AI Sensor Fusion for Intelligent Transportation

Ming Lei (雷鸣), PhD
Senior Platform Architect
ming.lei@intel.com

Intel Corporation

AI Computing Conference (AICC), Beijing, China
Nov. 29, 2023

Outline

- Roadside Sensing
  - Usage scenarios
  - System architecture
  - Traffic object attributes
  - Pros & cons of traffic sensors

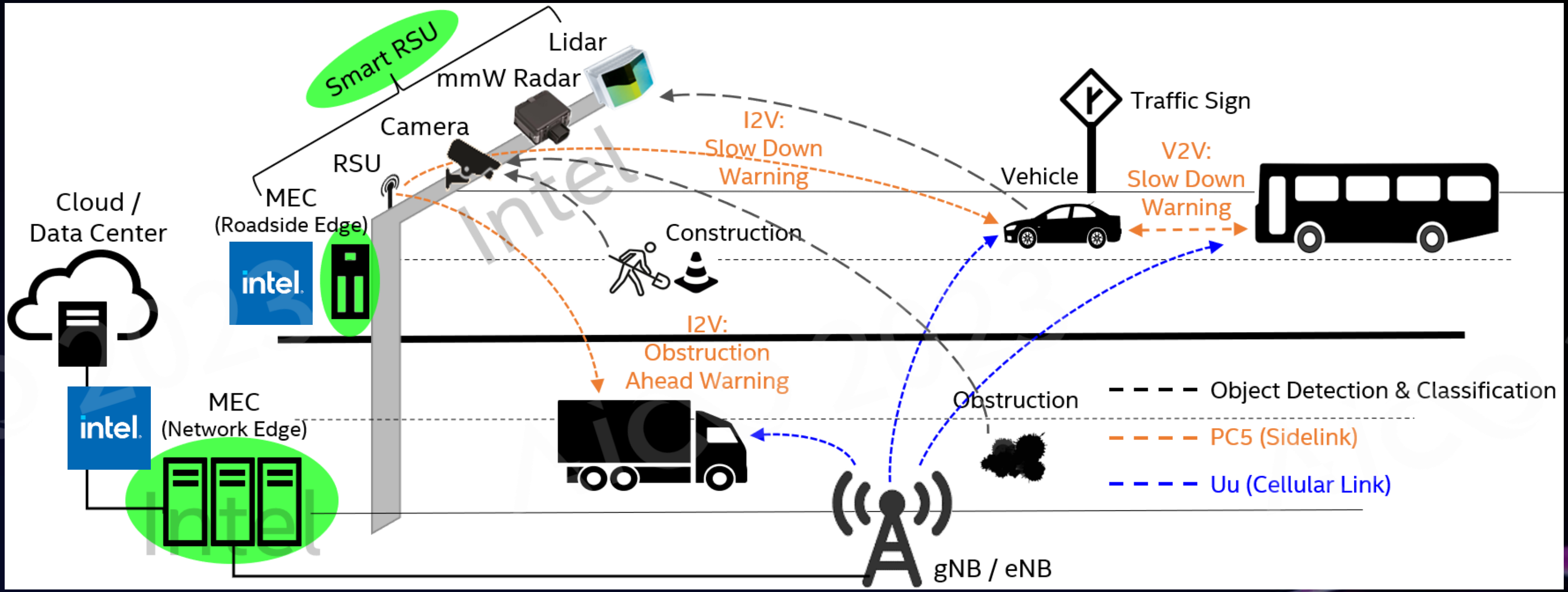
- Sensor Fusion (categorized by integration level)
  - Integrated & Distributed

- Sensor Fusion (categorized by involved sensor types)
  - "C+R": Camera + mmWave Radar
  - "C+L": Camera + Lidar

