

# 자연어 명령 기반 **unlabeled** 객체 파지를 위한 **AI** 로봇틱스 연구 아이디어 분석

## 1. 서론

최근 인간과 로봇 간의 협업이 다양한 분야에서 중요해짐에 따라, 인간의 자연어 명령을 이해하고 이를 기반으로 작업을 수행하는 로봇 팔의 능력은 그 중요성이 더욱 커지고 있다.<sup>1</sup> 특히, 음성 지시와 같은 직관적인 인터페이스를 통해 로봇이 사용자의 의도를 파악하고 다양한 객체를 정확하게 조작할 수 있다면, 가정, 산업 현장 등 일상 환경에서의 로봇 활용도가 크게 향상될 것으로 기대된다.<sup>1</sup> 실제로 향후 수십 년 안에 수많은 가정과 산업 분야에서 로봇 보조 기술이 채택될 것으로 예상되며, 이는 로봇이 더욱 유연하고 사용자 친화적인 상호 작용 능력을 갖추어야 함을 의미한다.<sup>1</sup> 이러한 맥락에서, 시각 정보와 언어 정보를 융합하여 로봇을 제어하는 연구가 활발히 진행되고 있으며, 이는 로봇이 복잡한 작업을 보다 효과적으로 수행할 수 있도록 돕는다.<sup>2</sup>

그러나, 기존의 많은 로봇 시스템은 특정 작업 환경이나 미리 정의된 객체에 국한되어 작동하는 경우가 많으며, 새로운 객체나 상황에 대해서는 광범위한 재학습 과정을 거쳐야만 한다.<sup>1</sup> 이는 로봇 시스템의 일반화 능력을 저해하는 주요 요인으로 작용하며, 다양한 실제 환경에서의 활용에 어려움을 초래한다. 반면, 인간은 사전 지식이 없는 새로운 물체를 마주하더라도 직관적으로 파악하고 적절한 방식으로 조작하는 뛰어난 능력을 보여준다.<sup>3</sup> 이러한 인간의 능력은 로봇 연구자들에게 지속적인 영감을 제공하며, 로봇이 보다 폭넓은 환경과 객체에 대해 유연하게 대처할 수 있는 지능적인 시스템을 개발하는 것을 목표로 한다.<sup>3</sup> 물류 센터에서의 패키지 분류 작업이나 공장에서의 부품 픽킹 작업과 같이, 다양한 로봇 응용 분야에서 객체 조작 능력은 핵심적인 요구 사항으로 자리 잡고 있다.<sup>3</sup> 따라서, **unlabeled** 객체를 포함한 다양한 환경에서 로봇이 자연어 명령에 따라 객체를 인식하고 파지할 수 있도록 하는 연구는 매우 중요하며, 이는 로봇 기술의 실질적인 발전과 광범위한 응용 가능성을 열어줄 것이다.

본 보고서는 자연어 명령 기반 **unlabeled** 객체 파지라는 사용자님의 연구 아이디어에 대한 심층적인 분석을 제공하고자 한다. 특히, 자연어 이해와 시각 정보 처리에 강력한 성능을 보이는 **CLIP (Contrastive Language-Image Pretraining)** 모델을 활용하여 자연어 명령과 시각 정보를 연결하고, 이를 기반으로 **manipulator arm**을 제어하여 객체를 파지하는 전체적인 파이프라인에 대한 다양한 접근 방식을 조사할 것이다. 또한, 사용자님께서 제시한 구체적인 질문들, 예를 들어 **segmentation**을 통한 파지의 장점, 깊이 카메라를 이용한 무게 중심(**COM**) 추정, 그리고 **COM** 외에 시각 정보를 활용한 파지 방법에 대한 답변을 포함하여, 사용자님의 연구 방향 설정에 실질적인 도움을 제공하는 것을 목표로 한다.

## 2. 자연어-시각 정보 연결을 위한 **CLIP** 모델 활용

CLIP (Contrastive Language-Image Pretraining) 모델은 이미지와 텍스트를 공통의 임베딩 공간으로 매핑하여, 이미지와 텍스트 간의 의미론적 유사성을 학습하는 심층 학습 모델이다.<sup>2</sup> 이 모델은 방대한 양의 이미지-텍스트 쌍 데이터셋을 이용하여 사전 학습되었으며, 이를 통해 다양한 시각적 개념과 그에 상응하는 언어적 표현을 이해하는 강력한 능력을 갖추게 되었다.<sup>4</sup> CLIP은 이미지 분류, 이미지 검색, 캡션 생성 등 다양한 컴퓨터 비전 태스크에서 기존의 단일 모드 모델들보다 뛰어난 성능을 보여주었으며<sup>2</sup>, 최근에는 로봇 팔 제어와 관련된 여러 작업에서도 특징 추출기로 활발하게 활용되고 있다.<sup>2</sup>

CLIP 모델은 자연어 명령을 기반으로 로봇 팔이 객체를 파지하는 작업을 수행하는 데 매우 유망한 도구로 여겨진다. 예를 들어, 사용자가 "빨간색 컵"이라고 명령했을 때, CLIP은 이 텍스트 명령의 의미를 이해하고, 로봇 팔에 장착된 카메라를 통해 얻은 이미지에서 "빨간색"이고 "컵"의 특징을 가진 객체를 식별할 수 있다.<sup>1</sup> 실제로 CLIP은 언어 기반 파지 감지<sup>2</sup>, 제로샷 객체 감지<sup>2</sup>, 언어 기반 로봇 탐색<sup>2</sup> 등 다양한 로봇 작업에 적용될 수 있음이 입증되었다. 특히, Robotic-CLIP과 같은 연구에서는 액션 데이터를 활용하여 CLIP 모델을 미세 조정함으로써, 로봇의 인식 능력을 더욱 향상시키는 방법을 제시한다.<sup>2</sup> 또한, CLIP-RT는 자연어 감독을 통해 로봇 정책을 학습하는 VLA (Vision-Language-Action) 모델로, 사전 학습된 CLIP 모델을 로봇 학습 영역으로 확장하여, 비전문가도 자연어만으로 로봇에게 새로운 조작 기술을 가르칠 수 있도록 한다.<sup>3</sup>

사용자님의 아이디어처럼, 만약 사용자가 "red cup"이라고 자연어로 명령을 내리면, CLIP은 먼저 이 텍스트 명령을 텍스트 임베딩이라는 벡터 공간의 한 점으로 변환한다. 동시에, 로봇 팔에 장착된 카메라를 통해 얻은 이미지 또한 이미지 임베딩이라는 벡터 공간의 또 다른 점으로 변환된다. CLIP 모델은 사전 학습 과정에서 이미지와 그에 대한 설명을 담은 수많은 데이터를 통해, 의미적으로 유사한 이미지와 텍스트는 임베딩 공간에서 서로 가까이 위치하도록 학습되었다. 따라서, "red cup"이라는 텍스트 임베딩은 이미지에서 실제 빨간색 컵의 이미지 임베딩과 높은 유사도를 가질 가능성이 높다. CLIP은 이러한 유사도를 계산하여 이미지 내에서 사용자가 원하는 "red cup"과 가장 유사한 객체를 식별할 수 있게 된다.<sup>1</sup> 더 나아가, Attention-Guided Integration of CLIP and SAM과 같은 연구에서는 CLIP과 SAM (Segment Anything Model)을 통합하여, 텍스트 기반 프롬프트에 따라 이미지에서 정확한 객체 마스크를 생성하는 파이프라인을 제안한다.<sup>4</sup> 이처럼 CLIP은 자연어 명령에 따른 객체 인식을 위한 강력한 도구이지만, 그 자체로는 로봇 팔의 조인트 제어나 엔드 이펙터의 직접적인 움직임을 생성하지 못한다는 점을 유념해야 한다. CLIP은 주로 시각적 인식 및 특징 추출에 사용되며, 실제 로봇 행동을 위해서는 CLIP의 출력을 기반으로 하는 별도의 제어 메커니즘이 필요하다. Robotic-CLIP과 CLIP-RT 연구는 이러한 CLIP의 한계를 극복하고, 자연어 명령을 로봇의 실제 행동으로 연결하기 위한 다양한 방법을 탐색하고 있다.

