Generative Model of Human Body Motion for Walking

Research Project
KIREPRO1PE

Mathias Bastholm
mbas@itu.dk

December 6, 2020

1 Introduction

short intro to the project

2 Background

Several tasks exist around the modelling of human motion and the given task usually dictates the shape of the model to a certain extent.

One such example is animating a human 3D mesh to conform with input from a game controller or equivalent keyboard input. This task is of particular interest for the video game industry. One such model is a phase-functioned neural network, wherein multiple networks are samples in a phase, as to mimic the repetitive nature of walking [1]. This approach can be extended to other tasks, such as playing basketball [2]. This approach values being able to run in real-time higher than animation quality. This makes it ideal for video games, but less ideal for other applications where the animations are typically produced in advance and as such there is no requirement for real-time production.

Another example is motion prediction, where the model is given a sequence of human poses and asked to generate future human poses. This is of particular interest for the autonomous car industry, but also has a potential impact for all industries that produce animations, which includes industries working with: movies, video games, entertainment, infotainment, education, etc. As auto-completing an animation could potentially speed up the time it takes to produce a finished animation. A common approach to this task is to use recurrent neural networks [3, 4]. Generative adversarial networks (GANs) are commonly used in image based generative tasks, such as generating human faces [5]. GANs have been shown to also be effective for human motion prediction [6].

In-painting of human motion can be considered a special case of motion prediction, where the model gets both the past and future motion, instead of only getting past motion. This could be used to help speed up the animation of humanoids for video games, movies, etc. It seems that data on this particular task is relatively sparse.

3 Method

3.1 Data

Training a model for human motion analysis requires a great deal of human motion data. There exists two efforts to help in this regard. First of the AMASS data set is an attempt to amass a large quantity of human motion data from motion capture [7]. Secondly the fairmotion library is a library that allows you to take motion from all of the heterogeneous formats currently available, including the

AMASS data set, and access all of it through one homogenous interface [8]. The fairmotion library requires both motion data and skeleton/model data of a human. As such, AMASS was used for the motion data and [9] for the skeleton/model data.

Only a subset of the AMASS data set was used. This sped up training by training on less data and it meant not having to stream in training data, as the subset used was small enough to fit into memory.

3.2 Models

detailed description of my models what makes them unqiue how they work, etc

3.3 Training

notes on how the training was performed

4 discussion

4.1 Results

a section presenting my results and comparing them to other results, might have to create some "artificial" results to get anything good to compare with, like using linear interpolation as a baseline, and final ik as "competing" solution, this is also a good place to show some concrete examples of what my models can do, using images

4.2 Further research

as with all papers that are things that i did not get to explore, this will detail where i think further energy should be exerted based on my findings.

5 Conclusion

References

1. Holden, D., Komura, T. & Saito, J. Phase-Functioned Neural Networks for Character Control. *ACM Trans. Graph.* **36.** ISSN: 0730-0301. https://doi.org/10.1145/3072959.3073663 (July 2017).
2. Starke, S., Zhao, Y., Komura, T. & Zaman, K. Local Motion Phases for Learning Multi-Contact Character Movements. *ACM Trans. Graph.* **39.** ISSN: 0730-0301. https://doi.org/10.1145/3386569.3392450 (July 2020).
3. Hu, J., Fan, Z., Liao, J. & Liu, L. *Predicting Long-Term Skeletal Motions by a Spatio-Temporal Hierarchical Recurrent Network* 2020. arXiv: 1911.02404 [cs.CV].
4. Jain, A., Zamir, A. R., Savarese, S. & Saxena, A. *Structural-RNN: Deep Learning on Spatio-Temporal Graphs* 2016. arXiv: 1511.05298 [cs.CV].
5. Karras, T., Laine, S. & Aila, T. *A Style-Based Generator Architecture for Generative Adversarial Networks* 2019. arXiv: 1812.04948 [cs.NE].
6. Ruiz, A. H., Gall, J. & Moreno-Noguer, F. *Human Motion Prediction via Spatio-Temporal Inpainting* 2019. arXiv: 1812.05478 [cs.CV].
7. Mahmood, N., Ghorbani, N., F. Troje, N., Pons-Moll, G. & Black, M. J. *AMASS: Archive of Motion Capture as Surface Shapes* in *The IEEE International Conference on Computer Vision (ICCV)* (Oct. 2019). https://amass.is.tue.mpg.de.
8. Gopinath, D. & Won, J. *fairmotion - Tools to load, process and visualize motion capture data* Github. 2020. https://github.com/facebookresearch/fairmotion.
9. Romero, J., Tzionas, D. & Black, M. J. Embodied Hands: Modeling and Capturing Hands and Bodies Together. *ACM Transactions on Graphics, (Proc. SIGGRAPH Asia). 245:1–245:17* **36** (Nov. 2017).