



**HAL**  
open science

## AIP: Adversarial Interaction Priors for Multi-Agent Physics-based Character Control

Mohamed Younes, Ewa Kijak, Richard Kulpa, Simon Malinowski, Franck Multon

### ► To cite this version:

Mohamed Younes, Ewa Kijak, Richard Kulpa, Simon Malinowski, Franck Multon. AIP: Adversarial Interaction Priors for Multi-Agent Physics-based Character Control. SIGGRAPH Asia 2022 - 15th ACM SIGGRAPH Conference and Exhibition on Computer Graphics and Interactive Techniques in Asia, Dec 2022, Daegu, South Korea. , pp.2, SIGGRAPH Asia 2022 Posters. 10.1145/3550082.3564207 . hal-03888489

**HAL Id: hal-03888489**

<https://inria.hal.science/hal-03888489v1>

Submitted on 8 Dec 2022

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License

# AIP: Adversarial Interaction Priors for Multi-Agent Physics-based Character Control

MOHAMED YOUNES, Inria, IRISA, University of Rennes, France

EWA KIJAK, University of Rennes, Inria, IRISA, France

RICHARD KULPA, University of Rennes, Inria, M2S, France

SIMON MALINOWSKI, University of Rennes, Inria, IRISA, France

FRANCK MULTON, University of Rennes, Inria, IRISA, M2S, France

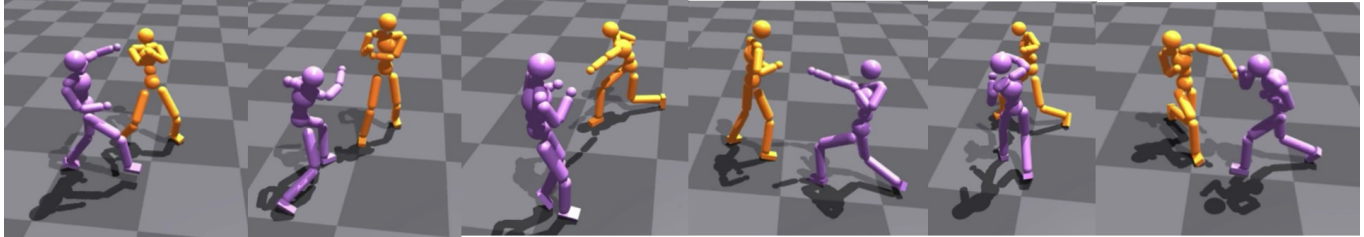


Fig. 1. Simulated shadowboxing interactions between two physics-based characters.

We address the problem of controlling and simulating interactions between multiple physics-based characters, using short unlabeled motion clips. We propose Adversarial Interaction Priors (AIP), a multi-agents generative adversarial imitation learning (MAGAIL) approach, which extends recent deep reinforcement learning (RL) works aiming at imitating single character example motions. The main contribution of this work is to extend the idea of motion imitation of a single character to interaction imitation between multiple characters. Our method uses a control policy for each character to imitate interactive behaviors provided by short example motion clips, and associates a discriminator for each character, which is trained on actor-specific interactive motion clips. The discriminator returns interaction rewards that measure the similarity between generated behaviors and demonstrated ones in the reference motion clips. The policies and discriminators are trained in a multi-agent adversarial reinforcement learning procedure, to improve the quality of the behaviors generated by each agent. The initial results show the effectiveness of our method on the interactive task of shadowboxing between two fighters.

CCS Concepts: • **Computing methodologies** → **Procedural animation; Adversarial learning; Multi-agent reinforcement learning.**

Additional Key Words and Phrases: Character Animation, Multi-Agent Reinforcement Learning, Adversarial Imitation learning, Physics-based Simulation, Motion Capture

## ACM Reference Format:

Mohamed Younes, Ewa Kijak, Richard Kulpa, Simon Malinowski, and Franck Multon. 2022. AIP: Adversarial Interaction Priors for Multi-Agent Physics-based Character Control. In *SIGGRAPH Asia 2022 Posters (SA '22 Posters)*, December 06-09, 2022. ACM, New York, NY, USA, 2 pages. <https://doi.org/10.1145/3550082.3564207>

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the owner/author(s).

