

Automated Pallet Pick-and-Place Manipulation Using Physics Based Simulation in NVIDIA Isaac Sim

School of Mechanical and Control Engineering
Handong Global University

Joungbin Choi

Automated Pallet Pick-and-Place Manipulation Using Physics Based Simulation in NVIDIA Isaac Sim

A Bachelor's Thesis

Submitted to the School of
Mechanical and Control Engineering of
Handong Global University

Joungbin Choi

December 2025

This certifies that the bachelor's thesis is approved.

Thesis Advisor: Ph.D. YoungKeun Kim

The Dean of Faculty: Ph.D. WonSang Ra

School of Mechanical and Control Engineering

Handong Global University

December 2025

Extended Abstract

Automated Pallet Pick-and-Place Manipulation Using Physics Based Simulation in NVIDIA Isaac Sim

In this study, we propose a palette automatic pick-and-place system that combines RGB-D geometric pose estimation techniques and deep object pose estimation (DOPE), a deep learning-based pose estimation model, in a physical simulation environment based on NVIDIA Isaac Sim. Pallets are widely used throughout logistics tasks in industrial sites, but in actual working environments, pallets are often not always arranged. This state of alignment not only degrades work efficiency but also requires manual intervention by workers, which increases the risk of safety accidents. Accordingly, the need for an automated system that can stably recognize and operate pallets even under various environmental conditions is constantly being raised. However, the process of building a large-scale real-world environmental dataset for palette pose estimation is time-consuming and expensive, and there are safety limitations during the experimental process.

To solve this problem, Isaac Sim is utilized to design and implement a full pick-and-place pipeline from recognition to robot manipulation with PhysX and high quality rendering in simulation environment. Two different approaches were applied to pose estimation of the palette. The first is a geometric pose estimation technique using a point cloud generated from an RGB-D camera, and the second is a DOPE deep learning-based pose estimation technique learned using label data generated automatically in a simulation environment.

In the geometric pose estimation technique, a color image and a depth image are simultaneously acquired using a virtual RGB-D camera placed on the top of the palette. A palette is detected by applying a YOLO-based object detection algorithm to the obtained RGB image, and the corresponding region is set as a Region of Interest (ROI). After that, depth information in the ROI is projected into a 3D space using the camera internal parameters, thereby creating a point cloud based on the camera coordinate system. The generated point cloud is converted into a world coordinate system and used for robot operation. DBSCAN-based clustering was applied to remove background area and noise, and the cluster including the most points was selected as the palette area. After that, principal component analysis (PCA) was used on the cluster to estimate the dominant geometric direction of the palette. Based on the main axis information obtained through PCA, the normal vector on the side of the pallet was calculated, and through this, a pick pose that could allow side collection of the adsorption method was calculated. The final calculated rotation matrix is converted into a quaternion form. The converting is for using it in the Isaac Sim environment and expressed as a six-degree of freedom pick pose.

In the deep learning-based pose estimation technique, pallets are randomly placed in various positions and directions in the Isaac Sim environment. To be a robust model in different environments the background and lighting conditions are also randomly changed. In this process, the keypoints and 6 degrees of freedom pose information of the pallets are automatically labeled and stored as a dataset. The DOPE model was trained using the generated synthetic dataset. The learned model receives a single RGB image as input to estimate the overall 6 degrees of freedom pose of the pallet. DOPE gradually refines the keypoint reliability map and the affinity field through a multi-stage network structure. After that, calculates the final pose by applying the PnP algorithm based on the extracted keypoints.

The pallet pose estimated through both methods is input to the Pick-and-Place controller and the RMPFlow-based motion controller and utilized for robot manipulation. The robot performs pallet access, side adsorption collector, lateral movement, and loading operations sequentially, and the entire manipulation process is executed stably and repeatably in a simulation environment. As a result of the simulation experiment, RGB-D based geometric pose estimation technique showed reliable pallet pose estimation performance without relying on large-scale training data.

In order to confirm the applicability of the verified algorithm in the simulation environment to the real environment, a real pallet Pick-and-Place experiment using the PiPER robot arm was conducted. As a result of applying the RGB-D-based geometric technique and the DOPE-based technique used in the simulation to the real environment under the same conditions, it was confirmed that the pallet recognition and manipulation performance were implemented at a level similar to the simulation results.

This research develops an autonomous pallet handling system using physics-based simulation to compare geometric pose estimation with deep learning-based methods. The system is tested in both simulation and real environments to evaluate its practical performance. Through these experiments this study shows that synthetic data generated in simulation can be effectively applied to real robotic pallet handling tasks in industrial environments.

차 례

| | |
|----------------------------------|----|
| Extended Abstract | i |
| I. 서론 | 5 |
| 1.1. 연구목표 | 5 |
| 1.2. 연구 필요성 | 5 |
| 1.3 연구 내용 및 구성 | 6 |
| II. 연구방법 | 7 |
| 2.1. 문제 정의 및 접근 방법 | 7 |
| 2.2. Isaac Sim 기반 시뮬레이션 환경 | 7 |
| 2.3 인지(자세 추정) | 8 |
| 2.3.1 RGB-D 기하학적 포즈 추정 기법 | 8 |
| 2.3.2 DOPE 딥러닝 기반 포즈 추정 기법 | 10 |
| 2.4 구동 실험 | 13 |
| 2.4.1 시뮬레이션 환경에서의 구동 실험 | 13 |
| 2.4.2 실제 로봇 환경에서의 구동 실험 | 13 |
| III. 실험결과 | 15 |
| 3.1 실험 환경 및 평가 기준 | 15 |
| 3.2 RGB-D 기하학적 포즈 추정 결과 | 15 |
| 3.3 DOPE 딥러닝 기반 포즈 추정 결과 | 16 |

| | |
|---|----|
| 3.4 RGB-D 기반 기법과 DOPE 기법의 비교 | 17 |
| 3.5 실제 로봇 실험 결과 | 18 |
| 3.6 RGB-D 기반 기법과 DOPE 기법의 비교 및 상호 보완성 | 19 |
| 3.7 시뮬레이션-현실 전이(Sim-to-Real)에 대한 고찰 | 19 |
| 3.8 연구의 한계 및 향후 연구 방향 | 19 |
| IV. 결론 | 21 |
| 참고문헌 | 22 |

I. 서론

1.1. 연구목표

본 연구의 목표는 NVIDIA Isaac Sim 기반의 물리 시뮬레이션 환경에서 RGB-D 기반 기하학적 포즈 추정 기법과 딥러닝 기반 DOPE(Deep Object Pose Estimation) 포즈 추정 모델을 활용하여 팔레트 자동 Pick-and-Place 시스템을 설계하고 구현하는 것이다. 특히 산업 현장에서 빈번하게 발생하는 팔레트 정렬 문제를 해결하고자 한다. 이를 위해 다양한 환경 조건에서도 신뢰성 있게 팔레트의 자세를 인식하고 이를 기반으로 안정적인 로봇 조작이 가능한 통합 시스템을 구축하는 것을 목표로 한다.

