라이다에 대한 물리적 적대적 공격 탐지를 위한 새로운 Feature로써의 강도 활용

박예지*, 조현수*, 최원석*, 이동훈*

*고려대학교 정보보호대학원

foqjszmffhblue@korea.ac.kr, hscho20@korea.ac.kr, beb0396@korea.ac.kr, donghlee@korea.ac.kr

Leveraging Intensity as New Feature to Detect Physical Adversarial Attacks Against LiDARs

Ye-Ji Park*, Hyunsu Cho*, Wonsuk Choi*, Dong Hoon Lee*

School of Cybersecurity, Korea University

요 약

최근 많은 연구에서 LiDAR 객체 탐지 모델에 대한 공격을 제시하고 있다. LiDAR는 기본적으로 여러 반사된 펄스 중 가장 강한 레이저 펄스를 인식하기 때문에, 이러한 공격들은 고강도 레이저 펄스를 이용하여 정상적인 포인트를 제거하고 3D 포인트 클라우드에 가짜 포인트를 주입한다. 이 가짜 포인트들은 탐지 모델의 오분류를 유발하는 가짜 객체가 된다. 실험을 통해 포인트 클라우드의 포인트 강도 정보를 조작하더라도 객체 탐지에는 큰 영향을 끼치지 않는다는 것을 발견했다. 본 논문에서는 공격자에 의해 생성된 가짜 포인트가 정상 포인트에 비해 높은 강도 수준을 나타난다는 점에 대해 초점을 맞춘다. 그래서 포인트의 강도 정보는 강한 레이저 펄스로 가짜 포인트를 생성하는 공격을 탐지하는 데 매우 유용하게 활용될 수 있다. 이 가정을 바탕으로, 객체의 강도 분포를 도출하고 이를 회귀 과정에 적용하여 가짜 객체를 식별하는 기법을 제안한다.

Ⅰ. 서론

최근 자율주행과 관련된 다양한 기술들이 운전자의 안전과 편의를 위해 빠르게 발전하고 있다. 고급 운전자 보조 시스템(ADAS; Advanced Driver Assistance System)은 자율주행 차량에 대한 기술 중 하나다. 자율주행 기술은 주변 환경을 인식해야 하며, 이를 위해 카메라, 레이더, 그리고 라이다(LiDAR; Light Detection And Ranging) 센서와 같은 다양한 센서들이 필요하다. 라이다는 주변 물체에 레이저 펄스를 발사하고, 반사되어 돌아오는 펄스를 수신한다. 라이다는 송수신 펄스 간의 ToF (Time-of-Flight)를 측정하여 물체와의 거리를 측정한다. 이러한 거리 측정 방식은 주변 환경의 구조를 인지하는 데 도움을 준다. 다른 센서들에 비해 높은 해상도를 가진 라이다는 자율주행 차량이 최소 레벨 3의 자율성을 달성하는 데 있어 필수적인 것으로 고려된다.

라이다는 공간 내의 이산 데이터 포인트들로 구성된 포인트 클라우드를 구축할 수 있다. 이 포인트들은 3D 형태나 객체를 나타낼 수 있으며 각 포인트의 위치는 카르테시안 좌표로 표현된다. 또한, 라이다는 반사된 펄스의 강도 값을 측정할 수 있다. 라이다에 의해 펄스가 발사되고 퍼져 나갈 때 여러 번의 반사 펄스가 돌아온다. 이에 따라 많은 수의 반사 펄스가 인식되며 모든 반사 펄스를 처리하기에는 주변 환경을 명확하게 인식하기 어렵다. 이러한 이유로 자율주행 라이다에는 특정 반사 펄스만 처리하는 세 가지 리턴 모드가 운영된다. Strongest 모드는 강도에 기반하여 가장 강한 반사된 펄스만 처리하고, Last 모드는 마지막으로 반사된 펄스만 처리한다. 마지막으로, Dual 모드는 가장 강한 펄스 및 마지막에 반사된 펄스를 처리한다. 그리고 디폴트 설정으로는 Strongest 모드가 사용된다.

