# 自动驾驶感知系列课程

审核：Tom Hardy

编辑人员：Tom Hardy、爱喝咖啡的小阿飞、陈伟、袁升、方志刚

# 自动驾驶感知系列

## 一、绪论

## 二、自动驾驶系统与传感器介绍

### 1、Camera

短焦、中焦、长焦相机，负责2D检测和识别；

### 2、鱼眼相机

超广角相机，利于环视检测，主要用在自动泊车或者环视检测上；

### 3、Lidar

三维空间数据采集，根据指定发射线，TOF方式返回对应xyz值，雨雪天效果不好；

### 4、Radar

毫米波雷达，通过指定波段的电磁波，TOF方式返回横向和纵向距离，雨雪天均可；

### 5、IMU

位姿计算，辅助定位

### 6、GPS

定位



### 7、超声波雷达

障碍物检测，倒车雷达

### 8、自动驾驶开源框架与仿真工具

1. Apollo
2. Autoware
3. AirSim
4. Udacity self-driving car
5. CARLA
6. LGSVL

## 三、自动驾驶感知模块

### 1、传感器标定

#### 相关综述

1. A Review of Data Fusion Techniques
2. A COMPREHENSIVE REVIEW OF THE MULTI-SENSOR
3. Multisensor data fusion：A review of the state-of-the-art

**单相机标定方法**

在理想相机成像模型(针孔成像模型)中,涉及到如下四个常用坐标系:

- 像素坐标系
- 图像坐标系
- 相机坐标系
- 世界坐标系

其中像素坐标系与图像坐标系由于坐标原点和单位的不同呈现出线性相关性(平移和缩放),图像坐标系与相机坐标系为比例关系(缩放),世界坐标系与相机坐标系则需要通过刚体变换(三维旋转和平移).

相机畸变:

- 径向畸变
    - 枕型畸变(由内向外凸出)
    - 桶型畸变(由外向内凹陷)
- 切向畸变

径向畸变 在相机制造过程中,很难保证镜头的厚度完全均匀,由于制造工艺的原因,通常为这种情况为中间厚、边缘薄,因而光线在远离透镜中心的地方,会发生更大程度的扭曲,这种现象在鱼眼相机(桶形畸变)中尤为明显.

$$\begin{bmatrix} x' \\ y' \end{bmatrix} = \left(1 + k_1 r^2 + k_2 r^4 + k_3 r^6\right) \begin{bmatrix} x \\ y \end{bmatrix}$$

这里 $x, y, x', y'$ 均为归一化平面下的坐标(焦距 $f$ 为1,即图像坐标系和相机坐标系相同)

- $r$为曲率半径,有$r^2 = x^2 + y^2$
- $k_1, k_2, k_3$ 为径向畸变系数,并且随着阶数的增加,$k$矫正的区域向外扩散
- $x, y, x', y'$ 分别为矫正前和矫正后的像素点坐标



切向畸变 切向畸变产生的原因在于相机在制造过程中,成像平面与透镜平面不平行,产生了透视变换,这里直接给出矫正公式

$$\begin{bmatrix} x' \\ y' \end{bmatrix} = \begin{bmatrix} 2p_1 xy + p_2\left(r^2 + 2x^2\right) \\ 2p_2 xy + p_1\left(r^2 + 2y^2\right) \end{bmatrix}$$ $p_1, p_2$ 为切向畸变矫正系数

两者联合矫正的公式为直接相加即可

$$\begin{bmatrix} x' \\ y' \end{bmatrix} = \underbrace{\left(1 + k_1 r^2 + k_2 r^4 + k_3 r^6\right) \begin{bmatrix} x \\ y \end{bmatrix}}_{\text{径向畸变}} + \underbrace{\begin{bmatrix} 2p_1 xy + p_2\left(r^2 + 2x^2\right) \\ 2p_2 xy + p_1\left(r^2 + 2y^2\right) \end{bmatrix}}_{\text{切向畸变}}$$

张正友标定法

具体流程

- 把被取的十个点的世界坐标(齐次坐标)进行转置
- 对单应性矩阵求解并优化
- 把六幅图的单应矩阵求解出来后求解出6向量(B矩阵).因为每个单应矩阵可以得到两个方程,通过循环对矩阵y赋值后,再对y进行正交分解即可得到6向量.进而得到相机的内参矩阵
- 先求解出相机的外参,然后对畸变系数进行求解,得到相机坐标(Xc, Yc, Zc)
- 调用函数对内参和畸变系数进行优化,并显示优化后的结果.然后根据优化后的结果求解外参矩阵
- 从旋转矩阵中分解出独立变量(三个坐标的转角),再得到平移矩阵,最后把它们和内参、畸变系数一起优化进行最终优化

### 单相机在线标定

1. A New Technique of Camera Calibration: A Geometric Approach Based on Principal Lines
2. Traffic Surveillance Camera Calibration by 3D Model Bounding Box Alignment for Accurate Vehicle Speed Measurement
3. Autocamera Calibration for traffic surveillance cameras with wide angle lenses

### 多相机标定

1. 相关paper 1、Calibration of Asynchronous Camera Networks: CALICO

   2、Infrastructure-based Multi-Camera Calibration using Radial Projections

   3、 Infrastructure-Based Calibration of a Multi-Camera Rig

   4、Leveraging Image-based Localization for Infrastructure-based Calibration of a Multi-camera Rig
2. 开源代码 https://github.com/idiap/multicamera-calibration https://github.com/strawlab/MultiCamSelfCal https://github.com/seed93/Calibration

### 鱼眼相机标定

1. 相关paper Calibration of fisheye camera using entrance pupil
2. 开源工程 https://github.com/sourishg/fisheye-stereo-calibration

