Check for updates

Journal of KIIT. Vol. 24, No. 10, pp. 175-185, Oct. 31, 2026. pISSN 1598-8619, eISSN 2093-7571
http://dx.doi.org/10.14801/jkiit.2026.24.10.175

실내 동적 환경의 타겟 모호성 완화를 위한 LiDAR 공간 주의집중 메커니즘

김건웅*[1], 고운별*[2], 정현준*[3], 이석훈*[4], 정동원*[5]

LiDAR-Guided Spatial Attention Mechanism for Target Ambiguity Mitigation in Indoor Dynamic Environments

Geonwoong Kim*[1], Unbyeol Ko*[2], Hyunjun Jung*[3], Sukhoon Lee*[4], and Dongwon Jeong*[5]

본 연구는 2026년도 과학기술정보통신부 및 정보통신기획평가원의 "SW중심대학사업" 지원을 받아 수행되었음 (2023-0-00065)

요 약

자율추종 로봇은 실내 동적 환경에서 특정 사용자를 지속 인식·추종해야 하는 서비스 로봇의 핵심 요소이나, 카메라 기반 인식은 거리 정보를 잃어 외형이 유사한 타인을 오인식하는 타겟 모호성 문제를 겪는다. 본 논문은 이를 완화하고자 LiDAR 거리 정보로 만든 가우시안 마스크를 OSNet Re-ID 유사도와 결합해 후보를 재순위화하는 공간 주의집중 메커니즘을 실제 로봇에 통합하여 실증한다. 실험 결과 오인식률은 26.7%에서 8.9%로 17.8%p 감소하였으나 경계선 수준의 유의성(p=0.051)에 그쳤고, ID 스위칭은 12에서 4회(-66.7%), FPS는 8.2% 저하되었다. 본 논문은 이론 수준에 머물렀던 메커니즘이 저전력 임베디드 환경 속에서 실시간으로 동작하며 오인식을 감소시킨다는 것을 실증하였다는 점에서 의의가 있다.

Abstract

Autonomous following robots must continuously recognize and follow a specific user in indoor dynamic environments, but camera-only recognition loses depth information, causing target ambiguity where a similar-looking bystander is misidentified as the registered user. This paper mitigates this by integrating, on a real robot, a spatial attention mechanism combining a LiDAR-based Gaussian mask with OSNet Re-ID similarity for candidate re-ranking. Experiments cut the false-target rate from 26.7% to 8.9% (17.8 points), though only at borderline significance (p=0.051); ID switches fell from 12 to 4 (-66.7%), and FPS dropped about 8.2%. This work confirms real-time, low-power operation and reduced false-target selection on a low-power embedded robot.

Keywords

2D LiDAR, spatial attention, person re-identification, target ambiguity, sensor fusion

* 국립군산대학교 소프트웨어학과(*[3,5] 교신저자)
- ORCID[1]: https://orcid.org/0009-0009-1812-9949
- ORCID[2]: https://orcid.org/0009-0008-9432-5573
- ORCID[3]: https://orcid.org/0000-0002-6717-1395
- ORCID[4]: https://orcid.org/0000-0002-3390-5602
- ORCID[5]: https://orcid.org/0000-0001-9881-5336

· Received: Sep. 16, 2026, Revised: Oct. 08, 2026, Accepted: Oct. 11, 2026

· Co-corresponding Author1: Hyunjun Jung
Dept. of Software Science & Engineering, Kunsan National University, 558, Daehak-ro, Gunsan, Jeollabuk-do, Republic of Korea
Tel.: +82-63-469-8911, Email: junghj85@kunsan.ac.kr
· Co-corresponding Author2: Dongwon Jeong
Dept. of Software Science & Engineering, Kunsan National University, 558, Daehak-ro, Gunsan, Jeollabuk-do, Republic of Korea
Tel.: +82-63-469-8911, Email: djeong@kunsan.ac.kr

Ⅰ. 서 론

최근 Physical AI 및 로봇 산업의 발전에 따라 카메라와 LiDAR 등 다양한 센서들을 활용하여 물리적 환경을 인지하고 스스로 판단 및 이동하는 로봇에 관한 연구가 활발히 진행되고 있다. 2024년 로봇산업 실태조사[1]에 따르면 국내 로봇 시장은 연평균 4.6% 이상의 성장률을 보여왔으며, 서비스 영역이 확장됨에 따라 위치 정보 기반 서비스에 대한 수요 역시 증가하고 있다. 로봇이 운용되는 환경이 점차 다양해지고 복잡해짐에 따라 동적 환경에서의 안정적인 객체 추적 및 실시간 고정밀 인식 그리고 장애물 회피 기술은 자율주행 기술의 상용화를 위한 핵심 과제로 다양한 연구가 활발히 진행되고 있다.

자율주행 시스템의 핵심 구성 요소는 센서를 활용한 환경 인지 기능이다. 인지 대상에 따라 신호등 및 차선과 같은 정적 객체 인식과 보행자 및 차량과 같은 동적 객체의 인식 등으로 구분할 수 있다. 신호등이나 차선과 같은 정적 인프라의 인식은 상대적으로 일정한 공간적·시각적 특성을 활용할 수 있는 반면, 보행자나 차량과 같은 동적 객체의 인식 및 추적은 객체의 위치와 외형이 지속적으로 변화하므로 높은 수준의 실시간성과 강건성을 동시에 요구한다. 특히 특정 개인을 지속적으로 추종해야 하는 서비스 로봇 응용 분야에서는 동적 객체의 정확한 식별과 추적이 핵심적인 기술로 자리 잡고 있다. 이에 본 논문은 자율주행 인지 기술 중에서도 객체 인식 및 추적 분야에 초점을 둔다.

그동안 다양한 비전 기반 객체 추적 시스템이 제안되고 활용되어 왔으나, 복잡한 배경이나 외형이 유사한 타겟이 혼재하는 실내 환경에서는 객체 간 교차, 폐색 및 ID(Identity) 스위칭 등 다양한 타겟 모호성 문제가 지속적으로 발생하고 있다[2]-[4]. 이는 3차원 공간을 2차원 영상으로 투영하는 카메라 센서의 특성상 객체 간 거리 및 깊이 정보를 직접적으로 확보하기 어렵다는 한계를 보여준다. 즉, 실제 공간에서는 서로 분리된 두 객체라도 카메라 시점에서는 겹쳐 보일 수 있으며, 실외 환경보다 상대적으로 협소하고 밀집된 실내 환경에서는 이러한 현상이 더욱 빈번하게 발생하며 로봇의 추적 신뢰도를 저하시킬 수 있다.

