatically segmented (Fig. 1c), then rebuilt with a fixed number of frames. The speed and acceleration data, as well as the corresponding directions along the x, y and z axis, were extracted from this rebuilt motion from: (i) the beginning of the throw, (ii) the maximum speed value and (iii) the end of the throw.

3 Results

The tests were made on a set of 13 people's data, consisting of 1300 throws in total. Different sets of joints have been used: hand (H), forearm (FA), arm (A), shoulder (S), these body parts being the most solicited during the motion.

¹ https://en.wikipedia.org/wiki/Bottle_flipping.

² <https://neuronmocap.com/>.

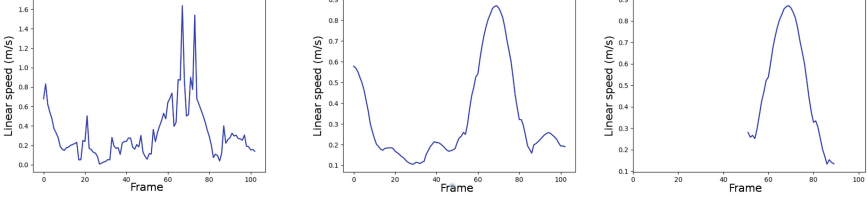


Fig. 1. (a) Speed of the captured motion through time of the right-hand of an user (b) Initial speed filtered (c) Extracted throwing part.

There are two main objectives here: (i) determine if the data can be partitioned and (ii) determine if it's possible to obtain a clustering based on the degree of success of the task.

3.1 Cohesion and Separation of the Clusters

This approach was based on the hypothesis that there are different types of motions that can be gathered in separable clusters. In this context, the computed metric is the Average Silhouette Score (ASS) [6]. The Silhouette Score (SS) gives a value indicating how a sample is well-fitted to the assigned cluster, compared to other clusters. The Average Silhouette Score (ASS) is the mean of every sample's SS. This value ranges from -1 to 1 , with 1 indicating that in average, every sample best belongs to their cluster, and 0 indicating that the clusters are overlapping. An ASS above 0.5 indicates that a reasonable structure is found in the data, while an ASS above 0.7 indicates that a strong structure is found [7]. Table 1 shows the results for the five best data combinations, with the highest ASS score being 0.7038 , suggesting that our data are separable.

Table 1. $\max(ASS)$ (for k varying from 2 to 10) for various joints and data combinations.

Data type	H	H, FA	H, FA, A	H, FA, A, S
BegMaxEnd Speed[x/y/z]	0.7038 ($k = 2$)	0.6734 ($k = 2$)	0.6677 ($k = 2$)	0.6650 ($k = 2$)
BegMaxEnd Speed Norm	0.5147 ($k = 2$)	0.3992 ($k = 2$)	0.3869 ($k = 2$)	0.3803 ($k = 2$)
BegMaxEnd Speed Norm,Dir[x/y/z]	0.3355 ($k = 2$)	0.3271 ($k = 2$)	0.2665 ($k = 2$)	0.2496 ($k = 2$)

3.2 Ground Truth Approach

The second hypothesis relies on the fact that each cluster corresponds to a success degree of the task. The k-means algorithm was run with $k = 2$ and $k = 3$, on the data. The clusters were then analyzed, in order to verify their contents in terms of motions leading to a “successful/failed” throw for $k = 2$

and leading to a “successful/almost successful/failed” throw for $k = 3$. For this experiment, the Adjusted Rand Index (ARI) was chosen [5], to verify if the obtained labeling was similar to the ground truth. This metric is a measure of the similarity between two data partitioning. This approach yielded low scores ($ARI \approx 0$), indicating that the labeling did not match the ground truth, showing that the speed combined to the partitioning strategy is not an indication of the degree of success of the task.

4 Discussion and Future Work

An approach of 3D captured motion analysis based on clustering was proposed. The goal is to assist the learner in their motion learning task, by allowing the expert to analyze the learner’s motion, and giving them a way to adapt the learning process. This approach was based on two hypotheses: (i) it is possible to find an explainable partitioning of the data, and (ii) it is possible to automatically separate the motions based on the degree of success of the task. Results shown in Sect. 3.1 suggest that the combination of the speed vectors on each axis is a good separation criterion. Having the best *ASS* values for the hand shows that the hand’s descriptors were the most significant. In addition, the most discriminant features were the speed at the throw moment, in both forward and up direction (regarding to the body of the person throwing), in terms of relative distance: 2.82 and 2.78, respectively (between 0.04 to 0.3 for the other speed values). Consequently, the motions were indeed separable, which validated the chosen indicators in terms of discriminant features. However, the results in Sect. 3.2 show that the *ARI* value is close to 0 in every case, suggesting a random data assignment. Hence, the current descriptors, as well as the considered approach, seem to be irrelevant to assess the degree of success of the performed motion. Considering the chosen features, it seems that the considered task does not have a significant variation. The chosen descriptors are low-level ones (kinematic/dynamic) [2], and using higher-level ones could allow separating the motions on more meaningful features, thus allowing a better analysis. The use of Dynamic Time Warping algorithm, computing a distance between time series [3], would provide another similarity measures between the motions. Performing recursive clustering on well separated clusters is another lead, as it could allow to determine more accurate learner profiles.

References

1. Hachaj, T., Ogiela, M.R.: Full body movements recognition - unsupervised learning approach with heuristic R-GDL method. *Digit. Signal Process.* **46**, 239–252 (2015)
2. Larboulette, C., Gibet, S.: A review of computable expressive descriptors of human motion. In: *Proceedings of the 2nd International Workshop on Movement and Computing, MOCO 2015*, pp. 21–28. ACM, New York (2015)
3. Morel, M.: Multidimensional time-series averaging: application to automatic and generic evaluation of sport gestures. Theses, Université Pierre & Marie Curie, November 2017. <https://hal.archives-ouvertes.fr/tel-01659753>

4. Nunes, J.F., Moreira, P.M.: Handbook of Research on Computational Simulation and Modeling in Engineering (2016)
5. Rand, W.M.: Objective criteria for the evaluation of clustering methods. J. Am. Stat. Assoc. **66**(336), 846–850 (1971)
6. Rousseeuw, P.J.: Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. J. Comput. Appl. Math. **20**, 53–65 (1987)
7. Struyf, A., Hubert, M., Rousseeuw, P.: Clustering in an object-oriented environment. J. Stat. Softw. **1**(4), 1–30 (1997)