Character Locomotion using Imitation Learning from Observations with Wavelets

João Carias
joao.carias@tecnico.ulisboa.pt

Instituto Superior Técnico
http://tecnico.ulisboa.pt
January 2021

This work presents a novel extension to the paradigm of Imitation Learning from Observations by integrating frequency information into the learning structure, in the context of 3D physically-based running animation. This kind of animation is very repetitive, thus it may be possible to improve the system by introducing frequency information. We proposed a Model for Wavelet Augmented Imitation Learning from Observations, by implementing a Wavelet Transform extraction component on top of DeepMimic, and conducted empirical tests to further develop and test our approach against the DeepMimic framework. Results show that the implemented model offered no benefit to the trained policy. Moreover, there was a dramatic increase in training time and the obtained policy had similar performance but was less resilient to external forces, compared to the base system. The code is available at https://github.com/jocarias/DeepMimic.

Main Concepts

Imitation Learning from Observations; Continuous Wavelet Transform; Convolutional Neural Network; Reinforcement Learning; DeepMimic.

1 Introduction

1.1 Motivation

In the context of 3D computer games, virtual characters play a fundamental role in the quality of immersion experience provided to the player. In order for a character to blend in with its surroundings, and thus providing a believable setting, it must show a natural and realistic animation - often as common forms of locomotion like walking and running.

Traditionally, these kinds of animations require specialized skills and are time-consuming to produce. Moreover, they are not able to react to unforeseen or unprepared interactions, which are bound to happen in complex scenarios, degrading the player's experience.

Physically-based animation [1, 2, 3, 4] promotes the correct behavior of objects and characters by exposing them to the laws of physics and thus, providing a realistic experience. A simple example is Ragdoll Physics, commonly used to animate a character's body in specific situations, like unconscious/death animation.

In recent times, important developments in Deep Reinforcement Learning and to the hardware which it is run on are allowing new efforts to emerge leading to the creation of a controller that is able to command the members of the simulated body, orchestrating a set of synchronized actions that produces an intended animation, reacting naturally and intelligently to unplanned interactions. In this context, locomotion animations, like walking and running, are an active area of intensive research and several approaches have been presented, which can be divided by the ones that need motion data [1, 2] and the ones that don't [3, 4]. In terms of the former, while achieving impressive results, some limitations need to be addressed, such as improving the flexibility of the animation time-wise and the robustness, given external forces.

It's also worth noting that improving locomotion of a virtual agent in a simulation may offer benefits, in some cases, to the control of robotics systems in the real world.

1.2 Problem

This work will focus on the development of a system that, on a human-like 3D character in a physics-based simulation, will be capable of generating realistic and robust running animation, from motion data.

The system characteristics we seek to improve are at the level of robustness, the ability to overcome external random forces, and the time it takes for the system to reach a realistic animation, one that agrees with a person's expectation. An relevant benefit of contextualizing this problem in a physics-based simulation is the inherent possibility of finding policies capable of recoveries from unexpected forces, making the locomotion more realistic.

1.3 Hypothesis

Motion data can be a flexible source of different locomotion types and styles. Imitation Learning from Observations - learning a policy with data from another agent performing a task - appears to us as the ideal framework from which a system can be based on, in order to learn simple or even complex motions. For these types of systems in this context, the input information is often a set of state variables (position, angles, velocities) with the current configuration of the character's body (a set of members/links connected by joints).

In this work, we will explore a new idea based on the observation that the animation that is our focus - running on even ground - is very repetitive, meaning every so often the same movement is executed just with minor adjustments

that are needed for a stable locomotion. Given the periodicity of the various movements, it is our belief that it can be processed in the frequency domain along the traditional time domain, adding otherwise hidden information that may help our objectives.

The Fourier Transform and its close variants have been used to better analyze periodic signals though it suffers from the inability to localize, with some precision, frequency in time, which will be a necessity for the types of signals that change abruptly. Wavelets, a more recent approach, have both good time and frequency localization. This rationale has guided good results, both in images and sound processing and particular speech recognition, though these types of signal are quite different among themselves and from the task at hand. Additionally, Wavelets have been used to analyze walking patterns [5, 6] and even to compress Human Motion Capture [7].

We believe that the use of Wavelets may bring additional unexplored advantages to assist in solving our problem so it is our interest to explore the possibility of using this mathematical tool to transform how the system processes data and understand its effects on tasks previously mentioned, advantages, if any, and its limitations.