이러한 센서의 근본적인 한계는 실제 로봇 환경에서도 확인된다. 연구에 앞서 비전 기반 객체 탐지[5]와 개인화 재식별(Re-ID, Person Re-Identification)[6] 알고리즘을 결합하여 객체 추종 자율주행 로봇을 구현하고 실제 환경에서 테스트를 진행하였다. 그 결과, 단일 비전 영상만을 활용하는 경우 유사한 색상의 의상을 착용한 타인을 등록된 사용자로 오인하여 그 대상을 추종하는 사례가 관찰되었다. 이는 앞서 언급한 타겟 모호성 문제가 실제 서비스 환경에서도 이러한 추적 오류를 유발할 수 있음을 보여준다.

이를 극복하기 위해, 2D LiDAR의 수평 거리 정보를 공간적 주의집중 가중치로 변환하여 타겟 모호성을 완화하는 아키텍처가 제안된 바 있다[7]. 핵심은 LiDAR 스캔으로 수평면상의 타겟 위치를 먼저 특정한 뒤, 이를 시각 데이터 분석 가중치로 융합하는 것이다. 고가의 3D LiDAR 대비 경량·경제적인 2D LiDAR를 도입함으로써 실내 서비스 로봇의 현실적 제약을 고려하면서도, 단일 비전 센서로는 어려운 수평 거리 기반 객체 분리 성능을 확보하고자 하였다. 그러나 해당 연구는 이론적 타당성 검토 단계에 머물렀으며, 실제 하드웨어 플랫폼에서의 실증은 한계로 남았다.

이에 본 논문은 앞서 언급한 아키텍처[7]를 실제 자율추종 로봇에 적용해 그 실효성을 검증한다. 구체적으로는 LiDAR 기반 거리 정보와 OSNet(Omni-Scale Network) Re-ID의 유사도 점수를 융합하여 후보군을 재순위화함으로써 유사 외형 객체에 대한 오인식을 억제한다. 본 논문의 주요 기여점은 다음과 같다. 첫째, 이론적으로 제시된 공간 주의집중 융합 메커니즘을 실제 로봇에 적용하고 그 실효성을 실증적으로 검증하였다. 둘째, 실제 환경 테스트에서 발생한 추적 실패 사례를 정량적으로 분석한 후, 제안 방법을 적용한 결과 오인식률이 26.7%에서 8.9%로 감소하였고, 17.8%의 개선 효과를 확인하였다.

본 논문의 구성은 다음과 같다. 2장에서는 비전 기반 재식별, 2D LiDAR 기반 실내 인지, 센서 융합 기반 다중 객체 추적, 트랜스포머 기반 희소 어텐

션, 임베디드 환경에서의 실시간 객체 탐지 및 이론과 실증 간극과 관련된 연구를 살펴본다. 3장에서는 기존 시스템 아키텍처와 제안하는 LiDAR-Re-ID 융합 공간 주의집중 메커니즘을 식 및 알고리즘과 함께 기술한다. 4장에서는 실험 환경과 정량적으로 평가한 결과를 제시하고, 5장에서는 결론 및 향후 연구 과제를 논한다.

II. 관련 연구

2.1 비전 기반 다중객체추적과 개인화 재식별

비전 기반 객체 추적은 외형 특징 추출과 칼만 필터를 활용한 프레임 간의 데이터 결합 방식으로 발전해 왔으나[8], 폐색이나 유사 외형 객체가 인접한 상황에서 ID 스위칭 오류가 발생하는 한계가 있다. 최근에는 사전 등록된 사용자의 영상에서 특징 벡터를 추출하여 특정 개인을 재식별하는 개인화 Re-ID 접근이 서비스 로봇 분야에 활용되고 있다[9]. OSNet[6]은 다중 스케일 특징을 옴니스케일 방식으로 학습하여 경량 임베디드 환경에서도 준수한 재식별 성능을 제공하는 대표적인 모델이며, Market-1501[10]은 다양한 카메라 각도와 가림 현상을 포함하여 재식별 모델 학습에 널리 활용되는 데이터셋이다. 그러나 비전 단독 방식은 2차원 영상에서 발생하는 거리 및 깊이 정보 손실을 해결하기 어려우며, 본 연구팀의 선행 실험에서도 유사한 색상의 의상을 착용한 타인을 등록 사용자로 오인식하는 사례가 관찰되었다.

2.2 2D LiDAR 기반 실내 객체 인지

2D LiDAR를 활용한 실내 객체 인지 연구도 다수 진행되어 왔다. DR-SPAAM[4]은 공간 주의집중과 자기회귀 모델을 결합하여 2D LiDAR 스캔만으로 실시간 인물 탐지를 수행하였으나, 외형 정보 없이 LiDAR 데이터만을 활용하기 때문에 유사한 움직임을 보이는 다수의 사람이 교차하는 상황에서 ID 스위칭이 발생할 수 있다. 또한 2D LiDAR기반 이동 로봇을 활용하여 낙상 감지 등 실내 활동 인식을 수행한 연구[11]도 있으나, 이는 특정 타겟을 지속적으로 추적하는 개인화 추종과는 목적이 다르다. 이러한 연구들은 LiDAR의 거리 정보를 활용한다는 장점이 있으나, 외형 기반 개인 식별과 결합되지 않아 특정 개인을 지속적으로 식별·추적하는 데에는 한계가 있다.

2.3 LiDAR-카메라 센서 융합 다중객체추적

LiDAR와 비전을 함께 활용하는 멀티모달 객체 추적 연구도 활발히 진행되고 있다. DeepFusionMOT[2]은 카메라와 LiDAR의 상이한 특성을 활용하는 심층 연관 메커니즘을 통해 추적 정확도와 처리 속도의 균형을 도모하였다. CAMO-MOT[3]은 외형 및 움직임 정보를 함께 활용하여 KITTI 데이터셋에서 낮은 수준의 ID 스위칭 성능을 보였다. 그러나 이러한 연구들은 주로 실외 자율주행 환경을 대상으로 하며, 3D LiDAR를 활용하는 경우가 많다는 점에서 저비용·경량 2D LiDAR와 개인화 재식별을 결합하는 본 연구와 차이가 있다.