### 3. Manipulator Arm 제어 및 파지 전략

CLIP 모델을 통해 사용자의 자연어 명령에 따라 객체를 식별했다면, 다음 단계는 식별된 객체를 실제로 파지하기 위해 **manipulator arm**을 제어하는 것이다. 이때 CLIP 모델로부터 얻은 인식 정보를 어떻게 활용하여 로봇 팔의 움직임을 결정하고, 어떤 방식으로 객체를 파지할 것인지에 대한 다양한 전략들을 고려해 볼 수 있다.

#### 3.1 CLIP 기반 인식 정보를 활용한 로봇 팔 제어 방법

CLIP을 통해 식별된 객체의 특징 정보를 이용하여 실제 로봇 팔의 제어를 위한 다양한 알고리즘들을 탐색할 수 있다. 주요 방법으로는 시각 서보잉, 강화 학습 기반 파지 전략, 그리고 3D 객체 포즈 추정 후 파지 계획 등이 있다.

##### 3.1.1 시각 서보잉 기반 제어

시각 서보잉 (Visual Servoing)은 카메라를 통해 얻은 시각적 피드백을 이용하여 로봇 팔의 움직임을 실시간으로 제어하는 방식이다.<sup>11</sup> 이 방법은 로봇이 목표 객체의 시각적 특징 (예: 이미지 상의 특정 점, 객체의 윤곽 등)을 추적하고, 이 특징들이 원하는 목표 위치에 도달하도록 로봇 팔의 움직임을 조정하는 방식으로 작동한다. 최근에는 DVSFN (Deep Visual Servoing Feature Network)과 같은 심층 학습 기반 특징 추출 네트워크를 활용하여 복잡한 시각 환경에서도 효율적인 특징 추출이 가능해졌다.<sup>11</sup> 이러한 네트워크는 YOLO (You Only Look Once)와 유사한 구조를 가지며, 스케일 불변 특징점과 객체 바운딩 박스를 실시간으로 추출할 수 있도록 설계되었다.<sup>11</sup> DVSFN은 LM-IBVS (Levenberg-Marquardt-based Image Visual Servoing) 컨트롤러와 통합되어 이미지 특징과 로봇 조인트 공간 간의 매핑을 생성하고, 목표 특징과 현재 특징의 차이를 이용하여 로봇 팔의 속도를 제어한다.<sup>11</sup> 시각 서보잉은 실시간 피드백을 통해 비교적 정확한 제어가 가능하다는 장점이 있지만, 복잡한 환경이나 객체의 갑작스러운 움직임에 취약할 수 있으며, CLIP 모델의 의미론적 정보를 직접적으로 활용하기보다는 주로 시각적 특징 점에 의존하는 경향이 있다.

##### 3.1.2 강화 학습 기반 파지 전략

강화 학습 (Reinforcement Learning)은 로봇 팔이 스스로 시행착오를 거치면서 환경과의 상호작용을 통해 최적의 파지 전략을 학습하는 방식이다.<sup>3</sup> 특히, 깊은 강화 학습 (Deep Reinforcement Learning, DRL) 알고리즘 (예: SAC - Soft Actor-Critic)은 심층 신경망을 활용하여 복잡한 상태 공간과 행동 공간을 처리할 수 있도록 한다.<sup>3</sup> 사용자님의 연구에서 CLIP을 통해 얻은 객체의 종류나 특징 정보를 강화 학습의 상태로 활용하고, 로봇 팔의 다양한 움직임 (조인트 각도 변화, 엔드 이펙터의 위치 및 방향 조절 등)을 행동으로 정의하여 강화 학습을 수행할 수 있다. 강화 학습은 다양한 형태와 재질을 가진 **unlabeled** 객체에 대한 일반화 능력을 향상시킬 수 있다는 큰 장점이 있지만, 학습에 많은 양의 데이터와 시간이 소요될 수 있으며, 특히 6 자유도 (6DoF) 로봇 팔의 복잡한 제어는 더욱 어려운 과제가 될 수 있다.<sup>3</sup> 따라서, CLIP 모델로부터 얻은 의미론적 정보를 강화 학습의

보상 함수 설계에 효과적으로 활용하는 방법이 중요하게 고려되어야 한다.

### 3.1.3 3D 객체 포즈 추정 후 파지 계획

세 번째 방법은 CLIP 모델을 통해 식별된 객체의 3D 포즈 (위치 및 방향)를 추정한 후, 이 추정된 정보를 기반으로 실제 파지 동작을 위한 구체적인 계획을 수립하는 것이다. 기존 연구에서는 3D 모델 기반 또는 깊이 점군 데이터 기반으로 객체의 3D 포즈를 추정하는 다양한 방법들이 사용되어 왔다.<sup>18</sup> 사용자님의 연구에서는 CLIP 모델로부터 얻은 객체의 특징 정보를 활용하여 3D 포즈 추정의 정확도를 더욱 높이는 방안을 모색할 수 있다. 예를 들어, CLIP이 식별한 객체의 종류에 따라 예상되는 일반적인 형태 정보를 활용하거나, CLIP이 주목하는 이미지 영역을 기반으로 3D 포즈 추정 알고리즘의 탐색 범위를 좁힐 수 있을 것이다. 이 방식은 파지 동작을 명시적으로 계획할 수 있다는 장점이 있지만, CLIP 모델로부터 정확한 3D 포즈 정보를 얻는 것이 어려울 수 있으며, 특히 unlabeled 객체의 경우에는 3D 모델이 존재하지 않을 가능성이 높으므로, 깊이 점군 데이터 기반의 포즈 추정 방법과의 연동이 중요하게 고려되어야 한다.

### 3.2 각 제어 및 파지 전략의 장단점 비교 분석

각각의 로봇 팔 제어 및 파지 전략은 고유한 장점과 단점을 가지고 있으며, 사용자님의 연구 목표와 환경에 따라 적합한 방법을 선택하는 것이 중요하다. 아래 표는 앞서 설명된 세 가지 주요 전략들을 데이터 요구량, 계산 복잡성, 실시간 제어 가능성, 일반화 성능 측면에서 비교 분석한 것이다.

