

ICS 43.020  
CCS R85

# T/SHV2X

团 体 标 准

T/SHV2X 1—2025

## 汽车驾驶自动化系统通用语料库 第 3 部分：语料数据采集

General corpus for automobile driving automation system—Part 3: data collection

2025 - 01 - 01 发布

2025 - 01 - 01 实施

全国团体标准信息平台

## 目 次

前 言 .....	II
引 言 .....	III
1 范围 .....	1
2 规范性引用文件 .....	1
3 术语和定义 .....	1
4 缩略语 .....	1
5 总体要求 .....	1
6 语料数据资源 .....	2
6.1 语料数据资源分类 .....	2
6.2 基本数据信息 .....	2
6.3 外部环境信息 .....	3
6.4 车辆信息 .....	3
6.5 时间戳信息 .....	3
6.6 采集任务信息 .....	3
6.7 坐标系信息 .....	4
6.8 IMU 数据采集内容 .....	4
6.9 图像数据 .....	4
6.10 视频信息 .....	5
6.11 雷达数据 .....	6
6.12 声音信息 .....	6
6.13 CAN 总线数据 .....	7
6.14 轨迹导航数据 .....	7
7 数据采集要求 .....	7
7.1 采集过程要求 .....	7
7.2 数据传输及存储要求 .....	7
7.3 采集数据质量要求 .....	7
8 数据采集安全 .....	7
8.1 数据接入端 .....	8
8.2 数据采集端 .....	8
8.3 数据存储端 .....	8
附 录 A (资料性) 语料数据资源参考指标 .....	9
附 录 B (资料性) 数据资源质量要求 .....	11
参 考 文 献 .....	12

## 前 言

《汽车驾驶自动化系统通用语料库》系列标准拟分为六个部分：

- 第1部分：总体要求；
- 第2部分：术语与定义；
- 第3部分：语料数据采集；
- 第4部分：语料数据清洗；
- 第5部分：语料数据标注；
- 第6部分：语料数据测试。

本文件为第3部分。

本文件按照GB/T 1.1—2020《标准化工作导则 第1部分：标准化文件的结构和起草规则》的规定起草。

请注意本文件的某些内容可能涉及专利。本文件的发布机构不承担识别这些专利的责任。

本文件由上海市车联网协会提出并归口。

本文件起草单位：上海库帕思科技有限公司、智能汽车创新发展平台（上海）有限公司、智己汽车科技有限公司、地平线征程（上海）科技有限公司、上海临港绝影智能科技有限公司、上海机动车检测认证技术研究中心有限公司、上海金桥智能网联汽车发展有限公司、北京一辅智行科技有限公司、上海优味网络科技有限公司、亿咖通（上海）技术有限公司、滴水智行科技有限公司、上海航盛实业有限公司。

本文件主要起草人：山栋明、黄海清、施佳樑、郭辉、周剑鸣、赵九花、贺锦鹏、蒋达夫、黄剑其、邵亚萌、徐春雷、解瀚光、朱雷、谭龙欢、于峰、滕添益、张帆、周轶、袁月明、李学根、张裕珍、曹宇、胡力旗、邓思文、李想、蔡雨辰、黄鹏飞、刘建业、李勋宏、马骏、李轶刚、刘壹青、田浩、郭晓宾、董连飞、范昌琪。

本文件首批承诺执行单位：上海库帕思科技有限公司、智能汽车创新发展平台（上海）有限公司、智己汽车科技有限公司、地平线征程（上海）科技有限公司、上海临港绝影智能科技有限公司、上海机动车检测认证技术研究中心有限公司、上海金桥智能网联汽车发展有限公司、北京一辅智行科技有限公司、上海优味网络科技有限公司、亿咖通（上海）技术有限公司、滴水智行科技有限公司、上海航盛实业有限公司。

## 引 言

人工智能是新一轮科技革命和产业变革的重要驱动力量,语料数据则是人工智能研究和应用不可或缺的资源,高质量的语料库更是人工智能赋能新质生产力的关键。

随着端到端模型整合度越来越高,对训练语料的规模和质量要求也成倍增加,相对而言,国内车企普遍面临高质量训练数据稀缺、分布分散、一致性差、处理效率低下,数据采集成本较高(全量采集的数据存储、计算成本巨大,且原始数据中高价值信息占比可能不足1%),单款车型上路数量较少、数据采集规模不足,不同厂商、不同车型采集数据分散且难以形成规模效应。