2.4 희소 어텐션 및 LiDAR 기하 정보 결합 메커니즘

트랜스포머 계열 연구에서는 멀티헤드 자기 주의집중이 전체 탐색 영역에 대해 연산을 수행함에 따라 불필요한 정보까지 처리하는 한계를 보완하기 위해 적응형 희소 어텐션을 적용하려는 연구가 이루어지고 있다[12]. 그러나 높은 연산 복잡도는 Jetson과 같은 저전력 임베디드 장치에서 실시간 처리를 어렵게 할 수 있다. LiDAR 융합 메커니즘[7]에서는 이러한 문제를 해결하기 위해 2D LiDAR의 수평 거리 정보를 가우시안 공간 주의집중 가중치로 변환하고, 임계값 이상의 유효 토큰에 대해서만 어텐션 연산을 수행하는 희소 연산 구조를 제안하였다. 해당 연구는 연산 복잡도를 $O(N^2)$(N: 전체 토큰 수)에서 $O(k^2)$(k: 유효 토큰 수, $k \ll N$)으로 감소시킬 수 있음을 이론적으로 제시하였으나, 실제 로봇 플랫폼에서의 실증과 개인화 재식별 시스템과의 결합은 수행하지 않았다. 본 연구는 해당 메커니즘을 실제 자율 추종 로봇의 Re-ID 파이프라인에 적

용하고 정량적으로 검증한다는 점에서 기존 연구와 차별화된다.

2.5 임베디드 환경에서의 실시간 객체 탐지 경량화 연구

서비스 로봇에 탑재되는 객체 탐지 모델은 정확도뿐 아니라 저전력 임베디드 하드웨어에서의 실시간성도 확보해야 한다. YOLO 계열 모델은 단일 단계(Single-stage) 구조로 이러한 요구를 충족해 왔으며, 최신 YOLOv7[13]은 학습 기법 개선을 통해 실시간 객체 탐지기 중 최고 수준의 속도-정확도 균형을 달성하였다. 이후 연구는 Jetson 보드에서 전력 모드별 처리 속도(FPS, Frames Per Second)를 비교하였다[14]. 이러한 연구들은 이론적 성능 지표만으로는 실제 엣지 디바이스에서의 동작을 예측하기 어렵다는 점을 보여주며, 본 연구가 이론적으로 제안된 희소 어텐션 구조를 저전력 Jetson 보드 위에서 직접 실증하고자 하는 문제의식과 맥을 같이한다.

2.6 이론적 설계와 실제 로봇 환경 간의 성능 차이

로봇 학습 및 인지 알고리즘 분야에서는 시뮬레이션이나 이론적 설계 단계에서 검증된 방법이 실제 로봇에 적용될 때 성능이 저하되는 현상이 반복적으로 보고되어 왔으며, 이를 시뮬레이션-실제 간극(Sim-to-real gap)이라 한다. 대표적인 조사 연구[15]는 이러한 성능 차이가 센싱과 구동 과정에서 발생할 수 있으며, 정교한 시뮬레이션이 실제 환경에서의 성능을 반드시 보장하지는 않는다고 지적한다. 이러한 문제는 주로 강화학습 기반 제어 분야에서 논의되어 왔으나, 실제 센서와 하드웨어를 사용하는 인지 알고리즘에도 적용될 수 있다. 특히 LiDAR 센서 노이즈, 카메라-LiDAR 캘리브레이션 오차 및 임베디드 하드웨어의 연산 제약 등은 이론적으로 설계된 메커니즘의 실제 성능에 영향을 미칠 수 있다. 따라서 본 연구에서는 LiDAR 기반 공간 주의집중 메커니즘[7]을 실제 자율 추종 로봇에 통합하고, 실제 환경에서의 성능을 정량적으로 검증한다.

III. LiDAR-Re-ID 융합 시스템 및 알고리즘

3.1 기존 시스템 구성

본 논문에서는 기존에 구현했던 비전 기반 객체 탐지 및 개인화 재식별(Re-ID) 기반 자율 추종 로봇 시스템을 바탕으로 연구를 진행하였다. 기존 시스템은 YOLO[5]를 활용해 프레임 내 사람 후보를 탐지하고, OSNet[6] 기반 특징 추출 모델을 통해 각 후보와 사전 등록된 사용자 간의 코사인 유사도를 계산한다. 이후 임계값 이상인 후보 중 가장 높은 유사도를 갖는 객체를 해당 프레임의 최종 추종 대상으로 선정한다.

기존 시스템은 단일 객체 환경에서 등록 사용자를 안정적으로 식별할 수 있었다. 그러나 다수의 사람이 존재하는 혼잡한 환경에서는 등록 사용자와 유사한 의상 색상이나 외형을 가진 타인을 오인식하는 문제가 확인되었다. 특히 유사한 색상의 의류를 착용한 사람이 카메라 프레임에 함께 등장할 경우, 이러한 오류가 반복적으로 발생하였다.

또한 기존 시스템은 매 프레임마다 탐지와 Re-ID 매칭을 독립적으로 수행하므로, 프레임 간 타겟의 위치 및 이동 연속성을 충분히 반영하지 못한다. 이에 따라 유사 외형 객체가 인접하거나 교차하는 상황에서 타겟이 다른 객체로 전환될 수 있다. 이는 카메라 단일 센서만으로는 외형 정보는 비교할 수 있으나, 후보 간 실제 거리와 공간적 분리 정도를 명확히 판단하기 어렵기 때문이다. 따라서 본 논문에서는 2D LiDAR 기반 공간 정보를 기존 Re-ID 결과와 융합하여 타겟 식별 성능을 향상하고자 한다.

3.2 LiDAR 융합 모듈의 통합

본 논문에서는 기존 시스템의 AI 계층에 선행 연구[7]에서 제안된 2D LiDAR 기반 공간 주의집중 메커니즘을 통합하였다. 2D LiDAR는 3D LiDAR보다 가볍고 경제적이므로, 비용과 탑재 무게에 제약이 있는 실내 서비스 로봇에 적합하다. 또한 카메라 단독으로 확보하기 어려운 수평 거리와 객체 간 공간 분리 정보를 제공할 수 있다.

