Learning to move in crowd

Jaedong Lee
Department of Computer Science and Engineering
Seoul National University
jaedong@mrl.snu.ac.kr

Jehee Lee
Department of Computer Science and Engineering
Seoul National University
jehee@mrl.snu.ac.kr

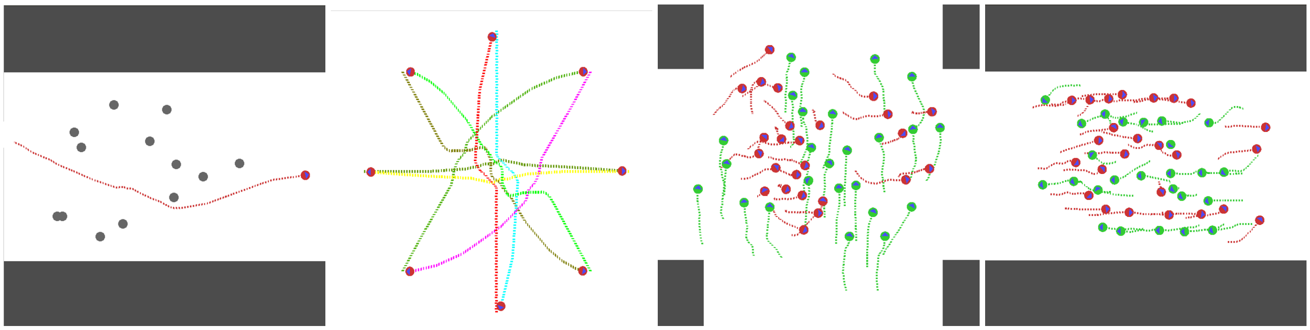

Figure 1: Experiment : Avoiding obstacles, Circle, Crossway, and Corridor

ABSTRACT

The main goal of the crowd simulation is to generate realistic movements of agents. Reproducing the mechanism that seeing the environments, understanding current situation, and deciding where to step is crucial point to simulating crowd movements. We formulate the process of walking mechanism using deep reinforcement learning. And we experiment some typical scenarios.

CCS CONCEPTS

•Computing methodologies → Animation; Crowd simulation; Reinforcement learning;

KEYWORDS

animation, crowd simulation, collision avoidance, reinforcement learning

ACM Reference format:
Jaedong Lee and Jehee Lee. 2018. Learning to move in crowd. In *Proceedings of SIGGRAPH '18 Posters, Vancouver, BC, Canada, August 12-16, 2018,* 2 pages.
DOI: 10.1145/3230744.3230782

1 INTRODUCTION

Crowd simulation has long been a popular topic in computer animation. Simulating the motions of multiple intelligent agents is widely used in virtual reality, media entertainment, and pedestrian dynamics. Through various studies, numerous kinds of models have been proposed for simulation. The two major categories are microscopic methodology and macroscopic methodology. Microscopic methodology, which we are concerned with, is a method for creating an individual's movement through interaction between an agent and his environments. Macroscopic methodology is a method for calculating crowd motion by considering all agents as one system. There were also studies in which several methodologies were mixed.

The common goal of all these studies is to reproduce the movement of the crowd as natural as possible. In order to reproduce realistic motion, many studies have been conducted on interaction between characters or obstacles. The most important factor in the interaction with other people is avoiding collision. A study of cognitive science shows that visual information seen by people through the eyes has a great impact on walking planning that the process of cognition, thinking, and behavior. Seeing the situation and taking steps to avoid colliding with others also follow that mechanism. Ondrej and his colleague proposed a vision-based algorithm having similar idea [Ondřej et al. 2010].

Due to the recent development of deep learning, deep reinforcement learning has also been actively studied. Reinforcement learning is one of the methodologies to solve the markov decision process. The goal of reinforcement learning is to find a policy that leads to the best action in the present situation, and taking into account the consequences on the distant future. And it is also known that neural networks can learn through various types of input values such as vision images, sequential data. Due to these features, we solve the problem of simulating crowd using visual information, neural networks, and reinforcement learning. In this work, we propose a methodology to apply reinforcement learning to simulating natural movements of crowd by reproducing the natural process of cognition, thinking, and behavior.

SIGGRAPH '18 Posters, Vancouver, BC, Canada
 978-1-4503-5817-0/18/08…$15.00
DOI: 10.1145/3230744.3230782

2 OUR APPROACH

We related the process of cognition-thinking-behavior to our simulation. We replace the cognitive process of a person by an agent receiving visual information from a simulation. We use discrete depth-map data as agent's visual information. A person's thinking-behavior process is related to its ultimate goal. Usually, a person's walking behavior is characterized by the destination where the person wants to go using an efficient path which will avoid collision with other people or obstacles. We formulated this process using reinforcement learning which is the process of finding a policy that maximizes the sum of the rewards. We can design the way of walking through the reward system. Once the optimal policy is learned, at every step the agent will think about this situation through the given visual information and select the best next action through the policy network.