为了实现可靠的汽车驾驶自动化,大量准确的语料数据采集至关重要。数据是自动驾驶系统的基础,它能够帮助系统理解复杂的交通环境、做出明智的决策,并不断优化性能。通过对各种外部环境信息、车辆自身状态信息以及驾驶行为数据的采集和分析,为自动驾驶技术的发展提供有力支持。

围绕“采、洗、标、测”一体化流程,制定《汽车驾驶自动化系统通用语料库》系列标准,本标准作为语料数据采集部分标准。通过本标准的制定,统一的语料数据资源的格式和要求,构建可复用的标准化训练数据集,促进汽车驾驶自动化系统语料资源高效流通利用,降低业内企业重复采集、处理和管理数据的巨额成本,提升本市汽车企业竞争力且推进汽车产业健康发展。为未来端到端技术架构发展所需的数据提供基础支撑和高质量发展的要素型支撑和关键性举措以及高水平对外开放的战略性支撑。

# 汽车驾驶自动化系统通用语料库 第3部分：语料数据采集

## 1 范围

本文件规定了汽车驾驶自动化系统通用语料库的总体要求、语料数据资源、数据采集要求以及数据采集安全等方面的要求。

本文件适用于企业、研究机构等对汽车驾驶自动化系统的研究、开发、维护、应用、评估。

## 2 规范性引用文件

下列文件中的内容通过文中的规范性引用而构成本文件必不可少的条款。其中，注日期的引用文件，仅该日期对应的版本适用于本文件；不注日期的引用文件，其最新版本（包括所有的修改单）适用于本文件。

GB 44497 智能网联汽车自动驾驶数据记录系统  
GB 44495 汽车整车信息安全技术要求  
GB 39732 汽车事件数据记录系统  
GB/T 39267 北斗卫星导航术语  
GB/T 19056 汽车行驶记录仪  
GB/T 17159 大地测量术语  
GB/T 15979 天气类型国标编码  
GB/T 14911 测绘基本术语  
JT/T 794—2019 道路运输车辆卫星定位系统 车载终端技术要求

## 3 术语和定义

下列术语和定义适用于本文件。

### 3.1

**语料数据资源 data resources**

以电子化形式记录和保存的具备原始性、可机器读取、可供社会化再利用的数据集合。

### 3.2

**数据表征模式 data representation modes**

计算机中存储和表达数据的形式或结构。

## 4 缩略语

下列缩略语适用于本文件。

BDS 北斗卫星导航系统 (BeiDou navigation satellite system)  
GNSS 全球卫星导航系统 (Global navigation satellite system)  
GPS 全球定位系统 (Global positioning system)  
IMU 惯性测量单元 (Inertial measurement unit)  
RTK 实时动态差分技术 (Real-Time Kinematic)

## 5 总体要求

### 5.1 语料数据资源

汽车驾驶自动化系统通用语料库的语料数据资源应满足多模态、同步性、多样性、一定规模和质量的要求。

### 5.2 语料数据采集

汽车驾驶自动化系统通用语料库的语料数据资源的采集方法应符合GB 44497、GB 39732、GB/T 17159、GB/T 14911以及《智能网联汽车时空数据安全处理基本要求》所规定要求。

### 5.3 语料数据安全

汽车驾驶自动化系统通用语料库数据资源的采集方应在组织、人员、制度、平台等方面实施安全策略、执行覆盖资源数据和语料产品全生命周期的安全控制措施，对车辆采集、存储和传输的数据进行加密、认证等保护措施，确保数据的保密性、完整性和可用性，应执行并满足GB 44495相关数据安全规定。

## 6 语料数据资源

### 6.1 语料数据资源分类

汽车驾驶自动化系统通用语料库的语料数据资源应包含数据采集车辆定向采集数据和量产车回传数据内容应包含基本数据信息、外部环境信息、车辆主体信息、时间戳设定内容、采集任务描述、激光雷达数据、毫米波雷达数据、摄像头数据、声音信息、轨迹导航数据、IMU数据采集内容等。

