SDAS: emantic Data Acquisition System for Minimizing Redundancy and Maximizing Diversity

Yeseung Park [*], **Hyunse Yoon** [*], **Jungwoo Huh, Jungsu Kim, Jeongwook Choi, Sanghoon Lee**

Yonsei University

pys940617, hsyoon97, gjwjddn9, integer, yunakevin16, slee@yonsei.ac.kr

Abstract

In this paper, we propose SDAS, a new motion data assessment and storage system designed to acquire new motion data with reduced redundancy and maximizing diversity. SDAS collects data in the field, retrieves the most similar data from the database in real-time, and provides visualization tools that allow for the comparison of differences between the capture data and the stored data. Through this system, researchers can efficiently build and manage a database. The demonstration video is available at https://youtu.be/vqW0uMDnZTw.

Introduction

Human motion research has attracted significant attention in various fields, including gaming, film, and the sports industry(Kim et al. 2023, 2024a). In response to this growing demand, many researchers are trying to acquire diverse motion data. When acquiring motion data, it is crucial not only to secure a large quantity of data but also to compose a diverse database without redundancy in terms of quality. However, current human motion databases contain an inherent redundancy issue, where identical or similar movements are stored multiple times (Park et al. 2020). This is an important issue, as increasing redundancy decreases the efficiency of model training. The straightforward way to address this issue is to assess not only the distribution of the data to determine the similarity between capture data and stored data but also to consider the semantic meaning of the motion. However, it is challenging to comprehensively consider these factors during the data acquisition process.

To address this, we propose a framework, tentatively named Semantic Data Acquisition System for Minimizing Redundancy and Maximizing Diversity(SDAS), which determines whether to store the new capture data based on various assessment criteria. SDAS consists of three main components: the Data Acquisition System, which merges human posture information obtained through an RGB camera into 3D pose data; a web-based Data Assessment Server that verifies the acquired data; and a Real-time Data Storage Server, where motion data, if decided to be stored, is saved and retrieved using an in-memory database. The user retrieves the stored data most similar to the capture data based on a semantic similarity metric, subsequently rendering it in real time for intuitive comparison. This allows for an intuitive examination of the semantic features of the data and the intrinsic meanings of the motions expressed through the LLM(Large Language Model).

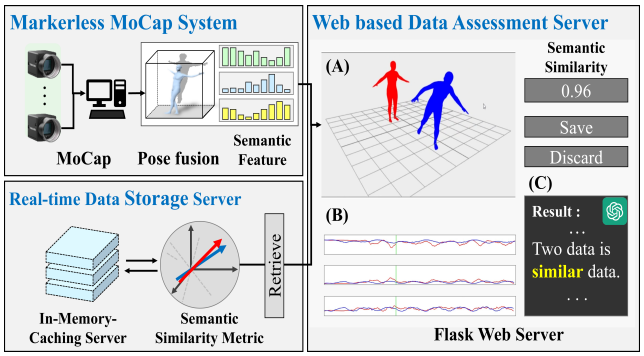

Figure 1: Overall framework of SDAS

System Design

Markerless MoCap System

To accurately capture human posture, we apply a multi-view 3D pose estimation algorithm. Our capture system consists of 4 RGB cameras, 2 client PCs to manage these cameras, and a master PC to control the entire system. First, the camera parameters of the 4 synchronized RGB cameras are calibrated using a checkerboard. When multi-view RGB images are provided, the parameters of the SMPL(Loper et al. 2023) model for each viewpoint are extracted to estimate the human pose. Other detailed implementation settings follow (Kim et al. 2024b).

Real-time Data Storage Server

The Real-time Data Storage Server performs data retrieval and storage functions. To facilitate real-time search and storage, we use Redis(Carlson 2013), an in-memory database with read-write speeds faster than those of disk-based RDBMS. The Redis server manages the data in the form of a dictionary containing 3D rotation data and the corresponding feature scores. Here, the 3D rotation data represents motion data in an $F * J * D$ format, while the feature

[*]These authors contributed equally.