### Lidar 和 Camera标定

1. Spatiotemporal Camera-LiDAR Calibration: A Targetless and Structureless Approach
2. An Extrinsic Calibration Tool for Radar, Camera and Lidar，项目地址：github.com/tudelft-iv/multi_sensor_calibration
3. LiDAR-Camera Calibration using 3D-3D Point correspondences
4. Online Camera-LiDAR Calibration with Sensor Semantic Information（在线标定）
5. Improvements to Target-Based 3D LiDAR to Camera Calibration
6. LiDAR and Camera Calibration using Motion Estimated by Sensor Fusion Odometry
7. Automatic extrinsic calibration between a camera and a 3D Lidar using 3D point and plane correspondences
8. A Novel Calibration Method between a Camera and a 3D LiDAR with Infrared Images（ICRA2020）
9. Unified Intrinsic and Extrinsic Camera and LiDAR Calibration under Uncertainties（ICRA2020）
10. Analytic Plane Covariances Construction for Precise Planarity-Based Extrinsic Calibration of Camera and LiDAR（ICRA2020）
11. 开源工程： https://github.com/ankitdhall/lidar_camera_calibration https://github.com/mfxox/ILCC https://github.com/beltransen/velo2cam_calibration https://github.com/swyphcosmo/ros-camera-lidar-calibration https://github.com/heethesh/lidar_camera_calibration https://github.com/TurtleZhong/camera_lidar_calibration https://github.com/XidianLemon/calibration_camera_lidar

### Lidar 和 Lidar标定

### Lidar和双目相机标定融合

1. Intersection Safety using Lidar and Stereo sensors

### Lidar和Radar标定

1. Extrinsic and Temporal Calibration of Automotive Radar and 3D LiDAR

### Lidar和事件相机标定

1. Calibration of Event-based Camera and 3D LiDAR

### Camera和Radar

1. Radar and vision sensors calibration for outdoor 3D reconstruction
2. Targetless Rotational Auto-Calibration of Radar and Camera for Intelligent Transportation Systems

### 激光雷达、Camera、毫米波雷达融合

1. Multiple Sensor Fusion and Classification for Moving Object Detection and Tracking.
2. An Extrinsic Calibration Tool for Radar, Camera and Lidar
3. Extrinsic 6DoF Calibration of a Radar – LiDAR – Camera System Enhanced by Radar Cross Section Estimates Evaluation

### IMU标定

### IMU与Camera标定

1. https://github.com/ethz-asl/kalibr/wiki/calibrating-the-vi-sensor

### IMU与Lidar标定

1. https://github.com/ethz-asl/lidar_align



### 深度相机、双目相机与激光雷达联合标定

具体流程:

- 通过计算相机和雷达对同一个标定板来确定两者的相对位姿,其中相机的位姿估计可以通过ArUco二维码来计算
- 将双目相机的点云和雷达点云通过上一步生成的相对位姿进行KD-tree匹配，通过ICP算法计算平移,Kabsch算法计算旋转

论文：LiDAR-Camera Calibration using 3D-3D Point correspondences 论文：[Automatic Extrinsic Calibration Method for LiDAR and Camera Sensor Setups](

### 传感器时间同步

### 传感器空间同步

## 2、2D目标检测

### 目标检测综述

1. A Survey of Deep Learning-based Object Detection
2. Object Detection in 20 Years: A Survey
3. An Overview Of 3D Object Detection
4. Point-Cloud based 3D Object Detection and Classification Methods for Self-Driving Applications: A Survey and Taxonomy
5. Automatic Target Recognition on Synthetic Aperture Radar Imagery: A Survey

**a、anchor-based**

**one-stage**

1. YOLOv3
2. YOLOv4
3. YOLOv5
4. SSD
5. RetinaNet

**two-stage**

1. Faster RCNN
2. Cascade RCNN
3. Mask RCNN

**b、anchor-free**

1. CenterNet
2. CornerNet
3. FCOS
4. NanoDet
5. ExtremeNet
6. FSAF
7. FoveaBox

**c、transformer**

1. DETR
2. Deformable DETR
3. UP-DETR



## 3、3D目标检测

### a、基于点云的三维目标检测算法

1. End-to-End Multi-View Fusion for 3D Object Detection in LiDAR Point Clouds
2. Vehicle Detection from 3D Lidar Using Fully Convolutional Network(百度早期工作)
3. VoxelNet: End-to-End Learning for Point Cloud Based 3D Object Detection
4. Object Detection and Classification in Occupancy Grid Maps using Deep Convolutional Networks
5. RT3D: Real-Time 3-D Vehicle Detection in LiDAR Point Cloud for Autonomous Driving
6. BirdNet: a 3D Object Detection Framework from LiDAR information
7. LMNet: Real-time Multiclass Object Detection on CPU using 3D LiDAR
8. HDNET: Exploit HD Maps for 3D Object Detection
9. PointNet: Deep Learning on Point Sets for 3D Classification and Segmentation
10. PointNet++: Deep Hierarchical Feature Learning on Point Sets in a Metric Space
11. IPOD: Intensive Point-based Object Detector for Point Cloud
12. PIXOR: Real-time 3D Object Detection from Point Clouds
13. DepthCN: Vehicle Detection Using 3D-LIDAR and ConvNet
14. Voxel-FPN: multi-scale voxel feature aggregation in 3D object detection from point clouds
15. STD: Sparse-to-Dense 3D Object Detector for Point Cloud
16. Fast Point R-CNN
17. StarNet: Targeted Computation for Object Detection in Point Clouds
18. Class-balanced Grouping and Sampling for Point Cloud 3D Object Detection
19. LaserNet: An Efficient Probabilistic 3D Object Detector for Autonomous Driving

