Paper ID: 631

Sensor-independent Pedestrian Detection for Personal Mobility Vehicles in Walking Space Using Dataset Generated by Simulation

Takahiro Shimizu[1], Kenji Koide[2], Shuji Oishi[2], Masashi Yokozuka[2], Atsuhiko Banno[2], and Motoki Shino[1]

[1]The University of Tokyo, Japan

[2]National Institute of Advanced Industrial Science and Technology (AIST), Japan

ICPR 2020 25th INTERNATIONAL CONFERENCE ON PATTERN RECOGNITION Milan, Italy 10 | 15 January 2021

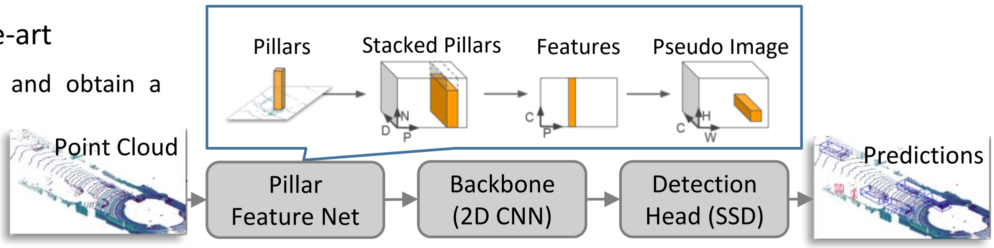
Abstract

- A pedestrian detection method based on PointPillars is proposed for personal mobility vehicles
- To make PointPillars sensor-independent, intensity features are replaced with **cosine local geometric features**
- The proposed CosPointPillars is trained on a dataset generated using **a realistic 3D LiDAR simulator**
- The combination of CosPointPillars and SimDataset enables us to **robustly detect nearby pedestrians** in real situations

Pedestrian Detection in a Walking Space

PointPillars: State-of-the-art

- Extract pillar-wise features and obtain a feature image in BEV
- Apply 2D CNN (SSD) to detect pedestrians



Problems

- Most of the existing detection models (including PointPillars) use as input reflectance intensity, whose characteristics largely depend on the sensor model
  - ➡ **Degraded accuracy on different sensors**
- Popular pedestrian datasets are recorded on roadway scenes and contain fewer number of nearby pedestrians
  - ➡ **Difficulty to detect nearby pedestrians**

Proposal

- Get rid of reflection intensity and introduce a new feature to compensate for the lost information
  - ➡ **Cosine features**
- Train the network on a large-scale pedestrian dataset dedicated for walking spaces
  - ➡ **Simulation-based dataset generation**

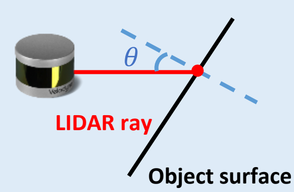
Alternative Feature for Reflection Intensity

- The reflection intensity ideally follow the Lambertian model that is composed of object reflectivity K_λ, incident angle θ, and distance d
- K_λ: depends on the wavelength of the LiDAR model
- d: can be easily encoded from the point coordinates
- θ: sensor-independent pure geometrical feature represents the local geometrical structure

Lambertian model

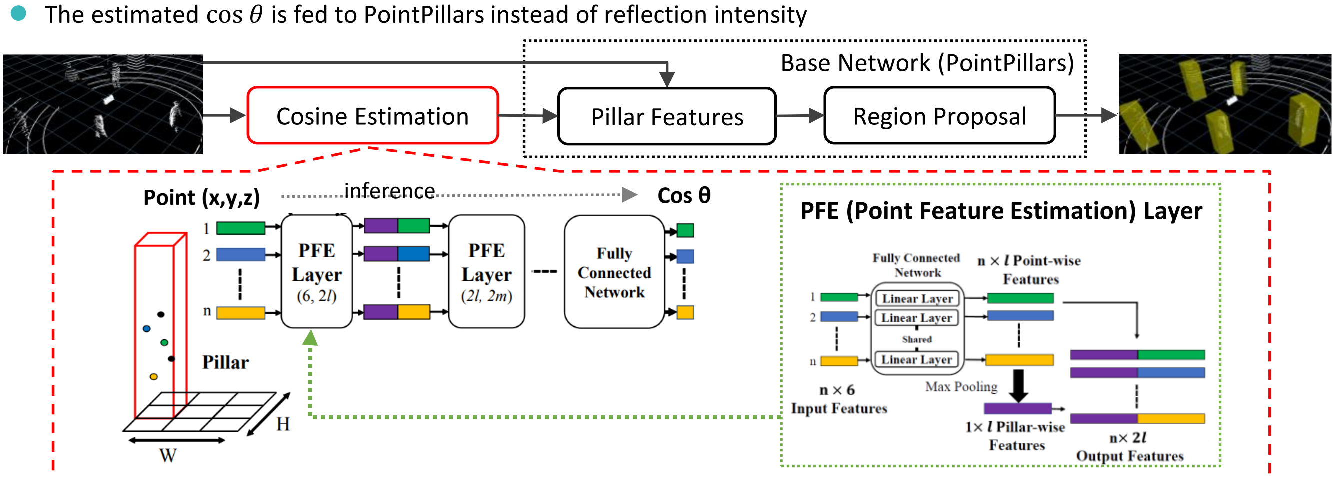
$$Intensity \propto \frac{K_\lambda \cos\theta}{d^2}$$

