Body-Scale-Invariant Motion Embedding for Motion Similarity

XIAN DU[1], CHUYAN QUAN[1], Ri Yu[1,2]

[1]Dept. of Artificial Intelligence, Ajou University, South Korea, [2]Dept. of Software and Computer Engineering, Ajou University, South Korea

2025 TAIPEI

PROBLEM

➢ Motion similarity analysis is vital in healthcare, sports science, and HCI

➢ Current 3D Human Pose Estimation (HPE) methods are inherently biased by body size and proportions variations

➢ We present a novel framework that learns body-scale-invariant motion embeddings directly from RGB video.

Characters of Varying Body Sizes

RELATED WORK

➢ 3D Human Pose Estimation (HPE) methods, such as MediaPipe, provide skeletal data that inherently confounds motion with anatomical variance.

➢ SAME (Lee et al. 2023) generalizes across skeletons but limits input formats, reducing applicability

➢ MocapNET (Qammaz et al. 2021) converts MediaPipe outputs to BVH, enabling consistent skeletal hierarchies

➢ Our framework combines data processing and embedding learning for scale-invariant motion similarity.

OVERVIEW

➢ Propose a framework for body scale invariant embeddings from RGB video

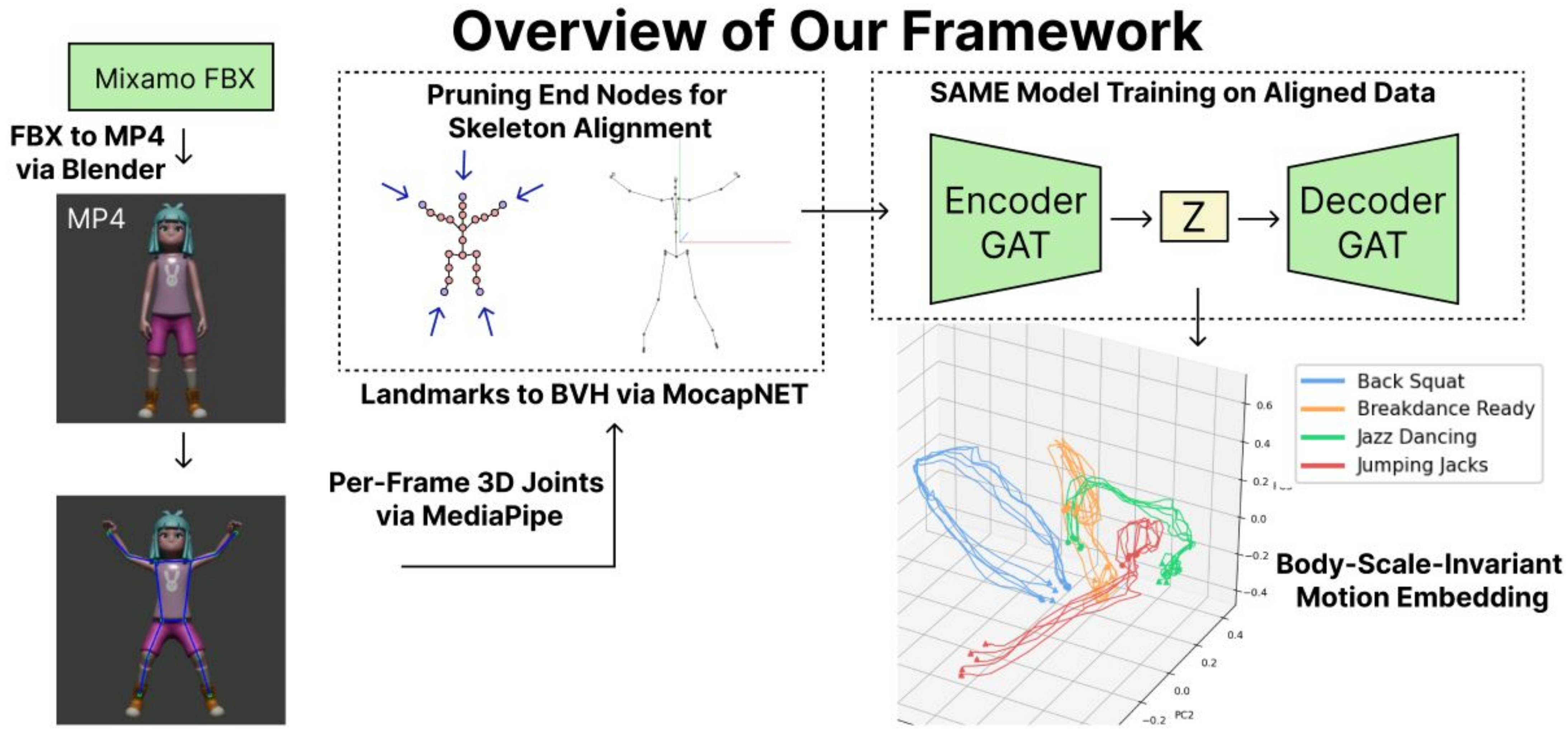
➢ Build a standardized dataset from diverse 3D characters with varied skeletal proportions