20. FVNet: 3D Front-View Proposal Generation for Real-Time Object Detection from Point Clou
21. Part-A^2 Net: 3D Part-Aware and Aggregation Neural Network for Object Detection from Point Cloud
22. PointRCNN: 3D Object Proposal Generation and Detection from Point Cloud
23. Complex-YOLO: Real-time 3D Object Detection on Point Clouds
24. YOLO4D: A ST Approach for RT Multi-object Detection and Classification from LiDAR Point Clouds
25. YOLO3D: End-to-end real-time 3D Oriented Object Bounding Box Detection from LiDAR Point Cloud
26. Monocular 3D Object Detection with Pseudo-LiDAR Point Cloud
27. Structure Aware Single-stage 3D Object Detection from Point Cloud（CVPR2020) 源代码
28. MLCVNet: Multi-Level Context VoteNet for 3D Object Detection（CVPR2020) 源代码
29. 3DSSD: Point-based 3D Single Stage Object Detector（CVPR2020） 源代码
30. LiDAR-based Online 3D Video Object Detection with Graph-based Message Passing and Spatiotemporal Transformer Attention（CVPR2020） 源代码
31. PV-RCNN: Point-Voxel Feature Set Abstraction for 3D Object Detection(CVPR2020) 源代码
32. Point-GNN: Graph Neural Network for 3D Object Detection in a Point Cloud（CVPR2020） 源代码
33. MLCVNet: Multi-Level Context VoteNet for 3D Object Detection（CVPR2020）
34. Density Based Clustering for 3D Object Detection in Point Clouds（CVPR2020）
35. What You See is What You Get: Exploiting Visibility for 3D Object Detection（CVPR2020)
36. PointPainting: Sequential Fusion for 3D Object Detection(CVPR2020)
37. HVNet: Hybrid Voxel Network for LiDAR Based 3D Object Detection（CVPR2020)
38. LiDAR R-CNN: An Efficient and Universal 3D Object Detector（CVPR2021)
39. Center-based 3D Object Detection and Tracking(CVPR2021)
40. 3DIoUMatch: Leveraging IoU Prediction for Semi-Supervised 3D Object Detection(CVPR2021)



**b、基于单目的三维目标检测算法**

1. Task-Aware Monocular Depth Estimation for 3D Object Detection
2. M3D-RPN: Monocular 3D Region Proposal Network for Object Detection
3. Monocular 3D Object Detection and Box Fitting Trained End-to-End Using Intersection-over-Union Loss
4. Disentangling Monocular 3D Object Detection
5. Shift R-CNN: Deep Monocular 3D Object Detection with Closed-Form Geometric Constraints
6. Monocular 3D Object Detection via Geometric Reasoning on Keypoints
7. Monocular 3D Object Detection Leveraging Accurate Proposals and Shape Reconstruction
8. GS3D: An Efficient 3D Object Detection Framework for Autonomous Driving
9. Accurate Monocular Object Detection via Color-Embedded 3D Reconstruction for Autonomous Driving
10. Task-Aware Monocular Depth Estimation for 3D Object Detection
11. M3D-RPN: Monocular 3D Region Proposal Network for Object Detection
12. Deconvolutional Networks for Point-Cloud Vehicle Detection and Tracking in Driving Scenarios
13. Learning Depth-Guided Convolutions for Monocular 3D Object Detection（CVPR2020）
14. End-to-End Pseudo-LiDAR for Image-Based 3D Object Detection（CVPR2020）
15. GrooMeD-NMS: Grouped Mathematically Differentiable NMS for Monocular 3D Object Detection(CVPR2021)
16. Delving into Localization Errors for Monocular 3D Object Detection(CVPR2021)
17. M3DSSD: Monocular 3D Single Stage Object Detector(CVPR2021)
18. MonoRUn: Monocular 3D Object Detection by Self-Supervised Reconstruction and Uncertainty Propagation(CVPR2021)
19. Categorical Depth Distribution Network for Monocular 3D Object Detection(CVPR2021)

**c、基于双目的三维目标检测算法**

1. Object-Centric Stereo Matching for 3D Object Detection
2. Triangulation Learning Network: from Monocular to Stereo 3D Object Detection
3. Pseudo-LiDAR from Visual Depth Estimation: Bridging the Gap in 3D Object Detection for Autonomous Driving
4. Stereo R-CNN based 3D Object Detection for Autonomous Driving
5. IDA-3D: Instance-Depth-Aware 3D Object Detection from Stereo Vision for Autonomous Driving（CVPR2020） 源代码
6. Disp R-CNN: Stereo 3D Object Detection via Shape Prior Guided Instance Disparity Estimation（CVPR2020) 源代码
7. DSGN: Deep Stereo Geometry Network for 3D Object Detection(CVPR2020) 源代码

**d、基于RGB-D的三维目标检测算法**

1. Frustum PointNets for 3D Object Detection from RGB-D Data
2. Frustum VoxNet for 3D object detection from RGB-D or Depth images

**e、基于Radar和RGB方式的三维目标检测算法**

1. CenterFusion: Center-based Radar and Camera Fusion for 3D Object Detection

**f、基于融合数据的三维目标检测算法**

1. MLOD: A multi-view 3D object detection based on robust feature fusion method
2. Multi-Sensor 3D Object Box Refinement for Autonomous Driving
3. Pseudo-LiDAR++: Accurate Depth for 3D Object Detection in Autonomous Driving
4. Improving 3D Object Detection for Pedestrians with Virtual Multi-View Synthesis Orientation Estimation
5. Class-specific Anchoring Proposal for 3D Object Recognition in LIDAR and RGB Images
6. MVX-Net: Multimodal VoxelNet for 3D Object Detection
7. Sensor Fusion for Joint 3D Object Detection and Semantic Segmentation
8. 3D Object Detection Using Scale Invariant and Feature Reweighting Networks
9. End-to-End Pseudo-LiDAR for Image-Based 3D Object Detection（CVPR2020） 源代码



