УДК 004.8

ГЕНЕРАЦИЯ MOTION CAPTURE ПО ТЕКСТОВОМУ ОПИСАНИЮ

Лапин М.В. (Университет ИТМО)

Научный руководитель – доцент, кандидат физико-математических наук Трифанов А.И. (Университет ИТМО)

Введение. Зачастую при обучении сложной модели с большим числом степеней свободы (примером может служить антропоморфный робот или другая сложная робототехническая система) может возникнуть разнообразный ряд сложных форм поведения, что довольно часто приводит к разным выбросам и артефактам. Для получения более стабильной и качественной работы модели разрабатывают различные подходы, одним из которых является имитационное обучение. Во многих работах [1, 2] на этапе обучения модели используют эталонные движения для уменьшения артефактов в поведении. Одним из самых популярных способов создания эталонного движения является motion capture (захват движения) [5]. Существует два основных вида систем захвата движения: маркерная и безмаркерная. Однако и в том и другом случае - это достаточно сложный и дорогостоящий метод. При этом сбор и хранение видео с размеченными данными может быть затруднен или подходящего эталонного действия для конкретной модели может не найтись в базе данных. Более того, данный способ сложно применить в инференсе, так как во время real-time выполнения крайне ограничены временные ресурсы на принятие решения.

Основная часть. Для решения данной проблемы предлагается модель, генерирующая motion capture на основе текстового описания или последовательности текстовых описаний. Для создания согласованных во времени кадров захвата движения была разработана модель, общая структура которой включает два основных блока:

1) Автокодировщик, состоящий из кодера и декодера и отвечающий за сжатое латентное представление видеоряда ключевых точек захвата движения. Модуль извлекает внутреннее векторное представление потока необработанных кадров и преобразует их с помощью пространственного и причинно-следственного авторегрессивного трансформера.
2) Предобученная языковая модель [6], используемая для получения векторных представлений слов из текста.

Для обучения представленной выше архитектуры был собран и размечен корпус данных с motion capture на основе базы данных CMU [7].

Выводы. Рассмотрены основные подходы при моделировании комплексных робототехнических систем. Проведен анализ методов создания motion capture. Была спроектирована и разработана архитектура для генерации клипов захвата движения на основе текстового описания.

Список использованных источников:

1. Peng, X.B., Abbeel, P., Levine, S. and Van de Panne, M., (2018) Deepmimic: Example-guided deep reinforcement learning of physics-based character skills // ACM Transactions On Graphics (TOG), 37(4), pp.1-14.
2. Peng, X.B., Berseth, G., Yin, K. and Van De Panne, M. (2017) Deeploco: Dynamic locomotion skills using hierarchical deep reinforcement learning // ACM Transactions on Graphics (TOG), 36(4), pp.1-13.
3. Mahmood, N., Ghorbani, N., Troje, N.F., Pons-Moll, G. and Black, M.J. (2019) AMASS: Archive of motion capture as surface shapes // In Proceedings of the IEEE/CVF international conference on computer vision (pp. 5442-5451).
4. Marian Bittner, Wei-Tse Yang, Xucong Zhang, Ajay Seth, Jan van Gemert, Frans C. T. van der Helm (2023) Towards Single Camera Human 3D-Kinematics // Sensors 2023, 23(1), 341.

5. Sharma, S., Verma, S., Kumar, M. and Sharma, L. (2019) February. Use of motion capture in 3D animation: motion capture systems, challenges, and recent trends // In 2019 international conference on machine learning, big data, cloud and parallel computing (comitcon) (pp. 289-294). IEEE.

6. Wang, H., Li, J., Wu, H., Hovy, E. and Sun, Y. (2022) Pre-Trained Language Models and Their Applications // Engineering.

7. Сайт базы данных захвата движений CMU [Электронный ресурс] – Режим доступа: http://mocap.cs.cmu.edu/search.php?maincat=3&subcat=2