방법	장점	단점	CLIP 활용 방안
시각 서보잉	실시간 제어 가능, 비교적 간단한 구현	환경 변화에 민감, CLIP의 의미론적 정보 활용 어려움	CLIP으로 목표 객체 식별 후 시각적 특징 추적
강화 학습	복잡한 전략 학습 가능, 일반화 능력 우수	학습 데이터 및 시간 요구, 6DoF 제어 어려움	CLIP으로 얻은 객체 정보를 상태/보상 함수에 활용
3D 포즈 추정 후 파지 계획	명시적인 파지 계획 수립 가능	CLIP으로부터 정확한 3D 포즈 정보 획득 어려움	CLIP으로 초기 포즈 추정 후 점군 데이터로 정교화

## 4. 깊이 점군 데이터를 이용한 인식 및 파지

사용자님께서서는 깊이 점군 데이터를 이용하여 객체를 인식하고 파지하는 연구에 대해 고민하고 계시며, segmentation 기반 방식과 직접 강화 학습을 적용하는 방식의 차이점,

그리고 무게 중심(COM) 외에 시각 정보를 활용한 파지 방법에 대해 질문하셨다.

#### 4.1 질문 1 답변: **Segmentation** 기반 파지의 장점 및 강화 학습과의 비교

##### 4.1.1 **Segmentation**의 역할 및 파지 성능 향상 기여도

이미지 또는 깊이 점군 데이터에서 관심 객체의 영역을 정확하게 분리해내는 **segmentation** 기술은 로봇 팔의 파지 성능을 향상시키는 데 여러 가지 중요한 장점을 제공한다.<sup>3</sup> 첫째, **segmentation**을 통해 객체의 형태, 크기, 방향 등과 같은 고유한 특징 정보를 보다 정확하게 파악할 수 있으며<sup>22</sup>, 이는 로봇이 안정적인 파지 포즈를 생성하고 계획하는 데 직접적인 도움을 준다. 예를 들어,<sup>22</sup>에서는 이미지 **segmentation**과 더불어 객체의 주요 특징점 (직선, 코너 등)을 검출하고 이를 통합하여 학습 데이터셋의 크기에 제한 없이 파지 포즈를 추론하는 방법을 제시한다. 이는 **segmentation**이 단순히 객체를 배경으로부터 분리하는 역할뿐만 아니라, 효과적인 파지 전략을 위한 핵심 정보를 제공할 수 있음을 시사한다. 둘째, 실제 작업 환경은 종종 여러 객체가 밀집되어 놓여 있는 클러터 (**clutter**) 환경인 경우가 많다. 이때 **segmentation**은 각 객체를 개별적으로 분리하여 인식하고, 로봇 팔이 원하는 특정 객체만을 정확하게 파지할 수 있도록 지원한다.<sup>3</sup> 셋째, **segmentation**은 객체의 가려진 부분이나 불완전한 정보를 보완하는 데에도 활용될 수 있다. 예를 들어, 깊이 정보를 활용한 **segmentation**은 2D 이미지 **segmentation**에 비해 객체의 3차원 형태를 더 잘 파악할 수 있도록 하며<sup>24</sup>, 이는 로봇이 부분적으로만 보이는 객체에 대해서도 적절한 파지 전략을 수립하는 데 도움을 줄 수 있다.

##### 4.1.2 깊이 점군 데이터 직접 활용 강화 학습의 특징 및 한계

깊이 점군 데이터를 **segmentation** 없이 직접 입력으로 사용하는 강화 학습 (**Reinforcement Learning**) 방식은, 입력 데이터로부터 특징을 추출하고 파지 전략을 학습하는 전 과정을 **end-to-end**로 처리할 수 있다는 특징을 가진다.<sup>25</sup> 이 접근 방식의 주요 장점은 로봇이 환경과의 능동적인 상호작용을 통해, 인간이 사전에 정의하기 어려운 복잡하고 미묘한 파지 전략까지 스스로 학습하고 발견할 수 있다는 점이다.<sup>3</sup> 그러나, 깊이 점군 데이터는 그 자체로 고차원적이고 노이즈를 포함할 가능성이 높아, 효과적인 강화 학습을 위해서는 매우 많은 양의 학습 데이터와 긴 학습 시간이 요구될 수 있다.<sup>3</sup> 특히, 로봇 팔의 6 자유도 (6DoF) 제어와 같이 복잡한 행동 공간을 학습해야 하는 경우에는 학습의 어려움이 더욱 증가한다. 여러 연구에서 깊이 점군 데이터를 직접 활용하는 강화 학습은 복잡한 관계를 모델링하는 데 유리하지만, 데이터 효율성이 낮고 학습 안정성이 떨어질 수 있다는 점을 지적한다. 예를 들어,<sup>25</sup>와<sup>25</sup>에서는 모방 학습 (**Imitation Learning**)과 강화 학습을 결합하여 초기 파지 전략을 빠르게 습득하고, 이후 강화 학습을 통해 이를 개선함으로써 데이터 효율성을 높이는 방법을 제안한다.

##### 4.1.3 SAM, MobileSAM 등 분할 알고리즘 활용의 이점

최근 등장한 SAM (**Segment Anything Model**)과 MobileSAM과 같은 강력한 분할 알고리즘들은 다양한 종류의 객체에 대해 매우 높은 품질의 **segmentation** 결과를

제공하며, 특히 **zero-shot** 또는 **few-shot** 학습 능력을 갖추고 있어, 학습 데이터에 없던 새로운 환경이나 객체에 대해서도 뛰어난 적응력을 보인다.<sup>4</sup> **MobileSAM**은 **SAM**의 기본적인 성능을 유지하면서 모델 크기와 연산량을 크게 줄여, 연산 자원이 제한적인 환경에서도 효율적인 **segmentation**을 가능하게 한다. 이러한 고성능 분할 알고리즘을 먼저 깊이 점군 데이터에 적용하여 관심 객체를 분리한 후, 분리된 객체 정보 (예: 마스크, 점군)를 입력으로 사용하여 강화 학습을 진행하면, 로봇이 학습해야 할 상태 공간의 복잡성을 크게 줄일 수 있으며, 결과적으로 보다 효율적인 학습이 가능하다.<sup>2438</sup>

연구에서는 **VFM (Vision Foundation Models)**을 활용한 **segmentation**이 강화 학습의 데이터 비효율성 문제를 효과적으로 해결하는 데 기여할 수 있음을 보여준다. **SAM**과 같은 모델은 방대한 데이터로 사전 학습된 강력한 표현 학습 능력을 바탕으로, 다양한 객체에 대한 정확한 **segmentation**을 제공함으로써, 강화 학습 에이전트가 객체 인식이라는 어려운 문제에 집중하는 대신, 실제 파지 전략 학습이라는 본질적인 목표에 더욱 집중할 수 있도록 돕는다.

