

에어로졸 상황에서 물체의 6 자유도 자세 추정을 위한 벤치마크 데이터셋 및 베이스라인 구축

6-DoF Object Pose Estimation in Aerosol Conditions: Benchmark Dataset and Baseline

양 희 진¹, 이 승 현¹, 김 태 주¹, 최 유 경^{1,*}
(Heejin Yang¹, Seunghyeon Lee¹, Taejoo Kim¹, and Yukyung Choi^{1,*})

¹Sejong University

Abstract: The research on 6-DoF pose estimation has been ongoing in several fields until recently. As deep learning advances, the corresponding 6-DoF pose estimation will be trained on a variety of data to make predictions after training. However, there is a lack of research utilizing datasets to perform 6DoF pose estimation in environments such as aerosols, which can be encountered in disaster environments. In this paper, we propose an RGB-D benchmark dataset consisting of paired normal and aerosol situations targeting aerosol environments to establish a baseline. We specifically address the overall pipeline that built the dataset proposed in this study, i.e., 3D model construction, sensor setup, data collection methods, 6-DoF pose annotation process for objects, and validation of those annotations. Using our proposed dataset, we apply previously studied 6-DoF pose estimation methodologies to benchmark and experiment on normal and aerosol situations. Our experiments show a significant performance degradation in aerosol situations, and further experiments to learn the aerosol environment show that performance can be improved. As such, the novel aerosol conditions demonstrate the difficulty of performing 6-DoF pose estimation and the need for future research.

Keywords: 6-DoF object pose estimation, aerosol condition, dataset, recognition, RGB-D

I. 서론

현실 세계에 존재하는 물체의 위치 및 방향 정보를 추정하는 6-자유도 자세 추정(6-DoF Object Pose Estimation)은 로봇틱스[1, 2, 3], 증강 현실[4, 5], 자율주행[6, 7, 8]과 같은 다양한 어플리케이션에서 사용된다. 이는 인식 대상 물체에 대한 3차원 정보를 인지하고 물체와 상호작용하기 위해 필요한 연구이다. 최근 딥러닝과 로봇 기술의 발전됨에 따라 실제 산업 환경에서 물류 분류와 운반 등의 작업에 다양한 로봇이 사람의 업무를 대체하고 있다[9].

정적인 작업 환경뿐만 아니라, 예측불허한 재난 환경에서도 사람을 대신하여 구조 및 비상 작업을 수행할 수 있는 시각 인지 능력 연구[10, 11, 12]와 지능형 로봇 개발[13]이 활발하다. 그러나 재난 현장은 화재, 폭발 등의 영향으로 조명 조건이 열악하거나 스팀, 먼지 등으로 인해 시야가 제한되는 이상 상황이 발생한다. 이때 RGB-D 센서를 통해 얻은 칼라(RGB) 정보가 제한적이며 깊이(depth) 정보에는 노이즈로 인한 품질이 저하되는 이상 상황에 직면하게 된다.

본 연구에서는 대기 중에 떠도는 미세한 입자로 인해 시야에 제약이 발생하는 에어로졸 상태를 이상 상황으로 정의하고, 에어로졸 환경에서의 6-자유도 자세 추정 문제를 해결하기 위한 데이터셋을 제안한다. 앞서 정의한 이상 상황에 처했을 때 기존 연구된 6-자유도 자세 추정 방법론[14, 15, 16]들은 학습 데이터 분포와 차이가 발생하여 강인하게 작동하기 어렵다. 이처럼, 기존의 6-자유도 자세 추정 연구에서 사용되는 데이터셋[17, 18, 19, 20, 21]은 에어로졸과 같은 이상 상황을 고려하지 않으므로 기존의 데이터셋을 활용하여 실제 재난 상황에 강인하게 작동하는 6-자유도 자세 추정 알고리즘을 연구하는 것에 어려움이 있다. 해당 문제를 해결하기 위해 본 논문에서는 에어로졸 상태를 포함하는 새로운 6-자유도 자세 추정 데이터셋을 제안한다. 본 연구에서 제안하는 데이터셋은 에어로졸이 발생 가능한 상황을 고려하여 일상생활에서 접근하기 쉬운 물체와 산업 현장에서 재난이 발생하였을 때 조작 대상이 되는 물체로 구성한다. 또한 해당 데이터셋에 기존 6-자유도 자세 추정에서 사용되는 단일 RGB와 RGB-D 기반의

* Corresponding Author

Manuscript received February 14, 2024;

양희진: 세종대학교 대학원생(hjyang@rcv.sejong.ac.kr, ORCID[®] 0009-0009-8452-1076)

이승현: 세종대학교 대학원생(shlee@rcv.sejong.ac.kr ORCID[®] 0000-0002-2498-5129)

김태주: 세종대학교 대학원생(tjkim@rcv.sejong.ac.kr ORCID[®] 0000-0001-9884-3177)

최유경: 세종대학교 부교수(ykchoi@rcv.sejong.ac.kr ORCID[®] 0000-0002-9970-0132)

※ 본 연구는 과학기술정보통신부 및 정보통신기획평가원의 정보통신방송혁신인재양성(메타버스융합대학원)사업 연구 결과로 수행되었음(HTP-2023-RS-2023-00254529). 또한 2022년도 정부(과기정통부)의 재원으로 한국연구재단의 지원을 받아 수행된 연구임(No. RS-2022-00144385).

표 1. 6-자유도 자세 추정 관련 데이터셋들의 비교. 제공하는 센서의 모달리티, 다양성(혼잡, 폐색, 에어로졸) 여부, 대상 물체 전체 개수, 대상 물체 범주.

