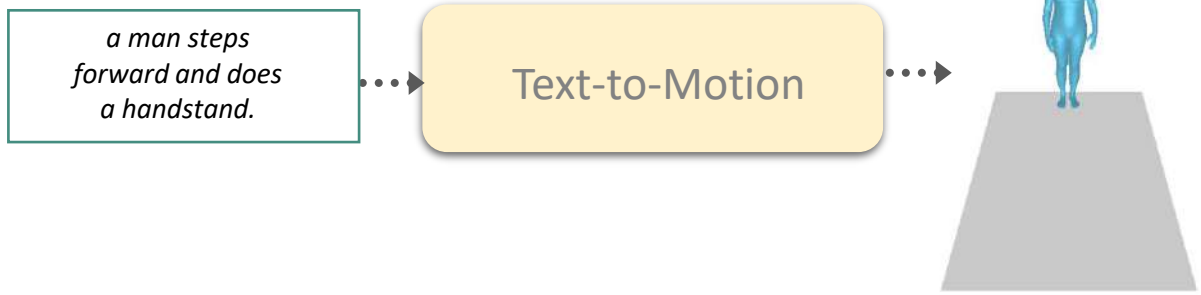
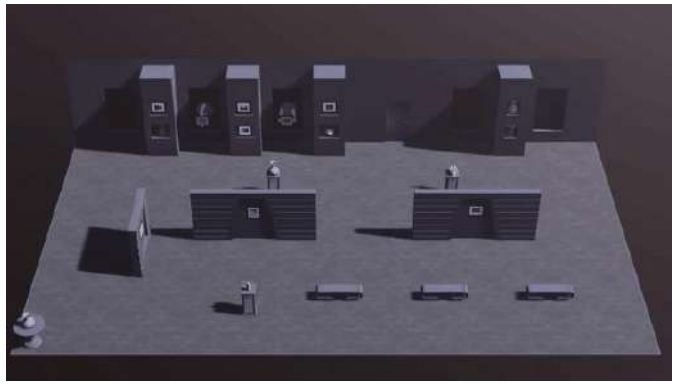
MotionScript: Natural Language Descriptions for Expressive 3D Human Motions

Payam Jome Yazdian, Rachel Lagasse, Hamid Mohammadi, Eric Liu, Li Cheng, Angelica Lim



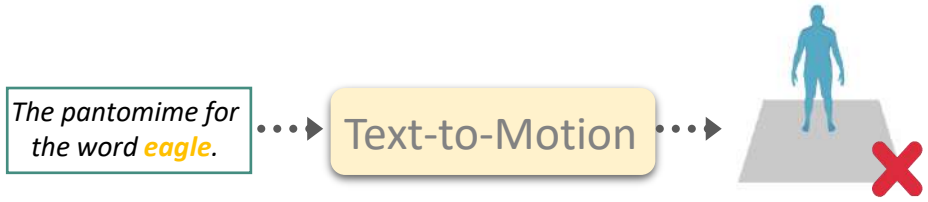


Virtual humans for robotics simulators

Animations for games, VFX

School of Computing Science, Simon Fraser University, Burnaby, BC, Canada
Dept. of Electrical and Computer Engineering, University of Alberta, Edmonton, Alberta, Canada

Text-to-Motion Models Struggle with Unseen Prompts



People or Animals Interactions

- Holding a baby's hands, teaching their first steps
- Attacked by a swarm of bees
- Walking a dog that suddenly gets away

Environmental Interactions

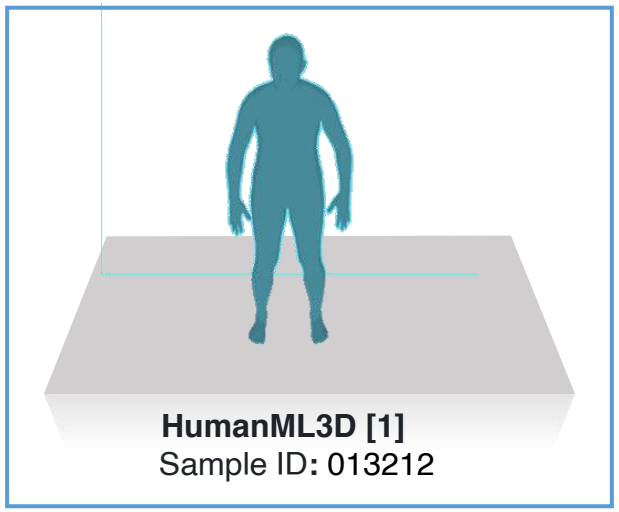
- Hanging clothes on a line, wind blowing them away
- Decorating a Christmas tree
- Sawing a block of wood when the saw suddenly gets stuck

Stylistic and Emotional Characters

- Soccer coach yelling at the team on the field
- Elderly woman doing water aerobics
- Eagle flying through the air

[1] Guo, Chuan, et al. "Generating diverse and natural 3d human motions from text." *CVPR*. 2022.

