EG UK Computer Graphics & Visual Computing (2018)
G. Tam and F. Vidal (Editors)

Character Animation Reconstruction from Content-Based Motion Retrieval

M. Wu [1,2] and Z. Xiao[1]

[1] National Centre for Computer Animation, Bournemouth University, UK
[2] Centre for Digital Entertainment, Bournemouth University, UK

Abstract

We present the initial design of a motion reconstruction framework for character animation which encompasses the use of supervised and unsupervised learning techniques for the retrieval and synthesis of new realistic motion. Taking advantage of the large amounts of Motion Capture data accumulated over the years, our aim is to shorten animation production times by providing animators with more control over the specification of high-level parameters and a user-friendly way of retrieving and reusing this data, applying clustering to organize the human motion database and Neural Networks for motion generation.

CCS Concepts

•***Computing methodologies*** → *Motion Capture; Machine learning; Motion processing;*

1. Introduction

Advancements in technologies have made Motion Capture (MoCap) a popular technique to capture human movements, with significant application in the production of movies, animations and video games. Although a large volume of human motion data has been collected over the years, animators rarely reuse it when creating new content. Studios, in fact, tend to capture new motion data due to their need for specific and stylized motion. Despite the apparent benefits and the more affordable MoCap technologies, costs of hiring are still high. By taking advantage of the available data instead, one could limit the need for extra captures. Key elements to the success of this approach are an efficient method for searching and retrieving motions from databases and a user-friendly data-driven motion synthesis system. We aim to combine these two techniques and propose the initial design of a novel framework, which removes the need for manual labeling of motion clips with the application of unsupervised learning algorithms, and provides the animator with more control over the generation of high quality animations using as inputs low dimensional, high-level parameters.

2. Related Work

An important step towards the data-driven synthesis of realistic animations is to have an efficient way of managing and retrieving motions from databases. Core element of the success of motion retrieval algorithms is the chosen motion feature representation. In [WFQ*16], the authors argue that a single low-level visual feature is usually unable to fully characterize all aspects of motion data. In their proposed method, they create a multiple visual feature set and then automatically select a compact and discriminative subset from it, thus improving the accuracy and speed of the retrieval process. Among other content-based motion retrieval methods, [LJH*18] proposes a generic retrieval scheme for heterogeneous MoCap data, using motion signatures to describe both high- and low-level characteristics of a subject in motion. Recently, a retrieval sketch-based approach [XTFX15] has been introduced, which selects the most discriminative feature component of a 2D Geometric Posture Descriptor as feature representation, and allows for the retrieval of motion sequences by sketching key postures.

With the increased popularity of deep learning algorithms, in the last few years many methods have been proposed that use Convolutional Neural Networks (CNN) [HSK16] for both motion retrieval and motion synthesis purposes. Other approaches involving Recurrent Neural Networks (RNN) [ZLX*18] highlight their suitability and potential for learning time-series data such as human motion, though they still present some shortcomings that [BBKK17] propose to avoid through a feed-forward temporal encoder of motion data used for both action classification and motion prediction.

3. Framework Overview

The diagram in Figure 1 shows the initial design of our proposed framework: in the *Training* stage, from a database of human motions, a relevant and discriminative feature vector is used for the purpose of clustering. Following the division into subgroups, we can consider these clusters as reasonable classes: a supervised learning model is then trained on these results and used for classification when new data is added to the database. The resulting learned motion representation is then used for motion retrieval. In the *Synthesis* stage, the animator should be able to specify, via a

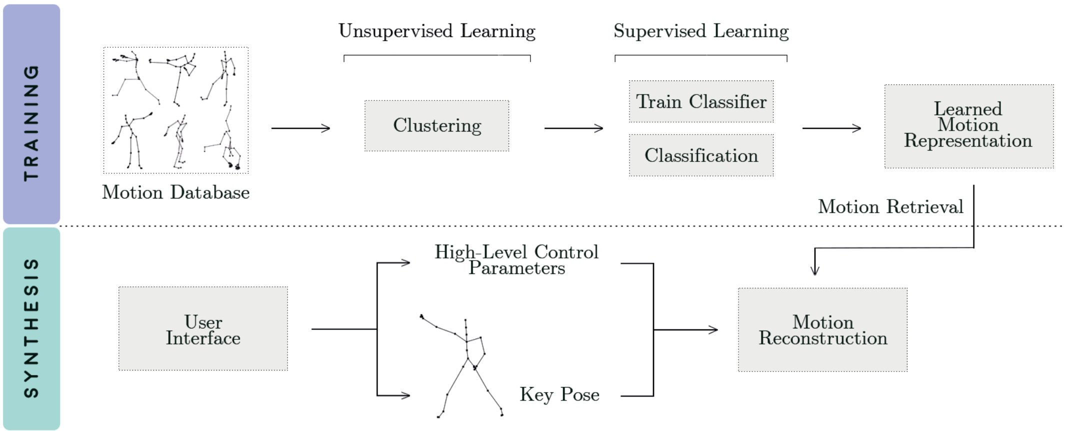

Figure 1: *Framework design: in the Training stage (top), clustering is performed over the motion database. Then, a supervised learning model is trained on the resulting partitions and used as a classifier when new data is added to the database. In the Synthesis stage (bottom), the user specifies high-level control parameters and provides key poses to retrieve motions and synthesize new character animations.*