θ: Incident angle
d: Distance
K_λ : Object reflectivity



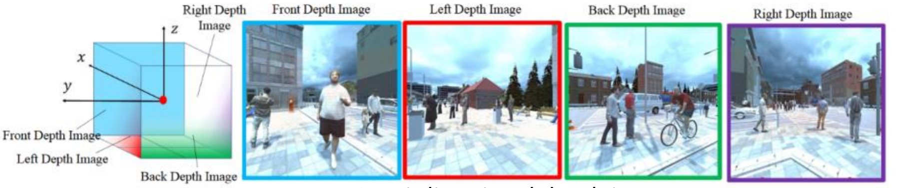
CosPointPillars

- A cosine estimation network, which estimates cosθ from point coordinates, is inserted just before PointPillars
- The estimated cos θ is fed to PointPillars instead of reflection intensity

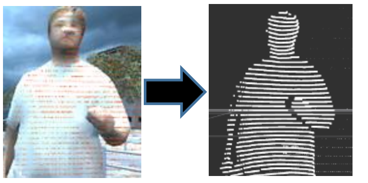


Simulation-based Pedestrian Dataset Generation

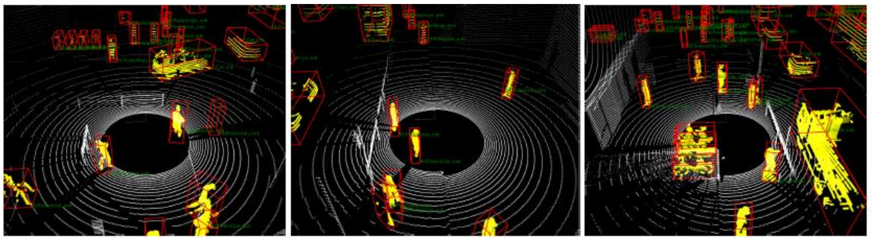
- Popular pedestrian datasets (e.g., KITTI) contain fewer nearby pedestrians since they are recorded on roadway scenes
- To apply CosPointPillars to real walking space environments, **we create a large-scale dataset using 3D LiDAR simulation**
- To obtain realistic point clouds, we developed a ray-casting-based 3D LiDAR simulator
- The quality of point clouds is very high compared to the ones generated from rough collision models in other simulators



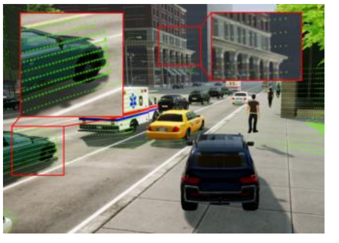
Omni-directional depth images



Realistic point cloud (Ray-casting)



Example of Simulation-based Pedestrian Detection Dataset (SimDataset)



AirSim (rough collision models)

- A pedestrian detection dataset that contains **over 22k frames and 120k pedestrian labels** is generated using simulation

Evaluation

KITTI Dataset

- We compared the proposed CosPointPillars with the original PointPillars trained with and without reflectance intensity
- The accuracy of **PointPillarge largely deteriorated when trained without reflectance intensity**
- **The accuracy got recovered by introducing cosine features** instead of reflectance intensity

Simulation-based Dataset (SimDataset)

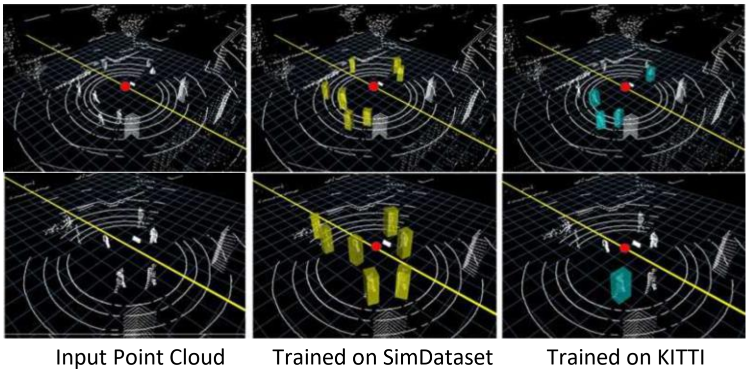
- We trained two CosPointPillars on KITTI and SimDataset, respectively
- The one trained on KITTI showed a low AP for pedestrians closer than 2.5m because KITTI contains fewer number of nearby pedestrians
- The network trained on SimDataset showed a **good AP in all the detection range**

Network	Reflectance	KITTI BEV AP [%] (IoU 0.5)		
		Easy	Moderate	Hard
PointPillars	w/	84.36	79.97	76.82
	w/o	80.22	75.92	74.48
CosPointPillars	w/o	82.35	77.29	75.94

Training Dataset	SimDataset BEV AP [%]		
	< 2.5m	≧ 2.5m	Total
KITTI	70.14	85.30	84.54
SimDataset	85.63	88.14	87.89

Pedestrian Detection in a Real Environment

- We ran two CosPointPillars models respectively trained on KITTI and SimDataset in a real walking space environment
- **The network trained on SimDataset successfully detected nearby pedestrians** while the other one failed to detect them



Input Point Cloud Trained on SimDataset Trained on KITTI

Training Dataset	F-measure		
	< 2.5m	≧ 2.5m	Total
KITTI	30.67	68.80	63.04
Sim Dataset	96.36	98.44	98.06