汽车驾驶自动化系统通用语料库的语料数据资源的数据种类、用途及数据来源详见表1。

表 1 语料数据资源的数据种类、用途及数据来源

序号	数据内容	用途	数据来源	量产车可选性
1	基本信息	记录语料数据资源包基本数据信息	软件记录	必选
2	外部环境信息	车辆采集发生时的外部环境信息，如天气等	软件记录或软件获取	必选
3	车辆主体信息	本车状态、本车位置、本车运动轨迹	车辆总线、IMU、BDS、GNSS+RTK	必选
4	时间戳设定信息	确保多传感器数据时间校准同步	BDS、GNSS、授时设备	必选
5	采集任务信息	系统性描述记录任务	手动记录、软件记录或软件获取	可选
6	车载传感器数据	感知道路交通参与者、交通设施、道路类型、车道标识标线、场景信息、目标物相对位置、驾驶员状态等原始数据	车载摄像头、毫米波雷达及激光雷达、软件获取或软件记录等	必选
7	声音信息	记录人车交互语音指令和车辆本体声音信息	车载麦克风	可选
8	轨迹导航信息	车辆行驶轨迹信息、高清地图数据、测量车辆三轴角速度及加速度	BDS、GNSS、IMU传感器、OTA等	必选

### 6.2 基本数据信息

基本数据信息应包括唯一编号、提交人姓名、提交人单位、提交人邮箱、语料数据资源名称、声明等信息，见表2。

表 2 语料数据资源包基本信息

序号	名称	字段名	说明
1	唯一编号	package_id	主要指语料数据资源的唯一id
2	提交人姓名	author_name	主要指语料数据资源提交人的姓名
3	提交人单位	author_unit	主要指语料数据资源提交人单位

4	提交人邮箱	author_email	主要指语料数据资源提交人邮箱
5	语料数据资源名称	dataset_name	主要指语料数据资源名称
6	声明	license	主要指支持在署名情况下可商用，如“Apache License 2.0”“Unsure”等

### 6.3 外部环境信息

外部环境信息应包括外部环境名称、环境类型、环境描述、天气、温度、湿度等，见表3所示。

表3 外部环境信息采集内容

序号	名称	字段名	说明
1	外部环境名称	env_name	主要指语料数据资源采集时外部的环境名称，如雨天真实环境采集。
2	环境类型	env_type	主要指语料数据资源采集时的环境类型，如真实环境、仿真环境。
3	环境/场景描述	env_info	主要指语料数据资源采集时的外部环境/场景的描述，如日常驾驶、城区路段驾驶等。
4	天气情况	weather_condition	主要指语料数据资源采集时的天气情况，满足GB/T 15979的天气类型。
5	环境温度	env_temperature	主要指语料数据资源采集时环境的温度值，单位为摄氏度（℃）。
6	环境湿度	env_humidity	主要指语料数据资源采集时环境的相对湿度，单位为百分比（%），手动记录或网络获取
7	道路类型	load_type	主要指语料数据资源采集时的道路类型，如城市道路、农村道路、高速公路等

### 6.4 车辆信息

记录采集车辆信息，应包括车辆序列号、车辆型号、车辆名称、车辆状态、车辆位置、车辆信息等，见表4所示。

表4 车辆信息采集内容

序号	名称	字段名	说明
1	车辆序列号	vehicle_seq	主要指采集车辆的序列号，用于追踪和识别
2	车辆型号	vehicle_type	主要指采集车辆的具体型号，如V6等
3	车辆名称	vehicle_name	主要指采集车辆名称，如V6_001等
4	车辆状态	vehicle_state	主要指采集车辆当前工作状态，如运行，停止，故障等
5	车辆位置	vehicle_position	主要指车辆在物理空间中的位置信息，位置描述或经纬度坐标
6	车辆信息	vehicle_info	主要指采集车辆改装信息，如传感器外参信息、软件版本号等

### 6.5 时间戳信息

对不同的数据进行采集时，应确保同一数据采集系统内部以及不同数据采集系统之间的时间的同步性，确保多传感器数据时间校准同步，时间戳信息见表5所示。

表5 时间戳信息

序号	名称	字段名	说明
1	时间戳名称	timestamp_name	例如：UNIX、POSIX、UTC
2	时间戳类型	timestamp_type	绝对时间、相对时间