## 4、语义分割、实例分割

**语义分割**

1. Deeplab系列

**实例分割**

1. Mask RCNN
2. TensorMask
3. DeepMask
4. Instance FCN
5. Mask Scoring RCNN
6. SOLO系列 参考：https://blog.csdn.net/jiaoyangwm/article/details/105491010
7. YOLACT
8. YOLACT++
9. PANet
10. GCNet
11. BlendMask 参考：https://zhuanlan.zhihu.com/p/102644623
12. CenterMask FCOS+mask head
13. PolarMask
14. MultiPath-Network

15. No-Local Netrual Networks
16. MaskLab+
17. Hybrid Task Cascade

**实时分割网络**

1. YOLACT
2. YOLACT++
3. SOLO
4. CenterMask

## 5、车道线检测

主要应用场景：车道保持和车道偏离预警

**综述**

1. A review of recent advances in lane detection and departure warning system

**传统车道线检测**

霍夫变换，进行直线检测，得出车道线方程，该种方式效果已经逐渐失去优势，被深度学习方式替换；

**近年来新的paper**

**深度学习车道线检测方式：**

1. Ultra Fast Structure-aware Deep Lane Detection（ECCV2020）

**深度学习车道线分割方式：**



首先说下分割方式的一些缺点

- 速度慢

因为分割是逐像素分类的，要对图像中每一个像素点进行分类。为了分割车道线要进行非常密集的计算，导致的结果就是速度比较慢。其实车道线像素其实只占图像很少一部分，想想也不需要进行这么舍本逐末的操作。

- 局部感受野

分割的另一个问题是感受野问题。因为分割一般是全卷积得到分割结果，而卷积基本上是比较局部的，所以每个像素的感受野有限。在其他分割问题中可能问题不大，但在车道线检测中，问题就很大了。由于我们关注的问题大多是上图这种语义线的检测，需要对全局有很好的感知才能实现良好的定位。比如在图1中，对于车道线的定位只有靠周围车流走向这种**全局信息**才能很好地定位。

虽然有些很好的工作，比如SCNN[1]使用不同方向上的特征传播实现信息传递，间接完成了增大感受野，增加全局信息的目标，但是速度更慢了。

1. LaneAF: Robust Multi-Lane Detection with Affinity Fields github
2. Multi-Class Lane Semantic Segmentation using Efficient Convolutional Networks
3. Multi-lane Detection Using Instance Segmentation and Attentive Voting
4. Robust Lane Detection via Expanded Self Attention
5. PINet：Key Points Estimation and Point Instance Segmentation Approach for Lane Detection github
6. Deep Learning Lane Marker Segmentation From Automatically Generated Labels Youtube
7. End to End Video Segmentation for Driving : Lane Detection For Autonomous Car

**深度学习回归曲线方程：**

1. PolyLaneNet: Lane Estimation via Deep Polynomial Regression github
2. End-to-end Lane Detection through Differentiable Least-Squares Fitting github

**3D车道线检测：**

1. 3D-LaneNet+: Anchor Free Lane Detection using a Semi-Local Representation
2. Gen-LaneNet: A Generalized and Scalable Approach for 3D Lane Detection github Datasets （ECCV 2020）
3. Semi-Local 3D Lane Detection and Uncertainty Estimation
4. 3D-LaneNet: end-to-end 3D multiple lane detection （ICCV 2019）

**时序类车道线检测：**

1. FastDraw: Addressing the Long Tail of Lane Detection by Adapting a Sequential Prediction Network （CVPR 2019）
2. RESA: Recurrent Feature-Shift Aggregator for Lane Detection （AAAI 2021）

**Transformer方式：**

1. End-to-end Lane Shape Prediction with Transformers
2. Detecting Lane and Road Markings at A Distance with Perspective Transformer Layers

**基于双目视觉和RGB-D方式：**

1. Real-time stereo vision-based lane detection system
2. Real-time Lane Marker Detection Using Template Matching with RGB-D Camera

**多传感器方式：**



1. Deep Multi-Sensor Lane Detection (IROS2018)
2. FusionLane: Multi-Sensor Fusion for Lane Marking Semantic Segmentation Using Deep Neural Networks github
3. Structure-Aware Network for Lane Marker Extraction with Dynamic Vision Sensor

**其它：**

YOLinO: Generic Single Shot Polyline Detection in Real Time

Keep your Eyes on the Lane: Attention-guided Lane Detection github

RONELD: Robust Neural Network Output Enhancement for Active Lane Detection github

CurveLane-NAS: Unifying Lane-Sensitive Architecture Search and Adaptive Point Blending （ECCV 2020 ）

Towards Lightweight Lane Detection by Optimizing Spatial Embedding （ECCV 2020）

Lane Detection Model Based on Spatio-Temporal Network with Double ConvGRUs

Heatmap-based Vanishing Point boosts Lane Detection

Synthetic-to-Real Domain Adaptation for Lane Detection

E2E-LMD: End-to-End Lane Marker Detection via Row-wise Classification

SUPER: A Novel Lane Detection System

Inter-Region Affinity Distillation for Road Marking Segmentation （CVPR 2020）

Learning Lightweight Lane Detection CNNs by Self Attention Distillation （ICCV 2019）

Driver Behavior Analysis Using Lane Departure Detection Under Challenging Conditions

Agnostic Lane Detection github

Enhanced free space detection in multiple lanes based on single CNN with scene identification

Robust Lane Detection from Continuous Driving Scenes Using Deep Neural Networks

Efficient Road Lane Marking Detection with Deep Learning

Multiple Lane Detection Algorithm Based on Optimised Dense Disparity Map Estimation

LineNet: a Zoomable CNN for Crowdsourced High Definition Maps Modeling in Urban Environments

LaneNet: Real-Time Lane Detection Networks for Autonomous Driving

EL-GAN: Embedding Loss Driven Generative Adversarial Networks for Lane Detection

Towards End-to-End Lane Detection: an Instance Segmentation Approach github

Lane Detection and Classification for Forward Collision Warning System Based on Stereo Vision