SA '22 Posters, December 06-09, 2022, Daegu, Republic of Korea

© 2022 Copyright held by the owner/author(s).

ACM ISBN 978-1-4503-9462-8/22/12.

<https://doi.org/10.1145/3550082.3564207>

## 1 INTRODUCTION

Physics-based character control is a very active field of research, as it enables to generate physically valid animation in complex interactive environments. In applications where a simulated character/agent has to interact with an environment that may involve other agents (simulated or real users), it has to react realistically to a large variety of unexpected situations. Here, "realistic" means that the motions should preserve a natural style, and the actions performed by the agent could have been performed by a real person in the same situation.

Previous works demonstrated the promising use of data-driven approaches based on deep RL in order to improve the quality of a character motions for a wide range of behaviors. Recent techniques include Generative Adversarial Imitation Learning (GAIL), which enables a character to have realistic behavior learned from large datasets of multiple motion clips [Peng et al. 2021]. This approach mainly uses an adversarial discriminator output as a reward instead of manually designing imitation rewards, such as measuring pose errors between a simulated and a synchronized reference motion. However, most of these works aim at controlling a single character, while behaviors involving interactions between multiple characters have not been widely explored.

We propose adversarial interaction priors (AIP), a method for imitating interactions of multiple characters from non-annotated motion clips, based on the Multi-Agent GAIL framework (MAGAIL) [Song et al. 2018], an extension of GAIL to the multi-agent setting. As input to the system, we use two non-annotated datasets: 1) a single-actor dataset containing motions of single actors performing a set of motions linked to a specific application, and 2) an interaction dataset containing a few examples of interactions between several actors for this application. Based on these datasets, our system trains control policies allowing each character to imitate the interactive skills associated with each actor. Similarly to AMP [Peng et al. 2021], the single-actor dataset is used by all agents

to provide them with the ability to imitate realistic motions, as well as a way to add user control through some defined objective. These individual motion capabilities provide each agent with a repertoire of behaviors that can be used to interact with other agents. The interaction dataset is used to train each agent on how to behave in different interactive situations that involve multiple agents. The interaction prior is therefore acting as a measure of similarity between the motions generated by the policies when the agents are interacting with each other, while the motion prior is measuring similarity between the motions produced by a character and the motions in the same dataset, independently of the other characters.

## 2 OUR APPROACH

Our method is based on MAGAIL [Song et al. 2018], where an individual discriminator is assigned to each agent  $i$ , and trained to recognize behaviors based on its role in the interaction dataset. A policy controlling each agent gets rewards from the associated discriminator. The original formulation of MAGAIL assumes having access to the actions of the demonstrations, which could be difficult to get when dealing with unstructured and unlabeled motion capture data. Previous works [Peng et al. 2021] tackled this problem by using transitions of observations contained in this unlabeled motion clip  $(o, o')$ , where  $o, o'$  are observations of an agent at time step  $t$  and  $t+1$  respectively. However, in the setting of interaction between multiple agents, the observation of each agent contains information about itself  $o_s$  as well as about other agents  $o_{-s}$ :  $o = \{o_s, o_{-s}\}$ . Since other agents are controlled by their own policies that change during training, the environment from the perspective of each agent becomes non-stationary. This is crucial for training the discriminators, as the observations of an agent might differ slightly from the demonstrations, because of this non-stationary environment. Hence, the policy could get lower reward because other agents changed their actions, as the full observation's transition could be never seen in the demonstration. Therefore, we propose to change the discriminator's input to omit observations about other agents for the next time step  $o'_{-s}$ , leading to  $(o_s, o_{-s}, o'_s)$ , without  $o'_{-s}$ , as the input for the individual discriminators. The intuition behind this choice is that we want each agent to perform actions that lead to similar interactive behaviors, as depicted by the demonstrations based on the current observation of other agents, no matter what their actions for the next time step. Therefore, the loss for the individual discriminator  $D_i$  for an agent  $i$ , is given by:

