

커넥터 조립을 위한 강화학습 기반의 탐색 궤적 생성 및 로봇의 임피던스 강성 조절 방법

Reinforcement Learning-based Search Trajectory Generation and Stiffness Tuning for Connector Assembly

김 용 건¹ · 나 민 우² · 송 재 복[†]

Yong-Geon Kim¹, Minwoo Na², Jae-Bok Song[†]

Abstract: Since electric connectors such as power connectors have a small assembly tolerance and have a complex shape, the assembly process is performed manually by workers. Especially, it is difficult to overcome the assembly error, and the assembly takes a long time due to the error correction process, which makes it difficult to automate the assembly task. To deal with this problem, a reinforcement learning-based assembly strategy using contact states was proposed to quickly perform the assembly process in an unstructured environment. This method learns to generate a search trajectory to quickly find a hole based on the contact state obtained from the force/torque data. It can also learn the stiffness needed to avoid excessive contact forces during assembly. To verify this proposed method, power connector assembly process was performed 200 times, and it was shown to have an assembly success rate of 100% in a translation error within ± 4 mm and a rotation error within $\pm 3.5^\circ$. Furthermore, it was verified that the assembly time was about 2.3 sec, including the search time of about 1 sec, which is faster than the previous methods.

Keywords: Reinforcement Learning, Robotic Assembly, Assembly Strategy, Connector Assembly

1. 서 론

최근 산업 현장에서는 대량 생산을 위해 로봇을 이용한 자동화 사례가 늘고 있다. 하지만 형상이 복잡하고 공차가 작은 부품에 대한 조립은 공정 속도가 느려 자동화가 어려운 실정이다. 특히 비정형 환경의 조립 작업에서는 오차가 발생하기 쉽고, 오차 보정에 많은 시간이 필요하다. 최근 비정형 환경에서 조립구의 위치를 파악하여 조립을 수행하기 위해 이미지를 이용한 강화학습 기반의 조립 방안이 연구되었다¹⁾. 해당 연구는 조립물과 조립구를 포함하는 이미지가 강화학습 네트워크

에 입력되어, 조립물과 조립구 사이의 오차를 식별하는 방법을 제안하였다. 하지만 이 방법은 위치제어를 기반으로 조립을 수행하여 미세한 오차에도 큰 힘이 발생하여 부품 사이에 재밍(jamming)이 발생하여 부품이나 로봇이 손상될 수 있다²⁾. 또한 이미지를 네트워크의 입력으로 하는 경우 연산시간이 증가하고, 오차에 대한 신속한 대응이 어렵다.

비정형 환경에서는 삽입구의 위치를 찾는 탐색 작업이 필수적이다. 일반적으로 나선형 또는 무작위의 움직임을 수행하여 홀 위치를 탐색한다^{3,4)}. 이러한 방식은 궤적을 따라 움직이는 중에 정의된 힘/토크 조건을 만족하면 삽입을 수행한다. 이러한 전략은 나선형 또는 무작위 방식으로 움직이면서 정의된 조건을 만족할 때까지 수행된다. 이 방법은 불필요한 움직임을 지속적으로 수행해야 하므로 조립시간이 오래 걸린다. 한편, 실제 현장에서 조립 자동화의 수요가 높은 커넥터에 특화된 조립 연구가 수행되었다⁵⁾. 해당 연구는 몰렉스 커넥터(molex connector)를 대상으로 하여 커넥터의 형상에 대한 수학적 모델을 통해 탐색 궤적을 생성하였다. 그러나 이 방법은 모델을 생성하는

Received : Jun. 6. 2022; Revised : Jul. 21. 2022; Accepted : Sep. 21. 2022

※ This research was supported by the MOTIE under the Industrial Foundation Technology Development Program supervised by the KEIT (No. 20008613)

1. Master Student, Mechanical Engineering, Korea University, Seoul, Korea (yonggeon0115@korea.ac.kr)

2. Ph.D. Student, Mechanical Engineering, Korea University, Seoul, Korea (jrmwl@korea.ac.kr)

† Professor, Corresponding author: Mechanical Engineering, Korea University, Seoul, Korea (jbsong@korea.ac.kr)

데 시간이 오래 걸리고, 각 커넥터의 특성에 맞추어 움직임을 일일이 정의해야 한다^[6]. 이러한 작업에는 전문적인 지식을 가진 숙련된 작업자가 투입되어야 한다는 단점이 있다.

조립을 반복적으로 시도하여 움직임을 최적화하는 강화학습을 이용한 연구가 수행되었다. [7]은 이산적(discrete)인 행동을 정의하여 상태정보에 따라 정의된 행동을 수행하며, [8]은 가상환경에서 학습된 결과를 빠르게 실제 환경에 적용한다. 하지만 두 방안 모두 조립 위치를 알고 있는 정형적인 환경에만 적용가능한 단점이 있다. 비정형 환경에 적용 가능한 방안이 제시되었다. [9]는 사용자의 시연을 이용하는 방안으로, 단순한 형태의 궤에만 적용했다는 한계가 있다.

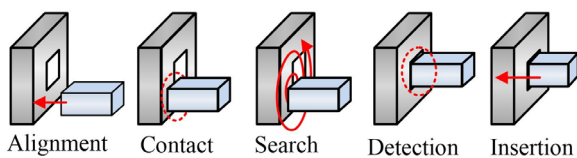
본 연구에서는 비정형 환경의 커넥터 조립작업에서 발생하는 힘/토크 정보를 이용한 강화학습 기반의 조립전략을 제안한다. 제안된 방법은 커넥터에 대한 힘/토크 정보 기반의 접촉 상태를 정의하고, 학습을 통해 커넥터의 구멍을 탐색하기 위한 빠른 탐색 궤적을 생성하여 조립시간을 단축시킬 수 있다. 또한 접촉 과정에서 과도한 접촉력이 발생하지 않도록 부품 사이의 로봇의 임피던스 제어에 사용되는 강성을 조절하여 부드러운 조립을 가능하게 한다.