LiDAR 센서는 카메라와 함께 로봇 전방의 유사한 관측 영역을 탐지하도록 장착하였다. LiDAR 데이터는 별도 스레드에서 지속적으로 수신되며, 최신 스캔 데이터는 버퍼에 유지된다. 수집된 좌표는 센서 간 위치 관계를 고려한 좌표 변환을 거쳐 카메라 이미지 좌표계로 매핑된다.

변환된 LiDAR 정보는 후보 객체 주변의 위치를 강조하는 가우시안 공간 마스크로 생성된다. 이 마스크는 YOLO 기반 탐지 결과 및 OSNet 기반 유사도 점수와 결합되며, 융합 재순위화 단계에서 최종 타겟 점수 산출에 활용된다. 즉, 외형 유사도뿐 아니라 기존 타겟과의 공간적 일관성도 함께 고려하여 추종 대상을 선정한다.

그림 1은 제안 시스템의 전체 파이프라인을 나타낸다. LiDAR 기반 공간 정보 처리와 카메라 기반 객체 탐지 및 특징 추출은 병렬적으로 수행되며, 이후 융합 재순위화 단계에서 두 정보를 결합하여 최종 타겟을 선정한다. 각 단계의 구체적인 식과 절차는 3.3절부터 3.6절까지 설명하며, 주요 기호는 표 1에 정리하였다.

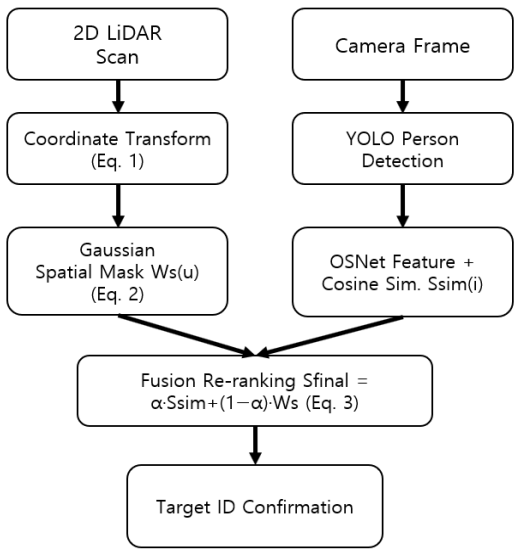


그림 1. 제안 시스템의 전체 파이프라인
Fig. 1. Full pipeline of proposed system

표 1. 주요 기호 정의
Table 1. Definition of key notations

Symbol	Meaning
r, θ	LiDAR measured distance, scan angle
x, y	Cartesian coordinates in the robot coordinate frame
u, v	Pixel coordinates in the image coordinate frame
μu	Horizontal image coordinate of the target center
$ub(j)$	Horizontal center of the bounding box of candidate j
σ	Standard deviation of the Gaussian mask
$Ws(u)$	Spatial attention mask weight
$Ssim(i)$	OSNet-based appearance similarity score of candidate i
$Sfinal(i)$	Final discrimination score of candidate i
a	Weighting hyperparameter between appearance similarity and distance signal
τ_sim	Similarity threshold for confirming a target candidate (τ_sim = 0.72)

3.3 좌표 변환 및 이미지 좌표계 매핑

2D LiDAR 센서로부터 획득한 스캔 데이터의 극좌표 (r, θ)는 로봇 좌표계의 직교 좌표 (x, y)로 변환된다. 이 변환은 식 (1)과 같이 정의된다.

$$\begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} r\cos\theta \\ r\sin\theta \end{bmatrix} \tag{1}$$

여기서 r은 LiDAR와 객체 간의 측정 거리, θ는 스캔 각도이다. 변환된 좌표는 카메라의 수평 시야각 및 캘리브레이션으로 이미지 좌표계의 수평축 u로 매핑되어, 비전 센서만으로는 파악하기 어려운 객체 간 거리 관계를 수평 평면상에서 분리한다. 본 논문에서는 수평 시야각(Horizontal Field of View, HFOV = 60.0°)과 LiDAR-카메라 간 각도 오프셋(ANGLE_OFFSET = 0.0°)을 캘리브레이션 파라미터로 사용하여 LiDAR 각도 θ를 수평 픽셀 위치 u로 변환한다.

다만 2D LiDAR는 단일 높이의 수평면만 스캔하므로, 이 변환으로 확보되는 정보는 객체의 수평 위치(u)에 한정되며 수직 방향(v) 정보는 제공하지 않는다. 따라서 3.4절의 공간 마스크도 수평축을 중심으로 정의되며, 수직 방향의 판별력 보완은 5장의 향후 과제와 연결된다.

3.4 가우시안 공간 주의집중 마스크 생성

LiDAR 투영점과 대응하는 타겟 수평 좌표 μu를 기준으로 가우시안 형태의 공간 주의집중 마스크를 생성한다. 마스크 가중치 W_s는 식 (2)와 같이 정의된다.

$$W_s(u) = \exp\left(\frac{-(u-\mu_u)^2}{2\sigma^2}\right) \quad (2)$$

여기서 표준편차 σ는 가우시안 분포의 범위를 결정하는 파라미터로, LiDAR 측정 거리와 타겟 크기를 고려하여 설정하며, 원거리 객체는 넓은 영역을, 근거리 객체는 좁은 영역을 고려하도록 설계하였다. 후보 i의 마스크 값은 박스의 수평 중심 $ub(i)$에서 식 (2)를 계산한 $W_s(i) = W_s(ub(i))$이다. [7]은 사전 신체 높이 정보로 수직 범위까지 포함한 2차원 마스크를 제안하였으나, 본 구현은 수평축 성분만 적용한다.

그림 2는 단일 다중 타겟 시나리오의 마스크 생성 방식을 보여주며, 타원은 수평 방향 마스크 범위를 나타낸다. 단일 타겟은 중심에서 마스크 값이 높고 주변으로 갈수록 감소하며, 다중 타겟에서는 LiDAR 거리 정보에 따라 유사 외형 객체에도 서로 다른 마스크 값이 부여되어 구분된다.