### 6.6 采集任务信息

在进行数据采集任务时，应针对工作任务进行系统描述，记录的内容见表6所示。

表6 采集任务内容

序号	名称	字段名	说明
1	任务ID	task_id	主键
2	任务名称	task_name	记录本次任务名称
3	任务描述	task_description	描述执行具体任务
4	任务流程	task_instruction	描述执行采集任务的过程或步骤

## 6.7 坐标系信息

汽车驾驶自动化中坐标系常见分为激光雷达坐标系、相机坐标系、车辆坐标系、站心坐标系以及WGS-84坐标系，见表7所示。

表7 汽车驾驶自动化训练中定义的坐标系类型

序号	坐标系类型	原点位置	X轴方向	Y轴方向	Z轴方向	作用
1	激光雷达坐标系	激光雷达设备中心或底部中心	车辆前方	车辆左侧	指向天空	主要用于3D物体检测、定位
2	相机坐标系	图像中心或相机镜头中心	图像右侧	图像下方	指向相机内部	主要用于2D目标检测
3	车辆坐标系	车辆后轴中心或前轴中心	车辆前方	车辆左侧	指向天空	主要用于获取车辆运动状态、相对位置
4	站心坐标系	定义的固定点，如停车场入口	东方	北方	指向天空	主要用于车辆导航、路径规划
5	WGS-84坐标系	位于地球质心	BIH1984.0的零度子午面和CTP赤道的交点	根据右手原则确定	BIH1984.0定义的协议地球极(CTP)方向	主要用于GPS全球定位系统

车辆坐标系应符合GB 39732-2020中对于车辆坐标系的定义。

## 6.8 IMU 数据采集内容

惯性测量单元是测量物体三轴角速度及加速度的装置。一般IMU包括三轴陀螺仪及三轴加速度计，部分IMU还包括三轴磁力计。IMU数据包括角速度零偏稳定性、加速度零偏稳定性、时间戳同步等。IMU采集数据应包括的信息内容和要求见表8所示。

表8 IMU 数据采集信息

序号	名称	字段名	说明
1	数据序列号	data_seq	-
2	时间戳	timestamp	主要用于数据同步和追踪
3	IMU编号	imu_sensor_id	主要指IMU传感器的标识ID，以区分多个IMU传感器采集数据
4	采集频率	collect_feq	单位: Hz
5	IMU姿态	orientation	表示车辆方向的四元数 (x,y,z,W)
6	线性加速度	linear_acceleration	表示沿各轴加速度的向量 (x,y,z)，单位通常为m/s
7	角速度协方差	angular_velocity_covariance	主要指角速度测量值的协方差矩阵，表示角速度测量的不确定性
8	线性加速协方差	linear_acceleration_covariance	主要指线性加速度测量值的协方差矩阵，表示线性加速度测量的不确定性
9	磁场强度	magnetic_field	表示环境中磁场强度的向量 (x,y,z)，单位通常为微特斯拉 (uT)
10	磁场强度协方差	magnetic_field_covariance	磁场强度测量值的协方差矩阵，表示磁场强度测量的不确定性

用于标识车辆横摆角速度和车辆侧倾角速度方向的车辆坐标系应符合GB 39732中对于车辆坐标系的定义。

## 6.9 图像数据

图像表征的数据，简称图像数据，是以像素矩阵的形式存储，每个像素点包含颜色信息，适用于图像的存储和处理。图像数据资源的指标和要求见表9所示。

表 9 图像数据资源的指标和要求

序号	指标或 (和) 要求		说明
1	类别		见表11
2	主题领域		参照《中国分类主题词表》(第二版)中的定义
3	数据资源内容		图像及对应文字说明或图像介绍
4	图像编码 (文件)	格式	宜JP(E)G、PNG等
5		位深度	8/24
6		分辨率	屏幕分辨率不低于1024×768，扫描图像的扫描分辨率不低于72dpi，确保能够清晰地分辨出远处的小标志和模糊的车道线边界
7	图像真实性 (如需要)		达到《图片真实性鉴定技术规范》(SF/T 0153-2023)中非负面的鉴定结果