2 Related work

Here, we present some of the relevant work in two different areas that are pertinent to this work. The Character Locomotion Animation section explores some of the work done in this area and pays particular attention to Deep Mimic, a state-of-the-art approach. Wavelets in Gait Analysis section presents some of the work done that is believed to provide a better understanding of their uses and capabilities in this work's context.

2.1 Character Locomotion Animation

Being an area of intense past and present research, much ground has been laid to this day. In the context of physics-based biped character, a diverse number of techniques are presently available to tackle this challenge.

Due to several recent developments in the field of Machine Learning - particularly in Deep Reinforcement Learning - some approaches have been gathering notoriety due to their impressive results. In [3] a character learns, using the PPO algorithm, a large set of locomotion styles with environment awareness.

Muscle-based locomotion [8] provides, to a character, a realistic foundation along with the proper range and reaction of movement of body members. For instance, in [9], the authors developed a generic control method compatible with different bipedal simulated creatures, where no motion data is needed. It consists of the optimization of muscle routing and control parameters resulting in the ability of

moving in a target direction at a target speed on irregular ground while being the focus of external perturbation.

Deep Mimic [1] is a framework for synthesizing physics-based character animation that fuses two distinct learning strategies - Imitation Learning from Observation and goal-direct Reinforcement Learning.

It is composed of several pre-existing components from different sources combined in such a way that, for a character model, it produces robust animations that closely follow the given motion data while able to achieve a goal, defined by a given reward function.

The system handles motion data from two different sources, motion capture (mocap) and keyframes. Mocap data, which has a duration between 0.5 and 5 seconds, is obtained from Carnegie Mellon University[1] and Simon Fraser University[2] Motion Capture Databases while keyframes data comes from artist-authored animations. The data was manually processed and retargeted to the different characters.

Following the authors's convention, the motion data is a sequence of target poses $\{\hat{q}_t\}$ and the control policy $\pi(a_t|s_t, g_t)$ is the mapping of the character's state s_t and goal g_t to an action a_t.

The state s represents the current configuration of the character's body, defined by several properties for each joint - position (x,y,z), quaternion rotation (w,x,y,z), linear and angular velocities (both x,y,z). Properties are computed in the local coordinate frame of the character, being the root (pelvis) at the origin and the facing direction being along the x-axis. There is also a phase variable ϕ due to the fact that the motion data's target poses change with time. Its value sits between 0 and 1, where 0 indicates the start and 1 the end of the motion where, in the case of a cyclic motion like walking and running, ϕ is reset to 0. An instance of the state vector starts with the phase variable followed by the root y value, then a sequence of position and rotation of each joint for all 15 joints, and ending with a sequence of the linear and angular velocities of each joint for all joints, totaling 197 variables.

The action contains a set of target angles that are transformed by proportional-derivative (PD) controllers into torques, which are applied to the model's joints. Targets for spherical joints and revolute joints are represented as axis-angle form and scalar rotation angles respectively.

2.2 Wavelets in Gait Analysis

Human gait analysis, the study of locomotion, is performed taking in consideration the characteristics of the data that supports it - low frequency shape and high frequency discontinuities [10].

[1] http://mocap.cs.cmu.edu
[2] http://mocap.cs.sfu.ca

For human muscle activity signals, wavelet analysis has also been extensively used for diagnostics in a clinical context [11]. Moreover, the differentiation of gaits from different people, is a useful feature in several different applications; for instance, the detection of early stages of Parkinson's disease [5].

Given the variability of options in terms of Wavelet analysis characteristics, it's important to find the appropriate ones for the task at hand and some studies have tried to shed some light on this, particularly for scale [12] and Mother wavelet [6]. The later work serves the purpose of investigating, in the context of detecting gait events using Continuous Wavelet Transform (CWT), the differences in performance of different mother wavelets for both hemiplegic and healthy individuals. The authors argue that being able to detect gait events is essential for several applications in the Human healthcare area like control mechanisms in drop foot correction devices, recognizing human activity and aiding the decision on rehabilitation strategies. When walking, two gait events - heel strike (HS) and toe off (TO) - are commonly regarded as the most relevant ones in a normal gait cycle, providing swing, stance and stride parameters information. Hence, these were the gait events with this type of locomotion that this work focused on. The source of the data for this study was provided by 16 individuals (3 of them were hemiplegic patients) using a wireless tri-axial accelerometer device on one of the lower legs, just below the knee.

