카메라-라이다 정합 모델에 대한 스케일링 공격

임이지[1], 최대선[2]
[1]숭실대학교 소프트웨어학과 석사과정
[2]숭실대학교 소프트웨어학과 교수
ezim@soongsil.ac.kr, sunchoi@ssu.ac.kr

Scaling attack for Camera-Lidar calibration model

Yi-JI IM[1], Dae-Seon Choi[2]
[1]Dept. of Software, Soongsil University
[2]Dept. of Software, Soongsil University

요 약

자율주행 및 robot navigation 시스템에서 물체 인식 성능향상을 위해 대부분 MSF(Multi-Sensor Fusion) 기반 설계를 한다. 따라서 각 센서로부터 들어온 정보를 정합하는 것은 정확한 MSF 알고리즘을 위한 필요조건이다. 다양한 선행 연구에서 2D 데이터에 대한 공격을 진행했다. 자율주행에서는 3D 데이터를 다루어야 하므로 선행 연구에서 하지 않았던 3D 데이터 공격을 진행했다. 본 연구에서는 스케일링 공격 기반 카메라-라이다 센서 간 정합 모델의 정확도를 저하시키는 공격 방법을 제안한다. 제안 방법은 입력 라이다의 포인트 클라우드에 스케일링 공격을 적용하여 다운스케일링 단계에서 공격하고자 한다. 실험 결과, 입력 데이터에 공격하였을 때 공격 전보다 평균제곱 이동오류는 56% 이상, 평균 사원수 각도 오류는 98% 이상 증가했음을 보였다. 다운스케일링 크기 별, 알고리즘별 공격을 적용했을 때, 10x20 크기로 다운스케일링 하고 lanczos4 알고리즘을 적용했을 때 가장 효과적으로 공격할 수 있음을 확인했다.

1. 서론

자율주행 기술은 지난 몇 년간 주목받고 있으며 급속도로 발전해왔다. 하지만 높은 수준(레벨 4 이상)의 자율주행을 위해선 정확하고 안전한 운전을 보장해야 한다.

이를 위해 카메라 및 라이다 등의 센서를 사용하고 있으며 센서는 주변 장애물들을 실시간으로 감지하고 충돌 회피와 같은 안전에 중요한 결정에 직접적인 영향을 미친다. 이때, 단일 센서에 의존하기보다 대부분 여러 개의 센서를 융합하여 사용하는 MSF (Multi-Sensor Fusion) 기반 설계를 하며 각 센서로부터 들어온 정보를 정합하여 분류나 객체 인식, 3D 모델링 등에 활용한다.

그러나 센서 간 정합이 제대로 되지 않으면 오차로 인해 성능 저하를 유발하므로 안전이 중요한 자율주행 기술에서는 더욱 치명적인 결과를 초래할 수 있다. 선행 연구에서 카메라를 공격하여 오분류, 객체 인식[1]에 대한 오류를 유도하거나 라이다에 스푸핑 공격을 하여 인식을 방해[2]하는 등 MSF 기반 모델에 대한 다양한 공격을 했지만 센서 간 정합에 대한 공격을 시도한 적은 없다.

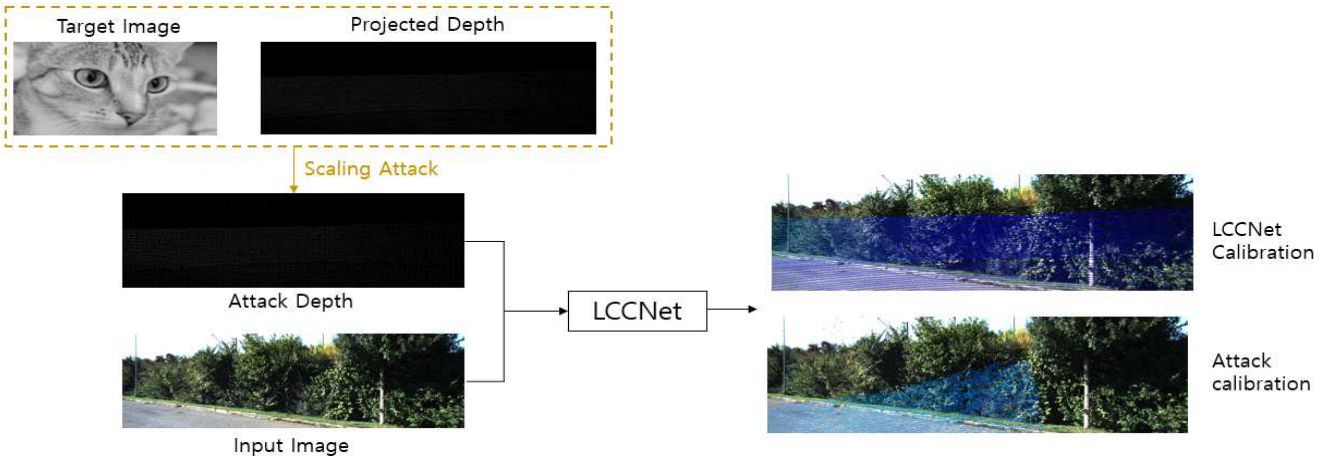
본 연구는 카메라-라이다 간 정합에 대한 오류를 유도하여 MSF의 근간을 공격하는 방법을 처음으로 제안한다. 라이다로부터 수집한 포인트 클라우드의 특징을 추출하여 스케일링 공격[3]을 적용하여 단일 센서에는 영향을 거의 주지 않고 카메라와 라이다 정합의 정확도를 낮춘다. 또한 스케일링 공격이 3D 데이터에도 적용할 수 있음을 보인다. 실험을 통해 가장 효과적인 다운스케일링 사이즈와 알고리즘을 찾는다.

