Visual Exploration of Human Motion Data

Petra Budikova(✉)[0000−0003−1523−1744], Daniel Klepac, David Rusnak, and Milan Slovak

Faculty of Informatics, Masaryk University, Brno, Czech Republic
budikova@fi.muni.cz

Abstract. Human motion data are beginning to appear in many application domains, which brings a need to develop user-friendly motion processing applications. One of important open challenges is the presentation of high-dimensional spatio-temporal motion data to end users in a way that is easy to understand and allows fast browsing and exploration of the motion datasets. For many applications such as computer-assisted rehabilitation or motion learning, it is also very desirable to visualize the differences between two motion sequences. In this paper, we present a publicly available software tool that provides the visualization functionality for individual motion sequences, comparison of two motions, and exploration of large motion datasets.

Keywords: Human motion data · skeleton sequences · visualization · multimedia exploration · explainability of similarity

1 Motivation

Human motion can be described by a sequence of skeleton poses, where each pose keeps 2D/3D coordinates of important body joints in a specific time moment. Such spatio-temporal skeleton data can be utilized in a number of application domains, ranging from gaming and sports to healthcare and security [12]. With the recent advances in human pose estimation from ordinary videos [4], a huge explosion of motion processing application can be expected in the near future. Consequently, efficient and effective tools are needed for different phases of motion data processing.

In this paper, we study the problem of motion data understanding from the user point of view. Motion data are a rich source of information, but in their raw form they are represented by long vectors of float numbers, which are completely uninteligible to humans. This is typically solved by displaying the source video recordings, when available, or creating animations of the skeleton data [13]. However, viewing the videos or animations is time-consuming and therefore not suitable for situations where users desire to quickly grasp the content of multiple motion sequences, e.g., when browsing or querying collections of motion data. Therefore, we propose to represent each motion by a single static image that captures the most significant poses and can be read at a glance. We focus especially on the visualization of so-called *motion actions*, which are short motions

with a clear semantic meaning, e.g. a jump, throw, or a cartwheel. While the motion data can be captured as a long unsegmented sequence, users are typically interested in retrieving and viewing only the short segments that contain some activity of interest. If needed, the longer motions can be represented as sequences of images for individual segments.

Apart from the motion data itself, it is also difficult to understand, and explain, how the similarity between two motion actions is measured. The concept of similarity is instrumental to all motion processing tasks, ranging from query-by-example searching to action classification, event detection, or computer-aided rehabilitation [12]. Although the similarity is often hidden in complex machine learning models such as neural networks, we need to understand it to be able to interpret, analyze, and optimize the machine learning techniques, and to improve our own movements in computer-assisted motion learning. Therefore, evaluations of motion similarity are another area that calls for easy-to-understand visualizations.

Yet another set of challenges appears when we extend our focus to large collections motion data. Let us consider some popular motion datasets, such as HDM05 [10], PKU-MMD [5], or NTU [6]. These contain thousands or tens of thousands of motion actions, accompanied by metadata that determine their semantic categories. After downloading such dataset, it is possible to find out the number of categories, their frequencies, and view the videos of some random samples. However, there is no efficient way to gain insight into what really happens in the individual motion sequences, how diverse the categories are, if there are any natural clusters, etc. Yet all this information is essential for designing the motion processing applications.

To answer all these challenges, we have created a new JavaScript library for motion data visualization and exploration. The MocapViz library[1] offers three mutually cooperating modules: visualization of individual short motions, visual explanation of differences between two motion sequences, and effective exploration of large motion data collections. The first two modules can be integrated within an arbitrary web presentation that utilizes motion data. The exploration module produces a complete web presentation of a given motion dataset, which allows interactive data browsing as well as detailed inspection of selected motions and their relationships. The functionality of all modules is demonstrated in two public web interfaces for the exploration of HDM05 and PKU-MMD datasets.

2 Preliminaries: Processing of Human Motion Data

Human motion is recorded as sequence $S = (P_1, \ldots, P_l)$ of skeleton poses P_i $(1 \leq i \leq l)$, where each pose $P_i \in \mathbb{R}^{j \cdot dim}$ represents the skeleton configuration estimated in time moment i and consists of $dim \in \{2, 3\}$ coordinates of j tracked *joints*. The number and position of the body joints and the dimensionality dim depends on the hardware or software tools used to acquire the data. We denote this as the *body model* and take it as one of the visualization inputs.