Table 1. Comparison of datasets for 6-DoF pose estimation. Modality of sensors provided, diversity (clutter, occlusion, aerosol), the number of total objects and category of target objects.

Dataset	Modality	Clutter	Occlusion	Aerosol	# Total Objects	Categories
LINEMOD[17]	RGB-D	O	X	X	15	Household
Occluded-LINEMOD[18]	RGB-D	O	O	X	8	Household
T-LESS[19]	RGB-D	O	O	X	30	Industry
HomebrewedDB[20]	RGB-D	O	O	X	33	Household
HOPE[21]	RGB-D	O	O	X	28	Household
Ours	RGB-D	O	X	O	4	Household, Industry

방법론[15, 16]을 적용하여 이상 상황에서의 강인성을 평가한다. 또한, 실험 결과에 대한 분석을 통해 에어로졸 환경에 대한 고려가 필요함을 입증한다.

II. 관련 연구

해당 절에서는 6-자유도 자세 추정 연구에 활용되고 있는 기존의 데이터셋들을 소개하고 각 데이터셋이 가지는 특성을 요약한다. 또한 각 데이터셋들에 대한 주요 정보는 표 1과 같이 정리된다.

LINEMOD[17]는 텍스처가 부족한 15개의 각 물체에 대응되는 3D 모델로 구성되어 있으며 각 물체가 여러 시점에 따라 달라지는 장면들에 대해 샘플링 된 영상들로 구성되어 있다. RGB-D 센서인 Microsoft Kinect v1을 사용하여 RGB 영상과 깊이 정보를 동시에 취득하였다. 6-자유도 자세 추정을 수행하기 위한 대상이 되는 물체는 중앙에 배치하고 해당 물체의 주위를 대상 외의 물체들로 둘러싸고 있는 장면으로 구성되어 있다. 연속적인 장면에 대해 학습 데이터와 평가 데이터로 분리하여 제공한다. 해당 데이터셋은 가려짐(occlusion)과 같이 자세 추정 시 하는 데 난이도가 높은 상황을 다루고 있지 않다.

OCCLUDED-LINEMOD[18]는 가려짐 현상이 발생했을 때 물체의 3차원 정보를 추정하기 어려운 문제를 해결하고자 한다. 기존의 LINEMOD 데이터셋을 기반으로 대상 물체 8가지에 대해 평가 데이터셋의 가려짐 현상을 중점으로 확장된 데이터셋이다.

T-LESS[19]는 텍스처가 부족하지만 뚜렷한 외관 정보를 가지고 있는 LINEMOD와 다르게 산업 현장에서 사용되는 텍스처가 없고 비슷한 외관을 가지는 전기 소품 30개로 구성된다. Primesense Carmine 1.09, Microsoft Kinect v2, Canon IXUS 950 IS와 같이 다양한 센서를 같이 사용하여 데이터를 취득하였으며 학습 데이터는 배경이 없는 장면으로 구성하였다. 평가 데이터는 대상 물체의 주위에 다른 물체들로 배치하여 학습 데이터에 비해 물체의 자세 정보를 예측하기 어려운 가려짐 및 혼잡(clutter) 상황을 연출하였다. 다양한 시점의 변화를 손으로 직접 카메라를 움직여 진행한 LINEMOD와 다르게 물체의 일정한 회전정도의 변화가 가능한 턴테이블(turn-table)과 특정 범위 내 정확한 기울기 조절이 가능한 지그(jig)에 각 센서들을 부착하여 촬영 환경을 구성하였다.

이 외에도 HomebrewedDB[20]는 선행 연구인 T-LESS에서 영감을 받아 데이터 촬영 환경을 유사하게 구성하였고 다양

한 조도 변화에 대한 강인성을 평가하고자 조명 환경을 다양하게 구성하였다.

HOPE[21]는 3D 물체에 대응되는 동일한 물체를 구매하는 게 어려운 문제를 해결하고자 비교적 시간이 지나도 외관이 변하지 않고 접근하기 쉬운 장난감들로 대상 물체를 구성하였다. 또한, 여러 장소에서 데이터를 취득하여 배경에 대한 다양성과 조도 변화를 주었으며 물체들의 배치를 다양하게 구성하였다.

이처럼 기존의 데이터셋은 물체에 대한 혼잡과 가려짐 등 물체 중심의 다양성과 배경 및 조도 변화와 같이 한정된 환경 다양성을 포함한다. 그러나 본 논문에서 정의한 이상 상황은 시야에 제한이 생기는 경우로, 기존 데이터셋에서 고려하지 않은 새로운 환경 다양성이 요구된다. 따라서, 에어로졸이라는 새로운 상황으로 환경 다양성을 확장하여 재난 상황에서 작동하는 6-자유도 자세 추정 알고리즘 연구가 필요하다.

본 연구진은 에어로졸이라는 새로운 상황을 제시하여 기존에 고려하지 않은 새로운 환경 조건에서의 6-자유도 자세 추정을 수행하기 위한 데이터셋을 구축하고 기존의 방법론들을 이용하여 강인성을 평가한다.