Text-to-Motion Models are Trained with Unaligned texts

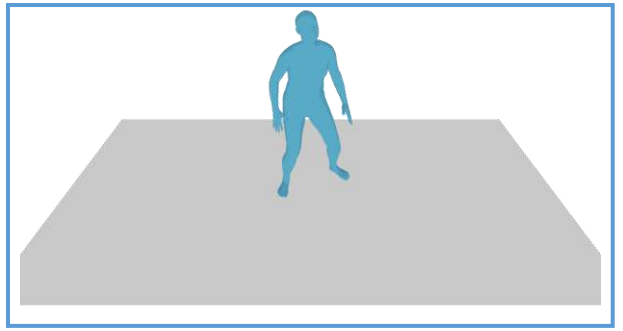


Human Annotations

- *The person was being a human monkey.*
- *Moving arms in a random pattern.*
- *Person makes move to scratch left armpit with left arm before dancing with arms waving.*

Our Solution:

MotionScript, an automatic framework that generates unlimited, perfectly aligned, low-level natural language captions from 3D human motion sequences.



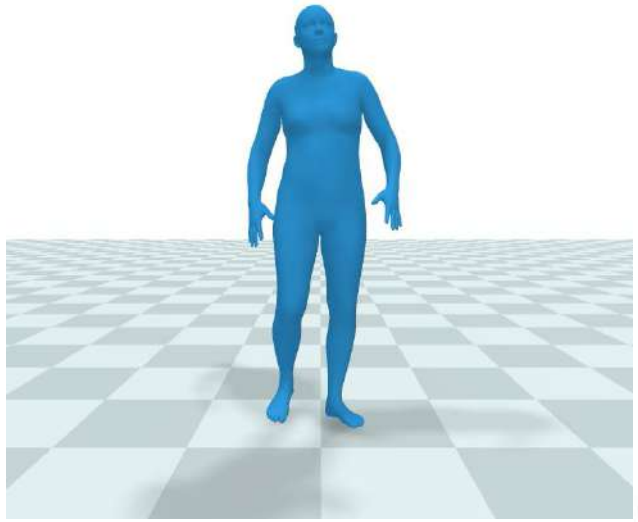
MotionScript

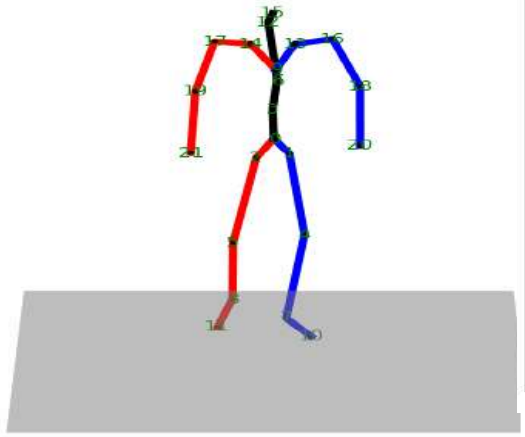
The person **moves** far **to the right** and at a normal speed while it is **moving forward** at a medium pace. In the seconds following, they **move** slightly **downwards**, and shortly thereafter, completely **turns clockwise**. The left elbow is rather bent and then, their left elbow **extends** super fast.

[1] Guo, Chuan, et al. "Generating diverse and natural 3d human motions from text." *CVPR*. 2022.

