Similarity Search in 3D Human Motion Data

Jan Sedmidubsky
Masaryk University
Brno, Czech Republic
xsedmid@fi.muni.cz

Pavel Zezula
Masaryk University
Brno, Czech Republic
zezula@fi.muni.cz

ABSTRACT

Motion capture technologies can digitize human movements into a discrete sequence of 3D skeletons. Such spatio-temporal data have a great application potential in many fields, ranging from computer animation, through security and sports to medicine, but their computerized processing is a difficult problem. The objective of this tutorial is to explain fundamental principles and technologies designed for searching, subsequence matching, classification and action detection in the 3D human motion data. These operations inherently require the concept of similarity to determine the degree of accordance between pairs of 3D skeleton sequences. Such similarity can be modeled using a generic approach of metric space by extracting effective deep features and comparing them by efficient distance functions. The metric-space approach also enables applying traditional index structures to efficiently access large datasets of skeleton sequences. We demonstrate the functionality of selected motion-processing operations by interactive web applications.

CCS CONCEPTS

• **Information systems** → *Similarity measures*; *Clustering and classification*; **Multimedia and multimodal retrieval**; • **Computing methodologies** → *Supervised learning by classification*;

KEYWORDS

motion capture data; 3D skeleton sequence; similarity search; subsequence matching; annotation; action detection; stream processing

ACM Reference Format:
Jan Sedmidubsky and Pavel Zezula. 2019. Similarity Search in 3D Human Motion Data. In *International Conference on Multimedia Retrieval (ICMR '19), June 10–13, 2019, Ottawa, ON, Canada.* ACM, New York, NY, USA, 2 pages. https://doi.org/10.1145/3323873.3326589

1 INTRODUCTION AND MOTIVATION

The availability of various kinds of capturing technologies enables digitizing human movements by tracking 3D positions of specific skeleton joints in time. The interest in capturing these spatio-temporal multimedia data is continuously growing and new application scenarios emerge in a variety of fields. For example, the data could be employed in military to virtually simulate a combat and conflict-resolving situations; in law-enforcement to identify suspicious subjects or events; at smart homes to detect anomalous behavior or body positions of elderly people; in sports to analyze ball games or to detect fouls; or in medicine to evaluate the progress in rehabilitation or to discover movement disorders as indicators for choosing suitable treatments.

A great application potential together with a growing availability of capturing devices indicate a fast increase of 3D motion data volume in the near future. Even though storing such quantity of data makes an issue, their intelligent management is a much more challenging problem. Content-based processing techniques, such as recognition, searching and annotation, are crucial to fully exploit the data potential and make the expensively recorded data more accessible, valuable and reusable. This requires employing completely different data-processing paradigms compared to the traditional domains such as text or images. With the explosion of deep-learning research, there is also a fast growth of research papers that process 3D motion data through neural networks. Such papers have been recently published in high-impacted conferences such as ACM Multimedia or CVPR, as well as journals like Multimedia Tools and Applications or IEEE Transactions on Pattern Analysis and Machine Intelligence.

The objective of this tutorial is to present the state-of-the-art approaches, principles and technologies developed for comparison, retrieval, recognition and annotation of 3D human motion sequences, as illustrated in Figure 1.

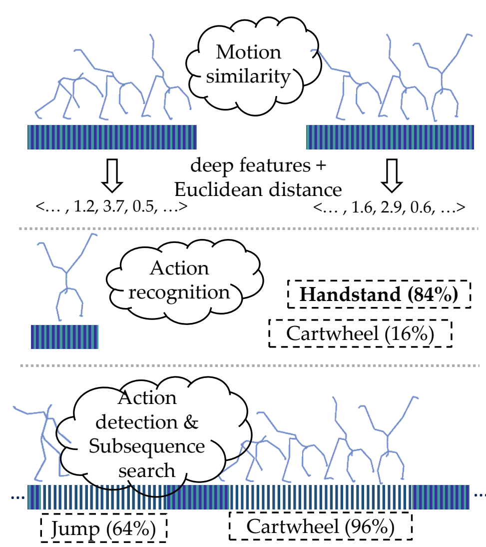

Figure 1: Illustration of similarity comparison, action recognition, action detection and subsequence search operations.