#### 4.1.4 데이터 효율성, 일반화 성능, 학습 난이도 측면에서의 비교 분석

특징	깊이 점군 데이터 직접 활용 강화 학습	분할 알고리즘 활용 후 강화 학습
데이터 효율성	낮음	높음 (분할 정보 활용으로 학습 용이)
일반화 성능	잠재적으로 높음 (다양한 환경 경험 필요)	높음 (사전 학습된 분할 모델의 일반화 능력 활용)
학습 난이도	높음 (복잡한 상태-행동 공간 학습)	중간 (분할 정보로 상태 공간 단순화)
장점	end-to-end 학습 가능, 복잡한 전략 스스로 학습 가능	학습 데이터 요구량 감소, 학습 속도 향상, 사전 학습 모델 활용 용이
단점	많은 데이터 및 긴 학습 시간 필요, 학습 안정성 확보 어려움	분할 알고리즘의 성능에 의존성 발생 가능

#### 4.2 질문 2 답변: 깊이 카메라를 이용한 무게 중심(COM) 추정 알고리즘

깊이 카메라로부터 얻은 깊이 점군 데이터를 이용하여 객체의 무게 중심 (**Center of Mass, COM**)을 추정하는 것은 로봇 팔이 객체를 파지하기 위한 기본적인 정보로 활용될 수 있다.



#### 4.2.1 깊이 점군 데이터로부터 COM 추출 방법론

깊이 점군 데이터는 객체 표면의 3차원 좌표 ( $x, y, z$ )들의 집합으로 표현된다. 가장 기본적인 방법은 이 점들의 좌표를 산술 평균하여 객체의 무게 중심을 추정하는 것이다.<sup>39</sup> Open3D 라이브러리와 같은 도구를 사용하면, 점군 데이터 객체에 내장된 함수를 통해 손쉽게 **centroid** (기하학적 중심)를 계산할 수 있다.<sup>40</sup> 그러나 실제 객체의 무게 중심은 기하학적 중심과 반드시 일치하지는 않으며, 객체의 밀도 분포에 따라 달라질 수 있다.<sup>39</sup>에서는 단순 평균 방식 외에도, 점군 데이터에서 이상치 (**noise**)를 제거한 후 평균을 계산하거나, **k-means** 클러스터링과 같은 알고리즘을 사용하여 점군 데이터를 여러 개의 클러스터로 분리한 다음, 각 클러스터의 **COM**을 개별적으로 계산하는 등 다양한 방법이 있음을 언급한다. 이는 단순히 모든 점의 평균을 취하는 것보다 더 **robust**한 **COM** 추정을 가능하게 할 수 있다.

#### 4.2.2 COM 추정 관련 기존 연구 및 알고리즘 소개

기존 연구에서는 깊이 점군 데이터를 활용하여 객체의 **COM**을 추정하기 위한 다양한 알고리즘들이 개발되어 왔다. 예를 들어, **M3dimStat()** 함수는 점군 또는 깊이 맵을 입력으로 받아 바운딩 박스, **centroid**, 점들 간의 거리, 점 개수 등의 다양한 통계 정보를 계산할 수 있는 기능을 제공한다.<sup>41</sup> 또한, **PCA (Principal Component Analysis)**를 이용하여 점군 좌표를 보정하고, 다양한 방향에서 **2.5D** 볼륨을 계산하여 객체의 **COM**을 추정하는 방법도 연구되었다.<sup>42,43</sup>에서는 항공기나 자동차 모델과 같이 객체의 재질 밀도 정보가 알려진 경우, 이러한 정보를 활용하여 더욱 정확하게 **COM**을 추정하는 방법을 제시한다. 그러나 사용자님의 연구에서는 **unlabeled** 객체를 대상으로 하므로, 재질 밀도에 대한 사전 정보 없이 깊이 점군 데이터만으로 **COM**을 효과적으로 추정할 수 있는 알고리즘에 초점을 맞추는 것이 더 적절할 수 있다.<sup>69</sup>에서는 스테레오 비전을 통해 얻은 점군 데이터를 이용하여 부유 입자가 있는 환경에서 객체의 **COM**을 추정하는 알고리즘을 제안하며,<sup>70</sup>에서는 **3D** 레이저 스캐너로 얻은 표면 점군 데이터로부터 내부 공간을 채우고 **COM**을 계산하는 방법을 소개한다.

#### 4.2.3 MoveIt2와의 연동 방안 및 고려 사항

로봇 모션 플래닝 및 제어를 위한 강력한 프레임워크인 **MoveIt2**는 깊이 카메라로부터 얻은 깊이 점군 데이터를 직접 활용하여 주변 환경을 인식하고, 로봇 팔의 움직임 계획 시 충돌을 방지하는 기능을 제공한다.<sup>44</sup> **MoveIt2**는 점군 데이터를 **Occupancy Map**이라는 3차원 격자 맵으로 변환하여 표현하며, 이를 통해 로봇 팔이 작업 공간 내의 장애물을 인식하고 안전한 경로를 계획할 수 있도록 한다. 사용자님의 연구에서는 깊이 점군 데이터로부터 추정된 객체의 **COM** 정보를 **MoveIt2** 환경에서 활용하기 위해, 추정된 **COM** 위치에 가상의 **collision object** (예: 작은 구)를 생성하고 이를 **planning scene**에 추가하는 방안을 고려해 볼 수 있다.<sup>45</sup> 이렇게 추가된 **collision object**는 로봇 팔이 파지 동작을 계획할 때 충돌을 회피해야 하는 대상으로 작용하게 된다. **MoveIt2**는 다양한 센서 플러그인을 지원하므로, 깊이 카메라로부터 얻은 점군 데이터를 직접 **MoveIt2** 시스템에

통합하는 것은 비교적 용이할 것으로 예상된다. COM 정보를 collision object로 추가하는 기본적인 방식 외에도, MoveIt2의 perception pipeline을 확장하여 사용자 정의 COM 추정 기능을 통합하거나, 추정된 COM 정보를 이용하여 파지 동작의 목표 위치를 설정하는 등 더욱 다양한 방식으로 연동할 수 있을 것이다.

#### 4.3 질문 3 답변: COM 외 시각 정보 기반 파지 대안

깊이 점군 데이터의 무게 중심 (COM)을 이용하는 방식 외에도, 순수하게 시각 정보만을 이용하여 로봇 팔이 객체를 파지할 수 있도록 하는 다른 대안적인 방법들이 존재한다.

##### 4.3.1 3D 바운딩 박스 기반 파지

깊이 점군 데이터로부터 객체의 3차원 바운딩 박스를 추정하고, 이 바운딩 박스의 중심점이나 특정 면의 중심점을 파지 대상으로 설정하는 방법이 있다.<sup>48</sup> 예를 들어, SU-Grasp와 같은 알고리즘은 3D 비전 기술과 듀얼 스트림 네트워크 (RGB 이미지와 깊이 이미지를 각각 처리)를 활용하여 객체의 정확한 바운딩 박스를 추정하고, 이를 기반으로 로봇 팔의 파지 동작을 계획한다.<sup>48</sup> 바운딩 박스 기반 파지는 COM 기반 방식에 비해 객체의 전체적인 크기와 대략적인 형태를 고려할 수 있다는 장점이 있지만, 객체의 세밀한 파지 포즈를 결정하는 데에는 한계가 있을 수 있다. 특히, 복잡한 형태의 객체나 특정 부분을 정확하게 파지해야 하는 경우에는 COM 기반 방식보다 성능이 떨어질 수 있다.