Advances in Vision-Based Lane Detection: Algorithms, Integration, Assessment, and Perspectives on ACP-Based Parallel Vision

Spatial As Deep: Spatial CNN for Traffic Scene Understanding （AAAI 2018 ）

Lane Detection Based on Inverse Perspective Transformation and Kalman Filter

VPGNet: Vanishing Point Guided Network for Lane and Road Marking Detection and Recognition

## 6、目标跟踪



**相关综述：**

1. DEEP LEARNING IN VIDEO MULTI-OBJECT TRACKING A SURVEY
2. Machine Learning Methods for Data Association in Multi-Object Tracking
3. Multiple Object Tracking A Literature Review
4. Tracking the Trackers An Analysis of the State of the Art in Multiple Object Tracking
5. MOT16: A Benchmark for Multi-Object Tracking
6. **Object tracking benchmark** [J]. TPAMI, 2015.
7. **Online object tracking: A benchmark** [C]// CVPR, 2013.

**检测类算法：**

**1、tracking-by-detection方式**（多目标）

主要针对目标检测算法和滤波类算法（多目标跟踪），yolo系列、SSD系列、anchor-free系列、two-stage系列等等。

**2、基于Siamese Networks（生成式，主要针对单目标）**

主要通过Siamese网络进行相似度匹配，主要操作为：首先手动选择初始图像中的目标，使用Siamese网络进行特征提取，然后以此特征为标准，遍历后面帧图像的每个位置，对每个位置进行特征提取，然后做比较，确定位置。

传统方式主要是一些特征提取+滤波类搜索算法。其中特征提取主要有：局部、全局特征、模板、直方图、binary pattern、PCA、sparse PCA、SR（sparse representation）、 discriminative model、generative model。

**跟踪类算法：**

1. 相关滤波和搜索类算法 SORT、Deep SORT、KF、UKF、EKF、CSK、KCF/DCF、CN、粒子滤波、马尔可夫链蒙特卡罗法、局部最优搜索、密集抽样搜索等。

**常用的一些跟踪算法汇总：**

| Method | Representation | Search | MU | Code | FPS |
|---|---|---|---|---|---|
| CPF [44] | L, IH | PF | N | C | 109 |
| LOT [43] | L, color | PF | Y | M | 0.70 |
| IVT [47] | H, PCA, GM | PF | Y | MC | 33.4 |
| ASLA [30] | L, SR, GM | PF | Y | MC | 8.5 |
| SCM [65] | L, SR, GM+DM | PF | Y | MC | 0.51 |
| L1APG [10] | H, SR, GM | PF | Y | MC | 2.0 |
| MTT [64] | H, SR, GM | PF | Y | M | 1.0 |
| VTD [33] | H, SPCA, GM | MCMC | Y | MC-E | 5.7 |
| VTS [34] | L, SPCA, GM | MCMC | Y | MC-E | 5.7 |
| LSK [36] | L, SR, GM | LOS | Y | M-E | 5.5 |
| ORIA [58] | H, T, GM | LOS | Y | M | 9.0 |
| DFT [49] | L, T | LOS | Y | M | 13.2 |
| KMS [16] | H, IH | LOS | N | C | 3,159 |
| SMS [14] | H, IH | LOS | N | C | 19.2 |
| VR-V [15] | H, color | LOS | Y | MC | 109 |
| Frag [1] | L, IH | DS | N | C | 6.3 |
| OAB [22] | H, Haar, DM | DS | Y | C | 22.4 |
| SemiT [23] | H, Haar, DM | DS | Y | C | 11.2 |
| BSBT [50] | H, Haar, DM | DS | Y | C | 7.0 |
| MIL [5] | H, Haar, DM | DS | Y | C | 38.1 |
| CT [63] | H, Haar, DM | DS | Y | MC | 64.4 |
| TLD [31] | L, BP, DM | DS | Y | MC | 28.1 |
| Struck [26] | H, Haar, DM | DS | Y | C | 20.2 |
| CSK [27] | H, T, DM | DS | Y | M | 362 |
| CXT [18] | H, BP, DM | DS | Y | C | 15.3 |



图片出自：Online Object Tracking: A Benchmark

KITTI上的跟踪类算法汇总：[http://www.cvlibs.net/datasets/kitti/eval_tracking.php](http://www.cvlibs.net/datasets/kitti/eval_tracking.php)