3.5 LiDAR 마스크와 Re-ID 유사도의 결합 재순위화

기존 논문[7]에서는 마스크 W_s를 CNN 특징맵 F에 곱해 가중 특징맵 F'을 산출하는 데 그쳤으나 $(F'(u,v) = F(u,v)\cdot Ws(u,v))$, 본 연구에서는 이를 개인화 재식별 파이프라인에 적용하여, 프레임 내 탐지된 각 후보 (i)의 OSNet 외형 유사도와 LiDAR 기반 결합한다. 최종 판별 점수는 식 (3)과 같이 정의된다.

$$S_{\text{final}}(\text{i}) = \alpha \cdot \text{S}_{\text{sim}}(\text{i}) + (1-\alpha) \cdot \text{W}_{\text{s}}(\text{i}) \quad (3)$$

여기서 α는 외형 유사도와 공간 정보 간의 상대적인 가중치를 조절하는 하이퍼파라미터이다. 본 연구에서는 외형 정보를 주요 판별 근거로 유지하면서 LiDAR 기반 공간 정보를 보조적인 신호로 활용하기 위해 α=0.7로 설정하였다. 각 후보의 최종 점수를 비교하여 임계값 이상이면서 가장 높은 점수를 갖는 후보를 최종 타겟으로 선정한다. 이를 통해 외형 유사도가 높은 후보라도 LiDAR 기반 공간 정보와의 일치도가 낮은 경우 최종 점수가 감소하여 오인식 가능성을 낮출 수 있다.

LiDAR 시야 밖에 위치하여 $W_s(i)$를 산출할 수 없는 후보의 경우 $S_{final}(i) = S_{sim}(i)$로 설정하여 기존의 비전 기반 판별 방식으로 전환한다. 이를 통해 LiDAR 정보가 확보되지 않는 상황에서도 시스템의 타겟 판별 기능을 유지할 수 있다.

3.6 LiDAR-Re-ID 융합 기반 타겟 판별 알고리즘

그림 3은 전역 어텐션과 제안하는 희소 어텐션의 연산 범위를 비교한 것이다. 전역 어텐션은 전체 토큰 간 관계를 계산하므로 토큰 수 (N)에 대해 (O (N^2))의 연산 복잡도를 갖는다. 반면 제안 방식은 LiDAR 공간 정보를 이용해 유효 영역만 선택하고, 해당 토큰 수를 (k)로 제한하여 연산 범위를 (O (K^2)) 수준으로 감소시키는 것을 목표로 한다.

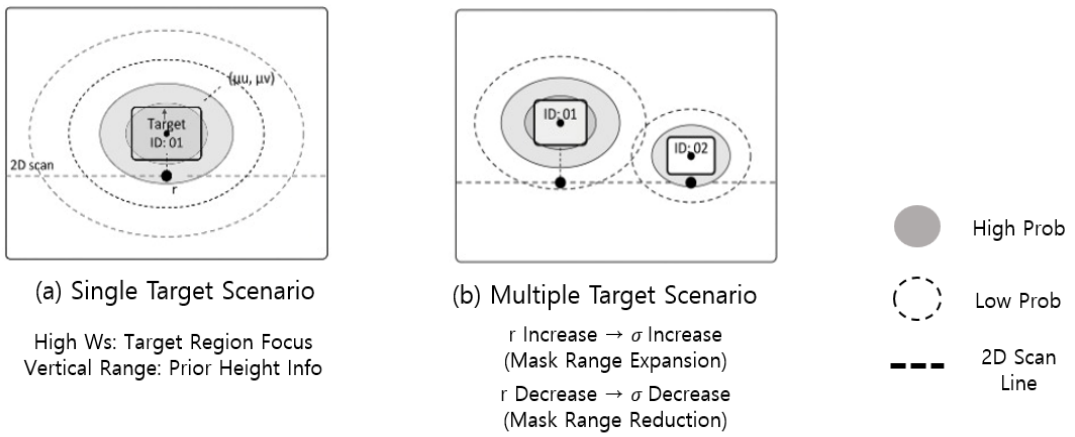


그림 2. LiDAR 거리 정보 기반 공간 마스크 생성 개념도
Fig. 2. Conceptual diagram of spatial mask generation based on LiDAR distance information

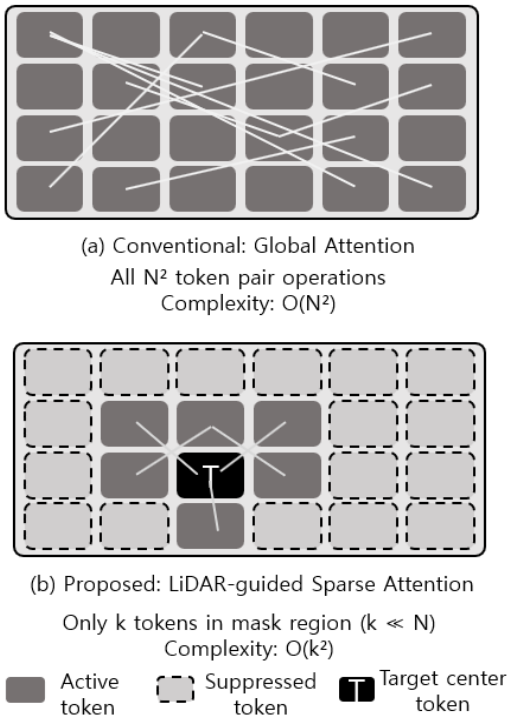


그림 3. 전역 어텐션과 제안 희소 어텐션 연산 범위 비교
Fig. 3. Comparison of global attention and proposed sparse attention computational range

유효 영역은 공간 마스크 값 ($W_s(u)$)가 임계값 이상인 수평 픽셀로 정의된다. 실내 서비스 로봇 환경에서는 사람 후보 수가 제한적이므로 일반적으로 (k ≪ N)의 관계가 성립한다. 따라서 불필요한 영역의 연산을 줄여 임베디드 환경에서의 실시간 처리에 기여할 수 있다.

다음은 앞서 설명한 좌표 변환, 마스크 생성, LiDAR-Re-ID 융합을 포함한 제안 알고리즘이다.