### 6.10 视频信息

视频表征的数据，简称视频数据，是以一系列序列帧的形式存储，每个帧可以是图像表征，适用于视频内容的存储和处理。汽车驾驶自动化视频类型详见10所示。

表 10 视频数据的指标和要求

序号	指标或 (和) 要求	说明
1	视频类型	见表11
2	数据资源内容	主要指视频及对应文字说明或视频介绍，如拍摄地点、时间、设备、照明条件、天气条件、道路条件等信息
3	视频分辨率	宜1920x1080及以上
4	视频帧率	宜20fps及以上
5	单一视频时间长度	宜10秒及以上
6	色彩空间标准	主要指选择合适的色彩空间，如 RGB (Red, Green, Blue) 或 YCbCr (Luminance-Chrominance)，并遵循相应的编码标准
7	视频文件格式	MP4/AVI等

视频数据来自车辆前视、侧视、后视摄像头的视频序列，用于物体检测、语义分割、道路检测、交通信号识别等。摄像头视频数据分类及说明见表11所示。

表 11 视频数据分类表

序号	类型	说明
1	前视视频	主要指从车辆前端拍摄，用于识别前方道路状况、车辆、行人、交通信号等
2	侧视视频	主要指从车辆侧面拍摄，监测盲点、辅助变道决策
3	后视视频	主要指从车辆后端拍摄，通常用于倒车时的辅助
4	环视视频	主要通过多个摄像头拼接，提供车辆周围360度的环境视图
5	红外视频 (如需要)	主要指使用红外摄像头拍摄，适用于夜间或光线不足的环境
6	立体视频 (如需要)	主要指采用两个或以上的摄像头从不同角度拍摄，用于深度感知和3D建模

视频数据场景信息包含的天气条件、照明条件和道路类型等多种数据语义特征，汽车驾驶自动化数据语义特征见表12所示。

表 12 视频数据语义特征表

序号	类型	说明
1	天气	主要包括晴朗、多云、阴天、雨天、雾天、雪天、沙尘暴、雷电等
2	时间	主要包括白天、夜晚、黎明、黄昏等
3	光照	主要包括光线较好、光线较弱、黑暗、过曝/眩光、大面积阴影、反射虚影等
4	道路	主要包括城市道路、高速公路、乡村道路、山区道路、施工路段、砾石道路、湿滑道路、冰雪道路、隧道、桥梁等
5	场景	主要包括高速、城区、郊区、高架上、高架下、路口、无车道路面、隧道内、匝道口、收费站、交警管制路段、施工区、事故现场、园区/加油站内、环岛、停车场等

6	交通密度	主要包括低密度、中密度、高密度等
---	------	------------------

相机水平视场角应能覆盖360°，应与汽车驾驶自动化数据集保持一致，多个视角内的目标可追踪，保持特征一致。

### 6.11 雷达数据

本文件规定的雷达数据包括激光点云数据和毫米波数据。

激光点云表征的数据，简称点云数据，通过雷达系统获取的三维空间信息集合，可通过发射激光束并接收反射回来的信号，精确测量物理的距离、形状和位置，构建三维环境模型。

激光点云数据可分为360°机械激光雷达和固态激光雷达，360°机械激光雷达一般用于真值数据采集。

360°机械激光雷达点云数据和固态激光雷达点云数据的指标和要求可参考附录A。

相关激光点云数据需满足以上指标和要求，如出现以下情形，则不适合作为激光点云数据，详见表13所示。

表 13 不适合作为点云数据资源的情形

序号	情形	说明
1	数据不完整	主要指缺少关键区域的点，如道路、行人或车辆
2	分辨率过低	主要是指点云过于稀疏，无法捕捉物体细节
3	精度不高	主要是指点的位置信息误差大
4	噪声过多	主要指包含大量错误或异常值
5	同步性差	主要指多传感器数据无法准确同步
6	动态范围小	主要指无法覆盖远近距离的物体
7	反射强度不一致	主要指反射强度信息不稳定或不可靠
8	光照影响大	主要指强光或逆光条件下产生阴影或反射
9	天气影响严重	主要指雨、雪、雾等恶劣天气条件下的数据
10	遮挡严重	主要指因树木、建筑物等遮挡无法获取完整场景
11	数据过时	主要指与当前环境变化较大，无法反映当前场景
12	法律和隐私问题	主要是指包含敏感信息，如车牌号、人脸等
13	冗余数据	主要指大量重复或非常相似的数据