We used the actor-critic, model-free reinforcement learning algorithm, Deep Deterministic Policy Gradient (DDPG) [Lillicrap et al. 2015], which is a version of reinforcement learning algorithm for continuous space. DDPG is a method to find the policy that maximizes the sum of cumulative reward by updating the policy, and minimizing the loss value using gradient descent.

We build a system to apply reinforcement learning to crowd simulation.

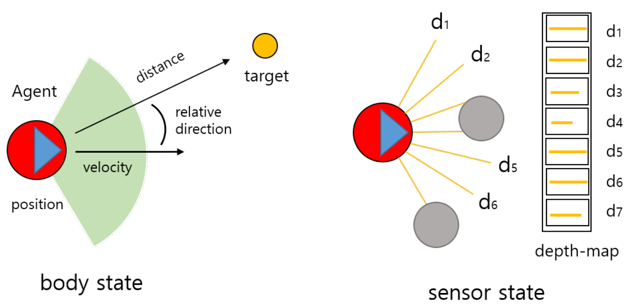

Figure 2: Agent state. Left is body state consists of position, velocity, distance to target, and relative direction to target. Right is sensor state consists of discrete depth-map. Each length of ray such as d1, d2... is the value of sensor. We take 3 frames of depth-map as input of sensor network.

First we define an agent which has two states(Figure 2), the first one is the body state composed of positions, velocities, and the relative direction to destination. The second one is the sensor states composed of the last three frames of depth-map collected on its field of view. And every agent use same policy network.

$$r = r_{target} + r_{collision}$$

$$r_{target} = -w_t d^2$$

$$r_{collision} = \begin{cases} -w_c o^2, & \text{if collision occurs} \\ 0, & \text{otherwise} \end{cases}$$

The actor network consists of two network which take the body state and sensor state as input of each network. As shown in Figure 3, body state, which are processed by two fully-connected layers, and sensor state, which are processed by one convolutional layers, merged into one fully-connected layer. The output of this network will be an action defined by two values : the change in velocity and direction, which are added to its current velocity and direction at every step. The critic network just added one action network to actor network.

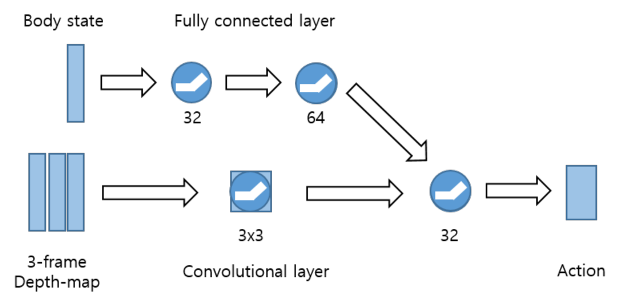

Figure 3: Actor network consists of two networks, body network and sensor network. Body state is 4 dimension body state data and sensor state is 3x20 dimension data. ELU activations are used for all hidden layers. Output is 2 dimension action data.

We design our reward system as the sum of two terms : r_{target} rewarding the agent for going toward the target and $r_{collision}$ encouraging it to avoid collision. d means the distance between agent and target, set w_t as 0.001. o means the distance between agent and obstacles. we set w_c as 0.01.

3 EXPERIMENTS

We experiment basic scenarios that reaching to the destination and avoiding obstacles. And we experiment three scenarios that circle, crossway, and corridor. We use one trained network that working on every scenario. It takes 2 hours for training and hyper parameters we don't referred are same to DDPG paper's setting. In circle scenario we test 4 times, 6, 8, 12, and 16 agents. Crossway scenario also test 4 times, 8, 16, 24, and 32 agents. In corridor scenario we test 4 times, 8, 16, 24, and 32 agents. And we will do more sophisticate experiments later. To do so, we have to tune network parameters, design reward, and some simulation parameters.

ACKNOWLEDGMENTS

This research was supported by the MSIT(Ministry of Science and ICT), Korea, under the SW Starlab support program(IITP-2017-0-00878) supervised by the IITP(Institute for Information & communications Technology Promotion).

REFERENCES

Timothy P. Lillicrap, Jonathan J. Hunt, Alexander Pritzel, Nicolas Heess, Tom Erez, Yuval Tassa, David Silver, and Daan Wierstra. 2015. Continuous control with deep reinforcement learning. *CoRR* abs/1509.02971 (2015). http://arxiv.org/abs/1509.02971

Jan Ondřej, Julien Pettré, Anne-Hélène Olivier, and Stéphane Donikian. 2010. A Synthetic-vision Based Steering Approach for Crowd Simulation. *ACM Trans. Graph.* 29, 4, Article 123 (July 2010), 9 pages. DOI: http://dx.doi.org/10.1145/1778765.1778860