본 연구에서는 시뮬레이션 환경에서 팔레트 인식부터 로봇 조작까지 이어지는 전체 파이프라인을 구현한다. 팔레트 인식을 위해선 RGB-D 기반 기하학적 포즈 추정 기법과 딥러닝 기반 포즈 추정 기법을 각각 적용하여 그 특성과 성능을 비교 분석한다. 또한 시뮬레이션 환경에서 검증된 알고리즘을 실제 로봇 시스템에 적용함으로써 시뮬레이션 기반 연구의 실제 환경에서의 적용 가능성을 확인하고자 한다.

1.2 연구 배경 및 필요성

물류 및 제조 산업 현장에서는 팔레트를 이용한 물품 적재와 운반 작업이 작업 공정 전반에 걸쳐 널리 활용되고 있다. 그러나 실제 작업 환경에서는 팔레트가 항상 일정한 방향이나 위치로 정렬되어 배치되지 않는 경우가 많다. 이러한 불규칙한 상태의 팔레트는 작업 효율을 저하시킬 뿐만 아니라 작업자가 직접 팔레트를 정렬하는 과정에서 안전 사고로 이어질 가능성도 존재한다.

팔레트를 수작업으로 정렬하는 과정은 반복적인 작업을 요구하며 장시간 작업 시 작업자의 피로도를 증가시키는 요인이 된다. 또한 작업 중 발생할 수 있는 충돌이나 부주의로 인해 산업 안전 측면에서도 위험 요소로 작용할 수 있다. 이와 같은 이유로 팔레트 작업 뿐만 아니라 다양한 산업 현장에서 물품 배치와 이송 작업의 자동화에 대한 요구가 점차 증가하고 있다. 팔레트 작업을 자동화하기 위해서는 로봇이 팔레트의 위치 뿐만 아니라 방향까지 정확하게 인식하고 이를 기반으로 안정적인 조작을 수행할 수 있어야 한

다.

포즈 추정을 위한 기존 연구에서는 주로 딥러닝 기반 기법이나 RGB-D 센서를 활용한 기하학적 접근 방식이 사용되어 왔다. 딥러닝 기반 기법은 복잡한 환경에서도 비교적 높은 인식 성능을 보이지만 모델 학습을 위해 대규모 데이터셋이 필요하며 데이터 수집과 라벨링 과정에 많은 시간과 비용이 소요된다는 한계가 있다.

반면 RGB-D 센서를 활용한 기하학적 방식은 깊이 정보를 직접 이용하여 물체의 3차원 구조를 복원할 수 있어 해석 가능성과 안정성이 높은 장점을 가진다. 그러나 센서 노이즈나 주변 환경 조건에 따라 성능이 저하될 수 있다. 더불어 실제 환경에서 로봇 실험을 반복적으로 수행하는 과정은 장비 비용과 안전 문제로 인해 현실적인 제약이 따른다.

이러한 한계를 보완하기 위한 대안으로 물리 기반 시뮬레이션 환경을 활용한 연구가 주목받고 있다. NVIDIA Isaac Sim은 고해상도 렌더링과 정확한 물리 엔진을 기반으로 RGB-D 센서 시뮬레이션을 포함한 실제 환경과 유사한 조건을 제공한다. 이를 통해 실제 로봇 실험 이전 단계에서 인식 및 조작 알고리즘을 효과적으로 개발하고 검증할 수 있다.

1.3 연구 내용 및 구성

본 연구에서는 팔레트 자동 Pick-and-Place 작업을 수행하기 위한 인식 및 조작 시스템을 구현하고 이를 시뮬레이션 환경과 실제 로봇 환경에서 단계적으로 검증한다. 팔레트 인식을 위해 두 가지 서로 다른 포즈 추정 기법을 적용하였다. 첫 번째 기법은 RGB-D 카메라 기반 기하학적 구조와 자세를 추정하는 방법이다. 두 번째 기법은 시뮬레이션 환경에서 자동 라벨링을 통해 생성된 데이터셋을 이용하여 학습한 DOPE 딥러닝 기반 포즈 추정 모델을 활용하는 방법이다.

시뮬레이션 환경에서 다양한 팔레트 배치 조건에 대한 실험을 수행하여 각 기법의 성능과 특성을 비교 분석하고 이후 PiPER 로봇 팔을 이용한 실제 Pick-and-Place 실험을 통해 시뮬레이션 기반 연구의 실제 환경에서의 적용 가능성을 검증한다.

II. 연구 방법

2.1 문제 정의 및 접근 방법

본 연구에서 다루는 문제는 산업 현장에서 무분별하게 배치된 팔레트의 자세를 인식하고 이를 기반으로 로봇이 팔레트를 안정적으로 집어 옮기는 자동 Pick-and-Place 작업이다. 팔레트는 구조적으로 직교 평면으로 이루어진 단순한 형상을 가지지만 실제 작업 환경에서는 위치와 방향이 일정하지 않으며 주변 물체 조명 변화 부분 가림 현상 등으로 인해 인식 난이도가 높다.

특히 팔레트를 측면에서 흡착 방식으로 집기 위해서는 단순한 위치 정보 뿐만 아니라 접촉 가능한 평면의 방향과 법선 벡터를 포함한 정확한 육자유도 자세 정보가 요구된다. 이에 본 연구에서는 팔레트 포즈 추정을 위해 RGB-D 센서를 활용한 기하학적 접근 방식과 딥러닝 기반 DOPE 포즈 추정 모델이라는 두 가지 상이한 방법을 적용하였다. 두 기법은 각각 해석 가능성과 데이터 효율성 그리고 인식 강건성 측면에서 서로 다른 특성을 가지며 이를 비교 분석함으로써 각 접근 방식의 활용 가능성을 평가한다.