毫米波数据，擅长测量目标的距离和速度，对金属物理比较敏感，在恶劣环境下也能较好工作。毫米波数据资源的指标和要求见表14所示。

表 14 毫米波数据的指标和要求

序号	指标或（和）要求	说明
1	频率范围	宜76GHz及以上
2	检测范围	宜30米及以上
3	角分辨率	宜1°~3°，±0.5°（长距离）/±1°（中距离）
4	距离分辨率	厘米到毫米级别，宜0.1米及以上
5	速度分辨率	宜0.1米/秒
6	更新率	宜20 Hz及以上
7	调制方式	FMCW
8	检测精度	距离：±0.05m，角度：±0.5°，速度：±0.1 m/s
9	文件格式	支持CAN/Ethernet等通信协议

### 6.12 声音信息

声音表征的数据，简称声音信息。是以音频和文字形式存储的音频对。声音信息应包含的内容见表15所示。

表 15 声音信息数据应包含的内容

序号	名称	字段名	说明
1	数据序列号	data_seq	
2	时间戳	timestamp	用于数据同步和追踪
3	采集频率	collect_feq	单位: Hz
4	语音指令	voice_instruction	人车交互中, 对车辆发出的语音信息
5	环境声音	voice_env	车辆驾驶过程中, 车辆发出和环境中的声音信息

### 6.13 CAN 总线数据

车辆为采集平台硬件基础, 车辆的驾驶行为和车辆状态数据是重要组成部分之一, 车辆总线可快速获取此类数据, 车辆总线数据可参考附件A。

### 6.14 轨迹导航数据

本文件规定的轨迹导航数据包括定位数据、轨迹数据等。轨迹导航数据的说明见表16所示。

表 16 轨迹导航数据资源说明

序号	指标或 (和) 要求	说明
1	类别	见表19
2	数据资源内容	主要指轨迹导航数据及对应文字说明, 且符合JT/T 794的要求
3	格式	主要指JSON或采集数据包
4	验收标准	主要指定位精度符合传感器精度标准, 包含时间数据, 且与视频数据保持时空一致性

轨迹导航数据的资源分类详见表17所示。

表 17 轨迹导航数据资源分类

序号	类型		说明
1	定位数据	GPS	主要包括定位精度、定位连续性、时间戳同步
2		IMU	主要包括角速度零偏稳定性、加速度零偏稳定性、时间戳同步
3	轨迹数据	轨迹点	主要包括位置精度、速度精度、时间戳同步

## 7 数据采集要求

### 7.1 采集过程要求

在数据采集开始前以及结束后, 应验证采集设备, 确保采集设备处于正常工作状态, 无故障, 且验证采集数据的合理性、正确性。

在将采集到的语料数据资源存入数据库前, 应检查其格式的统一性和数据的完整性。

### 7.2 数据传输及存储要求

语料数据资源的采集、传输、存储, 应考虑传输通道与存储系统等的整体性能, 以确保数据的完整性、准确性和安全性, 避免数据丢失、数据损坏等情况的发生。对数据的存储介质、存储能力、存储覆盖机制应执行和满足GB 44497以及GB/T 19056相关数据存储格式和要求, 同时还需要满足自然资源发〔2024〕139号《自然资源部关于加强智能网联汽车有关测绘地理信息安全管理的通知》的要求。

