6ZF-08

Pose Visualization and Feedback System Based on Pose Prediction

HUANG JIAYUN† Takahashi Shin‡
huang@iplab.cs.tsukuba.ac.jp shin@cs.tsukuba.ac.jp

1 Introduction

In recent years, young people have been on a healthy diet and doing bodyweight exercises on a regular basis. Without a coach at home to supervise them, many users struggle with their exercise accomplishments and are unable to evaluate their own movements.

To improve the training performance of bodyweight exercises, we build a system that predicts and visualizes future poses in real time. Users can see their future performance with several visualization methods, such as AR avatars. Our system also gives sound and voice feedback based on a pose prediction algorithm by comparing future poses with standard poses in future moment.

In this paper, we briefly present the design idea of the system, part of the system implemented up to date, and future implementation plan.

2 Related work

2.1 Pose estimation

Pose estimation is a task that detects human figures and gets their body data in images. Pose estimation could be explained as a skeleton-data-based algorithm [5] recognizing human pose and motion. Many pose estimation methods use sensors, depth cameras to track skeleton data.

2.2 Pose prediction

Based on data obtained through pose estimation, pose prediction predicts future poses according to a sequence data of previous poses. Wu [2] proposes a real-time pose prediction method – FuturePose which predicts and visualizes future poses in AR environments.

We implemented pose prediction based on the STS-GCN algorithm [1] which leverages the space-time representation to forecast joint coordinates in future. To provide real-time visual feedback, we consider showing future pose to the user through an avatar in an AR environment.

3 Design of our system

3.1 Pipeline of our system

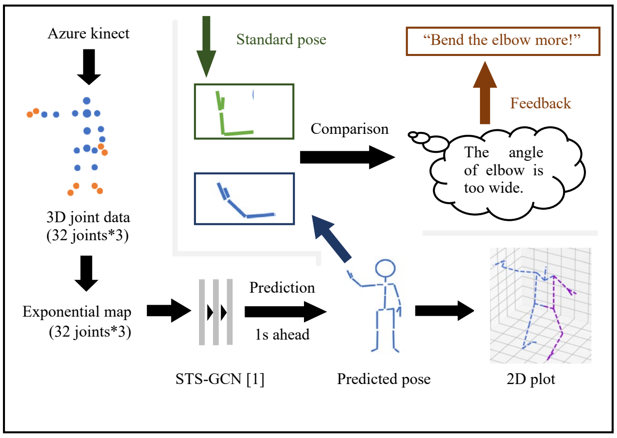

Fig 1. Flow chart

Our system predicts poses in real time, visualizes future poses, and gives audio feedback to users to improve their training performance of bodyweight exercises.

In our system, the observation data sequences of body joints are obtained through Azure Kinect. The joint data are converted into an exponential map [3], and then used as input to STS-GCN to predict future poses. Predicted poses are used for visualization and evaluation to help users with bodyweight exercises. Our system is implemented on a PC equipped with NVIDIA GeForce RTX 3090.

3.2 Prediction of future pose

Our system views joint data of 10 previous frames as the input sequence of STS-GCN. STS-GCN forecasts joint coordinates 25 frames ahead, that is, one second in the future. Once Azure Kinect began to take the shot, our system can visualize their future pose in real time every 10 frames.

As preliminary test, we recorded 100 frames of joint coordinates as the ground truth (GT) data of the user's movement and compared the GT data with the predicted poses forecasted by our system. Fig. 2 shows the pose of the 10th frame and the predicted pose one second ahead. The prediction performance was reliable to some extent while actions which didn't follow common movement rules were not expectable to be correctly predicted.

† University of Tsukuba, Master's program of Computer Science
‡ University of Tsukuba, Department of Computer Science

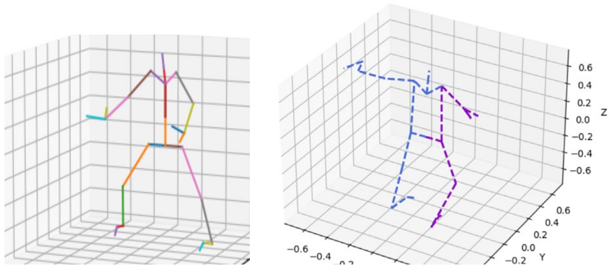

Fig 2. Current pose at the moment of 10th frame (left) and predicted pose 1s ahead (right).