As explained by the authors, the walking cycle can be divided into a double repetition of a sequence composed of a stance and a swing phase, which beginnings are indicated by HS and TO gait events, respectively. These events can be detected using CWT to process gait data, given the time-frequency connection between gait event and gait cycle, proved by previous studies that showed the effectiveness and stability of the CWT, even when subjected to disturbances.

3 Towards a Model for Wavelet Augmented ILFO

In this section we present a Model for Wavelet Augmented Imitation Learning from Observations (WAILFO). Finishing the chapter, there's a discussion of the results.

3.1 WAILFO

DeepMimic [1] applies Imitation Learning from Observations to obtain similar animations as those shown by the given motion data while under the constraints of a physics-based simulation. With the objective of developing a system capable of imitating, robustly, motion data of a running animation, with the same constraints, we based our system on DeepMimic, a fully functioning system.

The gist of our work is the integration of the Continuous Wavelet Transform coefficients, obtained from State data, into the system. By making use of the structure already in place, capable of finding good policies, the implementation complexity is reduced and the chances of success grow. Thus, the existing

neural networks are the target of the set of modules developed/used in this work. The added modules are a Memory Buffer (to save State data), a CWT (to transform several instances of the State data) and a CNN (that outputs scalogram features). These are connected by the following pipeline: a sequence of DeepMimic's State variables are saved in the Memory Buffer, which feeds into a Continuous Wavelet Transform (CWT) that, for each state variable, returns the respective coefficient matrix. Each coefficient matrix, like an image, is the input of a Convolutional Neural Network (CNN) that outputs high level scalogram features to DeepMimic's existing network, alongside the State Vector. The complete network can be seen in Figure 1.

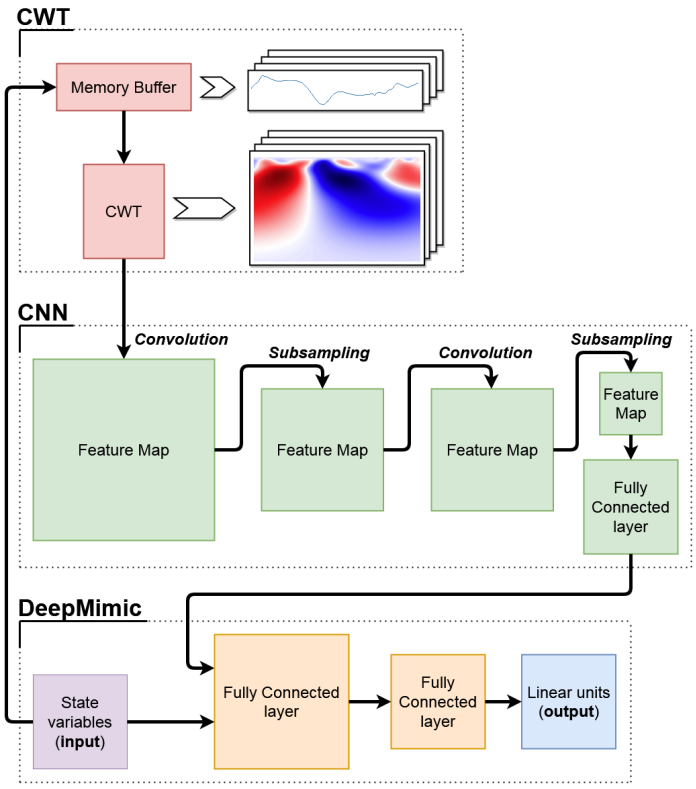

Fig. 1. Architecture of the policy network of the proposed WAILFO model. The modules of the new pipeline are in red (Memory Buffer and CWT) and green (CNN). Its output, alongside state variables, provide the input for the sequence of two fully connected layers. The final layer contains linear units that outputs target angles for the PD controllers of the joints. In the scalogram, which shows the CWT coefficients of the signal above it, the red color represents positive coefficients, blue color represents negative coefficients and white represents values close to zero.

It was expected that the new additions would represent a significant computational weight to the CPU on an already demanding load - state variables may need to be acquired more frequently, i.e. among policy updates; then stored/disposed in a first-in first-out (FIFO) buffer; the history of each variable is transformed by the CWT into a scalogram; all the scalograms are processed by the CNN. Of notice is the ability of DeepMimic to train with several parallel agents using different CPU threads. Since the state is unique to each agent, all the calculations mentioned previously have to be performed in each thread.