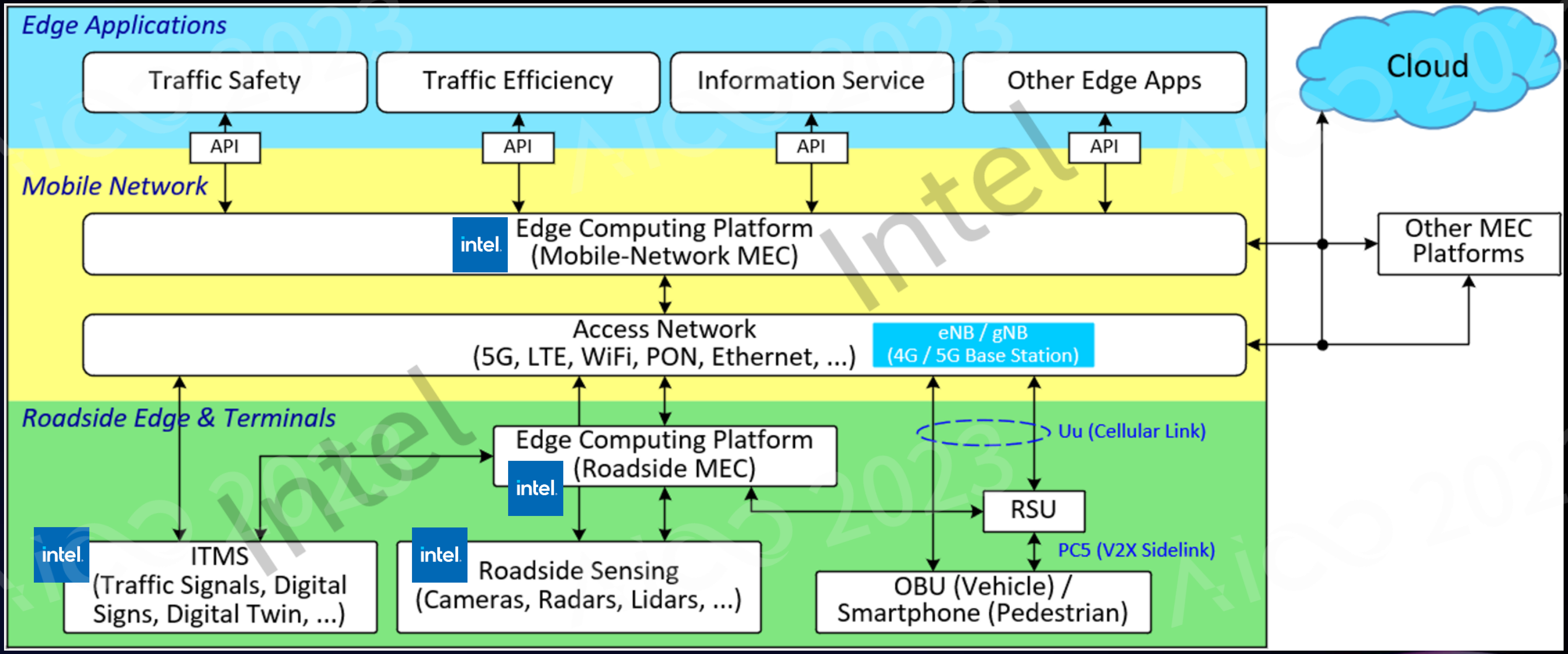
Roadside Sensing – Usage Scenarios

- **Roadside Sensing** mainly relies on three types of sensors: **cameras**, **mmWave radars**, and **lidars**

- **Edge Computing** (e.g., MEC deployed at roadside or mobile network edge) processes raw data from **Roadside Sensing**
  - **Deep learning** (e.g., object classification based on neural networks for video images or 3D point clouds)
  - **Traditional computer vision** (e.g., color space conversion for video images or clustering for 3D point clouds)
  - **Radar signal processing** (e.g., FFTs for range, velocity, and AoA estimation)



Roadside Sensing – System Architecture



Roadside Sensing – Traffic Object Attributes

RTPs (Road Traffic Participants)

RTP attributes defined in standards:
- Data Frame: "DF_ParticipantData"
  - YD/T 3709-2020
  - T/CSAE 53-2020

RTIs (Road Traffic Incidents)

RTI attributes defined in standards:
- Data Frame: "DF_RTEData"
  - YD/T 3709-2020
  - T/CSAE 53-2020
- Associated **RTI types**
  - GB/T 29100-2012

RTSs (Road Traffic Signs)