| | Method | Setting | Code | HOTA | DetA | AssA | DetRe | DetPr | AssRe | AssPr | LocA | MOTA | Compare |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | PC-TCNN | | | **80.90 %** | **78.46 %** | **84.13 %** | **84.22 %** | 84.58 % | **87.46 %** | 90.47 % | 87.48 % | **91.70 %** | |
| 2 | PermaTrack | | | 78.03 % | 78.29 % | 78.41 % | 81.71 % | 86.54 % | 81.14 % | 89.49 % | 87.10 % | 91.33 % | |
| 3 | CyberTrack | | | 77.98 % | 77.16 % | 79.70 % | 82.41 % | 85.19 % | 82.10 % | **91.93 %** | **87.70 %** | 89.58 % | |
| 4 | PC3T | | | 77.80 % | 74.57 % | 81.59 % | 79.19 % | 84.07 % | 84.77 % | 88.75 % | 86.07 % | 88.81 % | |
| 5 | EagerMOT | | code | 74.39 % | 75.27 % | 74.16 % | 78.77 % | 86.42 % | 76.24 % | 91.05 % | 87.17 % | 87.82 % | |
| | A. Kim and L. Leal-Taix'e: EagerMOT: 3D Multi-Object Tracking via Sensor Fusion. IEEE International Conference on Robotics and Automation (ICRA) 2021. | | | | | | | | | | | | |
| 6 | DEFT | | code | 74.23 % | 75.33 % | 73.79 % | 79.96 % | 83.97 % | 78.30 % | 85.19 % | 86.14 % | 88.38 % | |
| | M. Chaabane, P. Zhang, J. Beveridge and S. O'Hara: DEFT: Detection Embeddings for Tracking. arXiv preprint arXiv:2102.02267 2021. | | | | | | | | | | | | |
| 7 | DetecTrack | | | 73.54 % | 72.64 % | 75.25 % | 78.91 % | 82.18 % | 79.10 % | 88.61 % | 86.46 % | 85.52 % | |
| 8 | mono3DT | | code | 73.16 % | 72.73 % | 74.18 % | 76.51 % | 85.28 % | 77.18 % | 87.77 % | 86.88 % | 84.28 % | |
| | H. Hu, Q. Cai, D. Wang, J. Lin, M. Sun, P. Krähenbühl, T. Darrell and F. Yu: Joint Monocular 3D Vehicle Detection and Tracking. ICCV 2019. | | | | | | | | | | | | |
| 9 | OSN | | | 73.10 % | 77.35 % | 69.60 % | 83.05 % | 83.67 % | 78.68 % | 78.75 % | 86.72 % | 90.52 % | |
| 10 | CenterTrack | | code | 73.02 % | 75.62 % | 71.20 % | 80.10 % | 84.56 % | 73.84 % | 89.00 % | 86.52 % | 88.83 % | |
| | X. Zhou, V. Koltun and P. Krähenbühl: Tracking Objects as Points. ECCV 2020. | | | | | | | | | | | | |
| 11 | modat3D | | | 72.77 % | 74.09 % | 72.19 % | 78.13 % | 85.48 % | 74.87 % | 89.21 % | 87.16 % | 85.94 % | |
| 12 | TrackMPNN | | code | 72.30 % | 74.69 % | 70.63 % | 80.02 % | 83.11 % | 73.58 % | 87.14 % | 86.14 % | 87.33 % | |
| | A. Rangesh, P. Maheshwari, M. Gebre, S. Mhatre, V. Ramezani and M. Trivedi: TrackMPNN: A Message Passing Graph Neural Architecture for Multi-Object Tracking. arXiv preprint arXiv:2101.04206 . | | | | | | | | | | | | |
| 13 | SMAT | | | 71.88 % | 72.13 % | 72.13 % | 74.43 % | **87.33 %** | 74.77 % | 88.30 % | 87.19 % | 83.64 % | |
| | N. Gonzalez, A. Ospina and P. Calvez: SMAT: Smart Multiple Affinity Metrics for Multiple Object Tracking. Image Analysis and Recognition 2020. | | | | | | | | | | | | |
| 14 | TuSimple | | | 71.55 % | 72.62 % | 71.11 % | 76.78 % | 83.84 % | 74.51 % | 86.26 % | 85.72 % | 86.31 % | |
| | W. Choi: Near-online multi-target tracking with aggregated local flow descriptor. Proceedings of the IEEE International Conference on Computer Vision 2015.<br>K. He, X. Zhang, S. Ren and J. Sun: Deep residual learning for image recognition. Proceedings of the IEEE conference on computer vision and pattern recognition 2016. | | | | | | | | | | | | |
| 15 | CenterTrack+MTFF | | | 71.35 % | 74.58 % | 68.78 % | 77.95 % | 85.94 % | 71.14 % | 89.17 % | 86.78 % | 86.64 % | |
| 21 | MPMOT | | | 68.84 % | 72.17 % | 66.33 % | 76.13 % | 84.52 % | 69.16 % | 88.37 % | 86.39 % | 84.42 % | |
| 22 | MOTSFusion | | code | 68.74 % | 72.19 % | 66.16 % | 76.05 % | 84.88 % | 69.57 % | 85.49 % | 86.56 % | 84.24 % | |
| | J. Luiten, T. Fischer and B. Leibe: Track to Reconstruct and Reconstruct to Track. IEEE Robotics and Automation Letters 2020. | | | | | | | | | | | | |
| 23 | IMMDP | | | 68.66 % | 68.02 % | 69.76 % | 71.47 % | 83.28 % | 74.50 % | 82.02 % | 84.80 % | 82.75 % | |
| | Y. Xiang, A. Alahi and S. Savarese: Learning to Track: Online Multi- Object Tracking by Decision Making. International Conference on Computer Vision (ICCV) 2015.<br>S. Ren, K. He, R. Girshick and J. Sun: Faster R-CNN: Towards Real- Time Object Detection with Region Proposal Networks. NIPS 2015. | | | | | | | | | | | | |
| 24 | Quasi-Dense | | code | 68.45 % | 72.44 % | 65.49 % | 76.01 % | 85.37 % | 68.28 % | 88.53 % | 86.50 % | 84.93 % | |
| | J. Pang, L. Qiu, X. Li, H. Chen, Q. Li, T. Darrell and F. Yu: Quasi-Dense Similarity Learning for Multiple Object Tracking. CVPR 2021. | | | | | | | | | | | | |
| 25 | UDOLO | | | 68.41 % | 70.61 % | 66.83 % | 74.85 % | 85.24 % | 68.45 % | 91.83 % | 87.47 % | 80.57 % | |
| 26 | MASS | | | 68.25 % | 72.92 % | 64.46 % | 76.83 % | 85.14 % | 72.12 % | 81.46 % | 86.80 % | 84.64 % | |
| | H. Karunasekera, H. Wang and H. Zhang: Multiple Object Tracking with attention to Appearance, Structure, Motion and Size. IEEE Access 2019. | | | | | | | | | | | | |
| 27 | FTLMOT | | | 67.19 % | 73.82 % | 62.48 % | 77.56 % | 85.23 % | 65.20 % | 88.88 % | 86.52 % | 86.06 % | |
| 28 | MI-TIS MOT | | | 66.72 % | 63.87 % | 70.42 % | 68.98 % | 79.07 % | 74.23 % | 84.56 % | 83.61 % | 78.10 % | |
| 29 | JCSTD | | | 65.94 % | 65.37 % | 67.03 % | 68.49 % | 82.42 % | 71.02 % | 82.25 % | 84.03 % | 80.24 % | |
| | W. Tian, M. Lauer and L. Chen: Online Multi-Object Tracking Using Joint Domain Information in Traffic Scenarios. IEEE Transactions on Intelligent Transportation Systems 2019. | | | | | | | | | | | | |
| 30 | MDP | | code | 64.79 % | 63.04 % | 67.05 % | 66.18 % | 82.22 % | 69.61 % | 85.61 % | 84.24 % | 76.08 % | |
| | Y. Xiang, A. Alahi and S. Savarese: Learning to Track: Online Multi- Object Tracking by Decision Making. International Conference on Computer Vision (ICCV) 2015.<br>Y. Xiang, W. Choi, Y. Lin and S. Savarese: Subcategory-aware Convolutional Neural Networks for Object Proposals and Detection. IEEE Winter Conference on Applications of Computer Vision (WACV) 2017. | | | | | | | | | | | | |
| 31 | NOMT* | | | 64.77 % | 63.08 % | 67.04 % | 66.92 % | 79.28 % | 70.38 % | 83.14 % | 82.22 % | 77.91 % | |
| | W. Choi: Near-Online Multi-target Tracking with Aggregated Local Flow Descriptor . ICCV 2015. | | | | | | | | | | | | |
| 32 | SRK_ODESA(car) | | | 64.25 % | 74.87 % | 55.70 % | 78.62 % | 84.68 % | 62.10 % | 81.78 % | 85.85 % | 88.50 % | |
| | D. Mykheievskyi, D. Borysenko and V. Porokhonskyy: Learning Local Feature Descriptors for Multiple Object Tracking. ACCV 2020. | | | | | | | | | | | | |
| 33 | MOTBeyondPixels | | code | 63.75 % | 72.87 % | 56.40 % | 76.58 % | 85.38 % | 59.05 % | 86.70 % | 86.90 % | 82.68 % | |
| | S. Sharma, J. Ansari, J. Krishna Murthy and K. Madhava Krishna: Beyond Pixels: Leveraging Geometry and Shape Cues for Online Multi-Object Tracking. Proceedings of the IEEE International Conference on Robotics and Automation (ICRA) 2018. | | | | | | | | | | | | |
| 34 | PV3DMOT-V1 | | | 63.19 % | 57.70 % | 70.22 % | 71.85 % | 66.95 % | 75.47 % | 84.70 % | 82.44 % | 67.19 % | |
| 35 | MI-TIS MOT | | | 62.48 % | 60.12 % | 65.61 % | 62.69 % | 82.72 % | 68.76 % | 84.94 % | 84.19 % | 73.47 % | |
| 36 | mmMOT | | code | 62.05 % | 72.29 % | 54.02 % | 76.17 % | 84.89 % | 58.98 % | 82.40 % | 86.58 % | 83.23 % | |
| | W. Zhang, H. Zhou, Sun, Z. Wang, J. Shi and C. Loy: Robust Multi-Modality Multi-Object Tracking. International Conference on Computer Vision (ICCV) 2019. | | | | | | | | | | | | |
| 37 | PointTrackV2 | | | 61.26 % | 66.02 % | 58.11 % | 73.22 % | 76.00 % | 71.91 % | 70.39 % | 81.96 % | 81.91 % | |