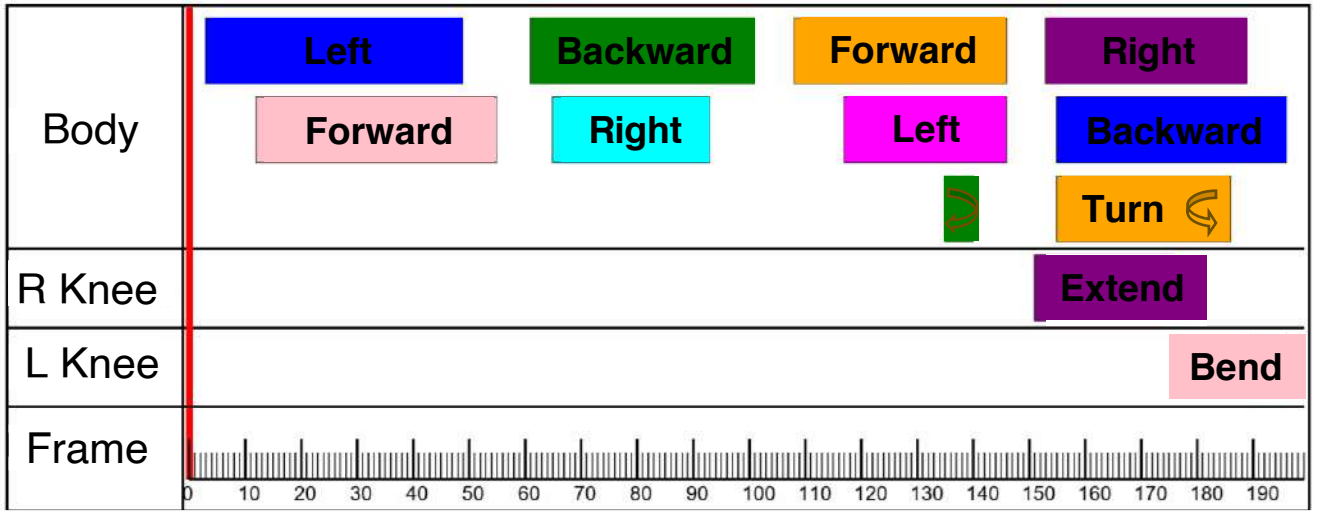
MotionScript: Motion to Text Conversion Algorithm