III. 데이터셋 취득

1. 3D 모델

대상 물체와 카메라의 기하학적 관계를 파악하기 위해 이미지 내에 존재하는 물체에 대응되는 3D 모델이 필요하다. 본 연구에서 대상 물체에 대한 3D 모델을 생성하기 위해 두 가지 광학 3D 스캐너를 사용한다. 텍스처에 상관없이 세밀한 메쉬(mesh) 형태의 데이터 추출이 가능한 HandySCAN BLACK[22]를 이용하여 대상 물체의 형태 정보를 취득하고, 물체의 형태를 바탕으로 Go!SCAN SPARK[23]를 이용하여 색감 정보를 취득하여 3D 모델을 구성하였다. 제한된 데이터셋의 3D 모델은 그림 1에서 확인할 수 있으며, LINEMOD의 대상 물체 2종과 금속 재질의 밸브 2종으로 구성하였다.



그림 1. 본 데이터셋에서 사용되는 3D 모델.

Fig. 1. The 3D models used in this dataset.

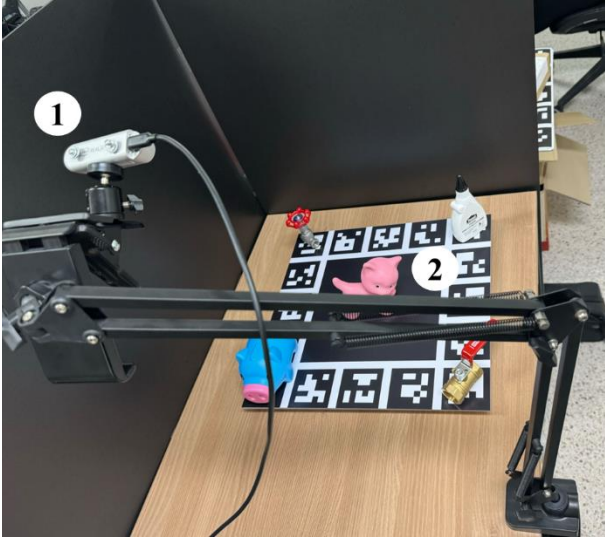


그림 2. 데이터 취득 환경. (1)높이 조절이 가능한 카메라 센서. (2)턴테이블

Fig. 2. Data acquisition environment. (1)Height-adjustable camera sensor. (2)Turntable.

2. 센서 및 데이터 취득 환경

2차원 영상 및 깊이 정보를 취득하기 위해 그림 2의 (1)과 같이 Intel RealSense D435i 센서를 이용한다. 해당 센서는 내부적으로 단일 RGB 센서와 스테레오 IR로 구성되어 있으며 IR projector로부터 투영된 적외선 점들의 패턴을 파악하여 깊이 정보를 구한다. D435i에서 기본 설정된 내부 파라미터를 사용하였다. 또한, 에어로졸 상황을 모사하는 경우 영상이 전체적으로 밝아져 영상의 색감이 자동으로 보정되는 현상을 방지하고자 안개 제거 알고리즘에서 활용되는 데이터셋[24, 25]과 동일하게 영상의 색감 및 밝기를 자동으로 보정해주는 white-balance와 auto-exposure를 모두 비활성화하였다.

물체의 여러 시점에 대한 학습 및 평가를 수행하기 위해서 다양한 위치에서 데이터를 취득해야 한다. 하지만 여러 각도에서 손으로 직접 촬영하는 것은 시간적 비용이 많이 들며, 부정

확한 카메라의 위치로 인해 유사한 시점에 대한 다양한 물체의 자세 정보를 평가하기 어렵다. 따라서, 비교적 효율적이며 물체마다 유사한 시점으로 6-자유도 자세 추정 알고리즘을 평가하기 위해 그림 2와 같이 TLESS와 유사한 환경을 구성하였다. 센서의 고도 범위를 0~90도의 내로 정의하였고, 회전판을 돌려 촬영함으로써 다양한 시점에 따라 물체가 변하는 모습을 촬영하였다.

3. 데이터셋 구성

본 연구에서 제안한 데이터셋은 그림 3과 같이 정상 상황과 에어로졸 상황에 대해 여러 시점에서 취득하였다. 같은 장면에 대해 두 가지 상황을 연출하기 위해 정상 상황을 캡처한 후, 스텝 생성기를 이용하여 에어로졸을 형성하였다. 해당 과정을 반복하여 4개의 물체에 대하여 각 200장의 대응되는 두 가지 영상을 취득하였다. 동일 장면에 대한 RGB 영상 및 깊이 정보 예시는 그림 3의 예시에서 확인할 수 있다. RGB 영상의 경우 에어로졸 상황에서 대상 물체에 대한 가시성이 크게 저하되며, 깊이 정보 또한 심한 노이즈가 발생하는 것을 확인할 수 있다. 이처럼 여러 고도 및 방위각에 따라 촬영을 진행하여 물체의 다양한 시점에 따른 대상 물체의 변화를 포착하였다.

단조로운 배경을 구성하기 위해 기준이 되는 대상 외 물체들은 외곽으로 배치하여 그림 3의 1행과 같이 학습 데이터를 구성했다. 평가 데이터는 그림 3의 2행과 같이 대상 물체가 주변의 물체로부터 둘러싸여 있는 장면들로부터 구성되어 있다. 이는 학습 데이터 보다 비교적 예측하기 어려운 상황에서 모델의 강인성을 평가하기 위한 것으로 기존의 단일 RGB와 RGB-D 기반의 6-자유도 자세 추정 방법론의 강인성을 평가하였다.