2.2 Isaac Sim 기반 시뮬레이션 환경

본 연구는 시뮬레이션 환경을 구축하고 설계된 알고리즘을 검증하는 절차를 거친다. 이유는 시뮬레이션 환경에서는 반복적인 실험을 진행할 때 안전성과 높은 반복성이 보장되기 때문이다. 실제 환경에서 설치하고 구현하는데에는 시간적 재정적 비용이 든다. 또한 검증되지 않은 알고리즘을 실행시키면 고가에 로봇 팔과 같은 기기들의 고장 가능성도 있기에 시뮬레이션 환경은 보다 많은 강점을 가지고 있다.

NVIDIA Isaac Sim은 실제 물리 특성을 반영한 PhysX 5라는 물리 엔진과 고해상도 물리 기반 렌더링 기능, 실시간 광선 및 경로 추적을 통한 포토리얼리즘 등 강력한 시뮬레이션 기술들을 제공한다. 또한 RGB-D 카메라와 같은 센서를 실제 환경과 유사한 방식으로 구성할 수 있어 인식 알고리즘 검증에 적합하다. 이미 내포된 물리 기법이 적용되고 여러 컨트롤러가 제공된 로봇들이 있기에 로봇 실험을 설계하고 시스템을 개발할 수 있다.

본 연구에서는 Isaac Sim 환경 내에 로봇 모델과 팔레트 객체 그리고 RGB-D 카메라를 배치하여 실험 환경을 구성하였다. 조명 조건과 배경 환경을 변경할 수 있도록 설정하여 다양한 작업 상황을 가정한 실험이 가능하도록 하였다. 이를 통해 시뮬레이션 환경이 실제 산업 환경을 일정 수준 모사할 수 있음을 확인하였다.

2.3 인지(자세 추정)

무작위 자세로 배치된 팔레트를 Pick-and-Place하기 위해선 자세를 인지하고 추정하는 단계가 필요하다. 본 연구에서는 인지 단계에서 두가지 자세 추정 기법을 적용하였다. 하나는 RGB-D 센서 정보를 활용한 기하학적 포즈 추정 기법이며 다른 하나는 딥러닝 기반 DOPE 포즈 추정 기법이다. 두 기법의 파이프라인은 다음 **Figure 1**과 같다.

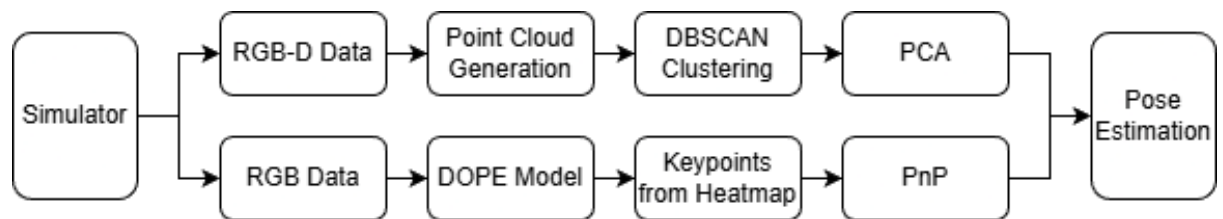


Figure 1. Pipeline of Two Methods

2.3.1 RGB-D 기하학적 포즈 추정 기법

RGB-D 기반 기하학적 포즈 추정 기법에서는 RGB-D 카메라를 팔레트 상부에 고정된 시점으로 배치하여 RGB 이미지와 깊이 영상을 동시에 획득한다. RGB 이미지에는 YOLO 기반 객체 검출 알고리즘[3]을 적용하여 팔레트를 검출하고 검출된 바운딩 박스를 관심 영역(Region of Interest, ROI)으로 설정한다. 이를 통해 불필요한 배경 영역을 제거하고 팔레트 영역에 대한 연산 효율을 향상시킨다.

팔레트 검출을 위해 본 연구에서는 Ultralytics에서 제안한 YOLOv8n 모델[4]을 사용하였다. 모델 학습에는 공개적으로 제공되는 오픈 데이터셋을 활용하였으며 팔레트 객체에 대한 바운딩 박스 정보를 기반으로 학습을 수행하였다. 모델의 학습에 사용된 주요 하이퍼파라미터는 **Table 1**에 정리하였다.

Table 1. Pallet Detecting Model Hyperparameter

| | |
|------------|------------------------|
| Model | YOLOv8n |
| Dataset | Open Dataset(Roboflow) |
| Epoch | 100 |
| Batch Size | 4 |
| Patience | 20 |

ROI 내의 깊이 영상은 카메라 내부 파라미터를 이용하여 3차원 공간으로 투영된다. 깊이 영상의 각 픽셀은 다음과 같은 관계식을 통해 카메라 좌표계의 3차원 점으로 변환된다. 다음은 픽셀(u, v)로 카메라 좌표계 상의 3차원 점 (X_c, Y_c, Z_c)로 변환 식이다. 여기서 f_x, f_y 는 초점 거리(focal length), c_x, c_y 는 주점(principal point)을 나타낸다.

$$X_c = \frac{(u - c_x) \cdot z}{f_x} \quad (1)$$

$$Y_c = \frac{(v - c_y) \cdot z}{f_y} \quad (2)$$

$$Z_c = z \quad (3)$$

생성된 포인트 클라우드는 이후 카메라 좌표계에서 월드 좌표계로 변환되어, 로봇 조작에 직접 활용 가능한 형태로 표현된다. 좌표계 변환 행렬은 다음과 같다.

$$\begin{bmatrix} X_w \\ Y_w \\ Z_w \\ 1 \end{bmatrix} = \begin{bmatrix} R & t \\ 0 & 1 \end{bmatrix} \begin{bmatrix} X_c \\ Y_c \\ Z_c \\ 1 \end{bmatrix} \quad (4)$$

생성된 포인트 클라우드는 바닥, 주변 물체, 센서 노이즈 등이 포함될 수 있다. 이러한 잡음을 제거하고 팔레트 구조만을 분리하기 위해 DBSCAN(Density-Based Spatial Clustering of Applications with Noise)[5] 알고리즘을 적용하였다. DBSCAN은 군집 개수를 사전에 정의할 필요가 없으며 밀도 기반으로 클러스터를 형성하기 때문에 불규칙한 구조를 가지는 포인트 클라우드 처리에 적합하다. DBSCAN 결과 중 가장 많은 포인트를 포함하는 클러스터를 팔레트 영역으로 선택하였다.