Input: 3D Motion

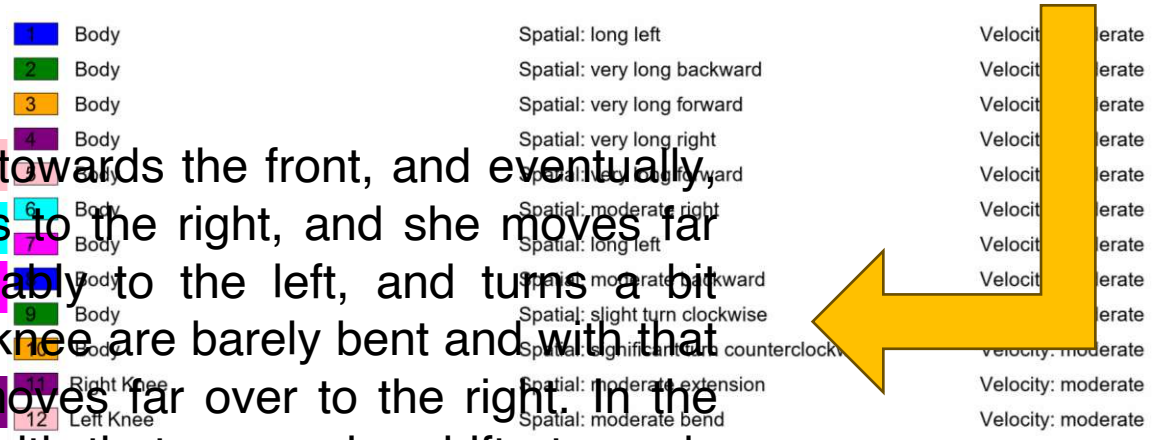




Dynamic Segment Detection Algorithm

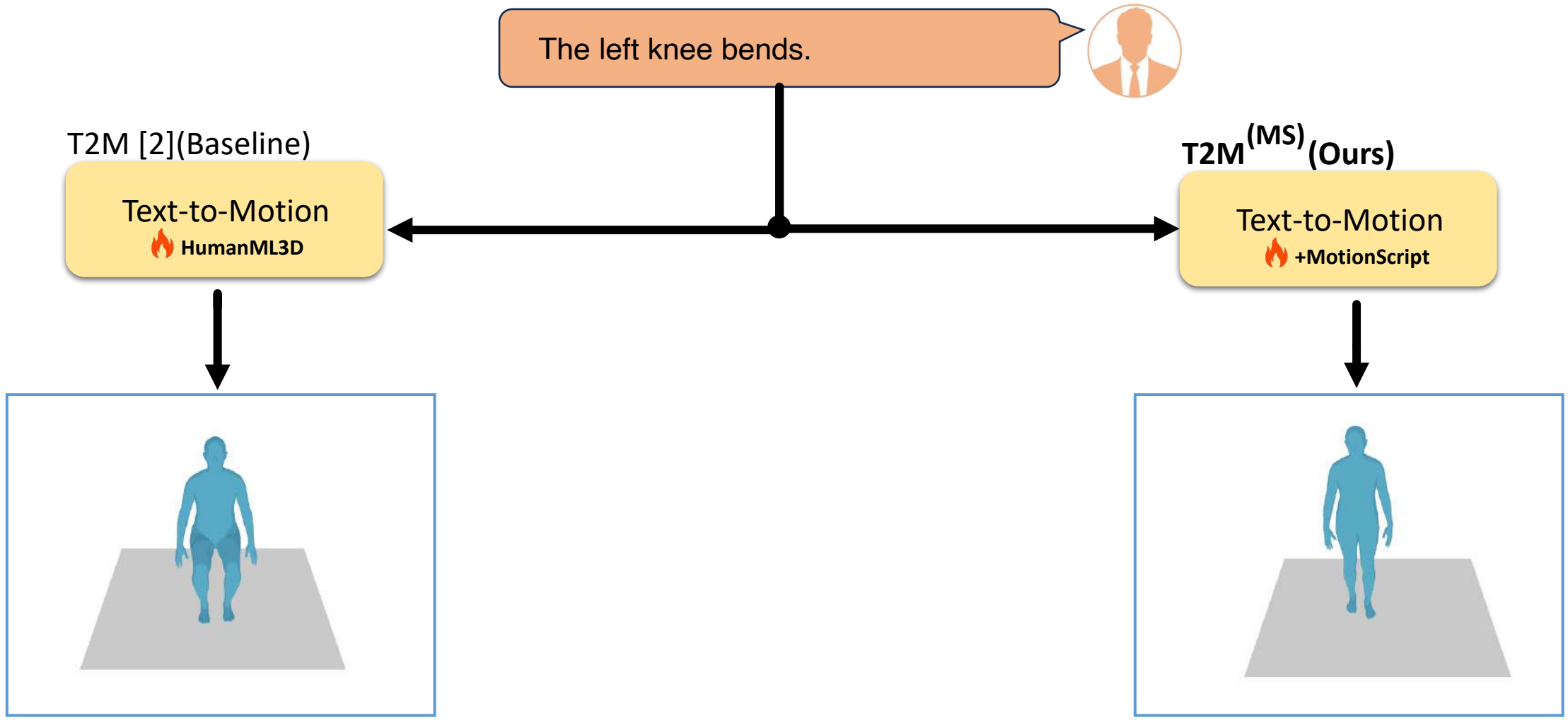


Motioncode Description



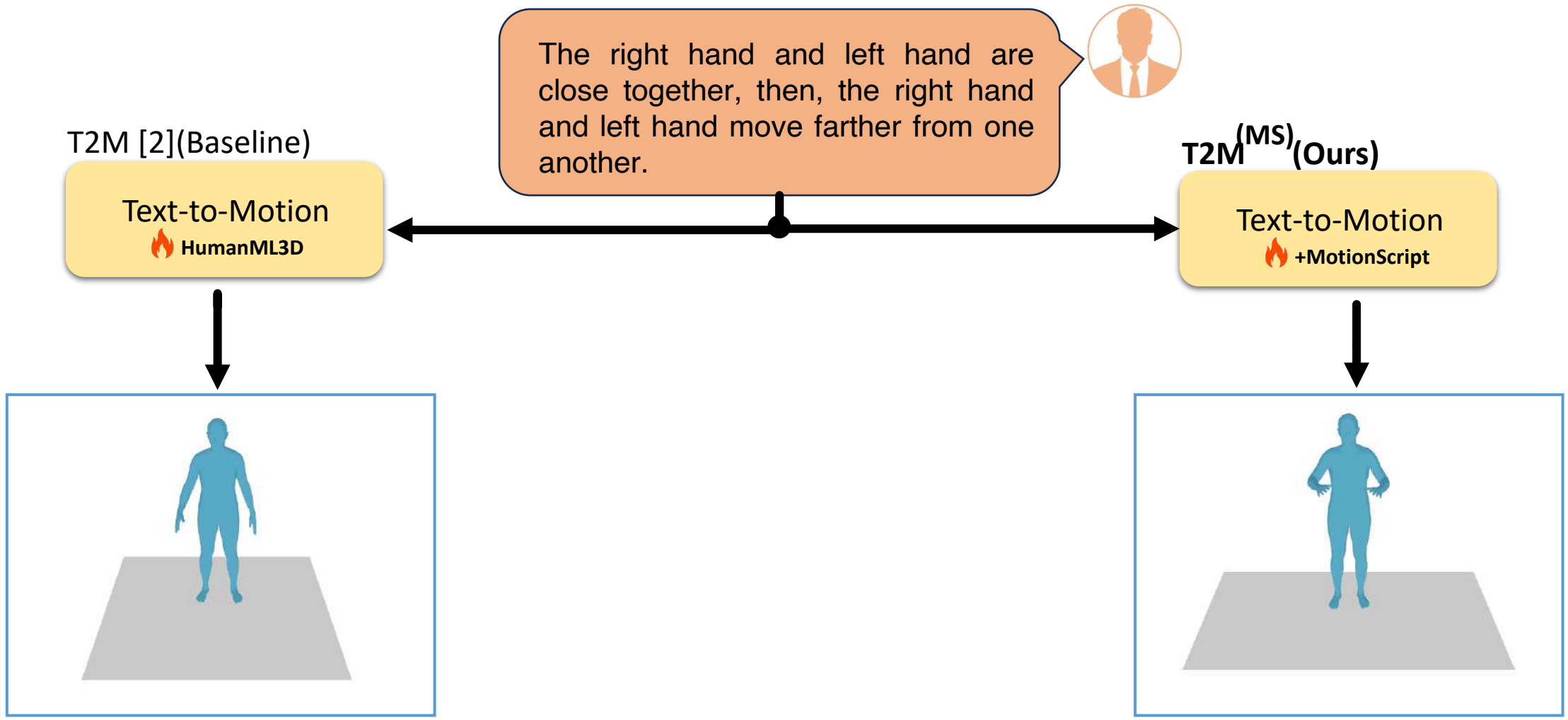
She moves noticeably to the left. She shifts far over towards the front, and eventually, she moves a great distance to the back. She moves to the right, and she moves far towards the front. Right after, she moves considerably to the left, and turns a bit clockwise. Shortly after, the right elbow and her right knee are barely bent and with that pose, the right knee extends. A second later, she moves far over to the right. In the meantime, she is moderately turned clockwise and with that pose, she shifts towards the back, and meanwhile, she turns significantly counterclockwise. Shortly after, the left knee is unbent. From this position, her left knee bends.