실제로 포토 다이오드와 레이저 다이오드를 사용하여 3D 포인트 클라우드 내의 가짜 포인트를 생성할 수 있다는 것을 입증했다[1]. 라이다에서 측정된 3D 포인트들이 딥러닝 모델에 의해 처리되기 때문에, 딥러닝 모델에 대한 적대적 공격도 가능하다[2, 3]. 레이저 다이오드를 이용해 강한 펄스를 방출함으로써 피해 라이다는 원래의 반사 펄스가 아닌 강한 반사 펄스, 즉 가짜 포인트만 처리한다. 스푸핑된 포인트들은 높은 강도의 레이저를 발사하여 다른 정상 포인트들을 제거함으로써 공격 대상 라이다에 객체로 인식되어 공격을 수행한다. 본 논문에서는 라이다의 객체 탐지 알고리즘이 강도 값을 수정하더라도 여전히 객체를 탐지할 수 있다는 것을 보여준다. 그리고 실험 결과를 통해 강도의 값이 라이다에 대한 적대적 공격을 탐지하는 유망한 Feature가 될 수 있다는 것을 알 수 있다. 이와 같은 실험 결과는 라이다에 대한 적대적 공격에 대응하는 방어 방법을 개발하는 데 도움이 될 것이다.

Ⅱ. 포인트 강도가 객체 탐지에 미치는 영향에 대한 실험

라이다의 리턴 모드에서 기본으로 설정된 Strongest 모드로 인해 공격자들은 원래 반사된 펄스보다 강한 레이저 펄스를 생성하여 3D 포인트 클라우드에 가짜 포인트를 주입할 수 있다. 이에 따라 공격자에 의해 생성된 가짜 포인트들은 정상 포인트보다 높은 강도를 가질 것이다. 높은 강도를 가진 가짜 포인트 세트가 정상 객체로 식별됨을 보여주기 위한 실험을 수행했다.

정상 데이터셋과 강도가 조작된 데이터셋에서 평균 정밀도(AP; Average Precision)를 평가한다. 이 평가를 위해 Kitti 데이터셋[4]과 PV-RCNN[5], PointPillar[6], PointRCNN[7], SECOND[8]이라는 딥러닝 3D 객체 탐지 모델을 선택했다. [표 1]은 강도 조정 배수와 객체 클래스에 해당하는 AP 결과를 보여준다. Car 클래스는 IoU (Intersection over Union)가 0.7이고, Pedestrian과 Cyclist 클래스는 IoU가 0.5이다. 관찰된 AP 값을 통해 Car 클래스에 속하는 객체들은 강도에 대한 변화가 거의 없다는 것을 확인하였다.

공격자가 가짜 포인트를 주입하기 위해 높은 강도의 펄스가 필요하다면, 가짜 포인트의 강도는 이러한 상한을 초과할 것이다. 즉, 라이다에서 기본으로 설정된 Strongest 모드에서 더 강한 레이저 펄스가 허용되기 때문에 물리적 적대적 공격을 달성하기 위해서 가짜 포인트는 정상 포인트보다

〔표 1〕 다양한 강도 조작을 하면서 PointPillar, PV-RCNN, PointRCNN, SECOND를 사용한 KITTI 데이터셋의 바운딩 박스 검출에 대한 AP(Average Precision) 측정 결과

Algorithms	Intensity	AP		
		Car	Pedestrian	Cyclist
PV-RCNN	Normal	89.4962	68.0819	81.3322
	1.2times	89.4589	66.8875	80.2891
	1.5times	89.4961	65.0686	74.8355
	2 times	89.3401	49.4466	63.0190
PointRCNN	Normal	89.6668	66.5387	77.1918
	1.2times	89.6960	64.9856	75.7956
	1.5times	89.6344	53.0288	72.5603
	2 times	89.3631	26.0176	57.8852
PointPillar	Normal	89.8089	62.4964	72.9961
	1.2times	89.7354	57.5781	69.3037
	1.5times	89.5119	29.2107	54.9482
	2 times	87.5852	6.3387	26.5147
SECOND	Normal	89.8999	66.3265	77.0911
	1.2times	89.8992	65.2369	77.1387
	1.5times	89.8740	63.4855	75.2616
	2 times	89.7453	60.9428	69.2813

높은 강도를 가져야 한다.