##### 4.3.2 Grasp Pose Detection

깊이 점군 데이터로부터 직접적으로 로봇 팔이 객체를 성공적으로 파지할 수 있는 6 자유도 (6DoF) 파지 포즈를 검출하는 방법도 효과적인 대안이 될 수 있다.<sup>49</sup> 이 접근 방식은 객체의 형태, 크기, 그리고 주변 환경과의 관계 등을 종합적으로 고려하여 최적의 파지 위치와 방향을 예측한다. 기존 연구에서는 GPD (Grasp Pose Detection) 라이브러리와 같이 점군 데이터로부터 다양한 파지 후보를 생성하고, 딥러닝 모델을 사용하여 각 후보의 파지 성공률을 예측하는 방법들이 개발되었다.<sup>49</sup> 또한, PointNet과 같은 심층 신경망 모델을 사용하여 깊이 점군 데이터를 직접 입력으로 받아 파지 포즈를 예측하는 연구도 활발히 진행되고 있다.<sup>52</sup> Grasp Pose Detection은 객체의 형태와 주변 환경을 고려하여 COM 기반 방식이나 바운딩 박스 기반 방식보다 더 정확하고 안정적인 파지 포즈를 생성할 수 있다는 장점이 있지만, 학습을 위해서는 많은 양의 데이터가 필요할 수 있으며, 특히 복잡한 클러터 환경에서의 성능 향상을 위한 지속적인 연구가 요구된다.

##### 4.3.3 End-to-end 학습 기반 파지

또 다른 대안적인 방법은 시각 정보 (RGB 이미지 또는 깊이 이미지)를 입력으로 받아 로봇 팔의 조인트 각도나 엔드 이펙터의 움직임과 같은 제어 명령을 직접 출력하는 end-to-end 모델을 심층 학습을 통해 학습하는 것이다.<sup>14</sup> 이 방식은 명시적인 객체



인식이나 3D 포즈 추정 단계를 거치지 않고, 입력 시각 정보로부터 파지 동작을 위한 최적의 제어 명령을 직접적으로 학습한다는 특징을 가진다. 예를 들어, CLIPort와 같은 연구에서는 사전 학습된 CLIP 모델의 강력한 의미론적 이해 능력과 TransporterNet의 뛰어난 공간적 추론 능력을 결합하여, 다양한 자연어 기반의 로봇 팔 조작 태스크를 end-to-end 방식으로 성공적으로 수행하는 것을 보여준다.<sup>66</sup> End-to-end 학습은 명시적인 중간 단계 없이 시각 정보로부터 직접 파지 동작을 학습할 수 있다는 장점이 있지만, 학습된 정책의 내부 작동 방식을 이해하기 어렵고, 학습 데이터에 없던 새로운 환경이나 객체에 대한 일반화 능력이 제한적일 수 있다는 단점도 존재한다. 따라서, CLIP과 같이 대규모 데이터로 사전 학습된 모델을 활용하여 end-to-end 학습의 단점을 보완하고, unlabeled 객체에 대한 파지 성능을 향상시키려는 연구가 활발하게 이루어지고 있다.

## 5. 연구의 주요 기여점 및 향후 연구 방향

제안된 연구는 자연어 명령을 이해하고, 깊이 점군 데이터를 효과적으로 처리하여 unlabeled 객체를 파지하는 통합 시스템을 개발하는 데 핵심적인 기여를 할 수 있다. 특히, CLIP 모델을 활용하여 자연어 명령과 시각 정보를 연결하고, 이를 기반으로 다양한 제어 및 파지 전략을 탐색하는 것은 기존 연구와 차별화되는 중요한 측면이다. 깊이 점군 데이터를 직접 이용한 강화 학습, SAM과 같은 최신 분할 모델을 활용한 파지, 그리고 COM 외의 시각 정보를 기반으로 하는 파지 전략 등 다양한 접근 방식을 비교 분석하고 실험적으로 검증하는 것은 학문적 가치와 실용적 의미를 모두 가진다.

본 연구를 더욱 심화시키기 위해 다음과 같은 추가적인 질문과 아이디어를 고려해 볼 수 있다. 첫째, CLIP 모델의 출력을 어떻게 하면 로봇 팔의 구체적인 움직임 (예: 조인트 각도 변화, 엔드 이펙터 속도)으로 효과적으로 변환할 수 있을까? 둘째, 다양한 형태와 재질을 가진 unlabeled 객체에 대한 파지 성공률을 더욱 높이기 위한 데이터 증강 기법 및 학습 전략은 무엇이 있을까? 셋째, 실제 로봇 작업 환경에서 발생할 수 있는 센서 노이즈, 객체의 미끄러짐, 예기치 않은 충돌 등의 불확실성을 고려하여, 보다 robust한 파지 시스템을 구축하기 위한 방법은 무엇일까?

본 연구와 관련된 최신 연구 동향을 살펴보면, Robotic-CLIP<sup>2</sup> 및 CLIP-RT<sup>3</sup>와 같이 CLIP 모델을 특정 로봇 작업에 맞게 미세 조정하거나, CLIP의 표현력을 활용하여 로봇의 행동 정책을 학습하는 연구가 활발히 진행되고 있다. 또한, Grasp-Anything<sup>68</sup> 데이터셋과 같이 자연어 명령 기반의 6DoF 파지 작업을 위한 대규모 데이터셋 구축 노력도 주목할 만하다. End-to-end 학습<sup>14</sup>과 강화 학습 기반 파지<sup>3</sup> 역시 지속적으로 발전하고 있으며, 특히 CLIP과 같은 사전 학습된 모델을 이러한 학습 방식에 통합하여 성능을 향상시키려는 시도가 늘어나고 있다.

## 6. 결론

본 보고서는 자연어 명령 기반 unlabeled 객체 파지라는 사용자님의 연구 아이디어에

대한 심층적인 분석을 수행하였다. 자연어 이해와 시각 정보 처리에 강력한 성능을 보이는 CLIP 모델과, 3차원 공간 정보를 제공하는 깊이 점군 데이터는 사용자님의 연구 목표를 달성하기 위한 핵심적인 도구가 될 수 있음을 확인하였다. 보고서에서는 CLIP 모델의 기본 원리 및 로봇 팔 파지 작업에의 적용 가능성을 분석하고, CLIP 기반 인식 정보를 활용한 다양한 로봇 팔 제어 및 파지 전략 (시각 서보잉, 강화 학습, 3D 포즈 추정)을 비교 분석하였다. 또한, 깊이 점군 데이터를 이용한 인식 및 파지 방법으로 segmentation 기반 방식과 직접 강화 학습 방식의 특징과 장단점을 비교하고, SAM과 같은 최신 분할 모델 활용의 이점을 제시하였다. 더불어, 깊이 카메라를 이용한 무게 중심 (COM) 추정 알고리즘과 COM 외에 시각 정보를 활용한 파지 대안 (3D 바운딩 박스, Grasp Pose Detection, End-to-end 학습)을 상세히 논의하였다. 마지막으로, 본 연구의 주요 기여 가능성을 평가하고, 향후 연구 방향 및 관련된 최신 연구 동향을 제시함으로써, 사용자님의 연구 주제 설정 및 심화에 실질적인 도움을 제공하고자 노력하였다.

향후 연구를 진행하면서, 사용자님께서 본 보고서에서 제시된 다양한 알고리즘 및 방법론들을 심층적으로 비교 분석하고, 실제 로봇 환경에서의 실험을 통해 각 접근 방식의 성능을 검증하는 것이 중요할 것이다. 또한, 인공지능 및 로봇틱스 분야의 최신 연구 동향을 지속적으로 주시하고, 사용자님의 구체적인 연구 목표와 환경에 가장 적합한 접근 방식을 탐색해 나가시기를 바란다.