4. Annotation

취득한 데이터에 대한 6-자유도 자세 정보를 어노테이션(annotation)하기 위해 오픈 소스로 제공되고 있는 6D-PAT[26] 툴을 사용한다. 6D-PAT은 각 이미지 내의 대상 물체에 대한 2D 점과 대응되는 3D 모델의 3D 점 사이의 관계를 고려하여 카메라의 자세를 구하는 PnP(perspective-n-point) 알고리즘을 기반으로 동작한다. 해당 알고리즘은 이론상 최소 3개의 2D-3D간 매칭 쌍으로도 동작하지만 좀 더 정확한 정보를 주기 위

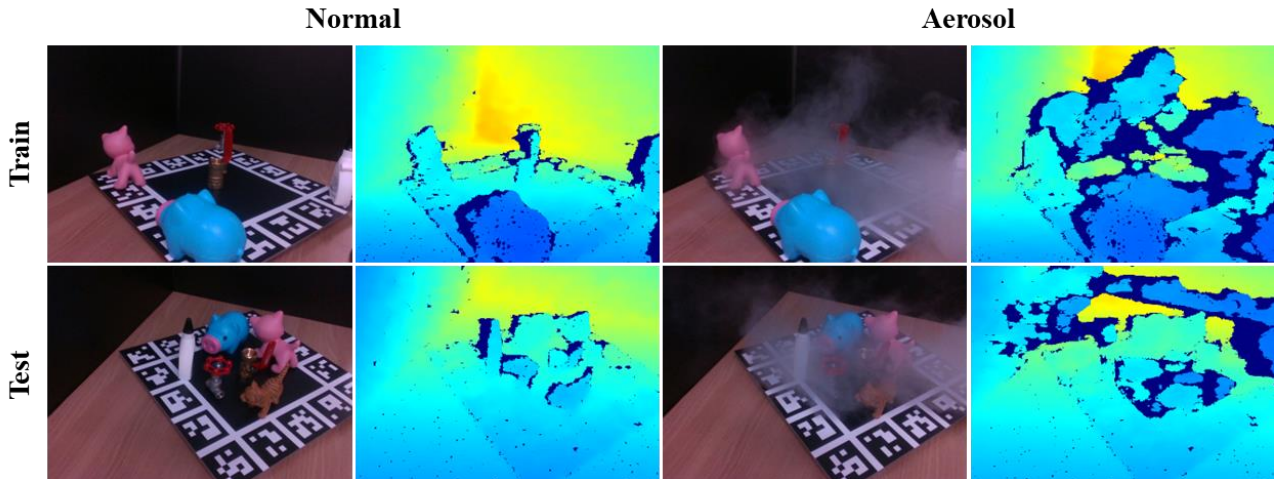


그림 3. 제안한 데이터셋의 예시. 학습 및 테스트 데이터셋에 대해 정상 상황과 에어로졸 상황으로 구성.

Fig. 3. Example of our proposed dataset. Compose normal and aerosol conditions for training and test datasets.

표 2. 각 물체 별 Annotation 오차(RMSE) 측정 결과.

Table 2. Annotation error (RMSE) measurement results for each object.

Object	RMSE(mm)
ball valve	6.3
cat	4.9
globe valve	3.4
glue	8.2
average	5.7

해 6D-PAT 에서는 최소 6개 이상의 대응 관계를 사용하도록 설계되었다. 6D-PAT을 사용하여 본 연구에 참여한 연구진들이 수동으로 각 프레임마다 2D-3D 매칭 점들을 구하여 물체에 대한 6-자유도 자세 정보를 어노테이션 하는 과정을 수행한다.

5. Annotation 오차 검증

본 연구에서 제안한 데이터셋의 실측(ground-truth) 자세를 검증하기 위해 깊이 정보를 이용하여 RMSE(Root Mean Squared Error)를 측정하였다. 먼저, 3D 모델을 실측 자세로 렌더링하여 구한 깊이 정보(d_r)와 D435i 센서로 측정한 깊이 정보(d_c)에 대한 오차(ε)를 구한다. 이때 대상 물체 영역만을 고려하였으며 오차는 식 (1)과 같이 나타낼 수 있다.

$$\varepsilon = d_c - d_r \quad (1)$$

식 (1)을 활용하여 물체 영역에 해당하는 모든 점에 대한 평균 오차를 측정한 후, 최종적으로 식 (2)를 이용하여 RMSE를 계산한다.

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=0}^{N-1} (\varepsilon_i)^2} \quad (2)$$

표 2의 정량적 결과를 통해 제안한 데이터셋이 모든 물체에 대해 평균적으로 약 6 mm 이내의 오차로 제공함으로써, 6-자유도 자세 추정 연구에 있어 에어로졸 상태에 대한 정확한 학습 및 평가가 가능함을 보였다.

IV. 벤치마크 및 실험

1. 베이스라인 방법론

우리는 에어로졸 상황에서 6-자유도 자세 추정을 수행하기 위해 제안한 데이터셋에서 실험을 진행하였으며 베이스라인으로 기존 연구인 ROPE[15]와 FFB6D[16] 방법론을 적용한다.

ROPE는 단일 RGB 영상을 사용하는 방법론으로, 가려짐에 강인한 데이터 증강(augmentation)기법을 제안한다. 두 단계(2-stage)로 진행되어 입력 영상과 증강된 영상에 대한 2개의 가지로 RPN(Region Proposal Network)을 통과한다. 이후 RoI(Region of Interest)에 대해 3개의 키포인트 헤드(keypoint head)를 이용한다. 각각의 키포인트 헤드에서 나온 키포인트에 대한 히트맵(heatmap)을 생성한다. 실측 자세 정보를 통해 손실 정도를 가우시안 분포(Gaussian Distribution)에 따라 학습을 진행한다.