팔레트 클러스터에 포함된 포인트들에 대해 주성분 분석(Principal Component Analysis, PCA)[6]을 수행하여 팔레트의 지배적인 기하학적 방향을 추정한다. PCA를 통해 얻어진

고유벡터 중 가장 큰 고유값에 대응하는 벡터는 팔레트의 주축 방향을 나타내며 이를 기반으로 팔레트 측면의 법선 벡터를 계산한다.

주성분을 찾기 위해 포인트들의 공분산이 계산되어야 한다. μ 가 포인트 클라우드의 평균 벡터일 때 공분산 행렬 C 는 다음과 같이 계산된다.

$$x'_i = x_i - \mu \quad (5)$$

$$C = \frac{1}{N} (X - \mu)^T (X - \mu) \quad (6)$$

공분산 행렬 C 에 대해 다음과 같은 고유값 분해를 진행한다.

$$Cv_i = \lambda_i v_i \quad (7)$$

여기서 λ_i 는 고유값(eigenvalue)이며 v_i 는 이에 대응하는 고유벡터(eigenvector)이다. 분산이 클수록 고유값도 커지며 대응하는 고유벡터는 팔레트의 장축 방향이 된다.

계산된 법선 벡터와 팔레트 중심 위치를 이용하여 측면 Pick 지점을 산출하고 최종적으로 회전 행렬을 구성한다. 이 회전 행렬은 Isaac Sim 환경에서 사용 가능한 쿼터니언 형태로 변환되어 6-DoF Pick 포즈로 표현된다.

2.3.2 DOPE 딥러닝 기반 포즈 추정 기법

DOPE(Deep Object Pose Estimation)는 단일 RGB 영상으로부터 물체의 전체 6자유도(6-DoF) 자세를 추정하기 위한 딥러닝 기반 포즈 추정 프레임워크이다. DOPE는 Tremblay 등 [1]에 의해 제안된 방법으로 합성 데이터만을 이용한 학습을 통해 실제 환경에서도 높은 일반화 성능을 달성할 수 있음을 보였다. 본 연구에서는 RGB-D 기반 기하학적 포즈 추정 기법의 한계를 보완하고 깊이 정보 없이도 팔레트 자세를 추정할 수 있는 학습 기반 접근법으로 DOPE를 도입하였다.

DOPE는 물체의 자세를 직접 회귀 방식으로 예측하지 않고 이미지 상에서 주요 키 포인트의 위치를 먼저 추정한 뒤 기하학적 계산 과정을 통해 최종 포즈를 산출하는 구조를 가진다. 이러한 방식은 카메라 내부 파라미터 변화에 비교적 유연하게 대응할 수 있으며 다양한 환경 조건에서도 안정적인 포즈 추정을 가능하게 한다.

DOPE 기반 포즈 추정 기법은 여러 가지 장점이 있다. 첫 번째로 단일 RGB 이미지를 입력으로 사용하여 물체의 6자유도 자세를 추정할 수 있기에 깊이 센서 사용이 제한적인 환경에서도 적용이 가능하다. 두 번째로 시뮬레이션 환경에서 생성된 합성 데이터

를 이용해 학습할 수 있어 실제 환경에서 요구되는 대규모 데이터 수집과 수작업 라벨링에 따른 시간과 비용 부담을 줄일 수 있다. 세 번째로 다중 객체가 동시에 존재하는 장면에서도 각 객체의 키포인트를 분리하여 추정할 수 있어 산업 현장과 같이 복잡한 작업 환경에 적합하다.

DOPE 네트워크는 다단계 완전 합성곱 신경망 구조를 기반으로 하며 특징 추출부와 포즈 추정 헤드로 구성된다. 입력된 RGB 이미지는 VGG19 네트워크의 초기 계층을 거쳐 특징 맵으로 변환되며 이후 각 단계에서 키포인트 신뢰도 지도와 Affinity 필드를 예측한다.

신뢰도 지도는 물체의 3차원 바운딩 박스를 구성하는 각 꼭짓점과 물체 중심점이 이미지 상에 존재할 확률 분포를 나타낸다. Affinity 필드는 각 꼭짓점에서 물체 중심을 향하는 방향 벡터를 표현하며 이를 통해 동일한 객체에 속한 키포인트들을 효과적으로 묶을 수 있다. 네트워크는 여러 단계를 거치면서 이전 단계의 출력과 이미지 특징을 함께 사용하여 키포인트 추정 결과를 점진적으로 보정한다.

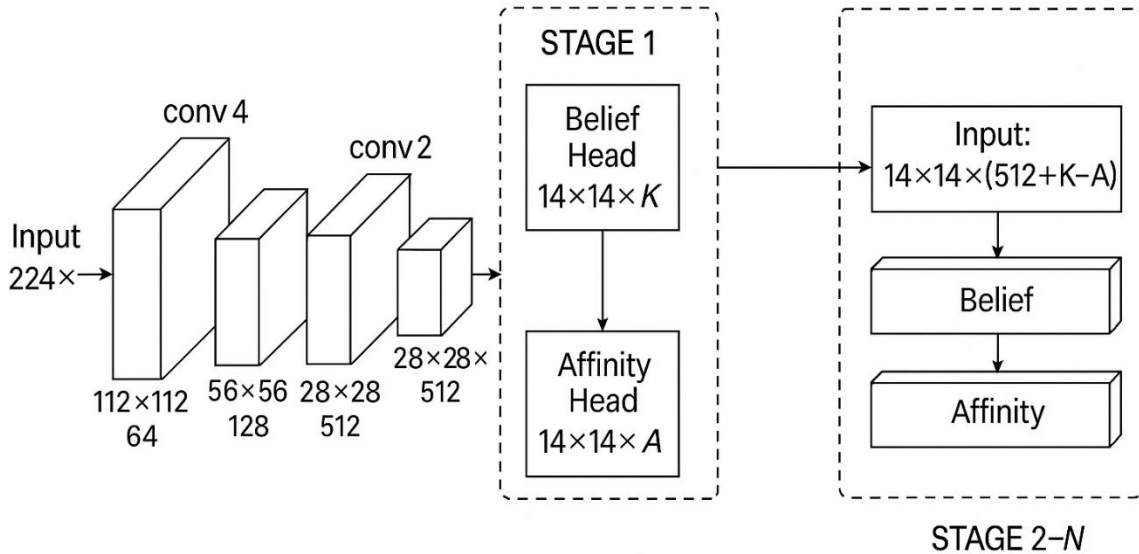


Figure 2. DOPE Network Structure

네트워크 추론 결과로부터 신뢰도 지도에서 임계값 이상의 국소 최대값(local maxima)을 탐색하여 키포인트 후보를 추출한다. 이후 Affinity 필드를 이용한 키포인트 그룹화 과정을 통해 동일한 물체에 속한 키포인트 집합을 구성한다. 이 과정은 다중 객체 환경에

서도 각 객체의 키포인트를 분리할 수 있도록 한다.