user-friendly interface, high-level control parameters, such as temporal and spatial constraints [HSK16], and provide a set of key poses for the retrieval of correlated motions from the database for character animation reconstruction.

3.1. Clustering

Critical to the success of a clustering algorithm is the choice of the right features: for human motion data, a logic-based feature such as the General Pose Feature (GPF) [WFQ*16], would be suitable for our purposes as it considers not only joints positions, but also velocity, acceleration and relative distance to other joints. In regards to clustering, various methods have been considered: among partitional clustering, k-means is the most popular and intuitive; however, it is sensitive to outliers and requires the specification a priori of the k number of clusters. Moreover, the random initialization of the clusters' centroids could lead to diverse results for multiple runs of the algorithm. The issue of the unknown number of motion clusters could be obviated by using hierarchical clustering instead, leading to a dendrogram representation of motions potentially suitable for describing their specialized subclasses. For heterogeneous motion sequences, a soft method like fuzzy clustering would theoretically be more suited since samples can belong to more than one cluster. Fuzzy c-means (FCM) and Gaussian Mixture Model (GMM) are the most widely used methods among the soft clustering group. However, only a comparison and evaluation through experiments can identify the best clustering method for motion data.

4. Conclusions & Future Work

We have presented an overview of our motion reconstruction framework, with the aim of giving animators more control over the generation of high quality animations and optimizing production times. We are especially focusing on how to handle unlabeled data similarly to [XWCH15] to avoid time-consuming manual labeling, which justifies our choice of using an unsupervised learning method (i.e. clustering) in the motion management phase, though contrary to their approach we plan to do everything offline. We also want to further investigate the best approach in deploying Neural Networks for motion synthesis to obtain a more robust system: Long Short-Term Memory (LSTM) networks seem well suited in predicting human motion as demonstrated in [ZLX*18]. However, since the network output is fed back into itself, the issues of error accumulation lead to the motion either becoming unrealistic or the network failing. Alternative approaches we would like to consider are Generative Adversarial Networks (GAN), an intersection of different networks or a modification of the LSTM network itself.

References

[BBKK17] BÜTEPAGE J., BLACK M. J., KRAGIC D., KJELLSTRÖM H.: Deep representation learning for human motion prediction and classification. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (2017), IEEE, p. 2017. doi:10.1109/CVPR.2017.173. 1

[HSK16] HOLDEN D., SAITO J., KOMURA T.: A deep learning framework for character motion synthesis and editing. *ACM Trans. Graph. 35*, 4 (2016), 138:1–138:11. doi:10.1145/2897824.2925975. 1, 2

[LJH*18] LV N., JIANG Z., HUANG Y., MENG X., MEENAKSHISUNDARAM G., PENG J.: Generic content-based retrieval of marker-based motion capture data. *IEEE Transactions on Visualization and Computer Graphics 24*, 6 (June 2018), 1969–1982. doi:10.1109/TVCG.2017.2702620. 1

[WFQ*16] WANG Z., FENG Y., QI T., YANG X., ZHANG J. J.: Adaptive multi-view feature selection for human motion retrieval. *Signal Processing 120* (2016), 691 – 701. doi:10.1016/j.sigpro.2014.11.015. 1, 2

[XTFX15] XIAO J., TANG Z., FENG Y., XIAO Z.: Sketch-based human motion retrieval via selected 2d geometric posture descriptor. *Signal Processing 113* (2015), 1 – 8. doi:10.1016/j.sigpro.2015.01.004. 1

[XWCH15] XIA S., WANG C., CHAI J., HODGINS J.: Realtime style transfer for unlabeled heterogeneous human motion. *ACM Trans. Graph. 34*, 4 (2015), 119:1–119:10. doi:10.1145/2766999. 2

[ZLX*18] ZHOU Y., LI Z., XIAO S., HE C., HUANG Z., LI H.: Auto-conditioned recurrent networks for extended complex human motion synthesis. In *International Conference on Learning Representations* (2018). URL: http://arxiv.org/abs/1707.05363. 1, 2