➢ Standardize motion data by aligning skeletons and retaining 27 key joints for SAME training

➢ Embeddings group motions by type and stay consistent across body shapes, enabling use in rehabilitation, sports, and HCI

METHODOLOGY

Overview of Our Framework



➢ Use 28 Mixamo characters with four motions each, rendered in Blender to MP4 videos

➢ Extract 3D landmarks from videos with MediaPipe; convert them to BVH using MocapNET

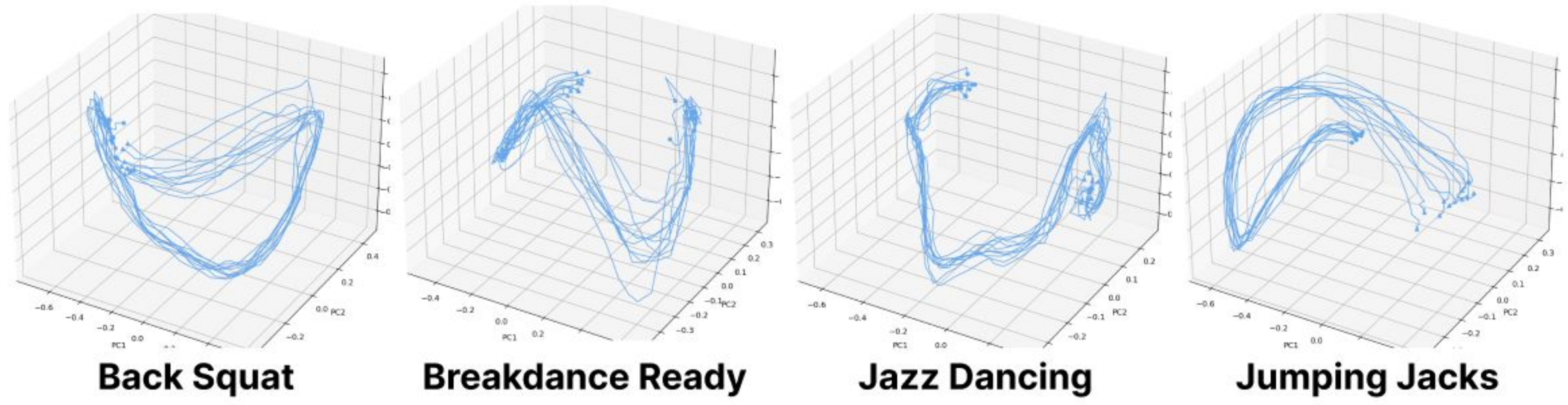
➢ Standardize skeletons by retaining 27 joints, unifying names, and pruning redundant nodes

➢ Train SAME model on aligned data for motion dynamics while minimizing anatomical bias

➢ Generate embeddings that preserve motion features and remain consistent across diverse body shapes

RESULTS

Motion-Specific Embedding Clustering



➢ Our model learns a decoupled motion representation, forming compact, separable clusters for each motion type that are independent of body scale and anatomical variation.

➢ In the embedding space, cyclic motions (e.g., Jumping Jacks) form regular, closed trajectories, while complex actions (e.g., Breakdance Ready) produce irregular yet separable paths.

➢ Quantitative results confirm robust temporal alignment, showing high DTW-aligned cosine similarity (0.92–0.95) for the same motion performed by individuals of dissimilar body proportions

➢ Qualitative analysis shows that minor deviations are localized to specific kinematic phases (e.g., shoulder abduction and spinal flexion), while the overall motion pattern is consistently captured, confirming the model successfully decouples kinematic motion patterns from anatomical variation

REFERENCES

[1] Adobe. Mixamo animation services. URL: https://www.mixamo.com

[2] Lee S., Kang T., Park J., Lee J., Won J. Same: Skeleton-agnostic motion embedding for character animation. In SIGGRAPH Asia 2023 Conference Papers (New York, NY, USA, 2023), SA '23, Association for Computing Machinery. DOI: 10.1145/3610548.3618206

[3] Qammaz A., Argyros A. A. Towards holistic real-time human 3d pose estimation using mocapnets. In British Machine Vision Conference (BMVC 2021), November 2021, BMVA, p. 418.