Training a text-to-motion (T2M) model with MotionScript captions



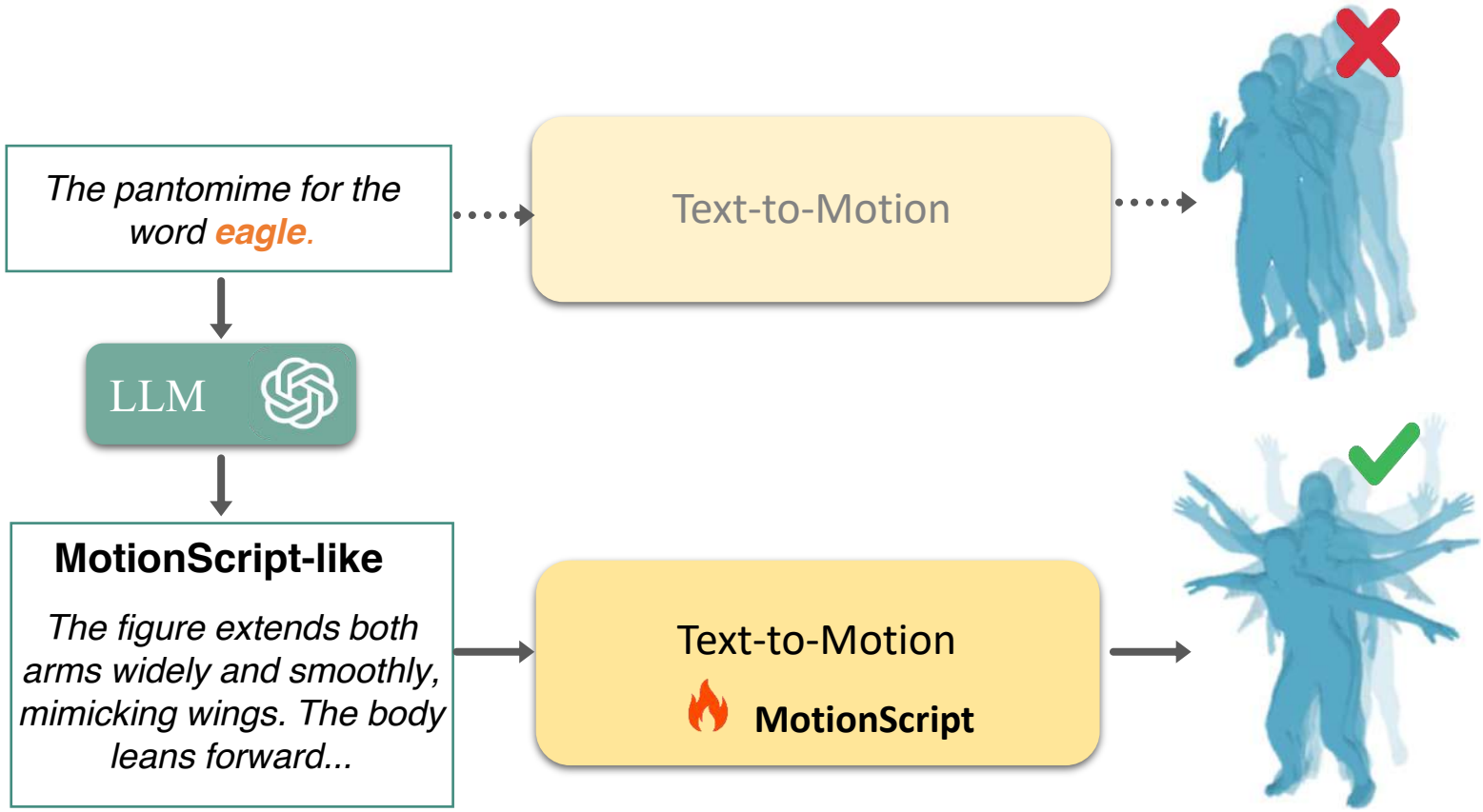
[2] Zhang, J., et al. Generating human motion from textual descriptions with discrete representations. CVPR, 2023.

Training a text-to-motion (T2M) model with MotionScript captions



[2] Zhang, J., et al. Generating human motion from textual descriptions with discrete representations. CVPR, 2023.

Using an LLM to convert OOD text to MotionScript-like captions

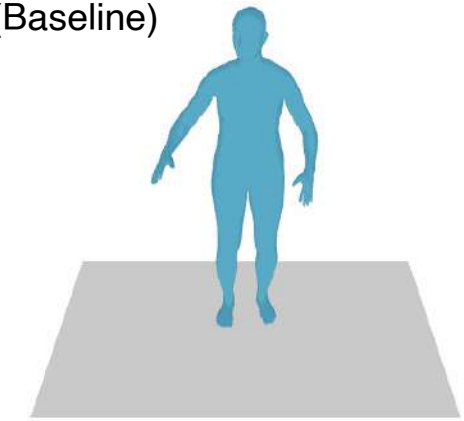


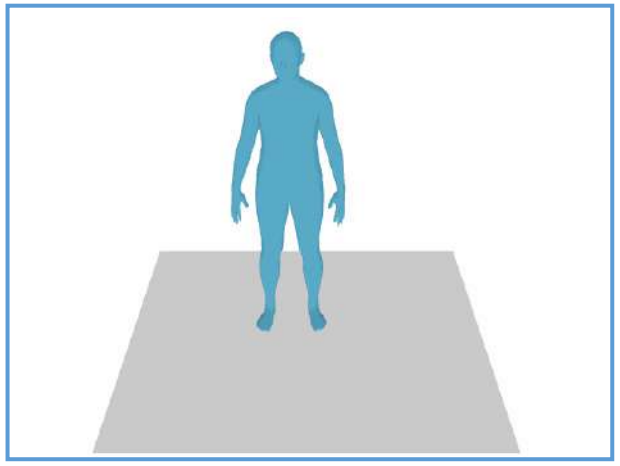
Text-to-Motion Results on OOD text prompts

T2M

Text-to-Motion
HumanML3D

(Baseline)