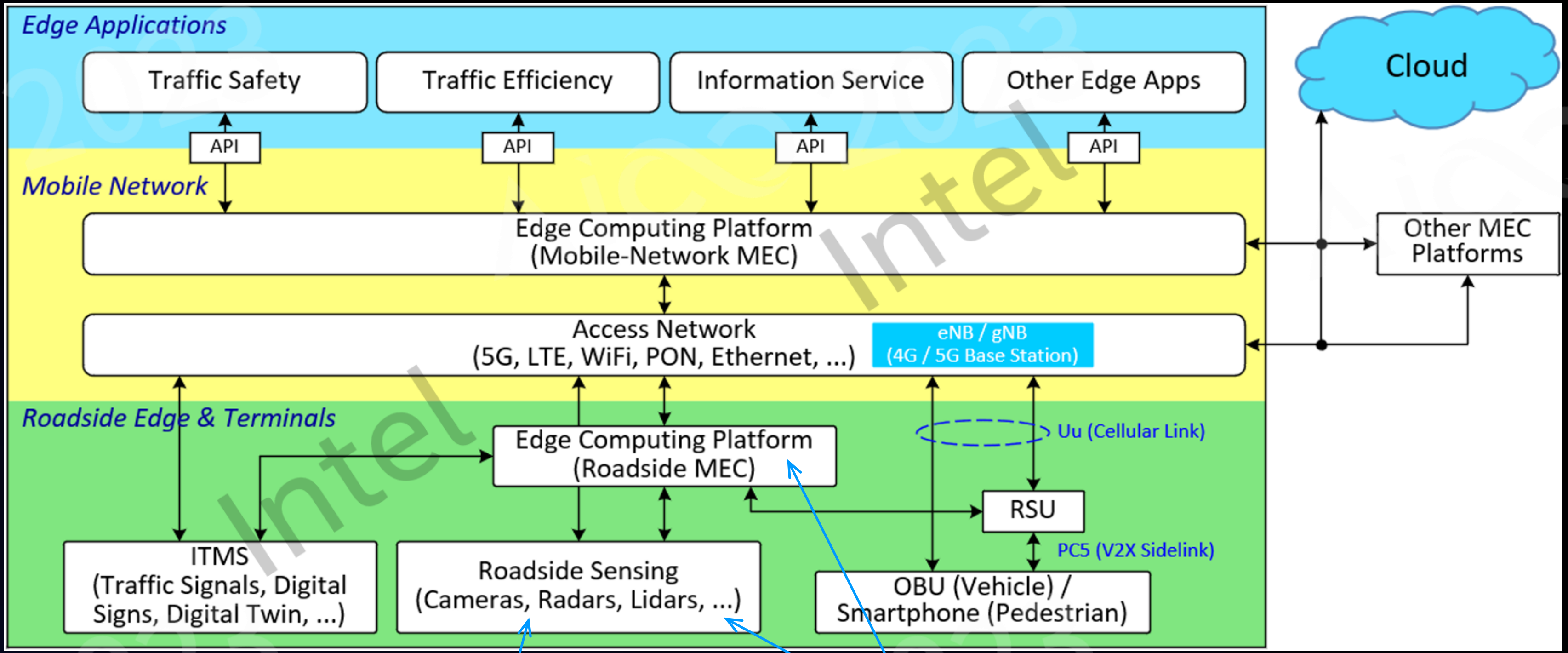
RTS attributes defined in standards:
- Data Frame: "DF_RTSData"
  - YD/T 3709-2020
  - T/CSAE 53-2020
- Associated **RTS types**
  - GB 5768.2-2009

Roadside Sensing – Traffic Sensors – Pros & Cons

Traffic Sensor Type (Strengths in detecting Traffic Object Attributes)	Advantages	Disadvantages
Camera (Object classification and estimation of pixel coordinates on a 2D plane)	1) Rich in details, excellent discernibility; 2) Can accurately capture the contour, texture, color distribution, and so on as well as generate semantic information which can facilitate the object classification under non-extreme ambient light conditions; 3) Can recognize static and plane traffic signs such as traffic lanes and zebra crossing; 4) High lateral resolution can be used to estimate relative lateral velocity; 5) Video and image processing technologies are relatively mature; 6) Relatively low cost and long product life.	1) Susceptible to ambient light conditions (e.g., low light at night, strong sunlight, and so on); 2) Susceptible to weather conditions (e.g., rain, snow, fog, haze, smoke, dust, and so on); 3) Relatively high workload of video analytics based on DL; 4) Lack of depth information, difficult for obtaining accurate 3D info; 5) Low positioning accuracy; 6) Decreased resolution at longer distance; 7) Difficult to estimate relative radial/longitudinal velocity.
mmWave Radar (Estimation of range, velocity and AoA)	1) Capable of estimating the range, velocity (relative radial/longitudinal velocity), micro-Doppler & AoA of the target through conventional signal processing (much lower computing power than that of DL); 2) Generates images with 4D radar or SAR, potential for object classification; 3) Relatively long range; 4) Performance is stable as range increases; 5) All weather conditions (e.g., rain, snow, fog, haze, smoke, dust, etc.); 6) Not affected by light conditions; 7) Relatively low cost and long product life.	1) Difficult for accurate object classification (sparse point cloud); 2) Difficult to distinguish stationary objects (e.g., stopped vehicles) from background (e.g., guardrails, manhole covers, etc.); 3) Difficult to detect pedestrians and small objects at long range; 4) Unable to recognize traffic signs and traffic signals (lights); 5) Detection accuracy of the object's lateral position is not high; 6) Serious lack of environmental details; 7) Ghost detections & false detections (high false alarm probability) caused by multipath, clutter, interference & noise; 8) Difficult for tracking.
Lidar (High-precision positioning and estimation of 3D dimensions of objects)	1) High ranging resolution and precision; 2) High angular resolution; 3) Wide field of view (FoV), 360° FoV for most mechanical lidars; 4) Dense enough data that AI inference (with NN model) can make use of; 5) Strong recognition abilities: capable of 3D imaging (obtaining accurate contours of pedestrians and even smaller objects) and estimation of 3D dimensions of the objects (length, width and height) and other info; 6) Accurate positioning of multiple targets; 7) Strong tracking ability; 8) Not affected by ambient light conditions.	1) Limited performance in recognizing information such as the colors of the objects, traffic signs and traffic signals; 2) DL inference on the 3D point cloud generated by lidar requires relatively high computing power; 3) Susceptible to certain weather conditions (e.g., rain, snow, fog); 4) Resolution & accuracy drop as range increases (not like radar); 5) Measurement affected by platform movement / vibration; 6) High cost.