[1] http://disa.fi.muni.cz/research-directions/motion-data/mocapviz/

The raw skeleton sequences are spatio-temporal data, which can be compared by sequence alignment methods. In particular, the Dynamic Time Warping (DTW) algorithm [9] is usually applied, since it takes into account the possible differences in speed of the compared movements. The algorithm finds optimal matching between the poses of the two compared motions, and computes the overall distance as the sum of distances of the mapped poses. However, the processing of raw skeleton sequences with DTW is rather expensive due to the high dimensionality of the skeleton data and the quadratic computation time of the DTW. Moreover, some complex relationships between motions may not be discovered by the DTW. Therefore, state-of-the-art motion processing techniques often represent motions by some derived features and learn complex similarity models by machine learning techniques, especially the neural networks [12]. These approaches provide very good application results, but the similarity computation is embedded in the learned model and cannot be easily explained.

The objective of our work is to visualize motion data and explain their similarity in a way that is easily understandable to humans. For the visualization of individual movements, the raw skeleton data representation is the most suitable, since it is the most detailed and semantically clear. For the comparison of two motion sequences, we utilize sequence alignment methods that can be intuitively explained over visualizations of the skeleton sequences.

3 Visualization of Single Motion Sequence

As discussed earlier, motion data are usually surveyed by watching the source video, if available, or watching the animated skeleton sequences. Sequences of stick figures are routinely used to represent motions in research papers, but these are created manually. In [3], the technique of *MotionCues* is proposed, which creates a single 3D figure representing the whole motion, with arrows expressing the movement of individual body parts. This visualization is compact and well understandable, but only suitable for very simple actions. The *Motion Belts* technique [15] is the most similar to ours: it draws selected key-poses on a timeline and uses pose coloring to express their orientation. However, the poses are sometimes clumped together, making it difficult to determine what is happening, and the use of moving viewpoint is not very intuitive.

The MocapViz motion visualization module represents each action by a single static image, which contains the most representative poses placed on a timeline (Figure 1). The keyposes are selected by a curve simplification algorithm and rendered as 2D stick figures. A few poses preceding each keypose are also drawn with a low opacity, which provides a better feeling of the movement. For each motion, a static camera position is chosen so that maximum information is shown; the camera is typically placed orthogonally to the motion direction. To make the poses easier to read, we use different colors for left/right body parts, and add artificial "nose" line that expresses the direction where the skeleton is looking. Furthermore, we also provide a bird-eye view of the motion in space, to allow better understanding of the spatial dimension that is lost in the 2D image.

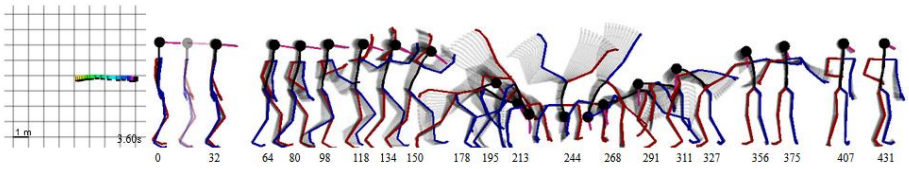

Fig. 1. Cartwheel motion from the HDM05 dataset visualized in a single image. The bird-eye view map on the left shows how much the person moved in space.

The MocapViz library can visualize any type of skeleton-based human motion data, provided that the appropriate body model is supplied. The most popular Vicon and Kinect body models are already included in the library. Noticeably, the visualizations are only able to display the movement of a single person. Interactions between several people are more difficult to depict because of the spatial relationships between the skeletons, and would require a different approach.

4 Understanding Motion Similarity

The evaluation of similarity between two motion sequences is the core concept of all motion processing tasks, and its explanation is vital for both researchers and common users who work with motion processing applications. A superficial understanding of motion similarity can be obtained by visually comparing the motion images presented in the previous section, but much more insight can be gained from a detailed analysis of the sequence mapping found by the sequence alignment methods.