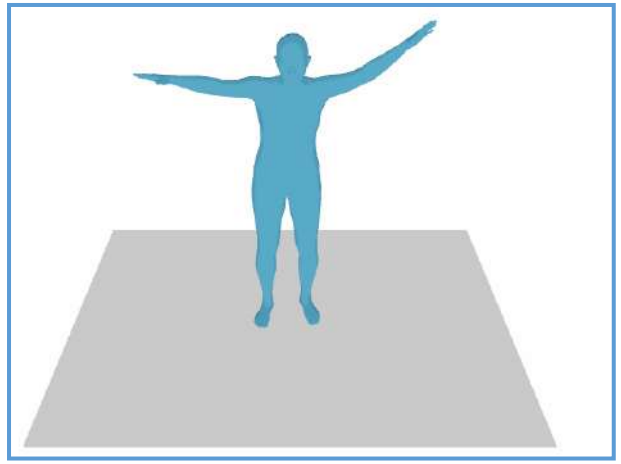
A person is attacked by a swarm of bees.

LLM

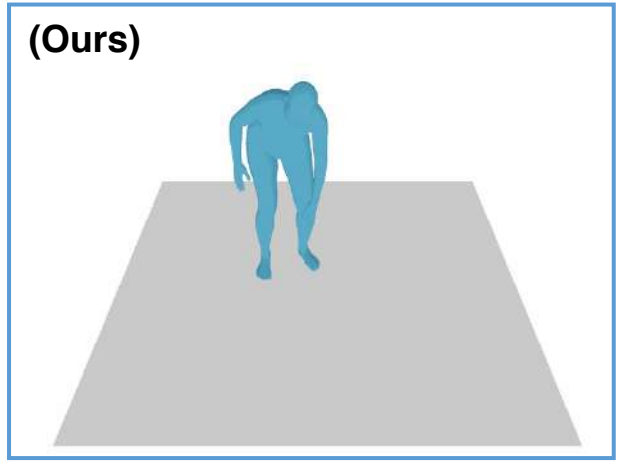
The person is surrounded and assaulted by a swarm of bees. Initially, the individual's arms are at their sides. Rapidly, both arms swing wildly around, attempting to swat the bees away. A moment later, the person …

T2M (MS)

Text-to-Motion
+MotionScript



(Ours)

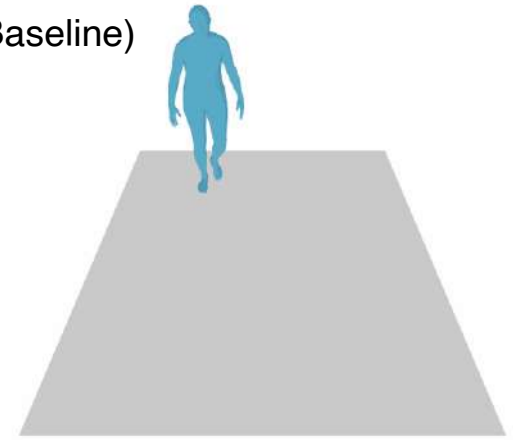


Text-to-Motion Results on OOD text prompts

T2M

Text-to-Motion
HumanML3D

(Baseline)



You are walking along but your dog keeps grabbing hold of your slipper.

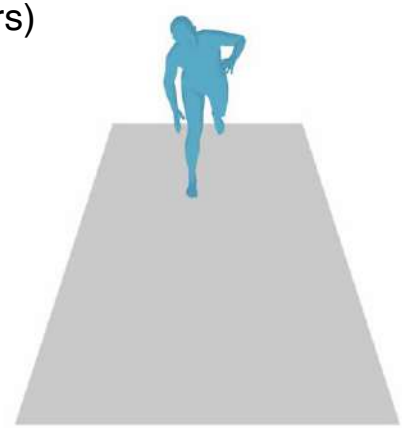
LLM

You are in a walking action. As you shift forwards, your right foot raises slightly and then drops back down rapidly. Moments prior, your left foot extends and plants firmly on the ground. Your dog, positioned near your …

T2M (MS)