FFB6D는 RGB-D 영상을 사용하는 방법론으로, 여전히 LINEMOD에서 최고의 성능을 보이고 있다. 2차원으로 투영된 깊이 영상을 3차원의 포인트 클라우드 형태로 차원을 하나 증가시켜 주어 RGB 영상과 함께 여러 개의 양방향 융합

표 3. 제안한 데이터셋에 대한 단일 RGB 기반의 방법론인 ROPE[15]의 ADD(-S) 정확도. 비대칭 물체에 대해 ADD, 대칭 물체('*로 표기)에 대해 ADD-S를 측정함. '→' 는 학습 환경과 예측 환경을 나타냄(N: Normal, A: Aerosol).

Table 3. ADD(-S) accuracy of ROPE [15], a single RGB-based methods on the proposed dataset. Measured ADD for asymmetric objects and ADD-S for symmetric objects (denoted by '*'). '→' denotes training and prediction environments (N: Normal, A: Aerosol).

Object	ADD(-S)			
	N→N	N→A	N+A→N	N+A→A
ball valve*	61.25	12.50	60.00	35.00
cat	73.81	45.24	70.24	52.38
globe valve*	45.00	7.50	31.25	15.00
glue*	85.00	31.25	95.00	71.25
average	66.27	24.12	64.12	43.41

(Bidirectional Fusion) 모듈을 통과시킨다. RGB 영상에서의 외관 정보와 깊이 영상의 기하학적인 정보를 상호 보완적으로 융합하는 효과를 주게 된다. 이후 융합된 표현력을 바탕으로 선행 연구로 진행됐었던 키포인트 기반의 방법론인 PVN3D[14]의 전반적인 파이프라인을 따른다. 추출된 특징으로부터 3D 키포인트를 탐지하고 이를 활용하여 최종적으로 6-자유도 자세 추정을 수행한다.

2. 평가지표

6-자유도 자세 정보인 회전(R) 및 이동(T) 정도를 평가하기 위해 일반적으로 많이 사용되는 ADD와 ADD-S를 사용한다. ADD(Average Distance)는 3D 모델에 예측된 자세와 실측 자세를 적용하여 변환한 뒤, 각 점들 사이의 유클리디언 거리(Euclidean distance)의 평균을 계산한다. 3D 모델의 점을 x_i , 해당 점들의 집합 $M = \{x_i \in \mathbb{R}^3\}$ 이 주어졌을 때, 아래의 식 (3)과 같이 계산할 수 있다. 이때, $[R^*|T^*]$ 와 $[R|T]$ 는 각각 실측 6-자유도 자세와 추정된 6-자유도 자세를 나타낸다.

$$ADD = \frac{1}{|M|} \sum_{x \in M} \|(Rx + T) - (R^*x + T^*)\|_2 \quad (3)$$

ADD-S는 대칭인 물체에 대한 평가 지표로 사용된다. ADD와 다르게 거리는 각 점과 다른 점 집합에 대해 가장 근접한 거리에 대한 평균을 계산하며 아래의 식 (4)로 나타낼 수 있다.

$$ADD-S = \frac{1}{|M|} \sum_{x_1 \in M} \min_{x_2 \in M} \|(Rx_1 + T) - (R^*x_1 + T^*)\|_2 \quad (4)$$

앞서 언급한 ADD(-S) 값을 활용하여 정확도를 측정하며, 3D 모델의 지름(diameter)의 10%보다 가까운 경우 정답으로, 그 외에는 오답으로 간주하여 정확도를 계산한다.

3. 구현 세부 사항

ROPE는 250 에포크(epoch), FFB6D는 25 에포크로 학습한다. 이때, 최적화 함수는 동일하게 Adam[27]을 사용한다. 학습 상황과 예측 상황에 대한 표기를 '→'로 하여, 두 방법론 모두 정상 상황을 학습하여 정상 상황(N→N)과 에어로졸 상황(N→A)에 대한 평가를 진행한다.

표 4. 제안한 데이터셋에 대한 RGB-D 기반의 방법론인 FFB6D[16]의 ADD(-S) 정확도. 비대칭 물체에 대해 ADD, 대칭 물체(*로 표기)에 대해 ADD-S를 측정함. '→' 는 학습 환경과 예측 환경을 나타냄(N: Normal, A: Aerosol).

Table 4. ADD(-S) accuracy of FFB6D[16], an RGB-D based methods on the proposed dataset. Measured ADD for asymmetric objects and ADD-S for symmetric objects (denoted by '*'). '→' denotes training and prediction environments (N: Normal, A: Aerosol).

Object	ADD(-S)	
	N→N	N→A
ball valve*	38.75	13.75
cat	65.48	38.10
globe valve*	47.50	7.50
glue*	65.00	40.00
average	54.18	24.84

4. 단일 RGB 기반 방법론

표 3는 본 데이터셋에서 ROPE를 사용하여 학습 및 예측을 수행한 정량적 결과를 나타낸다. N→N에 비해 N→A는 에어로졸 상황이 발생하면서 전체적으로 정확도가 평균 42.15% 하락한다. 그림 4의 ROPE에 대한 N→N과 N→A는 표 3의 실험 결과를 정성적으로 나타낸다. 특히, N→A의 cat은 N→N에 비해 비교적 윤곽 정보가 부족하여 키포인트 검출을 못한 결과를 보인다. 또한 glue에서는 키포인트를 검출했지만 크기를 다르게 예측한다. 이는 에어로졸 상황으로 인해 물체의 외형이 가려져 물체의 정확한 6-자유도 자세 추정이 어렵기 때문이다.