## 참고 자료

1. Grasping Any Object with Robotic Arms with Language Instructions, 4월 25, 2025에 액세스, <https://courses.grainger.illinois.edu/ece445zjui/getfile.asp?id=24452>
2. Robotic-CLIP: Fine-tuning CLIP on Action Data for Robotic Applications - arXiv, 4월 25, 2025에 액세스, <https://arxiv.org/html/2409.17727v1>
3. Review of Learning-Based Robotic Manipulation in Cluttered Environments - PMC, 4월 25, 2025에 액세스, <https://pmc.ncbi.nlm.nih.gov/articles/PMC9607868/>
4. (PDF) Attention-Guided Integration of CLIP and SAM for Precise ..., 4월 25, 2025에 액세스, [https://www.researchgate.net/publication/389392146\\_Attention-Guided\\_Integration\\_of\\_CLIP\\_and\\_SAM\\_for\\_Precise\\_Object\\_Masking\\_in\\_Robotic\\_Manipulation](https://www.researchgate.net/publication/389392146_Attention-Guided_Integration_of_CLIP_and_SAM_for_Precise_Object_Masking_in_Robotic_Manipulation)
5. Attention-Guided Integration of CLIP and SAM for Precise Object Masking in Robotic Manipulation - arXiv, 4월 25, 2025에 액세스, <https://arxiv.org/pdf/2502.18842?>
6. Fsoft-AIC/RoboticCLIP: [ICRA 2025] Robotic-CLIP: Fine ... - GitHub, 4월 25, 2025에 액세스, <https://github.com/Fsoft-AIC/RoboticCLIP>
7. CLIP-RT : Learning Language-Conditioned Robotic Policies from ..., 4월 25, 2025에 액세스, <https://clip-rt.github.io/>
8. (PDF) CLIP-RT: Learning Language-Conditioned Robotic Policies from Natural Language Supervision - ResearchGate, 4월 25, 2025에 액세스, [https://www.researchgate.net/publication/385510350\\_CLIP-RT\\_Learning\\_Language-Conditioned\\_Robotic\\_Policies\\_from\\_Natural\\_Language\\_Supervision](https://www.researchgate.net/publication/385510350_CLIP-RT_Learning_Language-Conditioned_Robotic_Policies_from_Natural_Language_Supervision)

9. [2502.18842] Attention-Guided Integration of CLIP and SAM for Precise Object Masking in Robotic Manipulation - arXiv, 4월 25, 2025에 액세스, <https://arxiv.org/abs/2502.18842>
10. Attention-Guided Integration of CLIP and SAM for Precise Object Masking in Robotic Manipulation | Request PDF - ResearchGate, 4월 25, 2025에 액세스, [https://www.researchgate.net/publication/388945661\\_Attention-Guided\\_Integration\\_of\\_CLIP\\_and\\_SAM\\_for\\_Precise\\_Object\\_Masking\\_in\\_Robotic\\_Manipulation](https://www.researchgate.net/publication/388945661_Attention-Guided_Integration_of_CLIP_and_SAM_for_Precise_Object_Masking_in_Robotic_Manipulation)
11. Robot Closed-Loop Grasping Based on Deep Visual Servoing ..., 4월 25, 2025에 액세스, <https://www.mdpi.com/2076-0825/14/1/25>
12. h2t.iar.kit.edu, 4월 25, 2025에 액세스, <https://h2t.iar.kit.edu/pdf/Vahrenkamp2008a.pdf>
13. AniruthSuresh/Real-Time-Visual-Servoing-Grasping: This ... - GitHub, 4월 25, 2025에 액세스, <https://github.com/AniruthSuresh/Real-Time-Visual-Servoing-Grasping>
14. Vision-Based Robotic Object Grasping—A Deep Reinforcement Learning Approach - MDPI, 4월 25, 2025에 액세스, <https://www.mdpi.com/2075-1702/11/2/275>
15. Review of Reinforcement Learning for Robotic Grasping: Analysis and Recommendations - Semantic Scholar, 4월 25, 2025에 액세스, <https://pdfs.semanticscholar.org/ed1c/6b6eb9b648cabd4f7ee3ecdaf4fc459e0483.pdf>
16. Review of Reinforcement Learning for Robotic Grasping: Analysis and Recommendations, 4월 25, 2025에 액세스, <http://www.iapress.org/index.php/soic/article/view/1797>
17. Learning to Grasp Objects with Reinforcement Learning - Bernoulli Institute for Mathematics, Computer Science and Artificial Intelligence, 4월 25, 2025에 액세스, [https://www.ai.rug.nl/~mwiering/Master\\_Thesis\\_Rik\\_Timmers.pdf](https://www.ai.rug.nl/~mwiering/Master_Thesis_Rik_Timmers.pdf)
18. GeorgeDu/vision-based-robotic-grasping: Related papers and codes for vision-based robotic grasping - GitHub, 4월 25, 2025에 액세스, <https://github.com/GeorgeDu/vision-based-robotic-grasping>
19. Automatic 3D Object Recognition and Localization for Robotic Grasping - SciTePress, 4월 25, 2025에 액세스, <https://www.scitepress.org/PublishedPapers/2021/105527/105527.pdf>
20. 3D Object Detection and Recognition for Robotic Grasping Based on RGB-D Images and Global Features - ResearchGate, 4월 25, 2025에 액세스, [https://www.researchgate.net/publication/319619008\\_3D\\_Object\\_Detection\\_and\\_Recognition\\_for\\_Robotic\\_Grasping\\_Based\\_on\\_RGB-D\\_Images\\_and\\_Global\\_Features](https://www.researchgate.net/publication/319619008_3D_Object_Detection_and_Recognition_for_Robotic_Grasping_Based_on_RGB-D_Images_and_Global_Features)
21. Robot Instance Segmentation with Few Annotations for Grasping - CVF Open Access, 4월 25, 2025에 액세스, [https://openaccess.thecvf.com/content/WACV2025/papers/Kimhi\\_Robot\\_Instance\\_Segmentation\\_with\\_Few\\_Annotations\\_for\\_Grasping\\_WACV\\_2025\\_paper.pdf](https://openaccess.thecvf.com/content/WACV2025/papers/Kimhi_Robot_Instance_Segmentation_with_Few_Annotations_for_Grasping_WACV_2025_paper.pdf)
22. Robotic Grasping of Unknown Objects Based on Deep Learning-Based Feature Detection, 4월 25, 2025에 액세스, <https://pmc.ncbi.nlm.nih.gov/articles/PMC11314913/>