본 논문의 구성은 다음과 같다. 2장에서 접촉 상태 기반의 커넥터 조립전략에 대해 소개한다. 3장에서는 강화학습 기반의 조립전략에 대하여 설명한다. 4장에서는 파워 커넥터의 조립을 통해 제안된 방법을 검증하였으며, 블라인드 탐색과의 비교를 수행하였다. 마지막으로, 5장에서 결론을 도출한다.

2. 접촉 상태 기반의 커넥터 조립

2.1 커넥터 조립전략

커넥터에 대한 일반적인 조립 전략은 [Fig. 1]과 같다. 먼저 비전 시스템을 기반으로 커넥터 구멍(connector hole)을 인식한 후 자세를 추출한다. 다음으로 로봇은 추출된 커넥터 구멍의 자세(즉, 위치와 방위)에 따라 과지한 커넥터를 커넥터 구멍에 정렬(alignment)시킨다. 이때 비전 시스템에 의한 오차로 인해 두 부품 사이에 상대 오차가 발생할 수 있다. 이는 커넥터의 삽입 실패를 야기하므로 미세한 정렬 오차를 줄이기 위해 힘/토크 정보를 이용한 탐색(searching)을 수행한다. 탐색은 미리



[Fig. 1] Blind search with spiral path for hole detection

정의된 궤적을 따라 커넥터와 커넥터 구멍 사이에 발생하는 접촉력을 모니터링하여 커넥터 구멍의 자세 정보를 얻는 과정이다. 로봇이 정의된 궤적을 따라 구동할 때, 구멍을 찾으면 순간적으로 접촉력이 감소하거나 미세한 동작이 발생한다. 이에 따라 탐색은 종료되고, 탐색 종료 시점의 로봇 자세를 정렬 자세로 정의한다. 마지막으로 정렬 자세 기준으로 삽입(insertion)을 수행하게 된다. 일반적으로 공차가 작은 부품의 삽입 시에 과도한 접촉력의 발생을 방지하기 위해 힘제어의 일종인 임피던스 제어를 사용한다^[10].

임피던스 제어는 로봇의 움직임과 접촉력 사이의 동적 관계를 이용하여 접촉력을 제어하는 방법이다^[11]. 임피던스 제어를 적용하면 로봇 조립 시에 안정적인 접촉을 구현함과 동시에 접촉력에 대해 적절히 반응할 수 있다. 위치제어를 기반으로 로봇을 구동하면 로봇은 궤적을 따라 구동된다. 이때 접촉이 발생하면, 로봇의 궤적과 더불어 임피던스 제어기에 의해 접촉력에 대한 위치 보상이 수행된다. 로봇에 입력되는 제어입력 x_c 는 다음 식과 같다.

$$x_c = x_d - x_F \tag{1}$$

여기서 x_d 는 로봇의 궤적으로부터 생성된 목표 위치 궤적, x_F 는 임피던스 제어기로부터 산출되는 접촉력에 따른 위치 보상으로 다음과 같다.

$$x_F = \frac{1}{B_s + K} F \tag{2}$$

여기서 B, K 는 로봇 말단에서 구현되는 감쇠계수 및 강성으로 임피던스 파라미터이다. F 는 로봇과 접촉 환경 사이에 발생하는 접촉력(힘과 토크)으로 로봇 말단에 부착된 힘/토크 센서를 통해 측정된다. 이때 강성을 조절하면 조립 시에 발생하는 오차에 대응하여 조립물 사이의 유연한 접촉을 구현할 수 있다. 따라서 본 연구에서는 접촉력에 대응하여 실시간으로 강성을 조절하여 조립을 수행한다.

2.2 접촉 상태 기반의 탐색 궤적 생성

기존의 탐색 방법은 정렬 오차를 고려하지 않는 형상(나선, 원형) 궤적을 사용하므로 조립시간이 길어진다. 이를 해결하기 위해 접촉 상태를 기반으로 탐색 궤적을 생성한다. 접촉 상태(contact state)란 커넥터와 커넥터 구멍이 특정한 자세로 접촉할 때 발생하는 접촉력의 양상이다. 본 연구에서 커넥터는 로봇에 파지되어 있고, 접촉력은 커넥터의 물체 좌표계 $\{O\}$ 를 기준으로 측정된 힘/토크를 의미한다. 이때 로봇의 툴 좌표계

{T}는 {O}와 동일하게 설정하였다. 예를 들어, [Fig. 2(a)]와 같이 커넥터가 구멍의 좌측에 접촉하면 {O}의 y축을 기준으로 양의 방향토크 $+\tau_y$ 가 측정된다. 반대로 구멍 우측에 접촉하면 $-\tau_y$ 가 측정된다. 한편 [Fig. 2(b)]와 같이 커넥터가 구멍 상단에 접촉하면 {O}의 x축을 기준으로 음의 방향토크 $-\tau_x$ 가 측정되며, 하단에 접촉하면 $+\tau_x$ 가 측정된다. [Fig. 2(c)]는 [Fig. 2(a)], [Fig. 2(b)]의 접촉 상태가 복합된 상태로 $-\tau_x$ 와 $+\tau_y$ 가 동시에 측정된다.