앞선 실험 내용을 바탕으로, 본 연구진은 에어로졸 상황의 성능 개선을 위해 추가적인 실험을 진행한다. 표 3의 N+A→N과 N+A→A는 ROPE에서 사용되는 배치 증강(Batch Augmentation)[28] 기법에 에어로졸을 추가하여 학습 및 평가한 결과를 나타낸다. N→N과 N+A→N을 비교해 보았을 때,

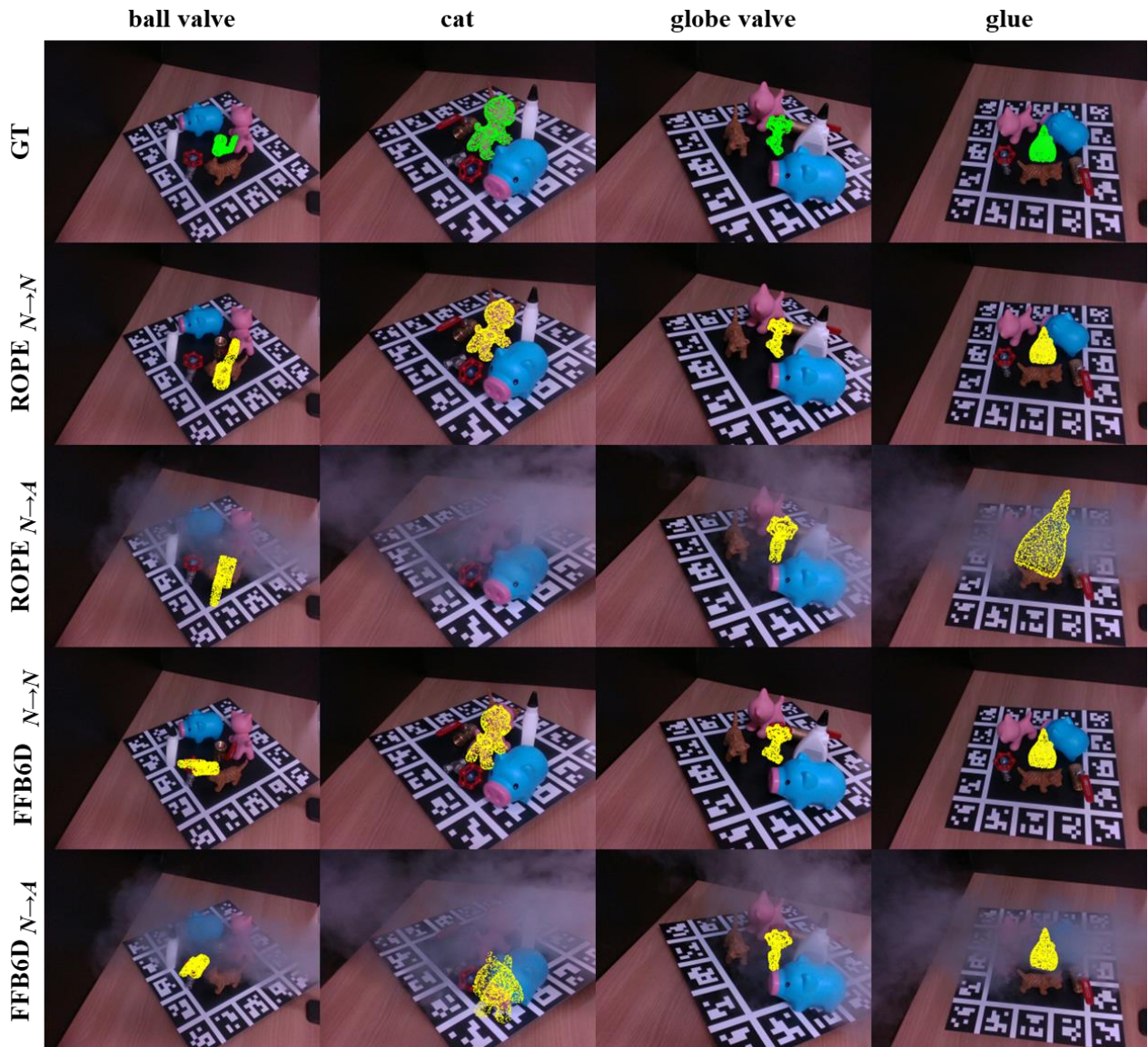


그림 4. 정상 상황(N) 학습 후 정상 상황(N)과 에어로졸 상황(A)의 ROPE 및 FFB6D 정성적 결과.

Fig. 4. ROPE[14] and FFB6D[15] qualitative results for normal and aerosol conditions.

에어로졸 상황을 학습할 경우 정상 상황에서 소폭의 성능 하락을 보인다. 하지만 $N \rightarrow A$ 와 $N+A \rightarrow A$ 를 비교해 보았을 때, 에어로졸 상황에 대한 예측을 수행했을 때 평균적으로 약 2배인 19.29%의 성능 개선을 보인다. 이를 통해 학습 과정에 에어로졸을 포함할 경우 성능 개선이 가능함을 확인하였다.