The visualization of the DTW mapping over skeleton sequences is studied in several existing research works. Malmstrom et al. [7] focus on angles of the joints making up individual body parts and visualize their differences in several graphs, which are detailed but difficult to understand for common users. In [2], color-coded bars are used to depict the development of motion in time. Urribarri et al. [14] focus on the visualization of time differences between the two compared motions. However, none of these techniques combines the visualizations of mapping with visualization of individual skeleton sequences, nor do they provide a combination of multiple views on the sequence differences.

In MocapViz, on the other hand, we strive to provide a comprehensive view on the dissimilarity of two movements. Therefore, we have designed several new visualizations that focus on different aspects of the motion data. The first two are shown in Figure 2, the other two examples are not included due to space restrictions but can be found on the web-page of the MocapViz library.

– *Overall similarity of motion sequences* (Figure 2-A): To visualize the complete pose-to-pose mapping of the compared sequences, we first represent each motion by the motion image presented earlier. To be able to draw the mapping among all poses and not just the key-poses depicted in the motion

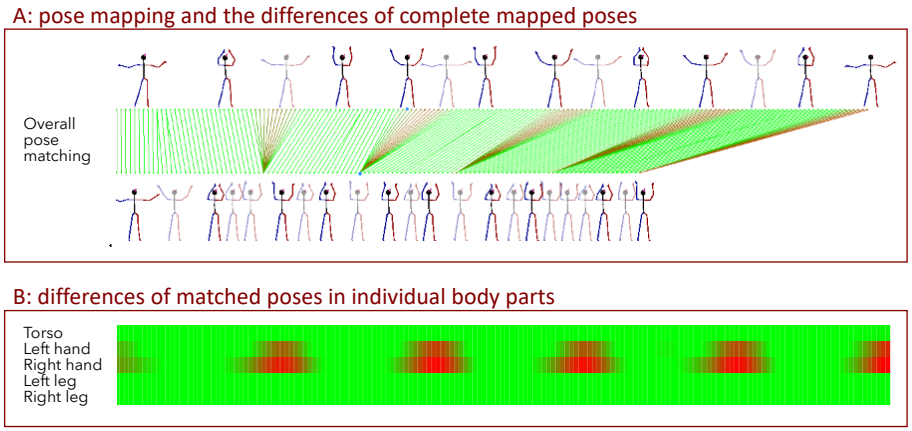

Fig. 2. Two views on the differences between two clapping motions from the HDM05 dataset. We can observe that the main differences occur when the actors move their hands apart – one of them claps faster and doesn't move hands far apart, the other one spreads his hands more.

image, we add a time-line of dots representing all the poses. The dots are connected by lines that express the optimal pose-to-pose mapping, colored on the red-green scale to express the closeness of individual mapped poses. Depending on user settings, the closeness of pose matching can be evaluated in the context of the specific two actions (thus highlighting even small differences in two similar motions) or in the context of the whole dataset (to better distinguish between minor and major differences).

- *Differences of matched poses in individual body parts* (Figure 2-B): Some pairs of motions may only differ e.g. in the movement of hands, while the legs are static or move in the same way. To highlight such situations, we visualize the closeness of pose mapping for individual body parts.
- *Detailed view on the matched poses*: For any two mapped poses, it is possible to view the detailed drawing of the poses and the computed differences between individual body parts.
- *Visualization of time alignment*: In this view, we detect and visualize the changes of speed in the compared motions, using the algorithm of [14].

In the current implementation, the optimal mapping between two skeleton sequences is determined by the DTW algorithm. However, the implementation is extensible, so the DTW distance can be seamlessly replaced by other sequence alignment algorithms. The only input required by the MocapViz module for visualization of movement differences are the two motion sequences to be compared, normalized according to user preferences. Data normalization is not part of the visualization procedure, since different approaches to position or orientation normalization may be suitable for individual use cases.

5 Exploration of Human Motion Datasets

The aim of multimedia exploration is to reveal the content of a whole multimedia collection, often totally unknown to the users who access the data. The exploration principles are mostly studied in the domain of image retrieval [8, 11]. To the best of our knowledge, only one similar technique exist for motion data [2]. However, this approach focuses on the level of individual poses, which is useful for understanding detailed variations of a small collection of movements (e.g. for gaming and animation applications) but not for the browsing of large collections. Therefore, we took inspiration from the image exploration interfaces and combined them with our methods for motion data visualization.

