

文章编号: 2095-2163(2022)08-0065-05

中图分类号: TP391

文献标志码: A

# 基于点云与图像融合的可行驶区域检测

邵哲钦, 黄影平, 郭志阳

(上海理工大学 光电信息与计算机工程学院, 上海 200093)

**摘要:** 可行驶区域检测是无人驾驶的前提, 相机和激光雷达是当前无人驾驶领域普遍采用的 2 类环境感知传感器, 不同传感器融合能够有效提升环境感知的鲁棒性和完备性。本文提出了一种基于激光雷达和相机信息融合的算法, 实现可行驶区域检测。首先, 将激光雷达点云投影至相机图像上, 并构建德洛内三角网络, 为每个离散的点云构建局部关系; 其次, 依据空间关系, 对点云进行障碍物与非障碍物的分类; 再以其所在图像的位置为聚类中心, 对图像进行超像素分割, 完成可行驶区域和障碍物之间的分割。使用 Kitti 公共数据集, 在几种不同场景和条件下进行实验, 获得了鲁棒的结果。

**关键词:** 传感器融合; 可行驶区域检测; 场景表达; 无人驾驶

## Drivable region detection based on point cloud and images fusion

SHAO Zheqin, HUANG Yingping, GUO Zhiyang

(School of Optical-Electrical and Computer Engineering, University of Shanghai for Science and Technology, Shanghai 200093, China)

**[Abstract]** Drivable area detection is the premise of unmanned driving. The camera and LIDAR are the most representative and widely used sensors in the field of unmanned driving. The fusion of different sensors can effectively improve the completeness of environmental perception. In this paper, an information fusion algorithm based on the LIDAR and camera is proposed to realize drivable area detection. Firstly, the proposed algorithm projects the LIDAR point cloud onto the camera image, and constructs the Delaunay triangle network to construct the local relationship for each discrete point cloud. After that, all point clouds are classified into obstacles and non-obstacles, and then the image is super-pixel segmented with the location of the image as the clustering center to complete the segmentation between the drivable area and obstacles and express the traffic scene. Kitti benchmark is used to experiments in several different scenarios and conditions, and robust results are obtained.

**[Key words]** sensors fusion; drivable region detection; scene representation; unmanned driving

## 0 引言

交通场景感知是智能汽车环境感知的基本任务, 汽车只有自主感知交通道路场景环境后, 才能进行车的移动控制决策。无人驾驶汽车主要通过激光雷达、相机、毫米波雷达、GPS 全球定位系统等探测车辆周围的环境信息。其中, 相机和激光雷达是智能汽车中运用最广泛的 2 类传感器, 能准确地感知物体的三维信息, 相机则能够获取环境丰富的纹理信息。现有的研究表明, 将相机与激光雷达两种传感器数据进行融合, 可以提高道路检测精度<sup>[1]</sup>。因此, 相机与激光雷达融合是目前无人驾驶领域的热点和难点问题之一。

Shinzato 等人<sup>[2]</sup>提出了一种基于传感器融合的鲁棒方法。该方法基于单目相机和 3D 雷达的空间

关系将雷达点云投影至相机图像之上, 使用一种障碍物分类方法<sup>[3]</sup>, 将雷达点云分为障碍物点和非障碍物点, 进而处理图像得到可行驶区域。Ren 等人<sup>[4]</sup>于 2003 年首次提出了超像素的概念, 超像素是指具有相似纹理、颜色、亮度等特征的相邻像素构成的有一定视觉意义的不规则像素块。Achanta 等人<sup>[5]</sup>于 2010 年提出了 SLIC 超像素算法, 即简单的线性迭代聚类, 首先将彩色图像转化为 CIELAB 颜色空间和 XY 坐标下的 5 维特征向量, 对 5 维特征向量构造距离度量标准, 对图像像素进行局部聚类。SLIC 算法能生成紧凑、近似均匀的超像素, 在运算速度、物体轮廓保持、超像素形状方面的综合性能上都获得较高评价。

本文先将雷达点云投影至图像中, 对点云进行德洛内三角剖分; 其次, 用两点之间的角度阈值区分

**作者简介:** 邵哲钦(1993-), 男, 硕士研究生, 主要研究方向: 基于雷达点云与相机图像融合; 黄影平(1966-), 男, 博士, 特聘教授, 主要研究方向: 智能汽车障碍物探测辨识、汽车电子系统仿真测试、视觉检测技术; 郭志阳(1993-), 男, 博士研究生, 主要研究方向: 车道线检测、障碍物识别。