접촉 상태는 커넥터의 형상에 따라 특정한 접촉 패턴을 보인다. 따라서 본 연구에서는 접촉 상태에 따른 자세 오차를 줄이도록 x축 및 y축의 위치 궤적과 z축의 방위 궤적을 생성하여 탐색 궤적으로 사용한다. 탐색은 커넥터와 구멍 사이의 자세 오차를 극복하는 과정으로, 커넥터의 위치는 {O}의 중심이다. 커넥터 구멍은 비정형 환경에 놓여 있으므로 구멍 자세는 알 수 없다. 탐색은 [Fig. 3]의 예시와 같은 순서로 수행된다. 먼저 접촉 상태 1에서 $-\tau_x$ 와 $+\tau_y$ 가 측정되므로, 녹색으로 표시한 점선과 같이 x축 및 y축의 음의 방향으로 궤적을 생성한다. 이를 통해 접촉 상태 2와 같이 위치 오차를 극복하는 과정에서

여전히 z축의 방위 오차가 존재한다. 이에 따라 토크 $-\tau_z$ 가 측정되고, 녹색으로 표시한 점선과 같이 z축의 음의 방향으로 방위 궤적을 생성하여 오차를 줄인다. 이러한 탐색 궤적을 따라 임피던스 제어를 기반으로 로봇을 구동하면 접촉 상태 3과 같이 커넥터와 구멍이 정렬된다. 해당 방안은 커넥터의 일반적인 형상인 사각 형상을 예시로 설명하였으나, 힘-토크 피드백이 발생 가능한 다른 형상에도 적용이 가능하다.

3. 강화학습 기반의 조립전략

앞서 설명한 접촉 상태 기반의 조립전략을 수행하려면 접촉 상태가 정의되어야 한다. 하지만 실제 접촉이 발생하면 많은 접촉 양상이 존재하므로 이를 모두 접촉 상태로 분류하여 대응하기가 어렵다. 이에 대응하기 위해 반복적인 시도를 통해 최적의 방안을 학습하는 강화학습(reinforcement learning)을 사용한다. 강화학습이란 에이전트(agent)의 현재 상태(state)에 따른 행동(action)을 취하여 산출되는 보상(reward)을 통해 의도한 방향으로 정책(policy)을 학습하는 과정이다. 이때 보상 함수(reward function)를 사용자가 설계하여 원하는 에이전트의 거동에 맞게 네트워크를 학습할 수 있다. 본 연구에서는 빠른 시간 내에 커넥터 구멍의 자세를 찾도록 로봇의 탐색 궤적을 생성함과 동시에 접촉력에 대응하여 적절한 강성을 출력하는 강화학습 기반의 조립전략을 개발하였다.

3.1 강화학습 구조

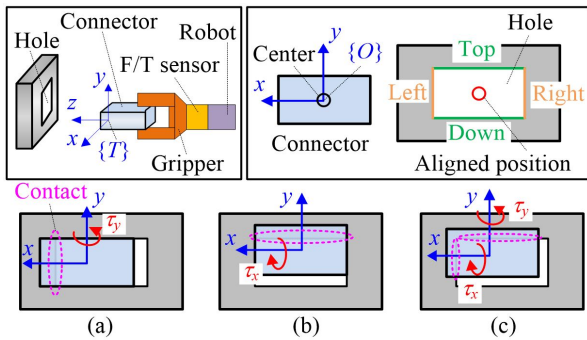
강화학습의 에이전트는 로봇이며, 상태는 조립 시 발생하는 접촉력과 조립 진행률이다. 접촉력은 -1과 1사이로 정규화된 힘 F_N 과 토크 τ_N 를 사용한다. 조립 진행률 d 는 커넥터의 총 삽입 깊이에 대해 로봇이 커넥터를 삽입한 거리의 비로 정의한다. d 가 0일 때는 탐색 단계, 0과 1 사이인 경우 삽입 단계, 1이면 조립 성공을 의미한다. 조립 진행률을 통해 학습 과정에서 조립 성공 여부를 판단할 수 있도록 하였다. 강화학습에 사용되는 현재 상태 s_t 는 다음과 같이 정의된다.

$$s_t = [F_{N,x} \ F_{N,y} \ F_{N,z} \ \tau_{N,x} \ \tau_{N,y} \ \tau_{N,z} \ d] \quad (3)$$

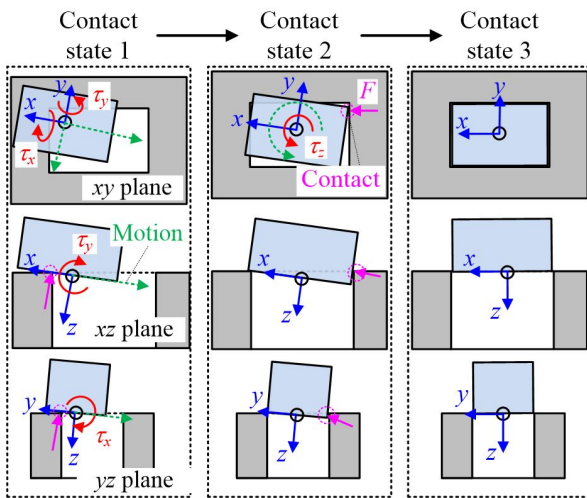
한편, 행동은 상대 자세와 강성으로 구분된다. 먼저 상대 자세 p 에 대한 행동 $a_{1,t}$ 는 다음과 같다.