## 7、深度估计

相关综述：

1. How do neural networks see depth in single images
2. Monocular Depth Estimation A Survey
3. Monocular Depth Estimation Based On Deep Learning：An Overview
4. 双目视觉的匹配算法综述（陈炎）
5. A Survey on Deep Learning Techniques for Stereo-based Depth Estimation

### 单目方式

技术博客：https://zhuanlan.zhihu.com/p/56263560

### a、传统方式

参考传统单目是怎么测距的：[https://zhuanlan.zhihu.com/p/135943895](https://zhuanlan.zhihu.com/p/135943895)

### b、深度学习方式

单目深度学习方式估计深度，一定程度上借鉴了双目的约束关系~

1. Monodepth1: Unsupervised Monocular Depth Estimation with Left-Right Consistency(CVPR2017)
2. Monodepth2: Digging into Self-Supervised Monocular Depth Prediction（ICCV2019)
3. GeoNet: Unsupervised Learning of Dense Depth, Optical Flow and Camera Pose（CVPR2018)
4. Unsupervised Monocular Depth Estimation with Left-Right Consistency（CVPR2017)
5. Single View Stereo Matching
6. Deep Ordinal Regression Network for Monocular Depth Estimation（CVPR2018)
7. Learning Depth from Monocular Videos using Direct Methods(CVPR2018)

### 双目方式

1. Unsupervised Learning of Stereo Matching(ICCV2017)
2. Weakly Supervised Learning of Deep Metrics for Stereo Reconstruction(ICCV2017)

## 8、SFM重建

## 9、后处理

## 10、多传感器融合



（1） 低层融合：直接将原始数据作为数据融合过程的输入，提供比单个源更精确的数据（更低的信噪比）；（2） 中级融合：融合特征或特征（形状、纹理和位置）以获得可用于其他任务的特征。这个层次也称为特征或特征层次；（3） 高级融合：这一级，也称为决策融合，以符号表示为源，并将它们结合起来，以获得更准确的决策。贝叶斯方法通常用于 这一级别；（4） 多层次融合：这个层次处理从不同抽象层次提供的数据（即，当一个度量与一个特征相结合以获得决策时）。

### 推荐三篇综述类paper

1. A COMPREHENSIVE REVIEW OF THE MULTI-SENSOR DATA FUSION ARCHITECTURES
2. A Review of Data Fusion Techniques
3. Multisensor data fusion: A review of the state-of-the-art