각 객체에 대해 추출된 2차원 키포인트와 사전에 정의된 물체의 3차원 모델 키포인트를 대응시킨 후 Perspective-n-Point(PnP)[7] 알고리즘을 적용하여 물체의 위치와 방향을 계산한다. 물체 좌표계에서 정의된 3차원 키포인트 $\mathbf{X}_i = [X_i, Y_i, Z_i]^T$ 는 회전 행렬 $\mathbf{R} \in SO(3)$ 과 이동 벡터 $\mathbf{t} \in \mathbb{R}^3$ 에 의해 카메라 좌표계로 변환되며 다음과 같이 표현된다.

$$\mathbf{X}_i^c = \mathbf{R}\mathbf{X}_i + \mathbf{t} \quad (8)$$

변환된 3차원 점은 카메라 내부 파라미터 행렬 \mathbf{K} 를 이용하여 이미지 평면으로 투영된다.

$$s \begin{bmatrix} u_i \\ v_i \\ 1 \end{bmatrix} = \mathbf{K}[\mathbf{R} \quad \mathbf{t}] \begin{bmatrix} X_i \\ Y_i \\ Z_i \\ 1 \end{bmatrix} \quad (9)$$

여기서 (u_i, v_i) 는 이미지 상의 2차원 키포인트 좌표이며 s 는 스케일 인자를 나타낸다. PnP 알고리즘은 관측된 2차원 키포인트 $\mathbf{x}_i = [u_i, v_i]^T$ 와 투영된 점 간의 재투영 오차[9]를 최소화하는 회전 행렬 \mathbf{R} 과 이동 벡터 \mathbf{t} 를 추정한다. 이 과정은 다음과 같은 최적화 문제로 정의된다.

$$\min_{\mathbf{R}, \mathbf{t}} \sum_i^n |x_i - \pi(\mathbf{R}\mathbf{X}_i + \mathbf{t})|^2 \quad (10)$$

본 연구에서는 키포인트 검출 과정에서 발생할 수 있는 오차에 대한 강건성을 확보하기 위해 RANSAC 기반 PnP 기법을 적용하여 이상치를 제거하였다. 이때 카메라 내부 파라미터와 물체의 실제 크기 정보를 함께 사용하였으며 최소 네 개 이상의 키포인트가 검출된 경우에만 최종 6자유도 포즈를 계산하였다.

DOPE 모델 학습을 위해 NVIDIA Isaac Sim[2] 환경에서 팔레트를 다양한 위치와 방향으로 무작위 배치한다. 그리고 배경, 조명, 카메라 시점을 랜덤화하여 RGB 이미지를 생성하였다. 이 과정에서 팔레트의 키포인트 위치와 6자유도 자세 정보는 자동으로 라벨링되며 JSON 형식의 정답 데이터로 저장된다.

이와 같은 합성 데이터 기반 학습 방식은 실제 환경에서 발생하는 조명 변화, 배경

복잡성, 시점 변화 등에 대한 강건성을 향상시키는 데 기여한다. 또한 시뮬레이션 환경에서 대량의 데이터를 안전하고 효율적으로 생성할 수 있어 실제 로봇 실험 이전 단계에서 포즈 추정 성능을 충분히 검증할 수 있다. 모델의 학습에 사용된 주요 하이퍼파라미터는 Table 2에 정리하였다.

Table 2. Hyperparameter of DOPE

| | |
|------------|----------------------|
| Model | DOPE |
| Dataset | Created in Isaac Sim |
| Epoch | 50 |
| Batch Size | 8 |
| Patience | 5 |

2.4 구동 실험

2.4.1 시뮬레이션 환경에서의 구동 실험

시뮬레이션 환경에서는 UR10 로봇 모델을 사용하여 팔레트 Pick-and-Place 작업을 수행하였다. 실험 환경에는 총 다섯 개의 팔레트를 무작위 각도로 배치하여 인식 난이도를 높였다. 각 팔레트에 대해 인지 단계에서 추정된 자세 정보를 기반으로 로봇의 집기 동작을 수행하였다.

로봇은 팔레트에 접근한 후 측면 흡착 방식으로 집기를 수행하고 지정된 위치로 이동하여 적재 동작을 완료한다. 로봇의 이동과 자세 제어에는 Pick-and-Place 컨트롤러와 RMPFlow 기반 모션 컨트롤러를 사용하였다. 모든 동작은 Isaac Sim의 물리 엔진을 기반으로 충돌 여부와 안정성을 고려하여 실행되었다.

2.4.2 실제 로봇 환경에서의 구동 실험

시뮬레이션 환경에서 검증된 인식 및 구동 로직은 실제 PiPER 로봇 팔과 ZED 2i Stereo Camera를 활용하여 실험을 수행하였다. 실제 환경에서는 시뮬레이션에서 사용한 인지 알고리즘과 제어 흐름을 동일하게 적용하였다.



Figure 3. PiPER Robot and Real Experiment Environment

PiPER 로봇 팔 1.5kg의 페이로드를 처리할 수 있는 경량 로봇 팔이다. 이 로봇 팔로 실험 환경의 제약을 고려하여 선택되었다. 이에 따라 실제 산업용 팔레트 대신 크기가 약 10배 축소된 팔레트 모형을 제작하여 실험을 진행하였다. 이를 통해 로봇 팔의 작업 범위와 적재 하중을 고려한 조건에서도 제안한 인식 및 구동 로직이 정상적으로 동작함을 확인하였다.