$$a_{1,t} = [p_x \ p_y \ p_z] \quad (4)$$

p_x, p_y 는 x축, y축 방향의 위치, p_z 는 z축의 방위이다. p 는 탐색 궤적에 해당하며, 커넥터와 구멍 사이의 자세 오차를 극복



[Fig. 2] Alignment error for assembly: (a) aligned to the left side of the hole, (b) aligned to the top side of the hole, and (c) aligned to the top left side of the hole



[Fig. 3] Searching process in assembly strategy

하는데 사용된다. 다음으로, 강성 K 에 대한 행동 $a_{2,t}$ 는 다음과 같다.

$$a_{2,t} = [K_x \ K_y \ K_z \ K_{rx} \ K_{ry} \ K_{rz}] \quad (5)$$

강화학습을 이용한 조립전략의 구조는 [Fig. 4]와 같다. 탐색을 위한 조립전략은 $a_{1,t}$ 를 $a_{2,t}$ 를 동시에 사용하고, 삽입을 위한 조립전략은 $a_{2,t}$ 만을 사용한다. 삽입 시에 $a_{1,t}$ 는 공차가 작은 커넥터와 구멍 사이에 불안정한 접촉을 유발하므로 사용하지 않는다. 이러한 행동은 20 ms의 주기로 로봇에 적용된다.

학습 과정에서 행동은 현재 상태에 대응하여 연속적으로 출력되어야 한다. 따라서 연속 공간에서 강한 학습 성능을 보이는 Soft Actor-Critic (SAC) 알고리즘을 이용하여 네트워크를 구축하였다^[12]. SAC는 액터-크리틱 구조 기반의 강화학습 알고리즘으로, [Fig. 5]와 같이 액터는 상태에 따른 행동을 출력하고, 크리틱은 출력된 행동을 평가한 후 Q값을 출력하여 최적의 정책을 학습한다. 이때 액터와 크리틱 모델은 5개의 전연결층 (fully-connected layer)으로 구성된다. 이때 활성화함수(activation function)는 정류 선형 유닛(rectified linear unit, ReLU)을 배치하였다. 액터의 출력층은 평균과 표준편차로 구성된 확률적 정

책 기반의 행동을 출력한다. [Fig. 4]에서 SAC 1은 탐색 궤적 생성을 위한 구조, SAC 2는 강성 조절을 위한 구조이다. 사전학습 모델(pre-trained model)은 3개의 전연결층을 갖는 신경망으로 구성되며, p_x 와 강성 K_z 을 출력한다. 삽입 방향의 위치 명령과 강성을 사전학습 모델을 이용하여 출력하지 않으면 학습이 느려지거나 실패할 수 있다. 따라서 삽입의 용이성과 제안된 네트워크의 적절한 학습을 위하여 사전학습 모델을 추가하였다.

3.2 보상함수 설계

상태, 행동에 따라 정책을 학습하려면 에이전트가 취한 행동에 따른 적절한 보상함수를 설계해야 한다. 본 연구에서는 앞서 설명한 SAC 1과 SAC 2의 보상함수를 각각 설계하여 사용하였다.

3.2.1 탐색 궤적 생성을 위한 보상함수

SAC 1은 $a_{1,t}$ 를 통해 탐색 궤적을 실시간으로 생성한다. 학습 과정에서 의도한 궤적이 생성되면 양의 보상을, 아니면 음의 보상을 주어 의도에 맞는 탐색 궤적을 학습한다. 탐색 궤적 생성을 위한 보상함수 $r_{1,t}$ 는 다음과 같다.

$$r_{1,t} = r_x + r_y \quad (6)$$

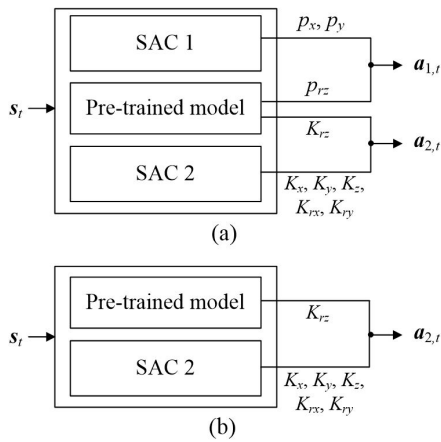
$$r_x = \begin{cases} +2, \text{sign}(\tau_y)\text{sign}(p_x) < 0 \\ -2, \text{sign}(\tau_y)\text{sign}(p_x) > 0 \end{cases} \quad (7)$$

$$r_y = \begin{cases} +2, \text{sign}(\tau_x)\text{sign}(p_y) > 0 \\ -2, \text{sign}(\tau_x)\text{sign}(p_y) < 0 \end{cases} \quad (8)$$

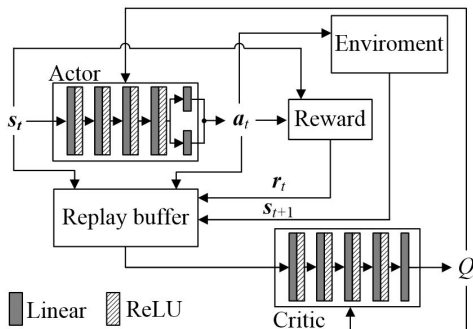
r_x, r_y 는 각각 x 축 및 y 축 방향의 위치 궤적에 대한 보상이다. 현재 접촉 상태의 측정 토크에 대해 의도한 궤적이 출력되면 +2, 그렇지 않으면 -2의 보상을 부여한다. 예를 들어, $+\tau_y$ 가 검출되었을 때, $-p_x$ 가 출력되면 +2의 보상을 부여한다. 한편 1회의 에피소드 종료 시점에 탐색이 성공하면 +50, 실패하면 -50을 추가로 부여한다.