The humanoid model is used to provide the embodiment for the system in the environment whose state information (this model's joints and phase variable) are fed to a network of 2 fully connected layers. By using the PPO algorithm, 2 similar networks are created in order to find the best policy - one used as actor, outputting the target angles for the PD controllers of the joints of the humanoid model, and one used as critic, returning the believed value of the current state. While being part of DeepMimic's capabilities, the goal input is ignored for this work alongside the ability to add terrain height information, since only regular terrain is used. The running animation, which is the target of the policy, is the one provided by mocap data already included in DeepMimic. In terms of Wavelet Transform type, the choices depend on the application. For our case - the need to analyze state signals - Continuous Wavelet Transform was chosen. This had the predicted drawback of being more computational intensive.

A reduced state was adopted, which contains the phase variable (1D), root linear velocity (3D), and the rotations from the right (4D) and left (4D) ankles, root (4D), right hip (4D), right knee (1D), right shoulder (4D), right elbow (1D), left hip (4D), left knee (1D), left shoulder (4D) and left elbow (1D) for a total of 36 variables.

Additionally, DeepMimic uses a strategy called early termination (ET) - stopping the learning episode when the character falls on the ground - that could interfer with the new model. To replace it, a new learning strategy, named Baby Walker, was introduced that allowed the character to remain in a straight position and floating in the air, by applying forces to the root joint.

3.2 Evaluation

In Figure 2 **a)** we can observe the performance difference between the base DeepMimic system and the WAILFO model. The base system reaches 0.91 while the new model does not surpass 0.88; as an additional comparison, the reduced state with Baby Walker system passes over 0.89. Figure 2 **b)** shows the training times of each system. The WAILFO model takes close to 2.4x more time than DeepMimic (or the reduced state) without bringing any improvements. Additionally, empirical tests show that the character using the policy is less resilient to external forces. All tests used an Intel Pentium G4560 CPU (Hyper-threading enabled) with 8GB RAM, running 2 agents (2 threads).

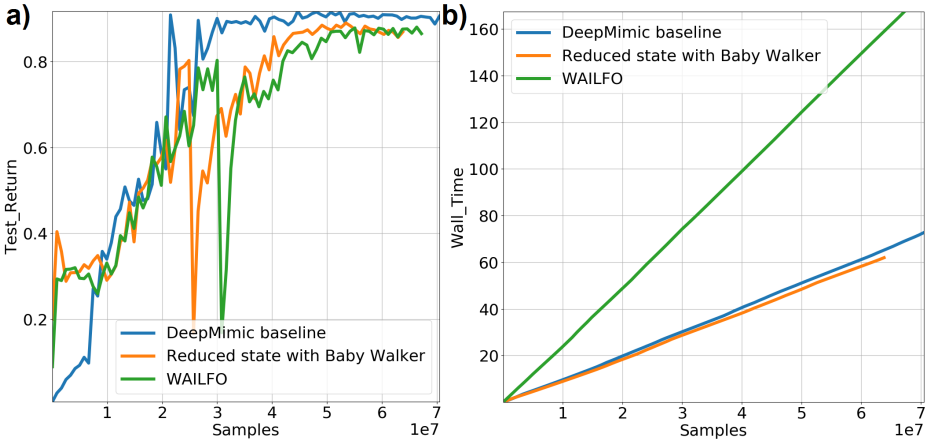

Fig. 2. Performance **a)** and training time **b)** comparison between DeepMimic baseline (blue), Reduced state with Baby Walker (orange) and the implementation of WAILFO model (green). *Wall_Time* is in hours.

4 Conclusions

The WAILFO model is a first instantiation of the novel approach, for this context, of introducing frequency information to the learning structure. The results are clear - the implemented model offered no observed benefits at the cost of training time and resilience to external forces. Some limitations may explain the results - the CWT calculations are demanding and the resulting scalograms, due to memory constraints, needed to be small and fewer; the CNN also increased the CPU load and there was the need for a simple network since it was not possible to use the GPU for the calculations. Despite this, we learned that some aspects of the implementation aren't advisable, namely saving scalograms in the replay buffer and calculating higher CWT scales (which we can consider wasted computational resources) or using a logarithmic scale distribution. We conclude that the broad approach is worth further investigation, whether with this particular model or other, but always with more computational power. Some compelling alternatives are introduced in the next section.