제시한 알고리즘 1은 다음과 같이 동작한다. Step 1에서는 LiDAR 스캔의 극좌표를 로봇 좌표계로 변환한 후 카메라 이미지 좌표계에 투영한다. Step 2에서는 YOLO를 이용하여 프레임 내 사람 후보를 탐지하고, 각 후보의 OSNet 특징과 등록 사용자 특징 간 코사인 유사도 S_{sim}을 계산한다. Step 3에서는 각 후보의 이미지 좌표를 기준으로 대응하는 LiDAR 투영점을 탐색하고 가우시안 공간 마스크 값 W_s를 산출한다. Step 4에서는 S_{sim}과 W_s를 가중합하여 S_{final}을 계산하고, 임계값을 만족하면서 가장 높은 점수를 갖는 후보를 최종 타겟으로 선정한다.

본 알고리즘은 프레임 내 탐지된 후보를 중심으로 융합 연산을 수행하므로 일반적인 실내 서비스 로봇 환경에서 후보 수가 제한적인 경우 연산 부담을 줄일 수 있다. 또한 LiDAR 기반 공간 정보를 활용하여 불필요한 영역의 연산을 제한하는 구조는 선행 연구[7]의 희소 연산 설계와 맥을 같이하며, Jetson과 같은 저전력 임베디드 환경에서의 실시간 처리에 기여할 수 있다.

알고리즘 1. LiDAR-Re-ID 융합 기반 타겟 판별 알고리즘
Algorithm 1. Target confirmation algorithm based on LiDAR-Re-ID fusion

Input: 2D LiDAR scan $S = \{(r_i, \theta_i)\}$, Camera frame I, registered feature f_user, similarity threshold τ_sim, weight α
Output: Confirmed target ID and tracking command

[Step 1: Coordinate Transform and Image Mapping]
for each (r_i, θ_i) in S do
  $x_i \leftarrow r_i \cos\theta_i$, $y_i \leftarrow r_i \sin\theta_i$
  $u_i \leftarrow$ Project(x_i, y_i, K_cam)
end for

[Step 2: Candidate Detection and Re-ID Similarity]
$\{b_j\} \leftarrow$ YOLO(I)
for each candidate b_j do
  $f_j \leftarrow$ OSNet(Crop(I, b_j))
  $S_{sim}(j) \leftarrow \cos(f_j, f_user)$
end for

[Step 3: Gaussian Spatial Mask Generation]
for each candidate b_j do
  $u_b(j) \leftarrow$ CenterX(b_j)
  $\mu_u(j) \leftarrow$ NearestProjection($u_b(j)$, $\{u_i\}$)
  $\sigma(j) \leftarrow$ ComputeSigma($r(j)$, target_size)
  $W_s(j) \leftarrow \exp(-(u_b(j) - \mu_u(j))^2/2\sigma(j)^2)$
end for

[Step 4: Fusion Re-ranking and Target Confirmation]
for each candidate j do
  $S_{final}(j) \leftarrow \alpha \cdot S_{sim}(j) + (1-\alpha) \cdot W_s(j)$
end for
$j^* \leftarrow$ argmax_j $S_{final}(j)$
if $S_{sim}(j^*) \geq \tau_sim$ and $S_{final}(j^*) \geq \tau_sim$ then
  return AssignTarget(j^*)
else
  return NotDetected

IV. 실험 및 평가

4.1 실험환경

실험은 키보드로 직접 제어하는 로컬 테스트 모

드에서 수행하였으며, 시스템은 AI 인지, 저수준 제어 및 하드웨어의 세 계층으로 구성된다. AI 계층은 YOLOv11 기반 객체 탐지와 OSNet 기반 유사도 매칭을 통해 타겟을 선정하며, 'S' 키로 사용자를 등록하고 '1/2/3' 및 'b/f' 키를 이용하여 시나리오와 조건을 전환한다. 저수준 제어 계층은 바운딩 박스 중심과 화면 중앙 간의 수평 오차를 기반으로 비례(P) 제어를 수행하여 조향각을 산출하고, 바운딩 박스 높이를 거리의 근사값으로 활용하여 접근 및 정지 여부를 결정한다. 하드웨어 계층은 Jetson 보드와 Arduino 기반 모터 드라이버로 구성된다.

Baseline은 OSNet 유사도만을 이용하여 타겟을 판별하며, 제안 방법은 여기에 3.2절의 LiDAR 융합 모듈을 추가하여 식 (3)에 따라 최종 타겟을 판별한다. 두 조건에서 공통으로 사용된 하드웨어 및 소프트웨어 환경은 표 2와 같다.

표 2. 실험환경
Table 2. Experimental environment

Item	Description
Board	NVIDIA Jetson Xavier NX
OS	Ubuntu 20.04
Framework	Python 3.8.10
Library	PyTorch 2.1.0a0 (NVIDIA build), torchvision 0.16.1, Ultralytics YOLO11n, torchreid 1.4.0
LiDAR	RPLiDAR (2D, model 24, firmware 1.29)
Camera	USB wide-FOV camera
Calibration	HFOV = 60.0°, ANGLE_OFFSET = 0.0°
Re-ID threshold/α	τ_sim = 0.72, α = 0.7

4.2 실험 시나리오 설계

본 논문에서는 유사도 수준을 2단계로 구분한 시나리오와 대조군 1개를 포함해 3개 그룹으로 실험을 설계한다. 고유사군은 등록 사용자와 동일 계열 색상 및 유사한 실루엣의 의상을, 중유사군은 색상 톤은 유사하나 동일하지는 않은 의상을 착용한 타인으로 구성하며, 대조군은 확연히 다른 색상의 의상을 착용한 타인으로 구성한다. 이는 외형이 명확히 다른 경우의 시스템 안정성을 확인하는 하한선 역할을 한다. 각 그룹에서 등록 사용자는 로봇 전방 1~1.5m 지점에 고정하고, 비교 대상 타인이 반대편에서 화면을 가로질러 지나가도록 하여 baseline·제안 시스템 각각에서 동일 조건을 반복 촬영하며, 반복 횟수는 그룹×시스템 조합당 15회로 설정한다.

표 3은 그룹별 실험 조건과 반복 횟수를 나타내며, baseline·제안 열의 숫자는 각 시스템으로 수행한 트라이얼 횟수를 의미한다.