3.2.2 강성 조절을 위한 보상함수

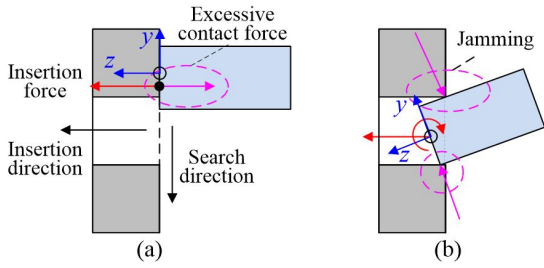
SAC 2는 $a_{2,t}$ 를 통해 강성을 실시간으로 조절하며, 탐색과 삽입 단계에 모두 적용된다. 탐색의 경우, [Fig. 6(a)]의 접촉 상태에서 강성이 높으면 과도한 접촉력을 받아 탐색이 중단된다. 삽입의 경우, [Fig. 6(b)]의 접촉 상태에서 강성이 높으면 y 축 방향으로 과도한 접촉력을 받아 재밍이 발생하여 삽입이 중단된다. 이를 방지하기 위해 과도한 힘을 받으면 강성을 낮추도록 보상을 부여한다. 또한 적절한 목표 힘을 유지해야 하는 경우에는 적절히 강성을 높이도록 보상을 부여한다. 이를



[Fig. 4] The overall structure of reinforcement learning-based assembly strategy: (a) searching phase, and (b) insertion phase



[Fig. 5] The structure for learning with the SAC algorithm



[Fig. 6] Undesirable contact states: (a) searching phase, and (b) insertion phase

반영하여 강성 조절을 위한 보상함수 $r_{2,t}$ 를 조립 진행률 d 에 따라 다르게 정의하였다.

$$r_{2,t} = \begin{cases} r_{s,z} + r_{h,x} + r_{h,y} + r_{h,rx} + r_{h,ry} & d = 0 \\ r_{h,z} + r_{s,x} + r_{s,y} + r_{s,rx} + r_{s,ry} & 0 < d < 1 \end{cases} \quad (9)$$

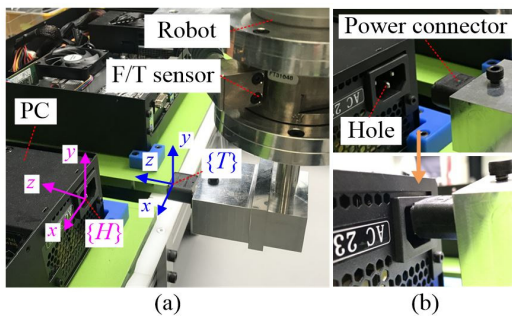
$$r_s = - ||F_N|| + K_N \quad (10)$$

$$r_h = - ||F_N|| - K_N \quad (11)$$

여기서 K_N 은 경험적으로 설정한 상한과 하한을 이용하여 $[-1.0, 1.0]$ 사이의 값으로 정규화한 강성이며, r_s 는 유연한 접촉을 구현한 경우의 보상이며, F_N 과 K_N 의 크기가 같고, 부호가 반대이면 최댓값을 갖는다. r_h 는 목표 힘을 유지한 경우의 보상이며, F_N 과 K_N 이 같을수록 최댓값을 갖는다. 탐색 단계($d=0$)이면 z 축 방향은 r_s 를 사용하여 유연한 접촉을 구현하고, 탐색을 위한 x 축 및 y 축은 r_h 를 사용하여 목표 힘을 유지하였다. 삽입 단계($0 < d < 1$)이면 z 축 방향은 r_h 를 사용하여 목표 힘을 유지하고, x 축 및 y 축은 r_s 를 사용하여 유연한 접촉을 구현하였다.

4. 실험 결과

강화학습 기반의 조립전략을 검증하기 위해 [Fig. 7(a)]의 실험 환경을 구성하였다. 로봇은 Universal Robots사의 UR5를, 힘/토크 센서는 ATI사의 Gamma를 이용하였다. 조립 대상



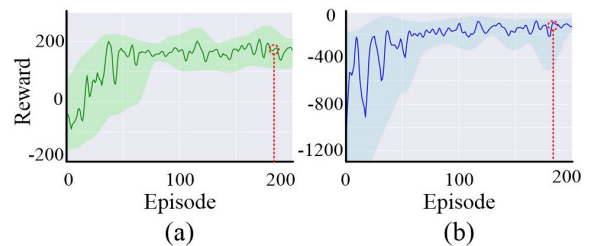
[Fig. 7] (a) Experimental setup, and (b) initial phase

물체는 PC 조립에 사용되는 파워 커넥터로 선정하였다. 로봇은 커넥터를 완벽히 파지한 상태이며, 파지된 커넥터의 움직임은 없는 것을 가정한다. 로봇의 툴 좌표계 $\{T\}$ 는 커넥터의 물체 좌표계 $\{O\}$ 와 동일하게 설정하였다.