During this work, some realizations about the base system were obtained. DeepMimic is a marvellous framework with advanced capabilities, very complete and optimized; the lacking of GPU support and the inability to resume training are the main shortcomings of an otherwise great system. In terms of learning assistance and as an alternative to early termination (ET), the Baby Walker strategy seems a simple and effective way to help the learning process and can be considered a valid Curriculum Learning procedure - learning a task with increasing steps of difficulty.

4.1 Future work

The most pressing change to the model's implementation is separating the CNN and the existing neural networks. The CNN is trained, independently, with an augmented mocap data (synthetic intermediary data with small random variations is introduced) to predict the respective value of the phase variable. Then the weights are saved and locked from further learning. When building the pipeline with DeepMimic, the output unit (along with its connections) is discarded and the last fully connected layer values are fed into the input placeholder of the existing networks. This brings several advantages: the CNN learns more useful features since it trained with small variations of a perfect running animation; during policy learning, the computational load is reduced since the CNN only processes the input and does no longer learn; an array with the values of the last fully connected layer is saved in the replay buffer instead of a set of scalograms (eliminating the memory limitations of the current implementation), allowing for more channels and a more detailed scalograms.

To further improve this new implementation, logarithmic scale distribution will bring faster CWT processing and more detailed to higher frequencies. It may also be worth checking if discarding (and, if possible, preventing the processing of) the COI area from the scalogram helps the learning. Finally, looking into other mother wavelets might reveal one more appropriate (with better performance) for this problem.

References

[1] Xue Bin Peng et al. "DeepMimic: Example-Guided Deep Reinforcement Learning of Physics-Based Character Skills". In: *CoRR* abs/1804.02717 (2018). arXiv: 1804.02717. URL: http://arxiv.org/abs/1804.02717.

[2] Soohwan Park et al. "Learning Predict-and-Simulate Policies From Unorganized Human Motion Data". In: *ACM Trans. Graph.* 38.6 (2019).

[3] Nicolas Heess et al. "Emergence of Locomotion Behaviours in Rich Environments". In: *CoRR* abs/1707.02286 (2017). arXiv: 1707.02286. URL: http://arxiv.org/abs/1707.02286.

[4] Thomas Geijtenbeek. "Animating Virtual Characters using Physics-Based Simulation". PhD Thesis. Utrecht University, 2013. ISBN: 978-94-6182-389-2.

[5] Yor Castaño-Pino et al. "Using Wavelets for Gait and Arm Swing Analysis". In: 2019-03. DOI: 10.5772/intechopen.84962.

[6] Ning Ji et al. "Appropriate Mother Wavelets for Continuous Gait Event Detection Based on Time-Frequency Analysis for Hemiplegic and Healthy Individuals". In: *Sensors* 19 (2019-08), p. 3462. DOI: 10.3390/s19163462.

[7] Philippe Beaudoin, Pierre Poulin, and Michiel van de Panne. "Adapting Wavelet Compression to Human Motion Capture Clips". In: *Graphics Interface 2007*. 2007-05, pp. 313–318.

[8] A.L. Cruz Ruiz et al. "Muscle-Based Control for Character Animation". In: *Computer Graphics Forum* 36.6 (2017), pp. 122–147. DOI: 10.1111/cgf.12863. eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/cgf.12863. URL: https://onlinelibrary.wiley.com/doi/abs/10.1111/cgf.12863.

[9] Thomas Geijtenbeek, A. Frank van der Stappen, and Michiel Panne. "Flexible Muscle-Based Locomotion for Bipedal Creatures". In: *ACM Transactions on Graphics* 32 (2013-11), 206:1–206:11. DOI: 10.1145/2508363.2508399.

[10] Kevin Quennesson, Elias Ioup, and Charles Lee Isbell. "Wavelet Statistics for Human Motion Classification". In: *AAAI*. 2006.

[11] Irene Koenig et al. "Wavelet analyses of electromyographic signals derived from lower extremity muscles while walking or running: A systematic review". In: *PLOS ONE* 13 (2018-11), pp. 1–15. DOI: 10.1371/journal.pone.0206549. URL: https://doi.org/10.1371/journal.pone.0206549.

[12] Carlotta Caramia, Cristiano De Marchis, and Maurizio Schmid. "Optimizing the Scale of a Wavelet-Based Method for the Detection of Gait Events from a Waist-Mounted Accelerometer under Different Walking Speeds". In: *Sensors* 19 (2019-04), p. 1869. DOI: 10.3390/s19081869.