가상환경에서 강화학습을 이용한 휴머노이드 로봇의 동적 장애물 회피에 관한 연구

여동현, Phuong Chu, Hoang Vu, 엄기현, 조경은
동국대학교 멀티미디어공학과
e-mail : cke@dongguk.edu(교신저자)

A Study on Dynamic Obstacle Avoidance of a Humanoid Robot Using Reinforcement Learning in Virtual Environment

Donghyeon Yeo, Phuong Chu, Hoang Vu, Kyhyun Um, Kyungeun Cho
Dept. of Multimedia Engineering, Dongguk University

요 약

본 논문에서는 휴머노이드 로봇이 더욱 지능적으로 보행할 수 있도록 강화학습을 적용하는 방법을 제안한다. 강화학습을 활용하면 로봇의 이동 경로에 동적으로 이동하는 장애물이 있을 경우 이를 인지하고 회피하여 목적지까지 문제 없이 이동할 수 있다. 실제 환경에서의 실험에 앞서 Unity3D 를 활용한 가상 환경에서 이를 구현하여 실험해보았다.

1. 서론

최근 사람의 업무를 대신해줄 수 있는 휴머노이드 로봇에 대한 연구가 활발히 진행되고 있으며, 이러한 로봇의 이동에 있어서 경로 탐색 기능은 매우 중요하다[1]. 로봇은 보행 도중 다양한 장애물들을 마주할 수 있으며, 이를 안정적으로 회피한다는 것은 상당히 어려운 문제이다[2]. 최단 경로 탐색 알고리즘은 로봇의 이동에 있어서 최단 경로를 보장하지만 이동 경로에 동적으로 위치가 변화하는 장애물이 있을 경우 목적지까지 도달하는 데 길막힘, 충돌 등의 문제가 발생할 수 있다.

본 논문에서는 이러한 문제점을 극복하기 위해 강화학습 알고리즘을 활용하였고, 실제 실험 환경과 동일하게 구현된 가상 환경에서 학습을 수행하여 이동에 관한 데이터베이스를 확보하였다.

2. 관련 연구

최단 경로 탐색 알고리즘에 대해서는 지금까지 이미 많은 연구가 진행되어 왔으며, A*, Dijkstra 등의 대표적인 최단 경로 탐색 알고리즘은 게임, 산업, 자동차, 로봇 등의 다양한 분야에서 응용되고 있다[3]. 하지만 최단 경로 탐색 알고리즘들은 이미 계산된 로봇의 최단 경로 도중에 동적 이동 장애물이 있을 경우 다시 경로를 계산해야 한다는 문제점이 있다.

강화학습은 발생 가능한 모든 경우의 수를 탐색하여 결과를 Q-Table 에 저장한다. 탐색 도중 목표에 가까운 행동을 취한 경우와 목표에 어긋나는 행동을 취한 경우를 기억하여 이후 같은 상황에 직면했을 때 목표에 가까운 행동을 취하게 된다[4][5]. 모든 경우에 대한 데이터베이스를 저장해놓는다면 로봇이 이동 도중 동적 이동 장애물에게 방해를 받아도 경로를 재탐색하지 않고 최단 경로로 이동할 수 있다.

3. 가상 환경에서의 실험 환경 구축

강화학습을 통한 학습 데이터베이스를 축적할 가상 환경은 맵의 좌표, 오브젝트들의 위치와 크기 등이 실제 실험 환경과 동일한 비율로 구성되어야 한다.

실험 공간 내에는 로봇, 가상 사람, 장애물, 침대, 소파, 가스렌지, 테이블 및 책상, 공 등이 있다. 이 중 로봇과 장애물은 상하좌우로 이동이 가능한 객체이며, 그 외에는 위치가 변하지 않는 정적 객체들이다. 장애물은 임의의 방향으로 항상 이동한다.

로봇의 임무는 초기 임의의 위치에서 공이 있는 위치로 이동하여 공을 잡은 후 가상 사람이 있는 곳으로 이동해 공을 전달하는 것이다.