- No one single type of traffic sensor can meet all requirements of roadway transportation, and that's where **Sensor Fusion** comes in.
- **Multimodal sensors** are not just **complementary**, but provide **redundancy**.

Sensor Fusion – Categories – ISF & DSF (by Integration Level)



ISF (Integrated Sensor Fusion): a.k.a. **AIO** (All-In-One), **Sensing** (1C+1R / 1C+1L) & **Computing** are integrated in one device

- **C:** Camera
- **R:** mmWave Radar
- **L:** Lidar

DSF (Distributed Sensor Fusion): **Sensing** (at least 4Cs / 4Cs+4Rs / 4Cs+2Ls for an intersection) & **Computing** are in multiple split devices connected via network

SoC Processors | iGPU | FPGA

SoC Processors | iGPU | dGPU

Sensor Fusion – "C+R" – Concept


©SHUTTERSTOCK.COM/SENSVECTOR

INTEGRATED SENSOR FUSION BASED ON 4D MIMO RADAR AND CAMERA

A Solution for Connected Vehicle Applications

Ming Lei, Daning Yang, and Xiaoming Weng

This article presents an integrated sensor fusion (ISF) solution based on the multiple-input, multiple-output (MIMO) radar, camera, and on-device computing. The MIMO radar is capable of estimating an object's attributes in four dimensions—range, velocity, azimuth angle, and elevation angle—which can be further used to estimate the length, width, and height of the object. The camera is responsible for object classification based on deep learning. The respective signal processing pipelines and the fusion of results are carried by the on-device computing platform. These two sensors complement each other very well in detecting and classifying traffic objects. Compared with existing sensor

Digital Object Identifier 10.1109/MVT.2022.3207453
Date of current version: 7 October 2022

38 1556-6072/22©2022IEEE IEEE VEHICULAR TECHNOLOGY MAGAZINE | DECEMBER 2022

M. Lei, D. Yang and X. Weng, ***"Integrated Sensor Fusion Based on 4D MIMO Radar and Camera: A Solution for Connected Vehicle Applications,"*** in IEEE Vehicular Technology Magazine, vol. 17, no. 4, pp. 38-46, Dec. 2022, doi: 10.1109/MVT.2022.3207453.

Impact factor of ***13.609*** (from the Thomson Reuters Journal Citation Reports)

Download:
https://ieeexplore.ieee.org/document/9913510

Sensor Fusion – "C+R" – Intel® SW RI

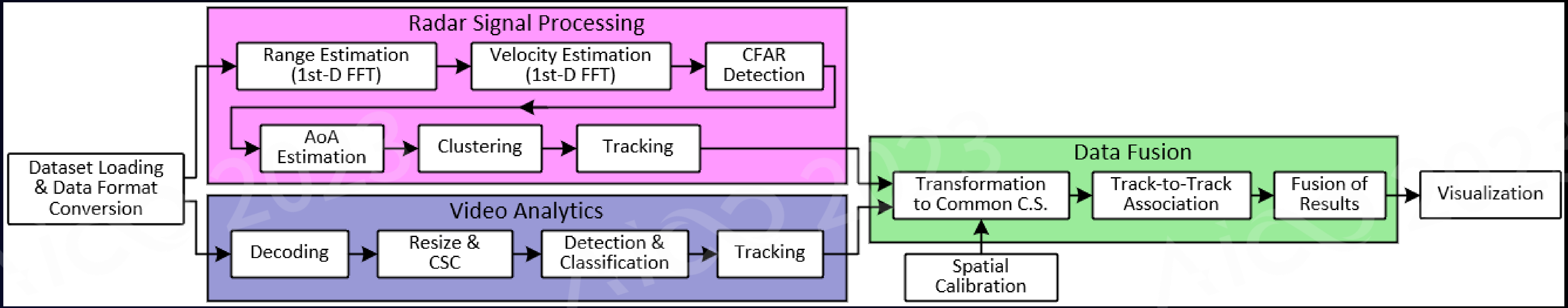
- Intel® SW RI (Reference Implementation): "Garnet Park" (codename)