**通讯作者:** 黄影平 Email: huangyingping@usst.edu.cn

收稿日期: 2022-02-08

障碍物与地面点;最后,使用超像素聚类,对点云分类结果在图像中进行补全,完成图像中可行驶区域检测。

## 1 方法

本文基于稀疏非结构化的三维激光雷达点云与相机图像融合,首先利用相机坐标与激光雷达的空间关系,进行雷达到图像的投影;其次,对投影至图像中离散无序的点云构建局部空间关系,并结合点云的三维度量来判断该点是否为障碍物;最后,在图像中以点云为核心,对图像所有像素点进行聚类,完成可行驶区域的划分。该方法包括5个处理步骤:

(1)传感器融合。将三维空间中的雷达点云数据投影至相机坐标系中。

(2)构造离散点云的局部空间关系。通过德洛内三角剖分,构建德洛内三角网络,找出点云中最接近的点,并以直线连接,构成三角网络。

(3)点云分类。通过三角网络中点的空间关系,将点云分类为障碍物与非障碍物。

(4)聚类。借鉴 SLIC 超像素算法中的聚类算法,在图像中以分类后的点云为中心进行聚类,将图像中所有像素点分为障碍物点与非障碍物点,对图像的分割结果进行补全。

(5)可行驶区域的提取。以视觉中心点、即图像最底部的中点为基准点,对可行驶区域做进一步分割。

### 1.1 坐标系融合

不同传感器有着不同的独立坐标系,必须把不同坐标系的数据变换到同一坐标系,找到同一时刻点云数据和图像中对应的像素点,才能实现融合。坐标变换主要分为2个步骤,拟做分述如下。

(1)根据相机的内参和外参,将雷达坐标系下的点云  $P_L = (X_L, Y_L, Z_L, 1)^T$  通过坐标变换到相机坐标系,得到  $P_C = (X_C, Y_C, Z_C, 1)^T$ , 对此可表示为:

$$P_C = R_{rect} T_{velo}^{cam} P_L \quad (1)$$

其中,  $R_{rect}$  是从原始相机到校正相机的旋转矩阵,  $T_{velo}^{cam}$  是变换矩阵,展开如式(2):

$$T_{velo}^{cam} = \begin{bmatrix} \alpha_{velo}^{cam} & t_{velo}^{cam} \ddot{o} \\ \zeta_{velo}^{cam} & \ddot{+} \\ \epsilon_{velo}^{cam} & 0 & 1 & \ddot{o} \end{bmatrix} \quad (2)$$

其中,  $R_{velo}^{cam}$  和  $t_{velo}^{cam}$  分别是三维激光雷达坐标系转换为相机坐标系的旋转矩阵和平移向量。

(2)为了得到像素坐标  $(U, V)$ , 将所有  $P_C$  代入公式(3):

$$\begin{bmatrix} \alpha_{C} \ddot{o} \\ \zeta_{C} Y_C \ddot{+} \\ \zeta_{C} Z_C \ddot{+} \\ \epsilon_{C} 1 \ddot{o} \end{bmatrix} = P_{rect} \begin{bmatrix} \alpha_{C} \ddot{o} \\ \zeta_{C} Y_C \ddot{+} \\ \zeta_{C} Z_C \ddot{+} \\ \epsilon_{C} 1 \ddot{o} \end{bmatrix} \quad (3)$$

其中,  $P_{rect}$  为矫正后的相机坐标系至像素坐标系的投影矩阵。

由于激光雷达的探测视野较大,有些点云会投影至图像外,直接剔除这部分图像外的点云,投影结果如图1所示。图1中,颜色越红,表示离激光雷达越近,越蓝则表示越远,将像素坐标  $(U, V)$  与  $(X_C, Y_C, Z_C)$  组成新的集合  $P = (X_C, Y_C, Z_C, U, V)^T$ 。



图1 激光雷达点云投影至图像上的结果

Fig. 1 Results of LIDAR point cloud projection on images

### 1.2 构造离散点云的局部空间关系

利用得到的  $P$  点的  $(U, V)$  坐标构造一个德洛内三角网  $G$ , 这是一系列相连但不重叠的三角形集合,以最近3个点形成一个三角形。这种网络拥有以下特性:

(1)最接近性。以最近的3个点形成一个三角形,且各三角形的边皆不相交。

(2)唯一性。无论从区域何处开始构建,最终都将得到一致的网络结果。利用这两点特性,可以保证所有点云都被包含在网络中,并能找出两两之间距离最接近的点云,将其作为三角形的顶点。利用德洛内三角网络将原本非结构化的离散点云数据结构化,其网络结构良好,数据冗余度小,大大减少障碍物分类的运行时间,其中所有小三角形顶点是上一步中投影得来的所有  $P$  点,如图2所示。