각 트라이얼의 판별 결과와 평균 유사도, 평균 마스크 값, 평균 최종 점수 및 프레임 처리 속도는 트라이얼 종료 시점에 터미널 로그를 기반으로 (Comma-Separated Values) 파일에 자동 기록하여 정량 분석에 활용하였다.

4.3 정량적 평가 지표 및 결과

평가 지표는 등록되지 않은 타인을 등록 사용자로 오인식한 비율인 오인식률, 트라이얼 중 추종 대상이 등록 사용자에서 타인으로 전환된 횟수를 전체 45회 트라이얼에 걸쳐 합산한 ID 스위칭 횟수, 및 LiDAR 마스크 적용에 따른 프레임 처리 성능(FPS)의 세 가지로 구성하였다.

표 3. 실험 시나리오 설계
Table 3. Experimental scenario design

Item / Group	Condition	Participants	Trials (baseline)	Trials (Proposed)
A. High-similarity	Same color family as registered user + similar silhouette	Registered user + 1 bystander	15	15
B. Medium-similarity	Similar color tone but not identical	Registered user + 1 bystander	15	15
C. Control	Clearly different color from registered user	Registered user + 1 bystander	15	15

표 4에 제시된 전체 평가 결과, 오인식률은 카메라 단독 시스템의 26.7%에서 제안 방법 적용 시 8.9%로 감소하여 17.8%의 개선을 보였다. ID 스위칭 횟수는 12회에서 4회로 감소하여 약 66.7%의 감소 효과를 확인하였다. 평균 FPS는 18.4에서 16.9로 약 8.2% 감소하였으나, LiDAR 융합 모듈을 추가한 이후에도 16.9 FPS의 처리 성능을 유지하였다.

표 4. 정량적 평가 결과
Table 4. Quantitative evaluation results

Metric \ Condition	baseline	Proposed (LiDAR fusion)
False-target rate (%)	26.7	8.9
ID switch count	12	4
Average FPS	18.4	16.9

전체 오인식률 차이의 통계적 유의성을 확인하기 위해, baseline과 제안 방법의 오인식 횟수(12/45 vs 4/45)를 두 시스템의 45회 트라이얼을 각각 독립 표본으로 간주하여 Fisher의 정확검정(Fisher's exact test, 양측)으로 비교하였다. 그 결과 p=0.051로 나타나, 통상적인 유의수준(α=0.05)에 근접하였으나 이를 충족하지는 못하였다.

표 5는 유사도 그룹별 오인식률을 비교한 결과이다. 고유사군(A)에서 33.3%의 가장 큰 개선폭이 확인되었으며, 중유사군(B)에서도 20.0%의 개선이 나타났다. 반면 대조군(C)에서는 baseline과 제안 방법 모두 오인식이 발생하지 않아, 외형이 명확히 구분되는 상황에서는 제안 방법에 따른 오인식 변화가 발생하지 않음을 확인하였다.

표 5. 유사도 그룹별 오인식률 비교
Table 5. False-target rate by similarity group

Metric \ Condition	baseline	Proposed (LiDAR fusion)
A. High -similarity	53.3(8/15)	20.0(3/15)
B. Medium -similarity	26.7(4/15)	6.7(1/15)
C. Control	0.0(0/15)	0.0(0/15)

4.4 결과 분석 및 논의

오인식률 감소폭은 특히 고유사군(A)에서 크게 나타났으며, 이는 외형 유사도가 높은 상황에서 LiDAR 기반 공간 정보가 Re-ID 판별을 보완하는 효과가 있음을 시사한다. 다만 4.3절에서 보고한 바와 같이 전체 오인식률 차이의 검정 결과는 p=0.051로 유의수준을 충족하지 못하였으며, 이는 그룹당 15회라는 제한된 표본 크기에 기인한 것으로 판단된다.

한편 평균 FPS는 18.4에서 16.9로 약 8.2% 감소하였다. 이는 LiDAR 마스크 계산에 따른 추가적인 연산 부담이 발생했음을 의미하지만, 제안 시스템에서도 16.9 FPS의 처리 속도를 유지하여 임베디드 환경에서의 실시간 추종 가능성을 확인하였다.

Ⅴ. 결론 및 향후 과제

본 논문은 이론적으로 제시되었던 2D LiDAR 기하학적 정보 결합 공간 주의집중 메커니즘[7]을 실제 자율추종 로봇의 개인화 재식별 파이프라인에 통합하고 그 실효성을 검증하였다. 유사 외형 타인 오인식 상황을 재현한 실험에서 제안 방법은 카메라 단독 시스템 대비 오인식률을 26.7%에서 8.9%로 17.8%(상대 약 66.7%) 감소시켰고, ID 스위칭 횟수도 12회에서 4회로 감소시켰다. 다만 전체 오인식률 차이의 검정 결과는 p=0.051로 유의수준 0.05를 충족하지 못해, 개선 효과를 통계적으로 확증하기에는 한계가 있다. 그럼에도 본 연구는 이론 단계에 머물렀던 메커니즘을 실제 로봇 플랫폼에서 정량적으로 검증했다는 점에서 의의를 갖는다.

본 연구는 단일 실내 환경의 정적 시나리오에서 조건별 15회의 제한된 반복으로 수행되어, 이러한 표본 수가 통계적 유의성 확보의 한계로 작용하였다. 향후에는 반복 횟수를 확대하고 다양한 혼잡도와 조명 조건에서 강건성을 검증할 필요가 있으며, 다중 후보 간 공간 관계를 모델링하는 그래프 신경망 기반 추적과 2D LiDAR가 제공하지 못하는 수직 정보를 보완하는 카메라 기반 높이 추정 모듈 통합을 진행할 예정이다.

References

[1] Korea Institute for Robot Industry Advancement

(KIRIA), "2024 Robot Industry Survey Report", KIRIA, Daegu, Korea, Dec. 2025.

[2] X. Wang, C. Fu, Z. Li, Y. Lai, and J. He, "DeepFusionMOT: A 3D Multi-Object Tracking Framework Based on Camera-LiDAR Fusion with Deep Association", IEEE Robotics and Automation Letters, Vol. 7, No. 3, pp. 8260-8267, Jul. 2022. https://doi.org/10.1109/LRA.2022.3187264.