23. Robot Instance Segmentation with Few Annotations for Grasping - arXiv, 4월 25, 2025에 액세스, <https://arxiv.org/html/2407.01302>
24. Grasping Novel Objects with Depth Segmentation - Stanford AI Lab, 4월 25, 2025에 액세스, <https://ai.stanford.edu/~ang/papers/iros10-GraspingWithDepthSegmentation.pdf>
25. Goal-Auxiliary Actor-Critic for 6D Robotic Grasping with Point Clouds - Proceedings of Machine Learning Research, 4월 25, 2025에 액세스, <https://proceedings.mlr.press/v164/wang22a/wang22a.pdf>
26. Active search and coverage using point-cloud reinforcement learning - arXiv, 4월 25, 2025에 액세스, <https://arxiv.org/pdf/2312.11410>
27. Experiments with Hierarchical Reinforcement Learning of Multiple Grasping Policies - IAS TU Darmstadt, 4월 25, 2025에 액세스, [https://www.ias.tu-darmstadt.de/uploads/Site/EditPublication/osa\\_ISER2016.pdf](https://www.ias.tu-darmstadt.de/uploads/Site/EditPublication/osa_ISER2016.pdf)
28. [1906.08989] Data-Efficient Learning for Sim-to-Real Robotic Grasping using Deep Point Cloud Prediction Networks - arXiv, 4월 25, 2025에 액세스, <https://arxiv.labs.arxiv.org/html/1906.08989>
29. Robotics Dexterous Grasping: The Methods Based on Point Cloud and Deep Learning, 4월 25, 2025에 액세스, <https://pmc.ncbi.nlm.nih.gov/articles/PMC8221534/>
30. A Fast 6DOF Visual Selective Grasping System Using Point Clouds - MDPI, 4월 25, 2025에 액세스, <https://www.mdpi.com/2075-1702/11/5/540>
31. Robotics Dexterous Grasping: The Methods Based on Point Cloud and Deep Learning, 4월 25, 2025에 액세스, <https://www.frontiersin.org/journals/neurorobotics/articles/10.3389/fnbot.2021.658280/full>
32. Data-Efficient Learning for Sim-to-Real Robotic Grasping using Deep Point Cloud Prediction Networks - ResearchGate, 4월 25, 2025에 액세스, [https://www.researchgate.net/publication/333971857\\_Data-Efficient\\_Learning\\_for\\_Sim-to-Real\\_Robotic\\_Grasping\\_using\\_Deep\\_Point\\_Cloud\\_Prediction\\_Networks](https://www.researchgate.net/publication/333971857_Data-Efficient_Learning_for_Sim-to-Real_Robotic_Grasping_using_Deep_Point_Cloud_Prediction_Networks)
33. Robotic Grasping based on Point Cloud Recognition and Deep Reinforcement Learning | Request PDF - ResearchGate, 4월 25, 2025에 액세스, [https://www.researchgate.net/publication/357319985\\_Robotic\\_Grasping\\_based\\_on\\_Point\\_Cloud\\_Recognition\\_and\\_Deep\\_Reinforcement\\_Learning](https://www.researchgate.net/publication/357319985_Robotic_Grasping_based_on_Point_Cloud_Recognition_and_Deep_Reinforcement_Learning)
34. Robotic Grasping based on Point Cloud Recognition and Deep, 4월 25, 2025에 액세스, [https://www.jstage.jst.go.jp/article/jsmermd/2021/0/2021\\_1A1-E14/\\_article/-char/en](https://www.jstage.jst.go.jp/article/jsmermd/2021/0/2021_1A1-E14/_article/-char/en)
35. On the Efficacy of 3D Point Cloud Reinforcement Learning - arXiv, 4월 25, 2025에 액세스, <https://arxiv.org/html/2306.06799>
36. FULL PAPER Hierarchical Reinforcement Learning of Multiple Grasping Strategies with Human Instructions - IAS TU Darmstadt, 4월 25, 2025에 액세스, [https://www.ias.informatik.tu-darmstadt.de/uploads/Publications/Publications/advanced\\_robotics\\_18osa.pdf](https://www.ias.informatik.tu-darmstadt.de/uploads/Publications/Publications/advanced_robotics_18osa.pdf)
37. Learning an End-To-End Spatial Grasp Generation Algorithm from Sparse Point Clouds - arXiv, 4월 25, 2025에 액세스, <https://arxiv.org/pdf/2003.09644>

38. Combining Teacher-Augmented Policy Gradient Learning with Instance Segmentation to Grasp Arbitrary Objects - arXiv, 4월 25, 2025에 액세스, <https://arxiv.org/html/2403.10187v1>
39. Finding the virtual center of a cloud of points. - Mathematics Stack Exchange, 4월 25, 2025에 액세스, <https://math.stackexchange.com/questions/195729/finding-the-virtual-center-of-a-cloud-of-points>
40. open3d.geometry.PointCloud - Open3D 0.19.0 documentation, 4월 25, 2025에 액세스, [https://www.open3d.org/docs/release/python\\_api/open3d.geometry.PointCloud.html](https://www.open3d.org/docs/release/python_api/open3d.geometry.PointCloud.html)
41. Calculating statistics on a point cloud or depth map, 4월 25, 2025에 액세스, [https://aurorahelp.zebra.com/aurorail/MILXSP4/MILXSP4HelpFiles/extfile/z\\_MIL\\_UG\\_3D\\_Image\\_processing/UserGuide/3D\\_Image\\_processing/Calculating\\_statistics\\_on\\_a\\_point\\_cloud\\_or\\_depth\\_map.htm](https://aurorahelp.zebra.com/aurorail/MILXSP4/MILXSP4HelpFiles/extfile/z_MIL_UG_3D_Image_processing/UserGuide/3D_Image_processing/Calculating_statistics_on_a_point_cloud_or_depth_map.htm)
42. Method for measuring the center of mass and moment of inertia of a model using 3D point clouds - Optica Publishing Group, 4월 25, 2025에 액세스, [https://opg.optica.org/ao/upcoming\\_pdf.cfm?id=470899](https://opg.optica.org/ao/upcoming_pdf.cfm?id=470899)
43. Method for measuring the center of mass and moment of inertia of a model using 3D point clouds - ResearchGate, 4월 25, 2025에 액세스, [https://www.researchgate.net/publication/365471490\\_Method\\_for\\_measuring\\_the\\_center\\_of\\_mass\\_and\\_moment\\_of\\_inertia\\_of\\_a\\_model\\_using\\_3D\\_point\\_clouds](https://www.researchgate.net/publication/365471490_Method_for_measuring_the_center_of_mass_and_moment_of_inertia_of_a_model_using_3D_point_clouds)
44. 3D Perception/Configuration Tutorial — moveit\_tutorials Indigo documentation - ROS 2, 4월 25, 2025에 액세스, [http://docs.ros.org/indigo/api/moveit\\_tutorials/html/doc/pr2\\_tutorials/planning/src/doc/perception\\_configuration.html](http://docs.ros.org/indigo/api/moveit_tutorials/html/doc/pr2_tutorials/planning/src/doc/perception_configuration.html)
45. Perception Pipeline Tutorial — MoveIt Documentation: Humble documentation, 4월 25, 2025에 액세스, [https://moveit.picknik.ai/humble/doc/examples/perception\\_pipeline/perception\\_pipeline\\_tutorial.html](https://moveit.picknik.ai/humble/doc/examples/perception_pipeline/perception_pipeline_tutorial.html)
46. Perception Pipeline Tutorial — MoveIt Documentation - PickNik Robotics, 4월 25, 2025에 액세스, [https://moveit.picknik.ai/main/doc/examples/perception\\_pipeline/perception\\_pipeline\\_tutorial.html](https://moveit.picknik.ai/main/doc/examples/perception_pipeline/perception_pipeline_tutorial.html)
47. moveit\_tutorials/doc/perception\_pipeline/perception\_pipeline\_tutorial.rst at master - GitHub, 4월 25, 2025에 액세스, [https://github.com/ros-planning/moveit\\_tutorials/blob/master/doc/perception\\_pipeline/perception\\_pipeline\\_tutorial.rst](https://github.com/ros-planning/moveit_tutorials/blob/master/doc/perception_pipeline/perception_pipeline_tutorial.rst)
48. Robotic Grasping Detection Algorithm Based on 3D Vision Dual-Stream Encoding Strategy, 4월 25, 2025에 액세스, <https://www.mdpi.com/2079-9292/13/22/4432>
49. Grasp Pose Detection in Point Clouds - Northeastern University, 4월 25, 2025에 액세스, <http://www.ccs.neu.edu/home/mgualti/2017-tenPas-GraspPoseDetectionInPointClouds.pdf>
50. [2502.16976] Task-Oriented 6-DoF Grasp Pose Detection in Clutters - arXiv, 4월