### 7.3 采集数据质量要求

采集数据的质量要求, 应满足准确性、完整性、一致性、可溯源性、代表性和可移植性, 具体要求可参考附录B。

## 8 数据采集安全

### 8.1 数据接入端

对于数据接入端，如传感器自带或外加加密模块，应采用硬件/软件加密后传输的方式增加数据安全性。

### 8.2 数据采集端

对于数据采集端的加密，包括数据采集接口自带或外加加密模块，应在采集端进行硬件/软件加密后显示或存储。

### 8.3 数据存储端

对于数据存储端，如语料数据资源存储器或存储平台，应采用以下方式进行加密：设置权限、源文件加密、自建格式等。

**附录 A**  
**(资料性)**  
**语料数据资源参考指标**

## A.1 360°机械激光雷达点云数据指标和要求

表 18 360°机械激光雷达点云数据指标和要求

序号	技术参数	说明
1	扫描原理	机械旋转
2	通道数	64线及以上
3	探测距离	0.3 ~ 200m (10%反射率)
4	测距准度	± 5cm(0.3 ~ 1m), 宜 ± 2cm(1 ~ 200m)
5	水平视场角	360°
6	水平角分辨率	0.2° (10 Hz), 0.4° (20 Hz)
7	垂直视场角	40° (-25° ~ +15°)
8	垂直视场角	0.167° (-6° ~ +2°) 1° (+2° ~ +3°, -14° ~ -6°) 2° (+3° ~ +5°) 3° (+5° ~ +11°) 4° (+11° ~ +15°) 5° (-19° ~ -14°) 6° (-25° ~ -19°)
9	扫描频率	宜10Hz及以上
10	传输方式	UDP/IP百兆以太网及以上
11	输出数据	距离、角度、反射率
12	点频	单回波: 1,152,000 点/秒及以上, 双回波: 2,304,000 点/秒及以上
13	噪声水平	距离误差 ± 2 ~ 3厘米, 角度误差0.01度到1度
14	时间同步准度	≤1 微秒
15	文件格式	PCD/LAS/XYZ

## A.2 固态激光雷达点云数据的指标和要求

表 19 固态激光雷达点云数据指标和要求

序号	技术参数	说明
1	扫描原理	一维转镜扫描
2	通道数	128线及以上
3	探测距离	1 ~ 210m (10%反射率)
4	测距准度	± 5cm
5	测距精度	3cm及以上
6	水平视场角	120°
7	水平角分辨率	远距 (>7.2 m) : 0.1° (10 Hz) 、宜0.2° (20 Hz) 近距 (0.5 ~ 7.2 m) : 0.2° (10 Hz) 、宜0.4° (20 Hz)
8	垂直视场角	25.4° (-12.5° ~ 12.9°)
9	垂直角分辨率	远距 (> 7.2 m) : 宜0.2° 近距 (0.5 ~ 7.2 m) : 宜0.8°
10	扫描频率	10Hz及以上
11	传输方式	UDP/IP百兆以太网及以上
12	输出数据	距离、角度、反射率
13	有效点频	单回波1,536,000点/秒及以上, 双回波3,072,000点/秒及以上
14	点云数据传输效率	单回波55.9/111.8 Mbps (平均值/峰值), 双回波111.8/223.6 Mbps (平均值/峰值)
15	噪声水平	距离误差 ± 2 ~ 3厘米, 角度误差0.01度到1度
16	时间同步准度	≤1 微秒
17	文件格式	PCD/LAS/XYZ