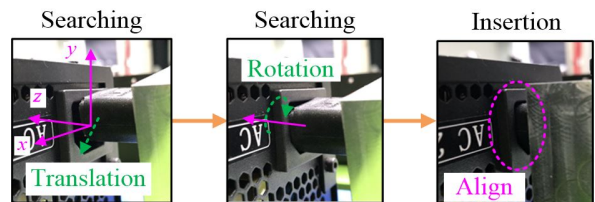
제안된 방법을 검증하기 위해 [Fig. 7(b)]와 같이 커넥터가 구멍 근처에 접촉한 후, 구멍의 좌표계 $\{H\}$ 로부터 툴 좌표계 $\{T\}$ 까지의 병진 및 회전 오차가 존재하는 상황에서 조립을 수행하였다. 이때 $\{H\}$ 는 3차원 공간 상에 고정된 좌표계이다. 병진오차는 $\{H\}$ 의 원점으로부터 $\{T\}$ 의 원점까지의 위치 오차를 의미하며, x 축 및 y 축 방향의 병진오차 e_x, e_y 를 ± 4 mm의 범위에서 무작위로 생성하였다. 회전오차는 $\{H\}$ 로부터 $\{T\}$ 까지의 각 축에 대한 회전오차로, x 축 및 y 축 기준의 회전오차 e_{rx}, e_{ry} 는 $\pm 6^\circ$, z 축 기준의 회전오차 e_{rz} 는 $\pm 15^\circ$ 의 범위에서 무작위로 생성하였다.

먼저, 총 200회의 에피소드에 걸쳐 강화학습 기반의 조립전략을 학습하였고, 결과는 [Fig. 8(a)], [Fig. 8(b)]와 같다. 학습 과정에서 탐색 궤적 생성 및 강성 조절을 위한 조립전략이 동시에 학습되며, SAC 1, 2에 대한 보상이 80회의 에피소드 수행 이후부터 점차 수렴하여 적절히 학습이 수행된 것으로 판단하였다. 본 연구에서는 180번째 에피소드의 가중치를 조립전략의 성능 평가에 사용하였다. [Fig. 9]는 파워 커넥터에 대한 조립 과정을 보여주는 사진이다.

오차는 정의된 범위 내에서 무작위로 생성하였다. 먼저, [Fig. 9]와 같이 로봇은 x 축 및 y 축 방향으로 병진 이동하며 구멍의 위치를 탐색한다. 커넥터가 구멍에 근접하면 z 축 기준의 회전을 통해 커넥터와 구멍을 정렬한다. 마지막으로, 커넥터와 구멍이 정렬되면 삽입을 수행한다. 이 과정에서 실시간으로



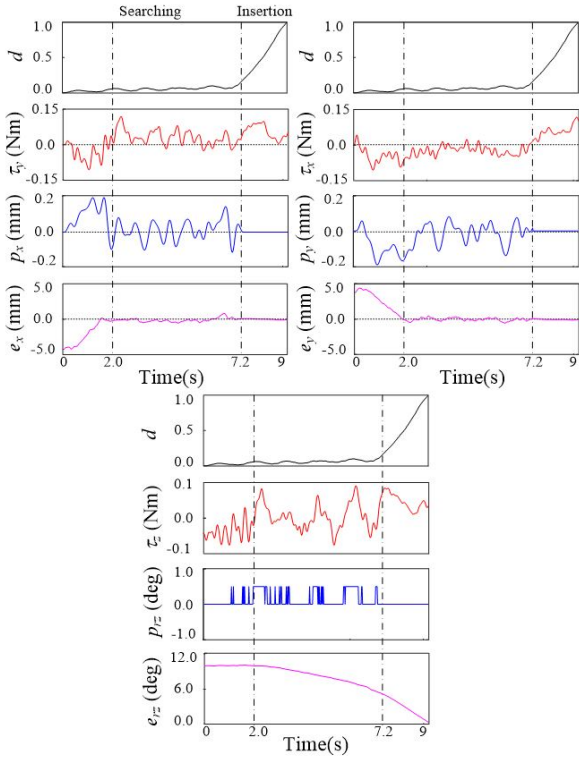
[Fig. 8] Sum of rewards per episode obtained while learning assembly strategy: (a) SAC 1, and (b) SAC 2



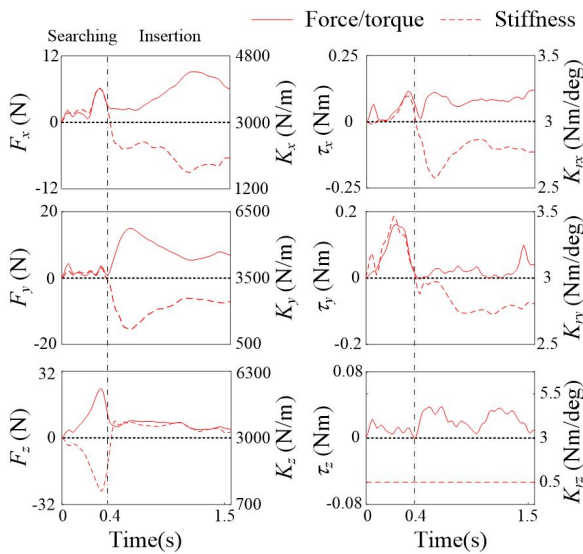
[Fig. 9] Power connector assembly process with learning results applied

측정되는 각 축의 접촉력이 50N을 초과하면 실패로 간주된다. 조립 과정에서 강화학습 기반의 조립전략을 통해 생성된 탐색 궤적(p_x, p_y, p_z)은 [Fig. 10]과 같다.