2. 관련 연구

2.1 카메라-라이다 정합

카메라-라이다 간 정합은 이미지의 픽셀에 대한 깊이 정보를 얻기 위해 라이다와 카메라의 데이터 동일한 좌표계로 변환하는 것이다. 카메라가 색감, 질감, 모양 정보를 캡처하는 동안 라이다는 이미지의 3D 구조적 정보를 캡처한다. 카메라-라이다 간 정합은 자율주행, 네비게이션, 로봇공학 등에서 3D 이미지 재구성을 위해 널리 사용된다.

RegNet[4]은 CNN을 이용하여 멀티모달 센서 간의 6-DoF 외부 매개 변수를 예측한다. Calibnet[5]



(그림 1) 카메라-라이다 센서 간 캘리브레이션 모델 공격 방법

은 실시간으로 카메라와 라이다 간의 6-DoF 변형을 추정하는 네트워크이다. LCCNet[7]은 외부 매개 변수를 실시간으로 추정하는 네트워크이다. 카메라의 RGB 이미지 특징과 Depth 이미지 간의 상관성을 나타내는 Cost volume을 사용한다.

2.2 스케일링 공격

Xiao[6]은 이미지 스케일링 알고리즘에 대한 공격을 제안했다. 이미지 스케일링은 컴퓨터 비전의 표준 절차이며 기계학습의 일반적인 전처리 단계이다. 스케일링 알고리즘은 보간을 통해 이미지를 축소한다. 제안된 연구에서 스케일링 알고리즘이 공격에 취약하다는 점을 보여주었다. 제안된 공격은 소스이미지를 교란하여 새로운 이미지를 생성하며 생성된 이미지를 스케일링 했을 시 대상 이미지와 일치하도록 한다.

3. 제안 방법

본 논문에서는 스케일링 공격 기반 카메라-라이다 센서 간 정합 모델 공격 방법을 제안하며 공격 과정은 그림 1과 같다. 입력 라이다 포인트 클라우드로부터 2차원 이미지 평면에 투영된 라이다의 포인트 클라우드를 스케일링 공격하여 공격 라이다 이미지를 생성한다. 생성된 공격 라이다 이미지는 기존의 라이다 이미지와의 차이를 거의 없게 하여 단일 센서에 영향을 미치지 않도록 한다. 포인트 클라우드와 대응되는 RGB 이미지와 공격 라이다 이미지를 LCCNet에 입력하여 정합 후 결과 정합 오류 유발을 유도한다.

4. 실험 및 실험 결과

4.1 실험 환경 설정

본 논문에서 제안한 카메라-라이다 센서 간 정합 모델 공격을 평가하기 위해 오픈 데이터 세트인 KITTI 주행 거리 측정 데이터 세트[7]를 사용했다. 데이터 세트는 각 센서 사이의 정합 매개 변수를 제공하며 그 중 카메라와 라이다 간의 정합 매개 변수를 ground-truth로 사용했다.

본 논문에서는 라이다와 카메라의 왼쪽 컬러 카메라 사이의 정합만 고려한다. 정합 모델은 데이터 세트의 20개 시퀀스(34350 프레임)를 학습시켰으며, 1개의 시퀀스(4541 프레임)로 공격을 수행했다. 최대 오보정 회전각은 20°이고 최대 오보정 이동 거리는 1.5m로 설정했다.

4.2 공격 라이다 이미지 생성

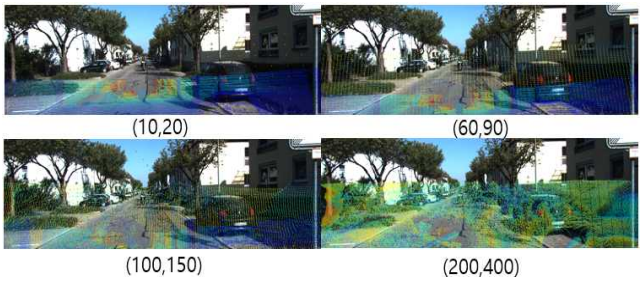
카메라 이미지 평면에 투영된 라이다의 포인트 클라우드를 투영하여 2차원 라이다 이미지를 생성한다. 라이다 이미지에 nearest, bicubic, bilinear, lanczos4, 4가지 스케일링 알고리즘과 10x20, 60x90, 100x150, 200x400 4가지 스케일링 크기를 사용한 스케일링 공격을 통해 공격 라이다 이미지를 생성했다.