III. 실험 결과

3.1 실험 환경 및 평가 기준

본 장에서는 제2장에서 설명한 RGB-D 기반 기하학적 포즈 추정 기법과 DOPE 딥러닝 기반 포즈 추정 기법의 성능을 평가한다. 모든 실험은 NVIDIA Isaac Sim 기반 물리 시뮬레이션 환경에서 수행되었으며 일부 실험은 실제 로봇 시스템을 이용하여 시뮬레이션 결과의 실제 환경 적용 가능성을 확인하였다.

시뮬레이션 환경에서는 팔레트를 다양한 위치와 방향으로 배치하여 인식 난이도를 조절하였으며 조명 조건과 배경 환경을 변화시켜 실제 산업에서의 다양성을 구현하였다. 인지 단계의 성능은 팔레트의 방향 오차를 기준으로 정량적으로 평가하였다.

또한 Pick-and-Place 작업의 수행 결과를 함께 평가하였다. 로봇 구동 성능 평가는 집기 과정에서의 성공 여부와 집기 이후 이동 및 적재 동작의 안정성을 기준으로 수행하였다. 집기 과정에서 팔레트가 정상적으로 흡착되었는지 여부와 적재 위치에 정확히 배치되었는지를 통해 각 실험의 성공 여부를 판단하였다.

3.2 RGB-D 기하학적 포즈 추정 결과

RGB-D 카메라를 통해 획득한 깊이 영상으로부터 생성된 포인트 클라우드는 바닥과 주변 구조물로 인한 잡음이 포함되어 있었다. DBSCAN 기반 클러스터링을 적용한 결과 팔레트 영역은 하나의 주요 클러스터로 안정적으로 분리되었으며 포인트 수가 상대적으로 적은 잡음 영역은 효과적으로 제거되었다. DBSCAN을 통해 분리된 팔레트 포인트 클러스터에 PCA를 적용한 결과 팔레트의 장축 방향이 일관되게 추정됨을 확인하였다. 특히 팔레트가 회전된 상태로 배치된 경우에도 주성분 벡터는 실제 팔레트 방향과 유사한 경향을 보였다.

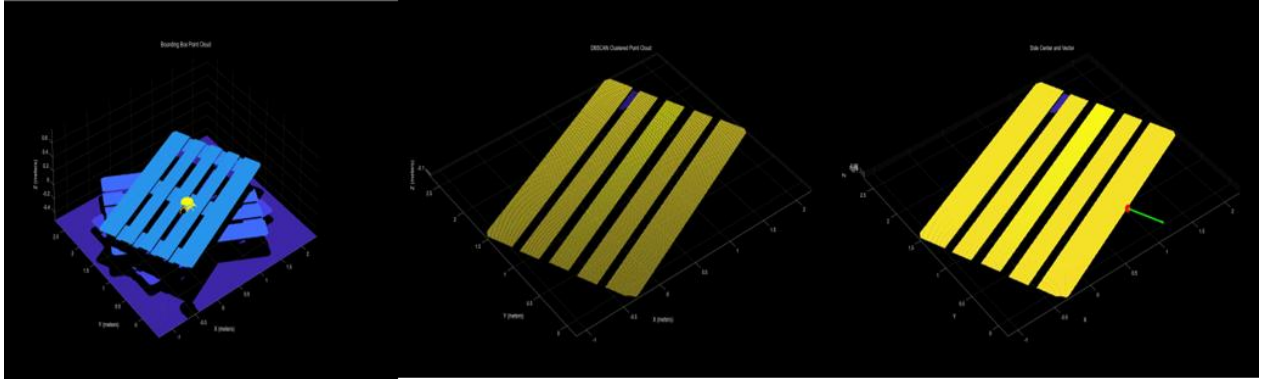


Figure 4. Raw Point Cloud(right) After DBSCAN(middle) the Vector after PCA(left)

이를 기반으로 계산된 팔레트 측면의 법선 벡터는 흡착 기반 Pick 동작에 적합한 방향을 제공했고 해당 법선 벡터를 이용해 생성된 Pick 포즈는 로봇 집기 동작에서 안정적인 접촉을 가능하게 하였다.

RGB-D 기반 기하학적 포즈 추정 기법은 깊이 정보를 직접 사용하기 때문에 팔레트의 공간적 구조를 비교적 명확하게 표현할 수 있다는 특징을 가진다. 결론적으로 수행한 실험을 통해 포인트 클라우드 생성 이후 DBSCAN을 적용한 클러스터링 과정과 PCA를 이용한 방향 추정 과정이 팔레트의 주요 기하학적 특성을 안정적으로 추출할 수 있음을 확인하였다.

특히 팔레트가 회전된 상태로 배치된 경우에도 클러스터링 결과가 비교적 일관되게 유지되었다. 또한 PCA를 통해 계산된 주성분 방향이 실제 팔레트의 방향과 유사한 경향을 보였다. 이러한 결과로부터 RGB-D 기반 기하학적 접근 방식은 별도의 학습 데이터 없이도 일정 수준 이상의 포즈 추정 성능을 확보할 수 있음을 확인할 수 있었다

3.3 DOPE 딥러닝 기반 포즈 추정 결과

Isaac Sim 환경에서 생성된 합성 데이터셋을 이용하여 DOPE 모델을 학습한 결과 다양한 배경과 조명 조건에서도 팔레트의 주요 키폰트가 안정적으로 검출됨을 확인하였다. 특히 팔레트가 부분적으로 가려진 상황에서도 일부 키폰트를 통해 포즈 추정이 가능하였다.

DOPE 네트워크로부터 추정된 키폰트에 PnP 알고리즘을 적용한 결과로는 팔레트의 위치와 방향이 일관되게 추정된다. RGB 영상만을 입력으로 사용함에도 불구하고 팔레

트의 전체 6자유도 자세를 추정할 수 있음을 확인할 수 있었다.