### 数据同步

在数据融合前，数据同步非常重要，在时序上保证对齐；

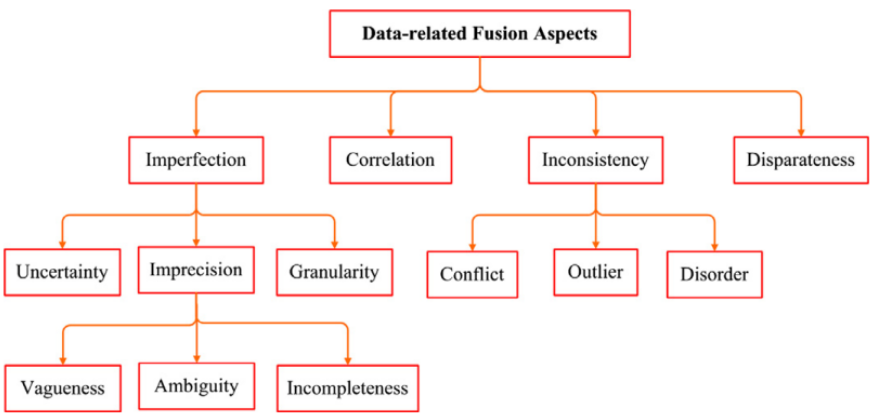


**Fig. 1.** Taxonomy of data fusion methodologies: different data fusion algorithms can be roughly categorized based on one of the four challenging problems of input data that are mainly tackled: namely, data imperfection, data correlation, data inconsistency, and disprateness of data form.

**数据融合的一些难点**

数据缺陷、离群点和噪声数据、数据形态的差异（点云、图像）、数据对齐/关联等；

**数据层面上的融合**

数据层面的融合意在结合多传感器原始数据，输出感知结果；

优点：对算力要求没那么高

缺点：太依赖各个传感器数据，如果有某个传感器数据失效，可能会导致任务失效；

**特征级别的融合**

特征级别上的融合对图像、点云等数据做特征拼接，使用DL方式来做结果上的检测；

**目标/决策层面上的融合**

结合多个结果，做融合投票，输出一个稳定的结果；

优点：结果输出比较稳定，不依赖单个传感器，如有失效，任务仍能有效工作；

缺点：需要处理多个传感器预测的结果，这些分支比较消耗算力；

**融合检测算法汇总**

1. CenterFusion: Center-based Radar and Camera Fusion for 3D Object Detection

# 四、深度学习模型部署

### 1、主流训练框架



pytorch

tensroflow

MXNet

CNTK

Caffe

PaddlePaddle

### 2、主流推理框架

推理框架的基本思路：训练框架的模型（torch、tensorflow等）首先转换为本框架下的模型格式或者中间模型格式（ONNX、NNEF等）然后进行量化推理~

TensorRT（速度优化比较好，英伟达）

OpenVINO（CPU支持良好，英特尔）

NCNN（代码结构相对直观，而且从实际使用经验上看更易于魔改，腾讯）

MNN（阿里）

TNN（腾讯）

MACE（小米）

MegEngine （天元，旷视）

Caffe（Fackbook）

**3、移动端推理框架**

TensorRT（速度优化比较好，英伟达）

OpenVINO（CPU支持良好，英特尔）

NCNN（代码结构相对直观，而且从实际使用经验上看更易于魔改，腾讯）

MNN（阿里）

TNN（腾讯）

MACE（小米）

MegEngine （天元，旷视）

Caffe（Fackbook）

**4、AI芯片汇总**

1. 地平线
2. 海思
3. 寒武纪
4. 英伟达
5. 汇顶科技
6. 平头哥
7. 联发科
8. 紫光展锐

# 五、自动驾驶定位模块



**综述类paper：**

1. A Survey of Autonomous Driving: Common Practices and Emerging Technologies

目前state-of-art的定位主要分为3种：

- **GPS/IMU融合** - Gps和IMU融合的方案是目前主流的定位方案，采用rtk 技术之后定位的精度可以达到厘米级别，目前遇到的问题是在高楼，隧道等信号不太好的地方定位误差会变大。
- **基于先验地图** - 基于先验地图的方案主要是实现把周围环境的信息保存下来，以高精度地图为例子，实现采集并存储高精度地图中的点云信息，然后基于ICP或者NDT算法进行点云匹配，匹配到的位置即是当前车辆的位置，好处是在Gps信号不好的时候也可以使用，缺点是运算量大，在周围环境发生变化时，如何更新地图。
- **SLAM** - slam方法的好处是不依赖于先验地图，可以在陌生环境（乡村或者一些环境变化迅速的场景）使用，缺点是计算量太大，算法做不到实时，基于视觉的slam精度也达不到要求。

## 1、基于视觉定位

- ORBSLAM && ORBSLAM2

## 2、基于点云定位

1. **单线激光雷达带反光板定位**

2. **多线激光雷达定位**

3. **点云重识别与重定位**
4. **配准(注册)传统方法**
    - ICP
    - NDT
5. **基于深度学习的点云重识别与重定位**
6. **未分类**

- 3D Point Cloud Registration for Localization using a Deep Neural Network Auto-Encoder
- PointNetLK: Point Cloud Registration using PointNet
- The Perfect Match: 3D Point Cloud Matching with Smoothed Densities
- Learning multiview 3D point cloud registration
- RPM-Net: Robust Point Matching using Learned Features
- Deep Global Registration
- 基于NDT+UKF的点云定位：(A portable three-dimensional LIDAR-based system for long-term and wide-area people behavior measurement)[https://github.com/koide3/hdl_localization]

## 3、基于融合方式定位

**基于卡尔曼的融合IMU、点云的滤波定位** 论文：[LINS: A Lidar-Inertial State Estimator for Robust and Efficient Navigation](