5. RGB-D 기반 방법론

표 4은 FFB6D에 제안한 데이터셋을 적용한 실험 결과이다. 기존의 FFB6D는 합성 데이터를 생성하여 모델을 충분히 학습시킨다. 본 논문에서 제안한 데이터에 대한 평가를 위해 동일한 방식으로 약 10,000장의 합성 데이터를 생성하여 학습을 진행한다. 그림 3과 같이 에어로졸이 발생할 경우 RGB 영상 뿐만 아니라 깊이 정보에도 노이즈가 발생하여 $N \rightarrow N$ 에 비해 $N \rightarrow A$ 는 정확도가 평균 29.34% 하락한다. 이는 RGB 영상과 깊이 정보 모두 노이즈가 발생함에 따라 두 모달리티를 동시에 고려하는 양방향 융합 모듈에 악영향을 주기 때문이다. 그림 4의 FFB6D에 대한 $N \rightarrow N$ 과 $N \rightarrow A$ 를 통해 표 4의 실험 결과를 정성적으로 확인할 수 있다. 에어로졸 상황에서 FFB6D는 glue를 제외하고 모두 옳지 않은 6-자유도 자세 추정을 수행하고 있다. 이는 RGB 영상과 깊이 정보 모두 노이즈가 발생함으로써 두 정보가 상호보완적인 역할을 하지 못하기 때문에 RGB-D 기반의 방법론도 에어로졸 상황에 강인하지 못하다는 것을 확인할 수 있다. 하지만 FFB6D는 단일 RGB 영상만을 이용하는 ROPE와 다르게 깊이 정보를 고려한다. 또한, 그림 4의 glue에 대한 정성적 결과를 통해 깊이 정보를 추가로 활용하는 RGB-D 기반의 방법론이 에어로졸이 발생하였을 때 크기 정보 추정에 도움을 주는 것을 확인하였다.

깊이 영상을 사용하는 FFB6D의 성능 개선을 위해 단일 RGB 영상 증강과 같은 기법을 사용할 수 있으나, 깊이 영상에 동일한 노이즈를 적용하는 것에는 어려움이 있다. 또한, 에어로졸 상황을 모사하는 깊이 영상을 합성하기 위해 추가적인 렌더링이 필요하다는 기술적 한계가 있다. 이러한 이유로 인해 RGB-D에 대한 추가적인 성능 개선 실험은 진행하지 않았다.

V. 결론

본 연구는 에어로졸 상황을 고려한 새로운 데이터셋을 제안한다. 해당 데이터셋은 같은 프레임에 대해 정상 상황과 에어로졸 상황을 쌍으로 제공하며, 이를 통해 단일 RGB 방법론과 RGB-D 방법론을 사용하여 정상 상황과 에어로졸 상황에 대한 6-자유도 자세 추정을 수행하였다.

본 연구에서 제안한 데이터셋에 깊이 정보를 보완하는 깊이 보완(Depth Completion), RGB에 대한 연무 제거(Dehazing)와 같은 여러 방법론들이 성능 개선을 위한 방향으로 활용될 것으로 기대한다. 더불어 본 연구자들은 6-자유도 자세 추정이 에어로졸과 같은 이상 상황 속에서 조작해야 하는 대상뿐만 아니라 다양한 물체에 대해 강인하게 작동하기 위한 데이터셋으로 확장하는 연구를 지속적으로 하고자 한다.

REFERENCES

[1] A. Collet, M. Martinez, and S. S. Srinivasa, "The moped framework: Object recognition and pose estimation for

manipulation," The international journal of robotics research, pp. 1284-1306, vol. 30, no. 10, Oct, 2011.

[2] J. Tremblay, T. To, B. Sundaralingam, Y. Xiang, D. Fox and S. Birchfield, "Deep object pose estimation for semantic robotic grasping of household objects," ArXiv preprint arXiv:1809.10790, 2018.

[3] M. Zhu, K. G. Derpanis, Y. Yang, S. Brahmabhatt, M. Zhang, C. Phillips, M. Lecce, and K. Daniilidis, "Single image 3d object detection and pose estimation for grasping," in Robotics and Automation (ICRA), 2014 IEEE International Conference on, IEEE, pp. 3936-3943, 2014.

[4] E. Marchand, H. Uchiyama, and F. Spindler, "Pose estimation for augmented reality: A hands-on survey," IEEE transactions on visualization and computer graphics, vol. 22, no. 12, pp. 2633-2651, 2016.

[5] E. M-Eppstein, "Project tango," in ACM SIGGRAPH 2016 Real-Time Live!, pp. 25-25, 2016.

[6] X. Chen, H. Ma, J. Wan, B. Li, and T. Xia, "Multi-view 3d object detection network for autonomous driving," Proceedings of the IEEE conference on Computer Vision and Pattern Recognition, pp. 1907-1915, 2017.

[7] A. Geiger, P. Lenz, and R. Urtasun, "Are we ready for autonomous driving? the kitti vision benchmark suite," Proceedings of the IEEE Computer Vision and Pattern Recognition (CVPR), IEEE, pp. 3354-3361, 2012.

[8] D. Xu, D. Anguelov, and A. Jain, "Pointfusion: Deep sensor fusion for 3d bounding box estimation," Proceedings of the IEEE Computer Vision and Pattern Recognition (CVPR), pp. 244-253, 2018.

[9] "New technologies to improve Amazon employee safety," 2021. [Online]. Available: <https://www.aboutamazon.com/news/innovation-at-amazon/new-technologies-to-improve-amazon-employee-safety/>.

[10] M. Lyu, Y. Zhao, C. Huang, and H. Huang, "Unmanned Aerial Vehicles for Search and Rescue: A Survey," Remote Sensing 15, no. 13: 3266, 2023.

[11] J. Nah, H. Kim, C. Kim, D. Kang, and J. Oh, "Intelligent CCTV identifying Missing Person based on Deep Learning," The 37th Institute of Control, Robotics and Systems Annual Conference (in Korean), pp. 726-727, 2022.