- Computing platform
  - Processor: Intel® Celeron® 6305E SoC Processor (with iGPU: integrated GPU)

- Software toolkits
  - Intel® Distribution of OpenVINO™ Toolkit
  - Intel® oneAPI Math Kernel Library (oneMKL)

SoC Processor

iGPU (in SoC)



Intel® Celeron® 6305E – A Single Processor for All Workloads

Radar signal processing

Traditional Computer Vision

Fusion of results

→ CPU

Intel® Celeron® SoC Processor		Parameter
CPU	# of Cores	2
	# of Threads	2
	Base Frequency	1.8 GHz
	Cache	4 MB
	Bus Speed	4 GT/s
Memory	Max Memory Size	64 GB
	Memory Types	DDR4-3200, LPDDR4x-3733
	Max # of Memory Channels	2
Processor Graphics	Processor Graphics	Intel® UHD Graphics
	Graphics Max Dynamic Frequency	1.25 GHz
	# of EUs	48
	Media	1 VDBOX
Expansion Options	PCIe for CPU	PCIe Gen 3 × 4 lanes
	PCIe for PCH	PCIe Gen 3 × 4 lanes
	Others	Thunderbolt™ 4 / USB 4 × 4

iGPU (Integrated GPU in SoC)

Deep Learning → # of EUs

Video Coding / Decoding → Media

Product link of Intel® Celeron® 6305E SoC:
https://ark.intel.com/content/www/us/en/ark/products/208072/intel-celeron-6305e-processor-4m-cache-1-80-ghz.html

Intel® Celeron® 6305E SoC Processor

Sensor Fusion – "C+R" – Intel® SW RI – Performance

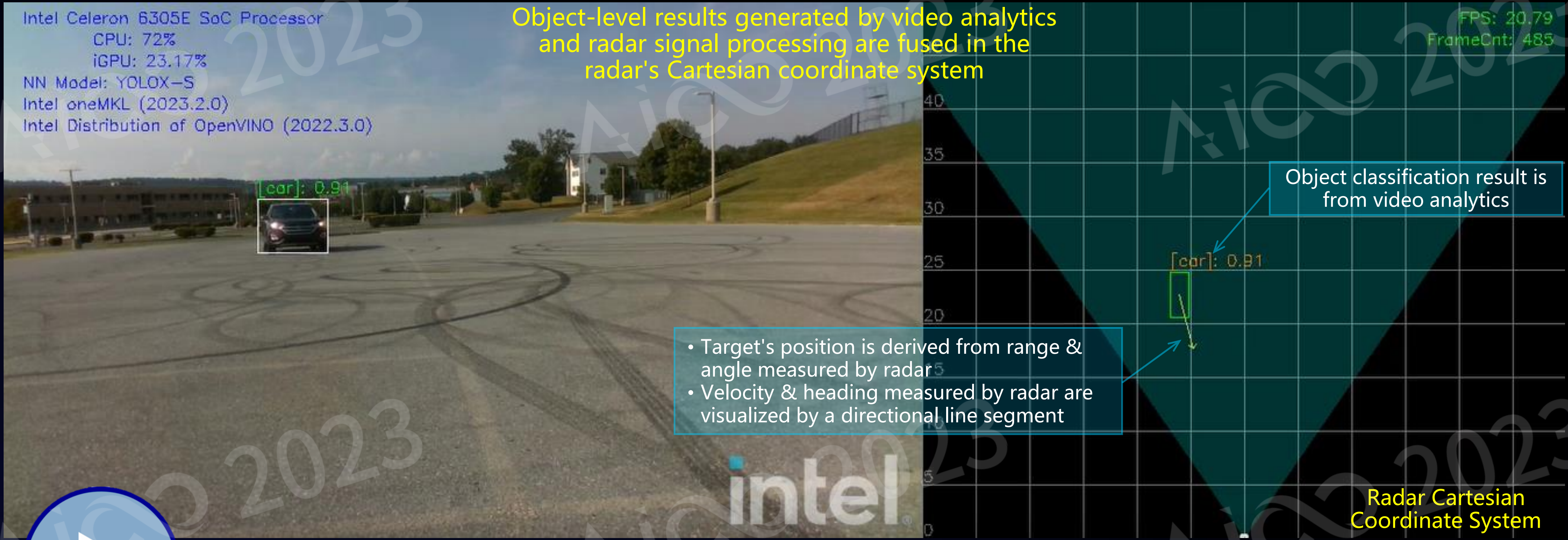
Hardware configuration	Host SoC processor	Intel® Celeron® 6305E SoC Processor (2 cores, 2 threads)
	iGPU	Intel® UHD Graphics 48 EUs, 1.25 GHz
	Memory	8GB, DDR4, 3200 MT/s
Software configuration	OS	Ubuntu 22.04
	Intel® Distribution of OpenVINO™ Toolkit	2022.3.0
	Intel® oneAPI Math Kernel Library (oneMKL)	2023.2.0
	NN model	YOLOX-S
Performance	Throughput (average)	22.61 FPS
	Processing latency (T_1-T_0) (average)	80.17 ms
	CPU loading rate (average)	85.3% (Full loading rate is 200% for 2 CPU threads)
	iGPU loading rate (average)	42.7% (Full loading rate is 100%)