The construction of exploration systems for large datasets usually comprises two steps. First, the large input collection has to be organized into a hierarchical tree structure, so that individual nodes of the tree can be visualized on a single screen. Next, the actual visualization needs to be designed, allowing intuitive browsing through the hierarchy and presenting each node in a way that conveys the maximum information about the relationships between individual objects within the node. For collections of motion sequences, we further find it important to incorporate the information about semantic categories of individual motions into the exploration interface. Let us recall that motion collections such as the HDM05, PKU-MMD, or NTU datasets contain short motion sequences sorted into semantic categories that determine the type of the motion (jump, run, etc.). For people who want to familiarize themselves with the dataset, it is also very relevant to see to what extent the semantic categories agree with the natural clustering of data as provided by the content-based distance measures (e.g., the DTW algorithm). Therefore, the additional objective of our exploration interface is to allow browsing by both the semantic categories and the content-based clusters, and to provide information about the semantic diversity of the content-based clusters.

Preparation of the hierarchical structure We process the input dataset in a top-down manner, gradually breaking the collection into smaller clusters of mutually similar objects. Sufficiently small clusters become the leaf nodes of the tree hierarchy, larger clusters give rise to subtrees. In particular, we utilize the hierarchical k-medoid clustering, which allows us to limit the number of subtrees for each internal node of the hierarchy. The parameter k was set to 10, with the maximum size of the leaf nodes being 20. During the construction of the hierarchical tree, we also collect some interesting statistics, such as the sizes of individual subtrees or the number of different semantic categories contained in each subtree. A separate hierarchy is also computed for each semantic category that contains more than 20 objects.

The computation of the hierarchical structures is performed off-line, using the MESSIF library for content-based data management [1], and the results are saved as a JSON file. If preferred, users can employ their own tools to produce the hierarchies in the defined format and submit them to the exploration interface.

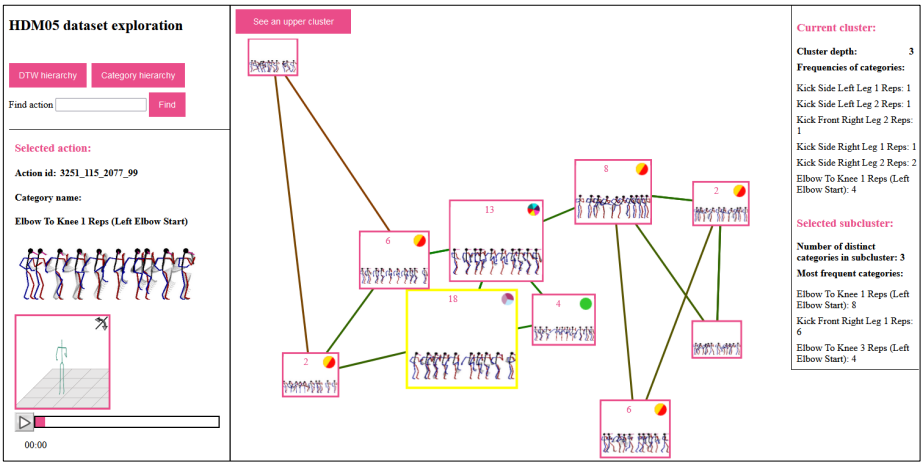

Fig. 3. Visual exploration of the HDM05 dataset.

Exploration interface The exploration interface is the third module of the MocapViz library. It was designed to allow easy orientation and browsing in the collection, and to provide rich information about individual motion objects and their relationships. The interface consists of three main parts (see Figure 3).

In the central part, users can browse the hierarchical tree structure (either complete or for a selected semantic category) and display individual nodes. In case of leaf nodes, all objects in the node are shown, whereas for inner nodes of the hierarchical tree, we show the medoids of the subtrees. The next level of the hierarchy is accessed by double-clicking on the subtree representative. The nodes are displayed using the force-based layout [8], which places the objects on the screen in such way that the more similar ones are close to each other and the more distant objects are placed further apart. Individual objects are represented by the motion images, edges between them are color-coded to express the level of similarity and upon clicking reveal the full visualization of the similarity between the two connected motions. The motion images representing subtrees of the hierarchy also contain information about the size of the respective subtree and a pictogram that expresses the subtree diversity in terms of semantic categories.