	Delay	Filp	Extend	Contract
Cossim	0.9594	0.6271	0.9452	0.9646
Ours	**0.9801**	**0.8973**	0.9448	0.9735

Table 1: Performance comparison table of similarity evaluation metrics

scores correspond to the three semantic features described earlier. F, J, D represents length of motion, joint, dimension, respectively. In the SDAS framework, motion data is efficiently stored and managed by defining the frame size as 240 frames, captured over 8 seconds at 30 FPS. The joint configuration follows the default format of 24 joints in the SMPL parameters.

The data retrieval function returns data similar to a given query. When a search request is made from the web-based data assessment server, the Redis cache server traverses the accumulated database and clusters the data using the K-NN algorithm. It then retrieves the comparison data that has the closest Semantic Similarity Metric distance to the query data. we propose a novel motion similarity evaluation method, the Semantic Similarity Metric, which emphasizes the semantic aspects of motion over traditional similarity evaluations. Conventional motion similarity evaluation metrics primarily focus on joint movements, effectively capturing the similarity of aligned motions(Park et al. 2021). However, they often fail to identify fundamentally similar motions that differ in timing and angle. To identify motions with fundamentally similar movements, it is more effective to utilize not only motion data that focuses on joint movements but also semantic features that capture the intrinsic motivations of human motion. We incorporate three semantic features(Stability, Liveliness, Attention) extracted from motion data, as proposed by (Lee et al. 2022), into the similarity assessment metric to develop a more semantically appropriate similarity assessment metric.

Let $\Theta = \{\boldsymbol{\theta}_t \in \mathbb{R}^{72} | t = 1, \ldots, T\}$ and $\mathcal{F} = \{\mathbf{f}_t = [f_t^{stab}, f_t^{live}, f_t^{attn}] \in \mathbb{R}^3 || t = 1, \ldots, T\}$ be a sequence of pose parameters and semantic features of a given motion, respectively. Then, the semantic similarity between the stored motion s and captured motion c is represented as:

$$\mathbf{S} = w_m \frac{\hat{\boldsymbol{\Theta}}^s \cdot \hat{\boldsymbol{\Theta}}^c}{\|\hat{\boldsymbol{\Theta}}^s\| \|\hat{\boldsymbol{\Theta}}^c\|} + w_f \frac{\hat{\mathbf{F}}^s \cdot \hat{\mathbf{F}}^c}{\|\hat{\mathbf{F}}^s\| \|\hat{\mathbf{F}}^c\|} \tag{1}$$

where $\hat{\boldsymbol{\Theta}} = [\boldsymbol{\theta}_1^\top, \boldsymbol{\theta}_2^\top, \ldots, \boldsymbol{\theta}_T^\top] \in \mathbb{R}^{72T}$ and $\hat{\mathbf{F}} = [\mathbf{f}_1^\top, \mathbf{f}_2^\top, \ldots, \mathbf{f}_T^\top] \in \mathbb{R}^{3T}$ denotes the flattened pose parameter and semantic feature vectors, respectively. We set w_m and w_f as a weight to balance the contribution of each term.

Web Based Data Assessment Server

The web-based data evaluation server is implemented using the Flask web framework, which allows various functions such as data verification, comparison, storage, and evaluation in a simple yet intuitive way. As mentioned above, to compare new capture data with stored data from different perspectives, the server provides key features such as motion visualization, semantic feature visualization, and LLM based feature analysis.

	Walk	Run	Throw	Punch
Accuracy	**1.000**	0.967	0.833	0.800
	Kick	Sit	Turn	Balancing
Accuracy	**1.000**	**1.000**	0.967	0.900

Table 2: Result of Top-1 Accuracy score

Motion data collected through the markerless Mocap system is visualized in the rendering window (Figure 1-A) and immediately sent to the data storage server as a query for data retrieval. The most similar data are then returned and visualized in the form of a fitted mesh through linear blend skinning. Users can qualitatively compare the similarity of the two different meshes in the rendering window.