System Configuration & Performance

SoC Processor

iGPU (in SoC)

Sensor Fusion – "C+R" – Intel® SW RI – Demo Video



Demo Video of Sensor Fusion "C+R" on Intel® Processor

Sensor Fusion – "C+R" – Intel® SW RI – Demo Video

Sensor Fusion – "C+L" – Solution White Paper

Solution

intel.

Lidar 3D Point Cloud Processing and Sensor Fusion based on Intel® Architecture for C-V2X

Abstract

In the applications of Cellular Vehicle-to-Everything (C-V2X), the traffic object and traffic environment information obtained by the Roadside Sensing Equipment is analyzed by the Roadside Edge Computing Equipment to generate various Infrastructure-to-Vehicle (I2V) messages which are sent by the Roadside Unit (RSU) to various Road Traffic Participants (including networked vehicles and vulnerable road users) through the wireless links to improve the traffic safety and efficiency. Lidar, as a Roadside Sensing Equipment, is being used more and more widely due to its excellent performance such as 3D imaging and precise ranging. The Roadside Edge Computing Equipment based on **Intel® Architecture** have shown superior performance in processing the 3D point cloud generated by lidar (whether it is deep learning or traditional computer vision). This paper introduces the **JHCTech® Roadside MEC Equipment** based on the **11th-Generation Intel® Core™ Processors** and **Intel® Distribution of OpenVINO™ Toolkit**, which is used to support the 3D point cloud processing based on deep learning and the Sensor Fusion for **Leishen® All-in-One Roadside Sensing Equipment** (Lidar and Camera). We provide cost-effective Roadside Sensing and Roadside Edge Computing solutions for the C-V2X industry.

Authors:

Ming Lei[1,a], Xinfeng Wang[2], Qing Xu[1], Zheng Yan[2,b], Jiegang Cheng[3,c], Shengxian Liu[1], Hao Huang[1]

Companies:

[1] Intel Corporation

[2] Leishen Intelligent System Co., Ltd.

[3] JHC Technology Development Co., Ltd.

E-mail:

[a] ming.lei@intel.com

[b] zhengyan@lslidar.com

[c] chengjiegang@jhctech.com.cn

Background: Rapid Development of China's Intelligent Transportation Infrastructure

In recent years, China has unveiled a number of industrial policies to support the development of Intelligent Transportation System (ITS), such as the "Guidelines on Developing Comprehensive National Transport Network", which emphasizes that the comprehensive deployment of traffic sensing systems is an important step in achieving intelligent transportation infrastructure.

As a new type of industry that integrates technologies such as automotive, semiconductor, wireless communications and transportation, the C-V2X has the great potential to improve the traffic safety and efficiency. With the rapid development of technologies such as artificial intelligence, edge computing and mobile networking, the functions and performance of the C-V2X are constantly being improved, and it will play an important role in the future ITS. China attaches great importance to the development of the C-V2X industry and has successively issued a series of policy documents such as the "Internet of Vehicles (Intelligent Connected Vehicles) Industry Development Action Plan" and "Intelligent Vehicle Innovation and Development Strategy". The Chinese industry has been actively promoting the Vehicle-Infrastructure Collaboration technical route built around the C-V2X technology.

The Roadside Infrastructure is an important part of the C-V2X industry. Its technologies and standards are continuously being improved, and the formation of related industry chains (including Roadside Sensing, Roadside Edge Computing and Roadside Communication, etc.) is also accelerating. The C-V2X Working Group under China's IMT-2020 (5G) Promotion Group has released a research report "Roadside Sensor Fusion based on Edge Computing" which systematically introduces the status quo of the technology and industry development [1].

Download

English Version

English version is also available:

https://www.intel.cn/content/www/cn/zh/internet-of-things/lidar-3d-point-cloud-perception-fusion-in-iov-en.html

解决方案

intel.