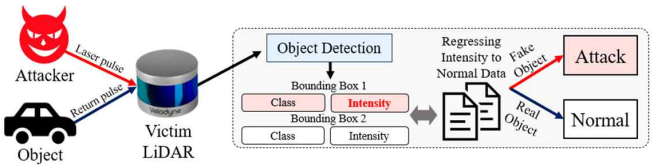
Ⅲ. 포인트 강도를 이용한 가짜 객체 탐지 방법

[그림 1]은 가짜 포인트를 사용하여 물리적 적대적 공격을 탐지하는 방법에 대한 주요 아이디어의 개요를 나타낸다. 강도 값이 특정 임계 값을 초과하는지 여부에 따라, 가짜 포인트나 물리적 적대적 공격을 탐지할 수 있다. 우선, 포인트 기반 3D 객체 탐지 모델인 PointRCNN을 사용하여 3D 바운딩 박스의 좌표를 얻고 이 바운딩 박스 내의 포인트들의 강도 값을 추출한다. 다음으로, 추출된 강도 값, 객체 클래스를 활용하여 합법적 데이터에 대한 회귀를 적용하여 공격자에 의해 생성된 가짜 객체를 식별한다. 또한, 동일 클래스의 다른 정상 객체들이 있을 때, 이 객체들의 강도를 비교하고 미리 정의된 임계값을 초과하는 강도 차이를 갖는 것들을 식별한다.

실제 객체들은 객체 표면의 입사각과 재료에 따라 포인트 간에 다양한 강도를 가져야 한다. 그러나 스푸핑된 포인트들은 높은 강도의 레이저를 방출하여 다른 정상 객체 포인트들을 제거해야 하므로 실제 객체들과 비교해 뚜렷하게 높은 강도를 나타내야 한다. 가짜 바운딩 박스 내의 포인트들의 강도가 균일한지 확인한다. Car, Pedestrian, Cyclist 클래스에 대해 라이다에서 수집한 강도를 기반으로 훈련하여 가짜 객체와 실제 객체를 분류한다.

Ⅳ. 결론

본 논문에서는 강도를 이용하여 라이다에 대한 물리적 공격을 탐지하는 새로운 접근법을 제안한다. 이러한 공격의 가능성은 라이다의 기본 리턴 모드와 3D 객체 탐지 모델이 사용하는 방법론의 본질적 특성에서 비롯된다. 이러한 고유한 특성은 공격자들이 3D 객체 탐지 결과를 조작함으로써 자율 시스템의 무결성을 위협할 수 있다. 종합적인 분석은 라이다 포인트 클라우드 내 객체들이 보여주는 강도 패턴에 초점을 맞춘다. 동일한 클래스 별 강도를 활용하여 가짜 객체를 식별하는 방법의 잠재적 활용 가능성을 제안한다. 목적은 제안된 탐지 방법을 실제 환경에서 철저히 검증하고 개선하는 것이다. 자율 주행 기술이 미래에 계속 발전함에 따라, 이 연구



〔그림 1〕 라이다에 대한 물리적 적대적 공격을 탐지하는 방법의 개요

가 안전을 보장하는 데 있어 매우 중요한 의미를 갖게 될 것이다.

ACKNOWLEDGMENT

이 논문은 2024년도 정부(과학기술정보통신부)의 재원으로 정보통신기획평가원의 지원을 받아 수행된 연구임 (No. 2020-0-00374, 무인이동체 공통핵심 보안기술 개발 연구)

참 고 문 헌

[1] Shin, Hocheol, et al. "Illusion and dazzle: Adversarial optical channel exploits against lidars for automotive applications." Cryptographic Hardware and Embedded Systems - CHES 2017: 19th International Conference, Taipei, Taiwan, September 25-28, 2017, Proceedings. Springer International Publishing, 2017.

[2] Cao, Yulong, et al. "Adversarial sensor attack on lidar-based perception in autonomous driving." Proceedings of the 2019 ACM SIGSAC conference on computer and communications security. 2019.

[3] Jin, Zizhi, et al. "Pla-lidar: Physical laser attacks against lidar-based 3d object detection in autonomous vehicle." 2023 IEEE Symposium on Security and Privacy (SP). IEEE, 2023

[4] Geiger, Andreas, et al. "Vision meets robotics: The kitti dataset." The International Journal of Robotics Research 32.11 (2013): 1231-1237.

[5] Shaoshuai Shi, Chaoxu Guo, Li Jiang, Zhe Wang, Jianping Shi, Xiao- gang Wang, and Hongsheng Li. 2020. Pv-rcnn: Point-voxel feature set abstraction for 3d object detection. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 10529 - 10538.

[6] Lang, Alex H., et al. "Pointpillars: Fast encoders for object detection from point clouds." Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2019.

[7] Shi, Shaoshuai, Xiaogang Wang, and Hongsheng Li. "Pointrcnn: 3d object proposal generation and detection from point cloud." Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2019.

[8] Yan, Yan, Yuxing Mao, and Bo Li. "Second: Sparsely embedded convolutional detection." Sensors 18.10 (2018): 3337.