In addition, semantic features are visualized to understand the inherent meaning of the motion data. Since these features represent the intrinsic nature of the data, they help identify its essential similarities. However, it can be challenging for users unfamiliar with motion data to analyze the inherent characteristics encapsulated in the semantic features. Therefore, we employ an LLM to provide an intuitive text representation of the semantic feature values (Figure 1-D). For LLM, we used ChatGPT-4.0(Achiam et al. 2023) and performed prompt engineering to allow LLM to explain the inherent meaning of the semantic feature values in a stable way. Finally, after reviewing the data from multiple perspectives, the evaluator can click the Save button to save the data to the storage server or discard it.

Experiment

To verify the validity of the proposed assessment metric, we conduct an experiment to compare how cosine similarity, a commonly used method for the assessment of motion similarity, and the proposed metric assess the similarity between the transformed motion and the original motion. In this experiment, we randomly select 10 original motions from the CMU Mocap database and create four types of reference motions for each: delaying the motion initiation, flipping the motion, contracting, and extending motion duration. Since these reference motions are modified without altering the original motion's meaning, both motions should be evaluated as similar. The more similar the motion, the closer the result is to 1.

As shown in Table 1, the conventional cosine similarity metric is robust against such transformations, except for the flipped motions. In contrast, our proposed evaluation metric, which focuses on the semantic movements of joints, effectively responds to all types of transformation, including flipping. In addition, in order to verify the retrieval performance of SDAS, a dataset was pre-configured to constitute 30 CMU Mocap data for each of 8 labels. Then, 30 new motions are captured to test whether the method retrieves the similar action from the pre-configured dataset. Table 1 shows the retrieval accuracy of each retrieval method. The Top-1 score calculation showed a high retrieval accuracy of over 0.8 for most motions, achieving the highest levels of accuracy in many labels, including "Walk" and "Kick."

Acknowledgments

This research was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT) (No. 2020R1A2C3011697, Contribution Rate: 50%) and Culture, Sports and Tourism R&D Program through the Korea Creative Content Agency grant funded by the Ministry of Culture, Sports and Tourism in 2024 (Project Name: Global Talent for Generative AI Copyright Infringement and Copyright Theft, Project Number: RS-2024-00398413, Contribution Rate: 50%).

References

Achiam, J.; Adler, S.; Agarwal, S.; Ahmad, L.; Akkaya, I.; Aleman, F. L.; Almeida, D.; Altenschmidt, J.; Altman, S.; Anadkat, S.; et al. 2023. Gpt-4 technical report. *arXiv preprint arXiv:2303.08774*.

Carlson, J. 2013. *Redis in action*. Simon and Schuster.

Kim, D.; Kim, T.; Lee, I.; and Lee, S. 2024a. Kinematic Diversity and Rhythmic Alignment in Choreographic Quality Transformers for Dance Quality Assessment. *IEEE Transactions on Circuits and Systems for Video Technology*.

Kim, J.; Kwon, B.; Kim, J.; and Lee, S. 2023. Mnet++: Music-driven pluralistic dancing toward multiple dance genre synthesis. *IEEE transactions on pattern analysis and machine intelligence*.

Kim, S.; Huh, J.; Park, Y.; Kim, J.; and Lee, S. 2024b. DanceMimic: Awaken Your Dancing Instinct through a Real-time Dance Imitation Capture System. In *Proceedings of the 32nd ACM International Conference on Multimedia*, 1–2.

Lee, K.; Park, Y.; Huh, J.; Kang, J.; and Lee, S. 2022. Self-updatable database system based on human motion assessment framework. *IEEE Transactions on Circuits and Systems for Video Technology*, 32(10): 7160–7176.

Loper, M.; Mahmood, N.; Romero, J.; Pons-Moll, G.; and Black, M. J. 2023. SMPL: A skinned multi-person linear model. In *Seminal Graphics Papers: Pushing the Boundaries, Volume 2*, 851–866.

Park, J.; Cho, S.; Kim, D.; Bailo, O.; Park, H.; Hong, S.; and Park, J. 2021. A body part embedding model with datasets for measuring 2D human motion similarity. *IEEE Access*, 9: 36547–36558.

Park, Y.; Jang, M.; Huh, J.; Lee, K.; and Lee, S. 2020. Data reduction using cluster sampling. In *2020 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC)*, 1274–1278.