The left and right panels of the exploration interface contain additional details about the selected action and cluster, respectively. The action details include a full motion image of the given action, its animation, and information about related actions. For clusters, we provide more detailed information about the distribution of semantic categories within the cluster.

6 Conclusions

Visualization of human motion data is an important part of creating insightful and user-friendly motion processing applications. The MocapViz library pre-

sented in this paper provides a unique set of techniques for visualizing human motion data, explaining their relationships, and exploration of large motion datasets. Its functionality is demonstrated in two public interfaces for the exploration of the HDM05 and PKU-MMD datasets. The exploration interfaces as well as the MocapViz library are available at http://disa.fi.muni.cz/research-directions/motion-data/mocapviz/.

References

1. Batko, M., Novak, D., Zezula, P.: MESSIF: metric similarity search implementation framework. In: Digital Libraries: Research and Development (DELOS). LNCS, vol. 4877, pp. 1–10. Springer (2007)
2. Bernard, J., Wilhelm, N., Krüger, B., May, T., Schreck, T., Kohlhammer, J.: Motionexplorer: Exploratory search in human motion capture data based on hierarchical aggregation. IEEE Trans. Vis. Comput. Graph. **19**(12), 2257–2266 (2013)
3. Bouvier-Zappa, S., Ostromoukhov, V., Poulin, P.: Motion cues for illustration of skeletal motion capture data. In: 5th Intl. Symp. on Non-Photorealistic Animation and Rendering (NPAR). pp. 133–140. ACM (2007)
4. Chang, S., Yuan, L., Nie, X., Huang, Z., Zhou, Y., Chen, Y., Feng, J., Yan, S.: Towards accurate human pose estimation in videos of crowded scenes. In: 28th ACM Intl. Conf. on Multimedia (MM). pp. 4630–4634. ACM (2020)
5. Liu, C., Hu, Y., Li, Y., Song, S., Liu, J.: PKU-MMD: A large scale benchmark for skeleton-based human action understanding. In: Workshop on Visual Analysis in Smart and Connected Communities (VSCC@MM 2017). pp. 1–8. ACM (2017)
6. Liu, J., Shahroudy, A., Perez, M., Wang, G., Duan, L., Kot, A.C.: NTU RGB+D 120: A large-scale benchmark for 3d human activity understanding. IEEE Trans. Pattern Anal. Mach. Intell. **42**(10), 2684–2701 (2020)
7. Malmstrom, C., Zhang, Y., Pasquier, P., Schiphorst, T., Bartram, L.: Mocomp: A tool for comparative visualization between takes of motion capture data. In: 3rd Intl. Symp. on Movement and Computing (MOCO). pp. 11:1–11:8. ACM (2016)
8. Mosko, J., Lokoc, J., Grosup, T., Cech, P., Skopal, T., Lánský, J.: Evaluating multilayer multimedia exploration. In: 8th Intl. Conf. on Similarity Search and Applications (SISAP). LNCS, vol. 9371, pp. 162–169. Springer (2015)
9. Müller, M.: Information retrieval for music and motion. Springer (2007)
10. Müller, M., Röder, T., Clausen, M., Eberhardt, B., Krüger, B., Weber, A.: Documentation Mocap Database HDM05. Tech. Rep. CG-2007-2, Universität Bonn (2007)
11. Nguyen, G.P., Worring, M.: Interactive access to large image collections using similarity-based visualization. J. Vis. Lang. Comput. **19**(2), 203–224 (2008)
12. Sedmidubsky, J., Elias, P., Budikova, P., Zezula, P.: Content-based management of human motion data: Survey and challenges. IEEE Access **9**, 64241–64255 (2021)
13. Sedmidubsky, J., Zezula, P.: Recognizing user-defined subsequences in human motion data. In: Intl. Conf. on Multimed. Retrieval (ICMR). pp. 395–398. ACM (2019)
14. Urribarri, D.K., Larrea, M.L., Castro, S.M., Puppo, E.: Overview+detail visual comparison of karate motion captures. In: 25th Argentine Congress of Computer Science (CACIC). Communications in Computer and Information Science, vol. 1184, pp. 139–154. Springer (2019)
15. Yasuda, H., Kaihara, R., Saito, S., Nakajima, M.: Motion belts: Visualization of human motion data on a timeline. IEICE Trans. Inf. Syst. **91-D**(4), 1159–1167 (2008)