<표 1> 스케일링 크기 별 공격 결과 비교

Scale size	E_t	E_R
base	0.125	0.003
(10,20)	0.285	0.148
(60,90)	0.259	0.212
(100,150)	3.607	0.251
(200,400)	7.141	0.197

<표 2> 스케일링 알고리즘별 공격 결과 비교

Scaling Algorithm	E_t	E_R
nearest	0.1339	0.1502
bicubic	0.2477	0.1403
bilinear	0.207	0.1309
lanczos4	0.285	0.148



(그림 2) 스케일링 크기 별 공격에 따른 정합 이미지 결과

4.3 카메라-라이다 정합 공격 성능평가

생성된 공격 라이다 이미지에 대한 공격 성능을 평가한다. 공격 성능평가는 스케일링 크기 별 공격과 스케일링 알고리즘별 공격 후 평균제곱 이동오류(E_t), 평균 사원수 각도 오류(E_R)를 측정한다.

스케일링 크기 별 공격 성능은 표 1, 그림 2와 같다. 공격을 하지 않은 네트워크의 성능은 표1의 base이며 평균제곱 이동오류가 0.125, 평균 사원수 각도 오류가 0.003으로 스케일링 공격을 했을 때보다 오차가 작으므로 스케일링 공격이 정합 성능 저하에 효과적임을 알 수 있다.

10x20 크기로 스케일링했을 경우, 평균제곱 이동오류가 0.285, 평균 사원수 각도 오류가 0.148로 오차율은 낮지만, 공격 라이다 이미지와 입력 라이다 이미지의 차이가 거의 없다. 다른 크기의 이미지들은 성능이 좋지만 공격 라이다 이미지에 타겟 이미지의 잔상이 투영되어 단일 센서에 영향을 줄 수 있으므로 10x20 크기가 가장 효과적인 스케일링 공격 크기임을 나타낸다.

공격 라이다 이미지 생성 시 4가지 스케일링 알고리즘을 준 가장 성능이 좋은 알고리즘을 찾기 위한 스케일링 알고리즘별 공격 성능은 표 2와 같다. lanczos4 알고리즘이 평균제곱 이동오류가 0.285, 평균 사원수 각도 오류가 0.148으로 오차율이 가장 높으므로 공격 성능이 가장 좋음을 나타낸다.

5. 결론

본 논문에서는 카메라-라이다 센서 간 정합 모델을 공격하는 방법을 제안한다. 실험을 통해 라이다에 대한 스케일링 공격은 정합 공격에 효과적임을 확인했으며, 스케일링 공격이 3D 데이터에 적용할 수 있음을 보였다. 스케일링 크기와 알고리즘별 성능 비교 실험을 통해 10x20 크기로 lanczos4 알고리즘을 사용한 스케일링 공격이 가장 성능이 뛰어난 조건임을 확인했다. 지금까지 카메라-라이다 센서 간 정합에 대한 공격 연구는 거의 없으며 본 연구를 통해 라이다 센서에 대한 공격으로 평균 77% 이상의 정합 오류를 유발하였다. Multi-Sensor Fusion 방법에서의 정합 공격을 방어하기 위한 연구도 필요할 것으로 보인다.

Acknowledge

이 논문은 2023년도 정부(과학기술정보통신부)의 재원으로 정보통신기획평가원의 지원을 받아 수행된 연구임 (No. 2021-0-00511, 엣지 AI 보안을 위한 Robust AI 및 분산 공격탐지기술 개발)

참고문헌

[1] Z. Xiong, H. Xu, W. Li and Z. Cai, "Multi-Source Adversarial Sample Attack on Autonomous Vehicles," in IEEE Transactions on Vehicular Technology, vol. 70, no. 3, pp. 2822-2835, March 2021.

[2] Y. Cao, C. Xiao, B. Cyr, Y. Zhou, W. Park, S. Rampazzi, Q. A. Chen, K. Fu, and Z. M. Mao, "Adversarial sensor attack on lidar-based perception in autonomous driving," in Proc. of the 2019 ACM SIGSAC Conf. on Computer and Communications Security, pp. 2267 - 2281, 2019.

[3] Quiring, E. Klein, D, Arp, D. Johns, M. & Rieck, K. "Adversarial preprocessing: Understanding and preventing image-scaling attacks in machine learning", USENIX Conference on Security Symposium, 1363-1380, 2020

[4] Schneider, N., Piewak, F., Stiller, C., & Franke, U. "RegNet: Multimodal sensor registration using deep neural networks." IEEE intelligent vehicles symposium (IV).1803-1810, 2017

[5] Iyer, G., Ram, R. K., Murthy, J. K., & Krishna, K. M"Calibnet: Geometrically supervised extrinsic calibration using 3d spatial transformer networks" IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), 1110-1117, 2018

[6] Lv, X., Wang, B., Dou, Z., Ye, D., & Wang, S. "LCCNet: LiDAR and camera self-calibration using cost volume network". In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2894-2901, 2021

[7] Geiger, A., Lenz, P., & Urtasun, R. "Are we ready for autonomous driving? the kitti vision benchmark suite." IEEE conference on computer vision and pattern recognition, 3354-3361,2012