## A.3 CAN总线数据指标和要求

表 20 车辆 CAN 总线数据指标和要求

序号	系统名称	字段名称	字段含义	备注	性能指标
1	状态反馈	vehidnmdig	车辆VIN	Hash加密	反馈频率50Hz
2		elecprkbrkstat	驻车状态	驻车开启的状态	反馈频率10Hz
3		trnlmpstat	转向灯状态	包含左转、右转、闪光灯	反馈频率50Hz
4		lightstat	车灯状态	包含大灯, 刹车灯等状态	反馈频率20Hz
5		tireprs	车辆轮胎胎压	车辆四个轮胎胎压	反馈频率20Hz
6		trnsshftlvrpos	车辆挡位反馈	当前挡位信息	反馈频率20Hz
7		brkpdlpos	刹车反馈	刹车命令/踏板百分比	反馈频率50Hz
8		thrpos	油门反馈	油门命令/踏板百分比	反馈频率50Hz
9		strwhang	方向盘角度	方向盘命令 方向盘转动角度、扭矩	反馈频率50Hz
10		actvehaccel	车辆加速度	车辆的瞬时加速度	反馈频率50Hz
11		vehodo	车辆里程	车辆行驶的总里程	反馈频率50Hz
12		dsttrvld	本次行驶里程	本次启动后车辆行驶里程	反馈频率50Hz
13		vehspdavg	平均行驶车速	本次里程平均速度	反馈频率50Hz
14		batvlt	电池电压	电池电压状态	反馈频率50Hz
15		hvdpltnmdrng	剩余里程	剩余可行驶里程	反馈频率50Hz
16		batsoc	充电状态	充电状态	反馈频率10Hz
17		offbrdhvblkchgcpplttm	充电信息	剩余充电时间	反馈频率10Hz
18		adsstatus	智驾激活状态	记录并统计智驾的激活状态	反馈频率50Hz
19		doorstat	车门开关状态	四门车/后备厢/引擎盖	反馈频率20Hz
20	线控	strwhctl	转向控制	转向范围-470°~470° 转速范围0~500° 独立使能	转速分辨率: 2°/s 转角分辨率: 0.01° 控制频率: 50Hz 延时小于300ms
21		thrctl	油门控制	油门范围0%~100%	分辨率0.01% 控制频率: 50Hz 延时小于300ms
22		brkpdctl	刹车控制	刹车范围0%~100% 刹车灯使能 独立使能	分辨率0.01% 控制分辨率: 50Hz 延时小于300ms
23		acclctl	加减速控制	控制范围-15m/s <sup>2</sup> ~ 15m/s <sup>2</sup>	分辨率0.01m/s <sup>2</sup> 控制分辨率: 50Hz 延时小于300ms
24		trnsshftlvrctl	挡位控制	PRNDL挡位随意切换	延时小于500ms
25		trnlmpctl	转向灯控制	左右停随意切换	延时小于500ms

车辆必须至少包含油门刹车控制或者加减速控制中的一种完整纵向控制类型, 上述信息为描述场景的必备要素, 是场景应用的基础信息。采集的场景数据将至少包含上述信息, 以满足后期场景数据处理的需求。

## 附录 B (资料性) 数据资源质量要求

### B.1 准确性

汽车驾驶自动化系统通用语料库数据资源的准确性要求应包括：

- (a) 数据形式的合理性；
- (b) 记录信息的准确性；
- (c) 文字与数据描述应准确、清晰、无歧义；
- (d) 应避免数据压缩等因素造成的数据不准确性。

### B.2 完整性

汽车驾驶自动化系统通用语料库数据资源应包含实现数据集预期用途、满足应用场景需要的信息，完整性要求应包括：

- (a) 数据结构与格式应支持数据元、标注信息、元数据的保存、传输与访问，避免信息空白、遗漏或丢失；
- (b) 对不可避免的缺失数据、离群值，给出明确的处理方式，证明应用场景的合理性，对相关数据元加以标识；
- (c) 数据包含的信息应支持数据标注与参考标准溯源的需要。

### B.3 一致性

汽车驾驶自动化系统通用语料库数据资源一致性应包括：

- (a) 内部一致性：
  - 1) 同一数据元在数据集生存周期不同阶段保持一致；
  - 2) 数据特征保持一致；
  - 3) 数据采集、预处理方法保持一致；
  - 4) 原始记录、中间记录与最终记录保持一致。
- (b) 外部一致性：
  - 1) 不同来源的数据在数据特征方面保持一致；
  - 2) 离群值可解释；
  - 3) 不同来源的数据在采集环节遵从相同法律法规、技术文件要求。

### B.4 可追溯性

汽车驾驶自动化系统通用语料库数据资源应具备证明数据可被追溯的验证证据，应包括：

- (a) 原始数据来源、元数据来源、合规性证明；
- (b) 数据采集活动记录；
- (c) 数据更新记录；
- (d) 数据质疑、审计、停用、纠正记录；
- (e) 数据服务异常与故障记录；
- (f) 数据维护与备份记录；
- (g) 数据更新记录。

### B.5 代表性

汽车驾驶自动化系统通用语料库数据资源的来源、多样性以及接近应用场景程度应符合相关技术文件要求。

### B.6 可移植性

汽车驾驶自动化系统通用语料库数据资源应允许在不同平台和系统下互通使用，数据质量不应随平台和系统发生变化。

### 参 考 文 献

- [1] 智能网联汽车时空数据安全处理基本要求
- [2] 自然资发〔2024〕139号 自然资源部关于加强智能网联汽车有关测绘地理信息安全管理的通知

全国团体标准信息平台