Text-to-Motion
+MotionScript

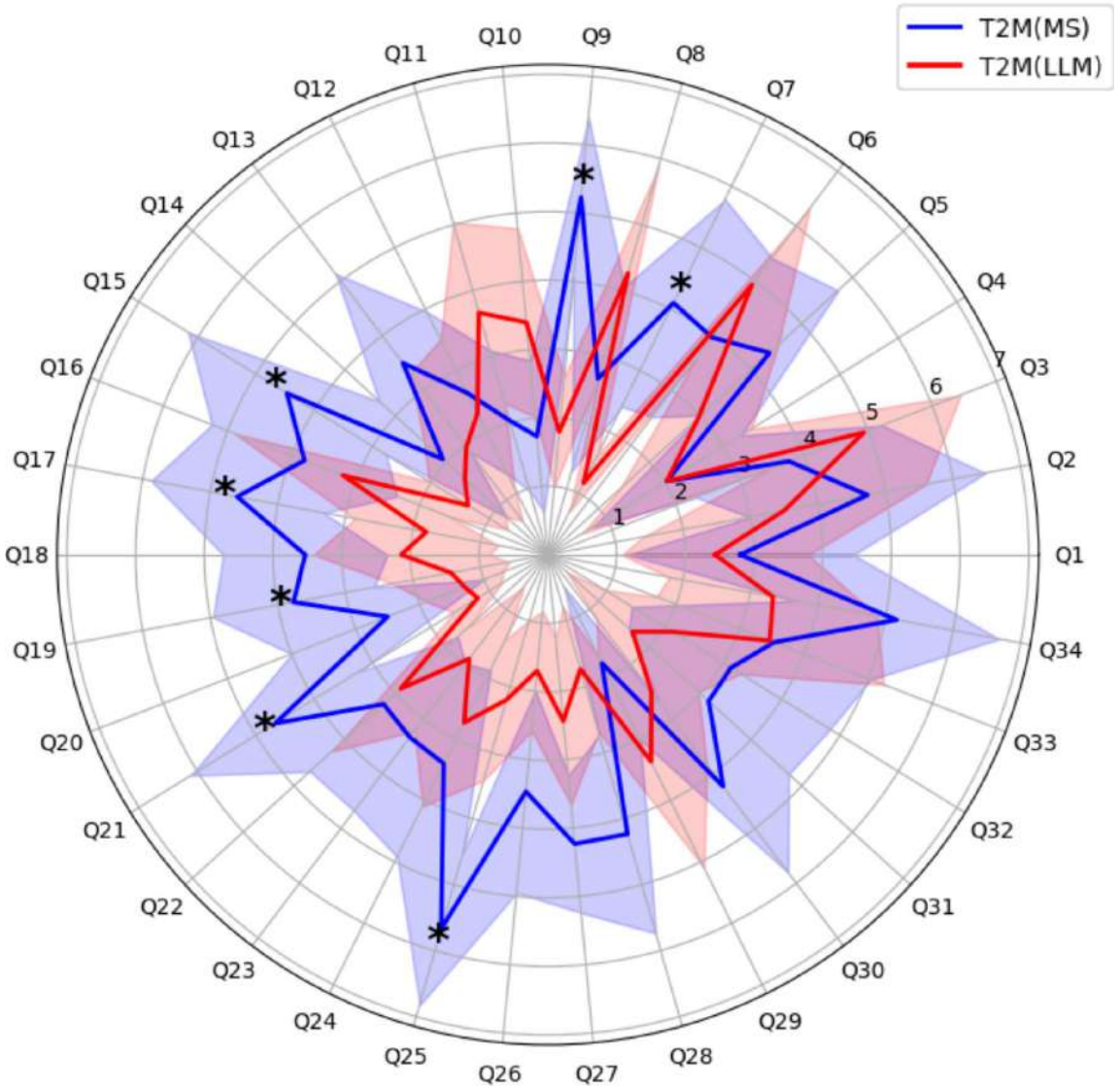
(Ours)



Human study shows strong preference for MotionScript on OOD texts

How well does video X match the prompt?

- 34 prompts, 30 participants
- Compared our T2M$^{(MS)}$ with a model trained with detailed captions generated by a language model, T2M$^{(LLM)}$
- Strong preference for our MotionScript approach



Experiment 2 Mean ± SD.* indicates significance.

MotionScript: Natural Language Descriptions for Expressive 3D Human Motions

Payam Jome Yazdian, Rachel Lagasse, Hamid Mohammadi, Eric Liu, Li Cheng, Angelica Lim

Key takeaways

- **First systematic framework** for translating 3D motion to structured natural language
- **Bridges LLM reasoning** with precise motion synthesis for out-of-distribution scenarios
- **Significant human preference** for MotionScript-augmented models over baselines

https://pjyazdian.github.io/MotionScript

*scan to visit the project page