[12] G. Kim and S. Kim, "Mechanism development of catamaran-type rescue robot with life-saving ring launcher," Journal of Institute of Control, Robotics and Systems (in Korean), vol. 28, no. 5, 414-419. 2022.

[13] J. Delmerico, S. Mintchev, A. Giusti, B. Gromov, K. Melo, T. Horvat, C. Cadena, M. Hutter, A. Ijspeert, D. Floreano, L. M-Gambardella, R. Siegwart, D. Scaramuzza, "The current state and future outlook of rescue robotics," Journal of Field Robotics, vol. 36, no. 7, 1171-1191, 2019.

[14] Y. He, W. Sun, H. Huang, J. Liu, H. Fan and J. Sun, "Pvn3d: A deep point-wise 3d keypoints voting network for 6dof pose estimation," Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp. 11632-11641, 2020.

[15] B. Chen, T. Chin and M. Klimavicius, "Occlusion-robust object pose estimation with holistic representation," Proceedings of the IEEE/CVF Winter Conference on Applications of Computer

Vision (WACV), pp. 2929-2939, 2022.

- [16] Y. He, H. Huang, H. Fan, Q. Chen and J. Sun, "Ffb6d: A full flow bidirectional fusion network for 6d pose estimation," Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp. 3003-3013, 2021.
- [17] S. Hinterstoisser, S. Holzer, C. Cagniart, S. Ilic, K. Konolige, N. Navab, and V. Lepetit. "Multimodal templates for real-time detection of texture-less objects in heavily cluttered scenes," Proceedings of the 2011 IEEE/CVF International Conference on Computer Vision (ICCV), IEEE, pp. 858-865, Nov. 2011.
- [18] E. Brachmann, A. Krull, F. Michel, S. Gumhold, J. Shotton and C. Rother, " Learning 6d object pose estimation using 3d object coordinates," Proceedings of the Computer Vision-ECCV 2014: 13th European Conference(ECCV), Zurich, Switzerland, vol. 2, no. 13, pp. 536-551, Sep. 2014.
- [19] T. Hodan, P. Haluza, Š. Obdržálek, J. Matas, M. Lourakis and X. Zabulis, "T-LESS: An RGB-D dataset for 6D pose estimation of texture-less objects," Proceedings of the IEEE Winter Conference on Applications of Computer Vision (WACV), IEEE, pp. 880-888, 2017.
- [20] R. Kaskman, S. Zakharov, I. Shugurov and S. Ilic, "Homebreweddb: Rgb-d dataset for 6d pose estimation of 3d objects," Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV), pp. 0-0, 2019.
- [21] S. Tyree, J. Tremblay, T. To, J. Cheng, T. Mosier, J. Smith and S. Birchfield, "6-DoF pose estimation of household objects for robotic manipulation: An accessible dataset and benchmark," Proceeding of the 2022 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), pp. 13081-13088, IEEE, Oct. 2022.
- [22] Creaform Inc. HandySCAN BLACK. 2024. [Online]. Available: <https://www.creaform3d.com/en/portable-3d-scanner-handyscan-3d>.
- [23] Creaform Inc. Go!SCAN SPARK. 2024. [Online]. Available: <https://get.creaform3d.com/lp-goscan-3d>.
- [24] C. Ancuti, C. Ancuti, M. Sbert and R. Timofte, "Dense-haze: A benchmark for image dehazing with dense-haze and haze-free images," In 2019 IEEE international conference on image processing (ICIP), IEEE, pp. 1014-1018, 2019.
- [25] C. Ancuti, C. Ancuti, R. Timofte and C. Vleeschouwer, "I-HAZE: A dehazing benchmark with real hazy and haze-free indoor images," Proceedings of the Advanced Concepts for Intelligent Vision Systems: 19th International Conference(ACIVS), Springer International Publishing, Sep. 2018.
- [26] 6D-PAT. (2.5.3). Accessed: Nov. 11, 2022. [Online]. Available: <https://github.com/florianblume/6D-pat>.
- [27] D. Kingma and J. Ba, "Adam: A method for stochastic optimization," arXiv preprint arXiv:1412.6980, 2014.
- [28] E. Hoffer, T. Ben-Nun, I. Hubara, N. Giladi, T. Hoeffler and D. Soudry, "Augment your batch: Improving generalization through instance repetition." Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition(CVPR), pp. 8129-8138, 2020.



양희진

2023년 세종대학교 지능기전공학부 스마트기기공학 졸업. 2023년~현재 동 대학원 인공지능융합학과 지능기전공학과 석박통합과정 재학 중. 관심분야는 컴퓨터비전, 로보틱스, 물체 자세 추정



이승현

2023년 세종대학교 지능기전공학부 무인이동체공학 졸업. 2023년~현재 동 대학원 지능기전공학과 석사과정 재학 중. 관심분야는 컴퓨터비전, 로보틱스, 물체 자세 추정



김태주

2021년 세종대학교 지능기전공학부 무인이동체공학 졸업. 2023년~현재 동 대학원 지능기전공학과 석사과정 재학 중. 관심분야는 컴퓨터비전, 로보틱스, 물체 자세 추정



최유경

2006년 숭실대학교 정보통신전자공학과 졸업. 2008년 연세대학교 전기전자공학부 석사, 2018 한국과학기술원 전기전자공학과/로봇학제전공 박사, 2018년~현재 세종대학교 지능기전공학과 부교수. 관심분야는 컴퓨터비전, 머신러닝, 로보틱스