车联网中基于英特尔® 架构的激光雷达 3D 点云处理与感知融合方案

摘要

在基于车联网（C-V2X）的车路协同应用中，路侧感知设备获取的交通目标与交通环境信息由路侧边缘计算设备进行分析处理，生成各类 I2V（Infrastructure-to-Vehicle）消息，并由路侧单元（RSU）通过无线链路发送给各类道路交通参与者（包括联网车辆和弱势交通参与者），用于提升交通安全与交通效率。作为路侧感知设备的激光雷达，由于其 3D 成像和精确测距等卓越性能，受到越来越广泛的应用。基于**英特尔® 架构**的路侧边缘计算设备，在处理激光雷达生成的 3D 点云中（无论是深度学习还是传统计算机视觉），都显示出了卓越的性能。本文介绍了基于**第 11 代英特尔® 酷睿™ 处理器**和**英特尔® OpenVINO™ 工具套件分发版**的**集和诚® 路侧 MEC 设备**，用于支持基于深度学习的 3D 点云处理和**镭神® 雷视一体机**（激光雷达与摄像头）的感知融合计算。我们为车联网产业界提供了高性价比的路侧感知和路侧边缘计算解决方案。

作者：

雷鸣[1,a] 王新锋[2] 徐青[1] 严征[2,b] 程加钢[3,c] 刘省贤[1] 黄昊[1]

公司：

[1] 英特尔（中国）有限公司

[2] 镭神智能系统有限公司

[3] 集和诚科技开发有限公司

E-mail:

[a] ming.lei@intel.com,

[b] zhengyan@lslidar.com,

[c] chengjiagang@jhctech.com.cn

背景：中国的智能交通基础设施的快速发展

近几年，国家有关部门出台多项产业政策支持智能交通发展，如《交通强国建设纲要》、《国家综合立体交通网规划纲要》和《国民经济和社会发展第十四个五年规划和 2035 年远景目标纲要》等，均强调全方位布局交通感知系统是实现交通基础设施智能化的重要步骤。

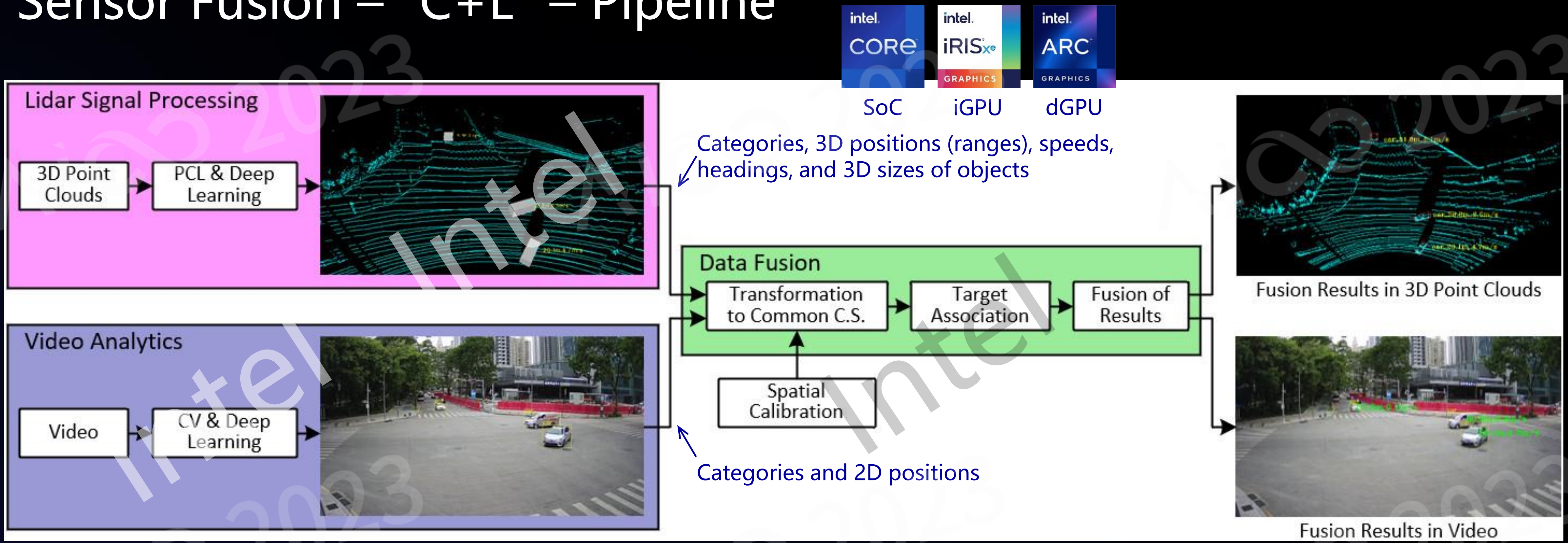
车联网作为半导体、智能计算、无线通信、汽车制造和交通运输等行业深度融合的新型产业形态，在提升交通安全与交通效率方面拥有巨大的潜能。随着人工智能、边缘计算和移动组网等技术的快速发展，车联网的功能与性能也不断的完善和提升之中，将在未来的智能交通系统中将发挥重要的作用。我国高度重视车联网产业的发展，相继出台了《车联网（智能网联汽车）产业发展行动计划》和《智能汽车创新发展战略》等系列政策文件。中国产业界一直在积极推动围绕 C-V2X 技术构建的车路协同技术路线。