ICMR '19, June 10–13, 2019, Ottawa, ON, Canada

ACM ISBN 978-1-4503-6765-3/19/06…$15.00
https://doi.org/10.1145/3323873.3326589

2 SPECIFIC OBJECTIVES

Intelligent processing of 3D human motion data requires to follow patterns used in real-life evolution and communication between species. There, recognition, learning and judgment presuppose an ability to categorize stimuli and classify situations by *similarity*, which is subjective and context-dependent. We model such similarity using the *metric space* approach [16] that assumes the availability of content-preserving *features* extracted from skeleton sequences and a distance function for their comparison. We present machine-learning approaches that learn effective features by encoding 3D skeleton sequences into 2D image representations on which convolutional neural networks (CNN) are trained [5, 11]. To better learn long-term temporal dependencies, the Long Short-Term Memory (LSTM) variant of recurrent neural networks is used and directly trained on raw 3D joint coordinates [6, 8, 15]. The quality of features can be enhanced by fusing CNN and LSTM architectures [9] or employing a triplet-loss learning [1] by providing training data as pairs of similar and dissimilar actions.

Once the features of actions are extracted, they can be indexed by any metric access structure. We present existing tree structures, hashing techniques, and pivot-permutation indexes [16]. We show how such index structures can speed-up recognition of actions, detection of actions within continuous skeleton streams, or searching for similar occurrences of query actions within a very long skeleton sequence. Specifically, we can use indexes to efficiently find similar actions with respect to a query and process them by k-nearest neighbor classifiers to recognize the query class [11, 13]. There are also approaches [2, 4] able to detect and recognize user-provided actions within unsegmented skeleton streams. Such stream annotation is more difficult as the beginnings and endings of actions are unknown and have to be determined. Some approaches detect actions even before they finish, which is suitable for early action detection [7] or future action prediction [3]. If no actions are provided in advance, subsequence search can be applied to find query-similar occurrences within a very long 3D skeleton sequence. The sequence is firstly partitioned into short segments that are then indexed to efficiently retrieve the ones similar to a query [10, 14]. In addition, indexing of segments of variable sizes enables localizing subsequences that are executed faster or slower [12].

3 INTERACTIVE DEMOS

The presented topics will include not only the state-of-the-art approaches, but also interactive web applications developed by the authors of this tutorial, such as subsequence search demo available online at http://disa.fi.muni.cz/mocap-demo/.

4 BRIEF BIOGRAPHY

Jan Sedmidubsky is a researcher of computer science at Masaryk University (Czech Republic) where he received the Ph.D. degree in 2011 along with the dean's and rector's prize for a distinguished dissertation thesis. His research activities are primarily concentrated on similarity processing of 3D human motion data. He is a co-author of about 40 research publications.

Pavel Zezula is a professor of computer science at Masaryk University (Czech Republic). His professional interests primarily concern multimedia content-based retrieval, large-scale similarity search, and big data analysis. He is a co-author of seminal similarity search structure, the "M-Tree", and the book "Similarity Search: The Metric Space Approach" by Springer US. He is also a co-author of more than 150 research publications with more than 6,000 citations.

ACKNOWLEDGMENTS

This research is supported by the Czech Science Foundation project No. GA19-02033S.