다양한 경우의 실험을 위해 임무를 완료할 때마다 가상 사람과 공의 위치를 임의로 변경한다.

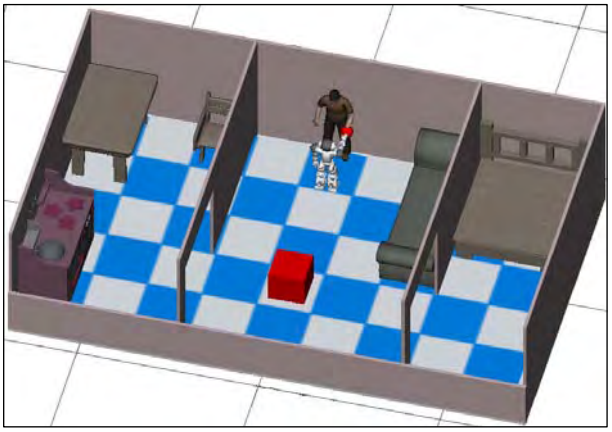
Q-Table 의 각 값은 로봇, 가상 사람, 장애물, 공의 위치와 로봇의 이동 가능한 방향에 대한 모든 경우의 수이다. 학습을 위한 Q-Table 은 다음과 같이 구성된다.

$$Q_{t+1}(s_t, a_t) = Q_t(s_t, a_t) + \alpha \times \left(r_{t+1} + \gamma \times maxQ_t(s_{t+1}, a_{t+1}) - Q_t(s_t, a_t)\right)$$

s_t, a_t 값은 로봇의 이동 전 위치와 이동 방향이고, s_{t+1}, a_{t+1} 은 현재 로봇의 위치와 이동 방향이다. r_{t+1} 은 로봇이 상태 s_t에서 액션 a_t을 수행 한 이후 관찰한 결과에 대한 보상 값이다. α는 학습률을 나타내고,

γ는 Discount factor 값이다.

다음 (그림 1)은 가상 환경에서 강화학습 알고리즘을 통해 학습 후 로봇이 임무를 수행한 화면이다.



(그림 1) 학습 후 로봇의 역할 수행

4. 결론 및 향후 과제

본 논문에서는 로봇이 보행 중 동적 이동 장애물을 실시간으로 회피하는 방법에 대해 연구하였고, 강화학습을 활용하여 경로를 재탐색하지 않고 이동하는 방법으로 접근해보았다.

차후 실제 휴머노이드 로봇을 사용하는 실험 환경에 학습한 데이터베이스를 이식하여 실제 로봇이 동적 이동 장애물을 피하고 목적지까지 도달하여 임무를 성공적으로 수행 가능하도록 구현 및 실험할 계획이다.

감사의 글

이 논문은 2012 년도 정부(교육과학기술부)의 재원으로 한국연구재단의 기초연구사업 지원을 받아 수행된 것임(2012R1A1A2009148).

참고문헌

[1] J. H. Park, “Moving Obstacle Collision Avoidance of a Mobile Robot Using Neural Network”, Proceedings of the 12th KACC, pp.1238-1241, 1997.

[2] K. J. Kim, “Numerical Performance Analysis of Obstacle Avoidance Method for a Mobile Robot”, The Journal of Korea Institute of Electronic Communication Sciences, Vol. 7, No. 2, 2012.

[3] Y. G. Ryu, “A Study on A* Algorithm Applying Reversed Direction Method for High Accuracy of the Shortest Path Searching”, The Journal of The Korea Institute of Intelligent Transport Systems, Vol.12, No.6, 2013.

[4] Y. Sung, “Human-Robot Interaction Learning using Demonstration-based Learning and Q-learning in a Pervasive Sensing Environment”, International Journal of Distributed Sensor Networks, 2014.

[5] Y. Sung, “Q-learning Reward Propagation Method for Reducing the Transmission Power of Sensor Nodes in Wireless Sensor Networks”, Wireless Personal Communications, 2013.