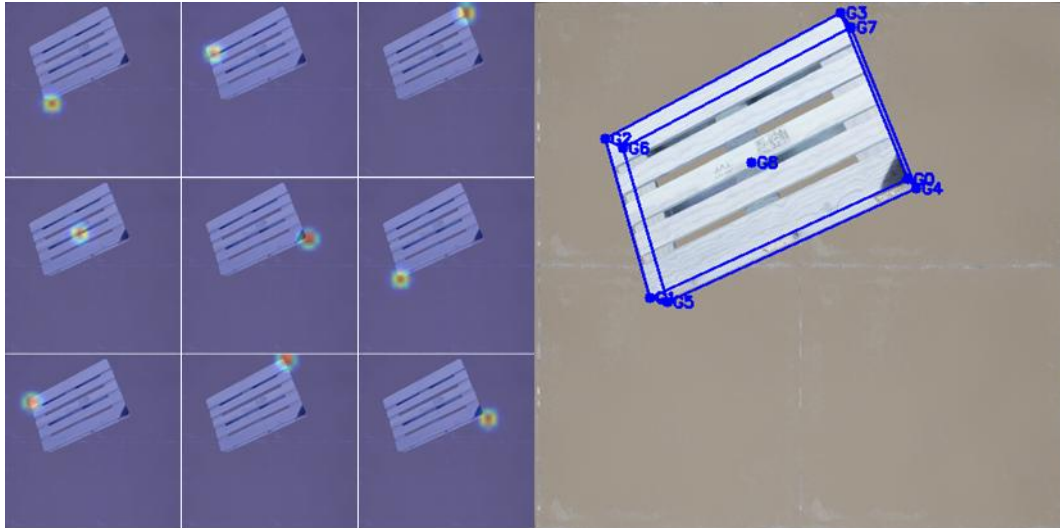


Figure 5. Predicted Heat Map and Keypoints Result

DOPE 딥러닝 기반 포즈 추정 기법은 단일 RGB 영상만을 입력으로 사용함에도 불구하고 팔레트의 6자유도 자세를 추정할 수 있다는 점이 강점이다. 시뮬레이션 환경에서 생성된 합성 데이터를 이용해 학습한 DOPE 모델은 다양한 배경과 조명 조건에서도 팔레트의 주요 키폰트를 안정적으로 검출하였다.

하지만 깊이 정보를 사용하지 않는 구조적 특성으로 인해 조명 변화가 크거나 배경이 복잡한 환경에서는 일부 키폰트의 검출 정확도가 낮아지는 한계가 있다. 이러한 결과는 RGB 영상 기반 딥러닝 기법이 환경 조건 변화에 따라 성능 편차가 발생할 수 있음을 보여준다.

3.4 RGB-D 기반 기법과 DOPE 기법의 비교

두 기법의 성능을 확인하기 위해서 Isaac Sim 시뮬레이션 상에서 100번의 랜덤한 자세의 팔레트에 대한 자세 추정을 진행하였다. 각도는 -80° 에서 80° 로 설정하였을 때 다음 Table 3와 같은 오차각이 나온다.

Table 3. The Performance of Two Pose Estimation Methods

| Method | Mean Error Angle [°] |
|----------------|----------------------|
| RGB-D Geometry | ± 2.308 |
| DOPE | ± 0.973 |

딥러닝 기반 DOPE가 RGB-D 기하학 기반 자세 추정보다 좋은 성능을 보인다.

3.5 실제 로봇 실험 결과

시뮬레이션 환경에서 검증된 포즈 추정 알고리즘을 PiPER 로봇 팔에 적용하여 실제 팔레트 Pick-and-Place 실험을 수행하였다. 시뮬레이션에서 사용한 동일한 포즈 추정 파이프라인을 실제 환경에 적용한 결과로 팔레트 인식 및 집기 동작이 안정적으로 수행됨을 확인하였다. 성능을 평가하기 위해 시뮬레이션과 실제 환경에서 50번에 시도를 하였고 place 단계에서 시뮬레이션에서는 ± 10 [cm], 실제 환경에서는 ± 1 [cm] 오차가 있으면 실패로 간주하였다.

Table 4. Pick-and-Place Accuracy

| 실험 환경 | 총 시도 횟수 | 성공 횟수 | 실패 횟수 | 성공률 |
|-----------|---------|-------|-------|-----|
| Isaac Sim | 50 | 49 | 1 | 98% |
| 실제 환경 | 50 | 45 | 5 | 90% |

Table 4 에 제시된 Pick-and-Place 실험 결과를 통해 시뮬레이션 환경과 실제 로봇 환경 간의 성능 차이를 확인할 수 있다. 이는 시뮬레이션 환경이 센서 노이즈나 물리적 불확실성이 상대적으로 적고 동일한 조건을 반복적으로 유지할 수 있기 때문으로 판단된다. 반면 실제 로봇 환경에서는 시뮬레이션에 비해 실패 횟수가 증가하였다. 이러한 차이는 실제 환경에서 센서 측정 오차나 세팅 된 카메라의 미세한 틀어짐과 같은 요인들이 복합적으로 작용한 결과로 볼 수 있다.

그럼에도 불구하고 실제 환경에서의 성공률이 높은 수준을 유지한 점은 시뮬레이션 환경에서 설계하고 검증한 포즈 추정과 구동 로직이 실제 로봇 시스템에서도 효과적으로 동작할 수 있음을 보여준다. 이는 물리 기반 시뮬레이션을 활용한 인식 및 제어

알고리즘 개발이 실제 산업 환경 적용을 위한 유효한 접근 방법임을 뒷받침하는 결과로 해석할 수 있다.

3.6 RGB-D 기반 기법과 DOPE 기법의 비교 및 상호 보완성

본 연구에서 제안한 두 가지 포즈 추정 기법은 각각 상이한 장단점을 가진다. RGB-D 기반 기하학적 기법은 깊이 정보를 활용하여 공간적 정확도가 우수한 반면 DOPE 기반 기법은 깊이 센서 없이도 포즈 추정이 가능하다는 장점을 가진다.

실험 결과 시뮬레이션 환경에서는 두 기법 모두 팔레트 Pick-and-Place 작업을 안정적으로 수행하였다. 다만 조명 조건이나 배경 복잡도 변화에 따라 각 기법의 성능 특성에는 차이가 나타났다. 이러한 결과를 통해 특정 기법 하나에만 의존하기보다는 작업 환경에 따라 두 기법을 선택적으로 적용하거나 함께 활용하는 방식이 보다 효과적일 수 있음을 확인하였다.