[3] L. Wang, et al., "CAMO-MOT: Combined Appearance-Motion Optimization for 3D Multi-Object Tracking with Camera-LiDAR Fusion", IEEE Transactions on Intelligent Transportation Systems, Vol. 24, No. 11, pp. 11981-11996, Nov. 2023. https://doi.org/10.1109/TITS.2023.3285651.

[4] D. Jia, A. Hermans, and B. Leibe, "DR-SPAAM: A Spatial-Attention and Auto-regressive Model for Person Detection in 2D Range Data", IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), Las Vegas, NV, USA, pp. 10270-10277, Oct. 2020. https://doi.org/10.1109/IROS45743.2020.9341689.

[5] R. Khanam and M. Hussain, "YOLOv11: An Overview of the Key Architectural Enhancements", arXiv preprint, arXiv:2410.17725, pp. 1-8, Oct. 2024. https://doi.org/10.48550/arXiv.2410.17725.

[6] K. Zhou, Y. Yang, A. Cavallaro, and T. Xiang, "Omni-Scale Feature Learning for Person Re-Identification", IEEE/CVF International Conference on Computer Vision (ICCV), Seoul, Korea, pp. 3701-3711, Oct. 2019. https://doi.org/10.1109/ICCV.2019.00380.

[7] G. Kim, U. Ko, H. Jung, S. Lee, and D. Jeong, "Spatial Attention Mechanism Combining LiDAR Geometric Information for Mitigating Target Ambiguity in Indoor Dynamic Environments", Proc. of KIIT Summer Conference, Jeju, Korea, pp. 933-937, Jun. 2026.

[8] N. Wojke, A. Bewley, and D. Paulus, "Simple Online and Realtime Tracking with a Deep Association Metric", IEEE International Conference on Image Processing (ICIP), Beijing, China, pp. 3645-3649, Sep. 2017. https://doi.org/10.1109/ICIP.2017.8296962.

[9] H. Ye, J. Zhao, Y. Zhan, W. Chen, L. He, and H. Zhang, "Person Re-Identification for Robot Person Following with Online Continual Learning", IEEE Robotics and Automation Letters, Vol. 9, No. 11, pp. 9151-9158, Nov. 2024. https://doi.org/10.1109/LRA.2024.3438042.

[10] L. Zheng, L. Shen, L. Tian, S. Wang, J. Wang, and Q. Tian, "Scalable Person Re-Identification: A Benchmark", IEEE International Conference on Computer Vision (ICCV), Santiago, Chile, pp. 1116-1124, Dec. 2015. https://doi.org/10.1109/ICCV.2015.133.

[11] M. Bouazizi, A. L. Mora, and T. Ohtsuki, "A 2D-Lidar-Equipped Unmanned Robot-Based Approach for Indoor Human Activity Detection", Sensors, Vol. 23, No. 5, Art. no. 2534, Feb. 2023. https://doi.org/10.3390/s23052534.

[12] F. Pan, L. Zhao, and C. Wang, "Adaptive Sparse Attention-Based Compact Transformer for Object Tracking", Scientific Reports, Vol. 14, No. 1, Art. no. 12256, May 2024. https://doi.org/10.1038/s41598-024-63028-5.

[13] C.-Y. Wang, A. Bochkovskiy, and H.-Y. M. Liao, "YOLOv7: Trainable Bag-of-Freebies Sets New State-of-the-Art for Real-Time Object Detectors", IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Vancouver, BC, Canada, pp. 7464-7475, Jun. 2023. https://doi.org/10.1109/CVPR52729.2023.00721.

[14] R. C. Câmara de M. Santos, M. Coelho Silva, and R. A. R. Oliveira, "Real-time Object Detection Performance Analysis Using YOLOv7 on Edge Devices", IEEE Latin America Transactions, Vol. 22, No. 10, pp. 799-805, Oct. 2024. https://doi.org/10.1109/TLA.2024.10705971.

[15] W. Zhao, J. P. Queralta, and T. Westerlund,

"Sim-to-Real Transfer in Deep Reinforcement Learning for Robotics: A Survey", IEEE Symposium Series on Computational Intelligence (SSCI), Canberra, Australia, Canberra, ACT, Australia, pp. 737-744, Dec. 2020. https://doi.org/10.1109/SSCI47803.2020.9308468.

저자소개

김 건 웅 (Geonwoong Kim)



2021년 3월 ~ 현재 : 국립군산대학교 소프트웨어학과 학사과정
관심분야 : 데이터베이스, 자율주행, 피지컬인공지능

고 운 별 (Unbyeol Go)



2022년 3월 ~ 현재 : 국립군산대학교소프트웨어학과 학사과정
관심분야 : 데이터베이스, 자율주행, 피지컬인공지능

정 현 준 (Hyunjun Jung)



2008년 3월 : 삼육대학교 컴퓨터과학과(학사)
2010년 3월 : 숭실대학교 컴퓨터학과(공학석사)
2017년 9월 : 고려대학교 컴퓨터·전파통신공학과(공학박사)
2017년 8월 ~ 2020년 8월 : 광주과학기술원 블록체인인터넷경제연구센터 연구원
2021년 3월 ~ 현재 : 국립군산대학교 교수
관심분야 : 블록체인, 데이터 사이언스, 센서 네트워크, 사물인터넷, 머신러닝

이 석 훈 (Sukhoon Lee)



2009년 2월 : 고려대학교 전자및정보공학부(학사)
2011년 2월 : 고려대학교 컴퓨터·전파통신공학과(공학석사)
2016년 2월 : 고려대학교 컴퓨터·전파통신공학과(공학박사)
2016년 3월 ~ 2017년 3월 : 아주대학교 의료정보학과 연구강사
2017년 4월 ~ 현재 : 국립군산대학교 소프트웨어학과 부교수
관심분야 : 사물인터넷, 메타데이터 레지스트리, 데이터 품질, 연합 학습

정 동 원 (Dongwon Jeong)



1997년 2월 : 군산대학교 컴퓨터과학과(학사)
1999년 2월 : 충북대학교 전자계산학과(석사)
2004년 2월 : 고려대학교 컴퓨터학과(박사)
2005년 4월 ~ 현재 : 국립군산대학교 교수
관심분야 : 데이터베이스, 시맨틱 서비스, 빅데이터, 사물인터넷, 피지컬인공지능, 엣지컴퓨팅, 지능형 융합 서비스