3.3 Input data format of STS-GCN

The input data format of STS-GCN is the exponential map [3] of H3.6M dataset [7] which maps a 3D joint coordinate vector describing the axis and magnitude of a three DOF rotation to the corresponding rotation. Our system converts a 32*3-size XYZ tuple of the user's joints in each frame (32 joints) to an exponential map tuple (32*3). We matched the skeleton structure defined by Azure Kinect with the joint structure of H3.6M following the human-joint kinematic tree [6] of H3.6M (Fig. 3 left). Our system utilizes this correspondence to convert the input Azure Kinect data into the input of STS-GCN.

tree
item (32)

name	id	parent	children
Hips	0	0	[2 7 12]
RightUpLeg	1	1	3
RightLeg	2	2	4
RightFoot	3	3	5
RightToeBase	4	4	6
Site	5	5	
LeftUpLeg	6	1	8
LeftLeg	7	7	9
LeftFoot	8	8	10
LeftToeBase	9	9	11
Site	10	10	
Spine	11	1	13
Spine1	12	12	[14 17 25]
Neck	13	13	15
Head	14	14	16
Site	15	15	
LeftShoulder	16	13	18
LeftArm	17	17	19
LeftForeArm	18	18	20
LeftHand	19	19	[21 23]
LeftHandThumb	20	20	22
Site	21	21	
L_Wrist_End	22	20	24
Site	23	23	
RightShoulder	24	13	26
RightArm	25	25	27
RightForeArm	26	26	28
RightHand	27	27	[29 31]
RightHandThumb	28	28	30
Site	29	29	
R_Wrist_End	30	28	32
Site	31	31	

Azure Kinect	H3.6M
PELVIS	0/11
SPINE_NAVEL	Unnecessary
SPINE_CHEST	12
NECK	13/14/16/24
CLAVICLE_LEFT	Unnecessary
SHOULDER_LEFT	17
ELBOW_LEFT	18
WRIST_LEFT	19/20/21
HAND_LEFT	22/23
HANDTIP_LEFT	Unnecessary
THUMB_LEFT	Unnecessary
CLAVICLE_RIGHT	Unnecessary
SHOULDER_RIGHT	25

Fig 3. Kinematic tree of 32-joint body (H3.6M)(left) and the part of joints annotations from Azure Kinect to H3.6M(right)

4 Conclusion

We have developed a system that predicts a single person's bodyweight exercise poses in future moments. The preliminary test shows that our system can predict future poses that follow common movement rules.

In future work, we will implement the pose evaluation and feedback. We plan to ask the user to do specific bodyweight exercises in advance and record the pose data as standard pose. With the comparison result of standard pose and prediction pose, our system will be able to evaluate the user's poses. We also plan to customize audio feedback with words or sounds.

In addition, to improve the transformation quality of exponential maps, we consider using Euler angles or quaternions to represent rotations of each joint and implement them as the transformation source of the exponential map instead of XYZ coordinates.

As for visualization, based on current 2D visualization methods, we plan to project 3D avatars of predicted poses on HoloLens.

5 References

[1] Sofianos T. (2021). Space-time-separable graph convolutional network for pose forecasting. In Proceedings of the IEEE/CVF International Conference on Computer Vision (pp. 11209-11218).

[2] E. Wu and H. Koike, "FuturePose - Mixed Reality Martial Arts Training Using Real-Time 3D Human Pose Forecasting With a RGB Camera," 2019 IEEE Winter Conference on Applications of Computer Vision (WACV), Waikoloa, HI, USA, 2019, pp. 1384-1392.

[3] C. Bregler and J. Malik. Tracking people with twists and exponential maps. Proceedings. IEEE Computer Society Conference on Computer Vision and Pattern Recognition (Cat. No.98CB36231), CVPR.1998.698581, pp. 8-15.

[4] Chengshuo Xia. VoLearn: An Operable Motor Learning System with Auditory Feedback. (UIST '21 Adjunct). Association for Computing Machinery, New York, NY, USA, 103–105.

[5] Duan, H., Zhao, Y., Chen, K., Lin, D., & Dai, B. (2022). Revisiting skeleton-based action recognition. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (pp. 2969-2978).

[6] S.Toyer. Human Pose Forecasting via Deep Markov Models. (DICTA), 2017, pp. 1-8.

[7] Homepage of H3.6M dataset: http://vision.imar.ro/human3.6m/description.php

† University of Tsukuba, Master's program of Computer Science
‡ University of Tsukuba, Department of Computer Science