조립 초기부터 2s까지 측정된 토크($+\tau_x, -\tau_x$)에 따라 위치 궤적($+p_x, -p_x$)이 생성되어 점차 위치 오차(e_x, e_y)를 줄인다. 이는 식 (7)-(8)에서 보상을 설계한 의도대로 궤적이 생성됨을



[Fig. 10] Search trajectory during the assembly process



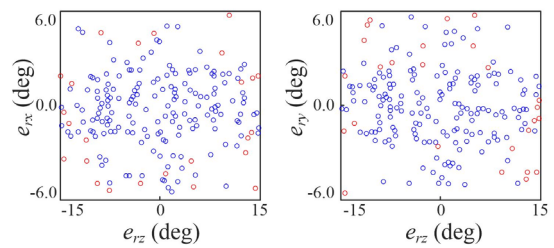
[Fig. 11] Force/torque and stiffness obtained during the assembly process

보여준다. 다음으로 위치 오차가 충분히 감소하면서 커넥터가 구멍 위치에 근접한다. 그 결과로 토크($+\tau_x$)가 측정되고 방위 궤적($+p_z$)이 생성되어 방위 오차(e_z)가 감소한다. 방위 오차는 점차 감소하여 7.2s 부근에서 커넥터와 구멍이 정렬되면 삽입이 시작되고, 최종적으로 조립을 완료한다.

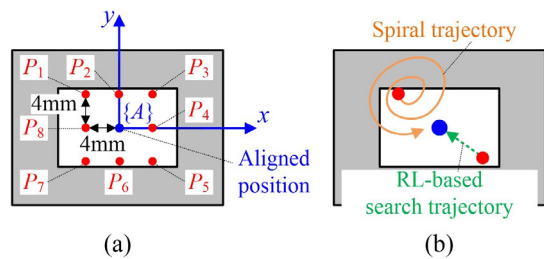
한편, 강화학습 기반의 조립전략을 이용하여 조절된 강성은 [Fig. 11]과 같다. z축 강성(K_z)의 경우, 조립 초기부터 0.4초까지의 탐색에서 측정된 힘($+F_z$)이 증가함에 따라 점차 감소한다. 반면에 이후의 삽입에서는 측정된 힘($+F_z$)이 적절한 접촉력을 유지하도록 강성(K_z)이 일정 수준으로 유지된다. 이와 마찬가지로 x축 및 y축 강성의 경우에도 식 (9)-(11)에서 보상을 설계한 의도에 부합하는 결과를 보여준다.

앞서 두 실험을 통해 강화학습을 기반으로 학습된 조립전략이 적절한 탐색 궤적을 생성하고, 강성을 조절할 수 있음을 보였다. 한편, 파워 커넥터 조립의 조립 성공률을 검증하기 위해 총 200회의 커넥터 조립을 수행하였다. 결과는 [Fig. 12]와 같다. 청색으로 표시한 점은 조립 성공, 적색으로 표시한 점은 조립 실패를 나타낸다. 총 200회의 조립 중에 27회를 실패하였다. 실패한 경우는 모두 e_{rx}, e_{ry} 는 $\pm 3.5^\circ$, e_{rz} 가 $\pm 7^\circ$ 의 범위를 초과하였다. 이는 비교적 큰 오차이므로 앞서 보상함수에 정의한 것과 다른 양상의 접촉이 발생하여 조립이 실패한 것으로 보인다. 하지만 해당 범위 이내에서는 100%의 조립 성공률을 보이는 데, 이는 보상함수를 통해 정의한 접촉 양상이 실제 커넥터와 구멍 사이의 접촉을 적절히 반영했음을 의미한다.

한편, 기존에 널리 사용되는 힘/토크 정보 기반의 블라인드 탐색(blind search)과 강화학습 기반의 탐색의 비교 실험을 수행하였다. 블라인드 탐색은 나선 궤적(spiral trajectory)을 이용



[Fig. 12] Experimental results of 200 trials of connector assembly



[Fig. 13] (a) Initial position with errors, and (b) search trajectory

[Table 1] Search time for blind search

Position 1	Position 2	Position 3	Position 4	Average
7.66s	Fail	6.02s	3.68s	
Position 5	Position 6	Position 7	Position 8	5.77s
4.18s	4.36s	7.0s	7.48s	

[Table 2] Search time for the proposed method

Position 1	Position 2	Position 3	Position 4	Average
2.14s	1.16s	1.0s	0.7s	
Position 5	Position 6	Position 7	Position 8	1.07s
0.98s	0.62s	1.08s	0.84s	

하였다. 나선 궤적은 회전은 고려하지 않으므로 두 방법의 비교 실험은 회전 오차는 없고, 최대 ±4 mm의 위치 오차가 있는 조건에서 실험을 수행하였다. 실험에 사용된 초기 위치는 [Fig. 13(a)]와 같고, 적색으로 표시된 8개의 점을 초기 위치로 하여 각 방법을 1회씩 수행하였다. 파워 커넥터 조립에 대한 탐색 시간의 산출 결과는 [Table 1], [Table 2]와 같다.

나선 궤적을 이용하면 8회 중 1회를 실패하였고, 강화학습 기반의 궤적은 조립을 모두 성공하였다. 탐색 시간은 각각 평균 5.77s, 1.07s로, 제안된 방법이 약 5배 정도 빠르게 탐색을 완료하였다. 나선 궤적은 [Fig. 13(b)]와 같이 생성된 궤적을 따라 더 긴 거리를 이동해야 하므로 시간이 오래 걸리고, 나선의 크기에 따라 탐색이 실패할 가능성이 높다. 하지만 강화학습 기반의 탐색 궤적은 접촉 상태를 기반으로 구멍을 향하는 궤적을 빠르게 생성할 수 있어, 더 빠른 탐색 시간을 확보할 수 있었다.