路侧基础设施是车联网新基建的重要组成部分，其技术与标准体系正在持续完善之中，而相关的细分产业链（包括路侧感知、路侧边缘计算和路侧通信等）也在加速形成。2020 年 8 月出台的《关于推动交通运输领域新型基础设施建设的指导意见》明确指出要让泛在感知设施深度覆盖交通运输行业。2021 年 9 月，中国 IMT-2020（5G）推进组所辖的 C-V2X 工作组发布研究报告《基于边缘计算的路侧感知融合系统研究》系统性地介绍了路侧感知融合的技术和产业发展现状[1]。

下载

[中文版](中文版)

Sensor Fusion – "C+L" – Pipeline



Pipeline of Sensor Fusion "C+L" (Camera + Lidar)

Source: M. Lei et al., "Lidar 3D Point Cloud Processing and Sensor Fusion based on Intel® Architecture for C-V2X," Intel Solution White Paper, Dec. 2021.

Sensor Fusion – "C+L" – All-in-One Roadside Sensing

Leishen® All-in-One Roadside Sensing Equipment (Lidar & Camera)

Lidar Model		Leishen® CH128X1
Ranging Method		ToF (Pulsed Lidar)
Laser Wavelength		905 nm
Number of Lines (Laser Beams)		128
Maximum Range		160m @ 10% (Reflectivity)
Range Resolution		±3 cm
Data Rate (Single Echo)		760,000 points/sec
Field of View (FoV)	Vertical	-18° - 7°
	Horizontal	120°
Angle Resolution	Vertical	0.125° (Central ROI Region) 0.25° (Side Regions)
	Horizontal	0.1° (5 Hz) 0.2° (10 Hz) 0.4° (20 Hz)

Lidar used to generate 3D point cloud

Source: M. Lei et al., "Lidar 3D Point Cloud Processing and Sensor Fusion based on Intel® Architecture for C-V2X," Intel Solution White Paper, Dec. 2021.

Sensor Fusion – "C+L" – Roadside MEC

Processor SKU	Intel® Core™ i7-1165G7	Intel® Core™ i7-1185GRE
CPU	4 cores, 8 threads Configurable TDP-up Freq.: 2.80 GHz	4 cores, 8 threads Configurable TDP-up Freq.: 2.80 GHz
Processor Graphics (iGPU)	Intel® Iris® Xe Graphics, 96 EUs, Graphics Max Dynamic Freq.: 1.30 GHz	Intel® Iris® Xe Graphics, 96 EUs, Graphics Max Dynamic Freq.: 1.35GHz
AI Performance (FP32)	1.996 TFLOPS (iGPU) 0.358 TFLOPS (CPU)	2.073 TFLOPS (iGPU) 0.358 TFLOPS (CPU)
AI Performance (INT8)	7.987 TOPS (iGPU) 1.433 TOPS (CPU)	8.294 TOPS (iGPU) 1.433 TOPS (CPU)
Memory	8GB, DDR4, Speed: 2400 MT/s	8GB, LPDDR4, speed: 4267 MT/s
BIOS	American Megatrends Inc. 2.21.1278	Intel Corporation TGLSFWI1.R00.4024.A01.2101201730

Configurations of the 11th-Gen Intel® Core™ processors (Tiger Lake)

Operating System	Ubuntu 20.04.1 LTS
Linux Kernel	5.8.0-43-generic
PyTorch	1.7.1+CPU
OpenVINO	2021.3

OS and Software Configuration (Used in Performance Evaluation)

JHCTech® Roadside MEC Equipment (KMDA-3301)
Specs: http://www.jhctechnology.cn/pro_show.php?id=91

- The maximum performance of the two processors is detailed in the product specifications:
  - Intel® Core™ i7-1185GRE Processor: link
  - Intel® Core™ i7-1165G7 Processor: link

Source: M. Lei et al., "Lidar 3D Point Cloud Processing and Sensor Fusion based on Intel® Architecture for C-V2X," Intel Solution White Paper, Dec. 2021.

Thank you !

Ming Lei, PhD
Senior Platform Architect
ming.lei@intel.com

Intel Corporation