3.7 시뮬레이션-현실 전이(Sim-to-Real)에 대한 고찰

본 연구에서는 NVIDIA Isaac Sim 기반 시뮬레이션 환경에서 설계하고 검증한 포즈 추정 알고리즘을 실제 PiPER 로봇 팔에 적용하여 실험을 수행하였다. 실험 결과 시뮬레이션 환경에서 구현한 인식과 조작 파이프라인이 실제 환경에서도 유사한 동작 특성을 보이는 것을 확인할 수 있었다.

이는 물리 기반 시뮬레이션 환경에서 생성된 데이터와 알고리즘 검증 과정이 실제 로봇 시스템 개발에 실질적인 도움이 될 수 있음을 보여준다. 특히 자동 라벨링을 활용한 합성 데이터 생성 방식은 실제 환경에서 발생할 수 있는 안전 문제와 실험 비용 부담을 줄이는 데 효과적인 방법으로 확인되었다. 또한 물리 엔진으로 실제와 비슷한 환경에서 실험하기에 실제 구현하기에도 효과적이었다.

3.8 연구의 한계 및 향후 연구 방향

현재 시뮬레이션 환경을 중심으로 팔레트 포즈 추정과 로봇 조작을 검증하였으며 실제 환경에서의 실험 규모는 제한적인 범위에 머물렀다. 또한 사용한 포즈 추정 구조는

팔레트 객체에 특화되어 있기 때문에 다양한 형태의 물체에 대한 일반화 성능에 대해서는 추가적인 검증이 필요하다.

향후 연구에서는 RGB-D 기반 기법과 DOPE 기반 기법을 결합한 포즈 추정 구조를 설계하여 보다 복잡한 산업 환경에서의 실험을 통해 시스템의 안정성과 적용 범위를 확장할 수 있을 것으로 판단된다. 더 나아가 도메인 랜덤화 기법을 적용하여 시뮬레이션 환경과 실제 환경 사이의 차이를 줄이기 위한 추가적인 연구가 필요하다.

IV. 결론

본 연구에서는 NVIDIA Isaac Sim 기반의 물리 시뮬레이션 환경을 활용하여 팔레트 자동 Pick-and-Place 시스템을 구현하고 그 유효성을 검증하였다. 산업 현장에서 널리 사용되는 팔레트는 작업 환경에 따라 정렬되지 않은 상태로 놓이는 경우가 많으며 이로 인해 작업 효율 저하 및 안전 문제가 발생을 방지하기 위해 자동화하는 시스템을 설계하였다.

우선 RGB-D 카메라를 이용한 기하학적 포즈 추정 기법을 적용하여 팔레트의 위치와 방향을 계산하였다. 깊이 영상으로부터 생성된 포인트 클라우드에 대해 DBSCAN 기반 클러스터링과 PCA를 수행함으로써 팔레트의 기하학적 특성을 분석하였으며, 이를 바탕으로 측면 Pick 포즈를 안정적으로 산출할 수 있었다. 해당 기법은 별도의 학습 데이터 없이도 비교적 일관된 성능을 보였다는 점에서 장점을 가진다.

또한 단일 RGB 영상만을 입력으로 사용하는 DOPE 딥러닝 기반 포즈 추정 기법을 함께 적용하여 깊이 정보가 제한된 환경에서도 팔레트의 6자유도 자세를 추정할 수 있음을 확인하였다. 시뮬레이션 환경에서 생성된 합성 데이터를 활용하여 모델을 학습함으로써 실제 환경에서 요구되는 데이터 수집 및 라벨링 부담을 줄일 수 있었다.

시뮬레이션 실험 결과, RGB-D 기반 기하학적 기법과 DOPE 기반 기법 모두 팔레트 Pick-and-Place 작업을 성공적으로 수행하였다. 더 나아가 시뮬레이션 환경에서 검증된 인식 및 조작 알고리즘을 실제 PiPER 로봇 팔에 적용한 결과는 실제 환경에서도 팔레트 인식 및 Pick-and-Place 동작이 안정적으로 수행됨을 확인하였다. 이 결과는 시뮬레이션 기반으로 설계된 알고리즘이 실제 로봇 시스템으로 충분히 확장 가능성을 보여준다.

본 연구는 물리 기반 시뮬레이션 환경을 활용하여 팔레트 자동 조작 시스템을 구현하고 기하학적 방법과 딥러닝 기반 방법을 비교, 분석하였다는 점에서 의의를 가진다. 향후 연구에서는 보다 다양한 산업 환경을 고려한 실험과 함께 두 포즈 추정 기법을 결합한 하이브리드 인식 구조를 적용함으로써 시스템의 강건성을 더욱 향상시킬 수 있을 것으로 기대된다.

참고문헌

- [1] J. Tremblay, T. To, S. Birchfield, and S. Tyree, “Deep Object Pose Estimation for Semantic Robotic Grasping of Household Objects,” *Proceedings of the Conference on Robot Learning (CoRL)*, pp. 306-316, 2018.
- [2] NVIDIA Corporation, “NVIDIA Isaac Sim: A Simulation Platform for Robotics,” <https://developer.nvidia.com/isaac-sim>, accessed 2024.
- [3] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, “You Only Look Once: Unified, Real-Time Object Detection,” *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 779-788, 2016.
- [4] G. Jocher et al., “YOLOv8: Ultralytics YOLO,” <https://github.com/ultralytics/ultralytics>, accessed 2024.
- [5] M. Ester, H. P. Kriegel, J. Sander, and X. Xu, “A Density-Based Algorithm for Discovering Clusters in Large Spatial Databases with Noise,” *Proceedings of the Second International Conference on Knowledge Discovery and Data Mining (KDD)*, pp. 226-231, 1996.
- [6] I. T. Jolliffe, *Principal Component Analysis*, Springer, 2nd Edition, 2002.
- [7] V. Lepetit, F. Moreno-Noguer, and P. Fua, “EPnP: An Accurate $O(n)$ Solution to the PnP Problem,” *International Journal of Computer Vision*, Vol. 81, No. 2, pp. 155-166, 2009.
- [8] R. Hartley and A. Zisserman, *Multiple View Geometry in Computer Vision*, Cambridge University Press, 2nd Edition, 2004.
- [8] NVIDIA, *RMPflow – Isaac Sim Documentation*, NVIDIA, 2025. [Online]. Available: <https://docs.isaacsim.omniverse.nvidia.com/5.1.0/manipulators/concepts/rmpflow.html> , accessed Dec. 23, 2025.