5. 결론

본 연구에서는 접촉 상태를 이용한 강화학습 기반의 조립 전략을 제안하였다. 이 방법은 미리 정의된 접촉 상태에 따라 구멍을 신속히 찾을 수 있는 탐색 궤적을 생성하고, 커넥터 삽입 과정에서 유연한 접촉을 구현하기 위해 강성을 조절한다. 이를 위해 SAC 알고리즘을 이용한 강화학습 구조를 통해 커넥터 조립을 위한 탐색 및 삽입 전략을 학습하였다. 이 과정에서 적절한 보상함수를 설계하여 실시간으로 측정되는 힘/토크 정보에 따른 접촉 양상을 반영하였다. 제안된 방법을 파워 커넥터 조립에 적용하여 학습을 수행하였고, 200회의 파워 커넥터 조립을 수행한 결과 허용 범위 이내의 오차에서 100%의 조립 성공률을 보였다. 또한 블라인드 탐색 방법과의 비교 실험을 통해, 제안된 방법이 블라인드 탐색보다 약 5배 빠른 1s 수준의 탐색 시간을 보였다.

향후 연구에서는 다양한 복잡 형상에 대해 대응이 가능함

을 추가 검증하고, 고정되지 않은 다양한 파지 상태에 대해서도 대응이 가능하도록 수행할 예정이다.

References

- [1] G. Schoettler, A. Nair, J. Luo, S. Bahl, J. A. Ojea, E. Solowjow, and S. Levine, "Deep Reinforcement Learning for Industrial Insertion Tasks with Visual Input and Natural Rewards," *2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, Las Vegas, NV, USA, pp. 5548-5555, 2020, DOI: 10.1109/IROS45743.2020.9341714.
- [2] D. E. Whitney, "Mechanical Assemblies: Their Design, Manufacture, and Role in Product Development," Oxford University Press, 2004, [Online], <https://www.globalspec.com/reference/69914/203279/4-k-further-reading>.
- [3] H. Park, J. H. Bae, J.-H. Part, M.-H. Baeg, and J. Park, "Intuitive Peg-in-Hole Assembly Strategy with a Compliant Manipulator," *IEEE ISR 2013*, Seoul, Korea, 2013, DOI: 10.1109/ISR.2013.6695699.
- [4] H. Park, J. Park, D.-H. Lee, J.-H. Park, M.-H. Baeg, and J.-H. Bae, "Compliance-based Robotic Peg-in-Hole Assembly Strategy Without Force Feedback," *IEEE Transactions on Industrial Electronics*, vol. 64, no. 8, pp. 6299-6309, August, 2017, DOI: 10.1109/TIE.2017.2682002.
- [5] F. Chen, F. Cannella, H. Huang, H. Sasaki, and T. Fukuda, "A study on error recovery search strategies of electronic connector mating for robotic fault-tolerant assembly," *Journal of Intelligent Robotic Systems*, vol. 81, pp. 257-271, February, 2016, DOI: 10.1007/s10846-015-0248-5.
- [6] W. Lian, T. Kelch, D. Holz, A. Norton, and S. Schaal, "Benchmarking Off-The-Shelf Solutions to Robotic Assembly Tasks," *2021 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, Prague, Czech Republic, pp. 1046-1053, 2021, DOI: 10.1109/IROS51168.2021.9636586.
- [7] G. Schoettler, A. Nair, J. A. Ojea, S. Levine, and E. Solowjow, "Meta-Reinforcement Learning for Robotic Industrial Insertion Tasks," *2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, Las Vegas, NV, USA, pp. 9728-9735, 2020, DOI: 10.1109/IROS45743.2020.9340848.
- [8] T. Inoue, G. De Magistris, A. Munawar, T. Yokoya, and R. Tachibana, "Deep reinforcement learning for high precision assembly tasks," *2017 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, Vancouver, BC, Canada, pp. 819-825, 2017, DOI: 10.1109/IROS.2017.8202244.
- [9] Y.-L. Kim, K.-H. Ahn, and J.-B. Song, "Reinforcement Learning based on Movement Primitives for Contact Tasks," *Robotics and Computer-Integrated Manufacturing*, vol. 62, April, 2020, DOI: 10.1016/j.rcim.2019.101863.
- [10] H.-C. Song, Y.-L. Kim, and J.-B. Song, "Guidance algorithm for complex-shape peg-in-hole strategy based on geometrical information and force control," *Advanced Robotics*, vol. 30, no. 8, pp. 552-563, February, 2016, DOI: 10.1080/01691864.2015.1130172.

- [11] N. Hogan, "Impedance Control: An Approach to Manipulation: Part III-Applications," *Journal of Dynamic Systems, Measurement, and Control*, vol. 107, no. 1, pp. 17-24, March, 1985, DOI: 10.1115/1.3140713.
- [12] T. Haarnoja, A. Zhou, P. Abbeel, and S. Levine, "Soft Actor-Critic: Off-Policy Maximum Entropy Deep Reinforcement Learning with a Stochastic Actor," *Int. Conf. on Machine Learning*, Stockholm Sweden, pp. 1861-1870, 2018, [Online], <https://proceedings.mlr.press/v80/haarnoja18b>.



김 용 건

2020 국민대학교 융합기계공학과(학사)
2020~현재 고려대학교 기계공학과(석사)

관심분야: 로봇 제어, 로봇 기반 조립, 강화학습



나 민 우

2017 고려대학교 기계공학과(학사)
2019 고려대학교 스마트융합학과(석사)
2019~현재 고려대학교 기계공학과(박사)

관심분야: 로봇 제어, 로봇 기반 조립, 3차원 측정



송 재 복

1983 서울대학교 기계공학과(공학사)
1985 서울대학교 기계공학과(공학석사)
1992 MIT 기계공학과(공학박사)
1993~현재 고려대학교 기계공학부 교수

관심분야: 로봇의 설계 및 제어, 협동로봇, 중력보상 로봇, AI 기반 로봇 매니플레이션