REFERENCES

[1] Andreas Aristidou, Daniel Cohen-Or, Jessica K. Hodgins, Yiorgos Chrysanthou, and Ariel Shamir. 2018. Deep Motifs and Motion Signatures. *ACM Transactions on Graphics* 37, 6 (2018), 187:1–187:13. https://doi.org/10.1145/3272127.3275038

[2] Said Yacine Boulahia, Eric Anquetil, Franck Multon, and Richard Kulpa. 2018. CuDi3D: Curvilinear displacement based approach for online 3D action detection. *Computer Vision and Image Understanding* 174 (2018), 57–69. https://doi.org/10.1016/j.cviu.2018.07.003

[3] Judith Butepage, Michael J. Black, Danica Kragic, and Hedvig Kjellstrom. 2017. Deep Representation Learning for Human Motion Prediction and Classification. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE, 6158–6166.

[4] Petr Elias, Jan Sedmidubsky, and Pavel Zezula. 2017. A Real-Time Annotation of Motion Data Streams. In *19th International Symposium on Multimedia*. IEEE Computer Society, 154–161.

[5] Sohaib Laraba, Mohammed Brahimi, Joelle Tilmanne, and Thierry Dutoit. 2017. 3D skeleton-based action recognition by representing motion capture sequences as 2D-RGB images. *Computer Animation and Virtual Worlds* 28, 3-4 (2017), e1782.

[6] Chaolong Li, Zhen Cui, Wenming Zheng, Chunyan Xu, and Jian Yang. 2018. Spatio-Temporal Graph Convolution for Skeleton Based Action Recognition. In *32nd International Conference on Artificial Intelligence (AAAI)*. AAAI Press, 7444–7452.

[7] Sheng Li, Kang Li, and Yun Fu. 2018. Early Recognition of 3D Human Actions. *ACM Trans. Multimedia Comput. Commun. Appl.* 14, 1s, Article 20 (2018), 21 pages.

[8] Jun Liu, Gang Wang, Ling-Yu Duan, Ping Hu, and Alex C. Kot. 2018. Skeleton Based Human Action Recognition with Global Context-Aware Attention LSTM Networks. *IEEE Transactions on Image Processing* 27, 4 (2018), 1586–1599.

[9] Juan C. Nunez, Raul Cabido, Juan J. Pantrigo, Antonio S. Montemayor, and Jose F. Velez. 2018. Convolutional Neural Networks and Long Short-Term Memory for skeleton-based human activity and hand gesture recognition. *Pattern Recognition* 76 (2018), 80–94.

[10] Jan Sedmidubsky, Petr Elias, and Pavel Zezula. 2016. Similarity Searching in Long Sequences of Motion Capture Data. In *9th International Conference on Similarity Search and Applications (SISAP)*. Springer International Publishing, Cham, 271–285.

[11] Jan Sedmidubsky, Petr Elias, and Pavel Zezula. 2018. Effective and efficient similarity searching in motion capture data. *Multimedia Tools and Applications* 77, 10 (2018), 12073–12094. https://doi.org/10.1007/s11042-017-4859-7

[12] Jan Sedmidubsky, Petr Elias, and Pavel Zezula. 2019. Searching for variable-speed motions in long sequences of motion capture data. *Information Systems* 80 (2019), 148–158. https://doi.org/10.1016/j.is.2018.04.002

[13] Jan Sedmidubsky and Pavel Zezula. 2018. Probabilistic Classification of Skeleton Sequences. In *29th International Conference on Database and Expert Systems Applications (DEXA)*. Springer International Publishing, Cham, 50–65.

[14] Jan Sedmidubsky, Pavel Zezula, and Jan Svec. 2017. Fast Subsequence Matching in Motion Capture Data. In *21st European Conference on Advances in Databases and Information Systems (ADBIS)*. Springer International Publishing, Cham, 50–72.

[15] Hongsong Wang and Liang Wang. 2017. Modeling Temporal Dynamics and Spatial Configurations of Actions Using Two-Stream Recurrent Neural Networks. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE, 3633–3642.

[16] Pavel Zezula, Giuseppe Amato, Vlastislav Dohnal, and Michal Batko. 2006. *Similarity Search: The Metric Space Approach*. Advances in Database Systems, Vol. 32. Springer-Verlag. 220 pages.