$$L_i = L_{int}(o_s, o_{-s}, o'_s) + L_{gp} \quad (1)$$

where  $L_{int}$  is the interaction loss which trains the individual discriminator to distinguish between the agent behavior and behaviors from interactive demonstrations,  $L_{gp}$  is the gradient penalty loss used in AMP which is crucial for training stability of the discriminators. Similarly to AMP, we include a common discriminator between all agents, for which the loss is given by:

$$L_{common} = L_s(o_s, o'_s) + L_{gp} \quad (2)$$

where  $L_s$  forces the discriminator to differentiate between agents behaviors and behaviors depicted in the single actor motion dataset. This is used for data augmentation in case the interaction dataset does not cover all possible situations. This is also a mean to use

different behaviors from those included in the interaction dataset, in case a control or constraint-specific reward is used for training the policies. At the current stage of this work, we do not consider such control or constraint-specific rewards, but only rewards defined by the discriminators.

The policies are trained using a combination of the multi-agent proximal policy optimization algorithm [Yu et al. 2021], and MAGAIL: at each time step  $t$ , each agent receives an interaction-reward from the interaction prior  $r_{int}^t = r_{int}(o_s, o_{-s}, o'_s)$  and a style reward from the motion prior  $r_s^t = r_s(o_s, o'_s)$ , computed from the discriminators output, as described in [Peng et al. 2021]. The rewards are then used in a linear combination to optimize each policy. The overview of the approach is depicted in the supplementary material.

## 3 RESULTS AND DISCUSSION

We applied this work to shadow-boxing with two fighters. In this combat sport exercise, the fighters minimize contacts to avoid injury. As we used marker-based motion capture, shadow-boxing prevent from potentially dangerous contacts with rigid markers. A 70-second interaction clip was recorded with the two fighters, along with 6 minutes single actor clips of boxing-specific motions (see the supplementary video).

Some snapshots of the simulation results are given in figure 1. The supplementary video shows some simulated sequences, which are compared to motion capture data. The simulated agents learned various interactive behaviors. They move around each other with light steps until they find openings for attacks, get close to each other and keep their guard up, respond fast to attacks by either dodging to the back or to the side for counter-attacks. Distinct roles and motion styles for each agent are also noticed.

Our future direction involves imitating reference motions with consistent physical contact between actors, as well as defining constraint-specific rewards to control the interactions. We also plan to evaluate the ability of this system to handle a wide variety of interactions.

## ACKNOWLEDGMENTS

With the support of ANR in the framework of the PIA EUR DIGISPORT project (ANR-18-EURE-0022)

## REFERENCES

- Xue Bin Peng, Ze Ma, Pieter Abbeel, Sergey Levine, and Angjoo Kanazawa. 2021. AMP: Adversarial Motion Priors for Stylized Physics-Based Character Control. *ACM Trans. Graph.* 40, 4, Article 144 (2021), 20 pages. <https://doi.org/10.1145/3450626.3459670>
- Jiaming Song, Hongyu Ren, Dorsa Sadigh, and Stefano Ermon. 2018. Multi-Agent Generative Adversarial Imitation Learning. In *Advances in Neural Information Processing Systems*, Vol. 31. <https://proceedings.neurips.cc/paper/2018/file/240c945bb72980130446fc2b40fbb8e0-Paper.pdf>
- Chao Yu, Akash Velu, Eugene Vinitzky, Yu Wang, Alexandre Bayen, and Yi Wu. 2021. The Surprising Effectiveness of PPO in Cooperative, Multi-Agent Games. <https://doi.org/10.48550/ARXIV.2103.01955>