- 25, 2025에 액세스, <https://arxiv.org/abs/2502.16976>
51. Physics-Based Self-Supervised Grasp Pose Detection - MDPI, 4월 25, 2025에 액세스, <https://www.mdpi.com/2075-1702/13/1/12>
  52. LieGrasPFormer: Point Transformer-based 6-DOF Grasp Detection with Lie Algebra Grasp Representation - mediaTUM, 4월 25, 2025에 액세스, <https://mediatum.ub.tum.de/doc/1720412/d0rg0kp16l5shyo6ag683v2hi.lieGrasPFormer.pdf>
  53. 6D Pose Estimation of Industrial Parts Based on Point Cloud Geometric Information Prediction for Robotic Grasping - MDPI, 4월 25, 2025에 액세스, <https://www.mdpi.com/1099-4300/26/12/1022>
  54. A Graph-Based SE(3)-invariant Approach to Grasp Detection | website for Edge Grasp Network - Haojie Huang, 4월 25, 2025에 액세스, [https://haojhuang.github.io/edge\\_grasp\\_page/](https://haojhuang.github.io/edge_grasp_page/)
  55. Using Geometry to Detect Grasp Poses in 3D Point Clouds - Khoury College of Computer Sciences, 4월 25, 2025에 액세스, [http://www.ccs.neu.edu/home/atp/publications/grasp\\_poses\\_isrr2015.pdf](http://www.ccs.neu.edu/home/atp/publications/grasp_poses_isrr2015.pdf)
  56. atepas/gpd: Detect 6-DOF grasp poses in point clouds - GitHub, 4월 25, 2025에 액세스, <https://github.com/atepas/gpd>
  57. yudhisteer/Robotic-Grasping-Detection-with-PointNet - GitHub, 4월 25, 2025에 액세스, <https://github.com/yudhisteer/Robotic-Grasping-Detection-with-PointNet>
  58. End-to-End Learning of Semantic Grasping - Proceedings of Machine Learning Research, 4월 25, 2025에 액세스, <http://proceedings.mlr.press/v78/jang17a/jang17a.pdf>
  59. Neural End-to-End Learning of Reach for Grasp Ability with a 6-DoF Robot Arm, 4월 25, 2025에 액세스, [https://personalrobotics.cs.washington.edu/workshops/mlmp2018/assets/docs/19\\_CameraReadySubmission\\_Neural\\_End\\_to\\_End\\_Learning\\_of\\_Reach\\_for\\_Grasp\\_Ability\\_with\\_a\\_6\\_DoF\\_Robot\\_Arm\\_Using\\_Augmented\\_Reality\\_Training\\_Final.pdf](https://personalrobotics.cs.washington.edu/workshops/mlmp2018/assets/docs/19_CameraReadySubmission_Neural_End_to_End_Learning_of_Reach_for_Grasp_Ability_with_a_6_DoF_Robot_Arm_Using_Augmented_Reality_Training_Final.pdf)
  60. Learning End-to-End 6DoF Grasp Choice of Human-to-Robot Handover using Affordance Prediction and Deep Reinforcement Learning, 4월 25, 2025에 액세스, [https://huangjuite.github.io/socially\\_aware\\_handover.pdf](https://huangjuite.github.io/socially_aware_handover.pdf)
  61. End-to-End Learning of Semantic Grasping - Google Research, 4월 25, 2025에 액세스, <https://research.google/pubs/end-to-end-learning-of-semantic-grasping/>
  62. Learning to Grasp and Regrasp using Vision and Touch, 4월 25, 2025에 액세스, <https://sites.google.com/view/more-than-a-feeling>
  63. Robotic Grasping of Novel Objects using Vision, 4월 25, 2025에 액세스, <http://graphics.cs.cmu.edu/nsp/course/16899-s10/papers/visionAndGrasping/visionAndGraspingNGShapelJRR.pdf>
  64. A Brief Survey on Leveraging Large Scale Vision Models for Enhanced Robot Grasping, 4월 25, 2025에 액세스, <https://arxiv.org/html/2406.11786v1>
  65. [2009.12606] Grasp Proposal Networks: An End-to-End Solution for Visual Learning of Robotic Grasps - arXiv, 4월 25, 2025에 액세스, <https://arxiv.org/abs/2009.12606>
  66. CLIPort, 4월 25, 2025에 액세스, <https://cliport.github.io/>
  67. CLIPORT: What and Where Pathways for Robotic Manipulation - Proceedings of



Machine Learning Research, 4월 25, 2025에 액세스,

<https://proceedings.mlr.press/v164/shridhar22a/shridhar22a.pdf>

68. Language-Driven 6-DoF Grasp Detection Using Negative Prompt Guidance - European Computer Vision Association, 4월 25, 2025에 액세스,

[https://www.ecva.net/papers/eccv\\_2024/papers\\_ECCV/papers/02933.pdf](https://www.ecva.net/papers/eccv_2024/papers_ECCV/papers/02933.pdf)

69. (PDF) Development of an Algorithm for Volumetric Reconstruction and Estimation of the Center of Mass of Solid Cohesive in Environments with Suspended Particles - ResearchGate, 4월 25, 2025에 액세스,

[https://www.researchgate.net/publication/376832927\\_Development\\_of\\_an\\_Algorithm\\_for\\_Volumetric\\_Reconstruction\\_and\\_Estimation\\_of\\_the\\_Center\\_of\\_Mass\\_of\\_Solid\\_Cohesive\\_in\\_Environments\\_with\\_Suspended\\_Particles](https://www.researchgate.net/publication/376832927_Development_of_an_Algorithm_for_Volumetric_Reconstruction_and_Estimation_of_the_Center_of_Mass_of_Solid_Cohesive_in_Environments_with_Suspended_Particles)

70. Research on the center of mass calculation method for dynamic tilting of terracotta warriors from the mausoleum of Emperor Qin Shihuang - SPIE Digital Library, 4월 25, 2025에 액세스,

<https://www.spiedigitallibrary.org/conference-proceedings-of-spie/13549/135491W/Research-on-the-center-of-mass-calculation-method-for-dynamic/10.1117/12.3058807.pdf>