图2 德洛内三角网络

Fig. 2 Delaunay triangle network

### 1.3 障碍物点分类

本文采用基于点云三维坐标的障碍物分类方法,该方法简单高效,仅利用点的  $(X_c, Y_c, Z_c)$  数据,即可完成分类,达到可行驶区域的初步分割。利用德洛内三角网  $G$ ,取所有三角形的每条边所连接的 2 个点  $p$  和  $q$  为一对,使用公式(4)即可将障碍物与非障碍物节点区分开。此方法不用设定任何高度阈值,便可将点分类为障碍物与非障碍物点。函数  $Obst(p)$  在公式(4)中定义,如果函数  $Obst(p)$  返回正值 1,则将节点  $p$  分类为障碍物;如果函数返回 0,则将其分类为非障碍物。这里用到的数学公式可写为:

$$Obst(p) = \begin{cases} 1 & \frac{p[y_c] - q[y_c]}{\|p - q\|} > \cos \theta \\ 0 & \text{其他} \end{cases} \quad (4)$$

其中,  $p[y_c] - q[y_c]$  为这两点之间的高度差;  $\theta$  是一个阈值, Shinzato 等人<sup>[2]</sup>使用该方法得到  $77^\circ$ 、为最优分类角度阈值,故直接取  $77^\circ$ ;  $\|p - q\|$  是向量  $(p - n)$  长度,即  $p, q$  两点的距离。

分类结果如图 3 所示,该方法将铁轨、汽车、树木上的点云都分类为障碍物点,而道路、2 条铁轨之间、以及草坪等比较平坦的地方分类为非障碍物点。



图 3 障碍物点分类结果

Fig. 3 Obstacle points classification results

### 1.4 基于超像素算法的聚类

超像素是由一系列位置相邻,且颜色、亮度纹理等特征相似的像素点组成的小区域,可将一个小区域内具有相似特性的像素聚集起来,形成一个大元素、即一个超像素。使用超像素算法总能将原有物体的边界分割出来,利用该算法,对可行驶区域的检测结果进行补充,进一步细化障碍物的边界。常规超像素算法初始时,通过人为设置需要将图像分割成的超像素个数  $K$ ,并将聚类中心均匀置于图像中。若设置  $K$  太小,则分割结果不一定能够覆盖物体所有的边界;若设置  $K$  过大,虽然能更好地将物体轮廓覆盖,但会加深过分割的程度,一个物体会被

分成许多区域,故而就要在后期合并相似区域来消除过分割的影响。

本文借鉴了 SLIC 算法中用到的聚类方法,在其基础上进行改动,取消了迭代过程。本文的方法无需设定超像素的个数  $K$ ,而是直接将图像中所有分类后的点云作为聚类中心,以点云的数量作为  $K$  值。此外,由于只将图像分为障碍物与非障碍物两类,故只要将同一类区域合并,即可很好地消除过分割的负面影响,同时又能够将障碍物边界精准地分割出来,简单又高效。具体步骤如下:

(1) 初始化聚类中心(种子点)。将分类后的点云对应的图像位置的点设为聚类中心点(种子点)。

(2) 计算距离度量。在每个种子点的  $2S \times 2S$  范围内搜索所有像素点,对于每个搜索到的像素点,分别计算这个像素点和该种子点的颜色距离与空间距离,推导得到的数学公式分别如下:

$$D_c = \sqrt{(l_j - l_i)^2 + (a_j - a_i)^2 + (b_j - b_i)^2} \quad (5)$$

$$D_s = \sqrt{(x_j - x_i)^2 + (y_j - y_i)^2} \quad (6)$$

结合颜色距离与空间距离,得到距离度量  $D$ ,计算方法见式(7):

$$D = \sqrt{\frac{\frac{\partial D_c}{\partial N_c} \frac{\partial D_c}{\partial \theta}}{\frac{\partial N_c}{\partial \theta}} + \frac{\frac{\partial D_s}{\partial N_s} \frac{\partial D_s}{\partial \theta}}{\frac{\partial N_s}{\partial \theta}}} \quad (7)$$

式(5)~式(7)中,  $D_c$  表示颜色距离;  $D_s$  表示空间距离;  $N_s$  是类内最大空间距离。

进一步,研究给出的数学定义表示式为:

$$N_s = S = \text{sqrt}(N/K) \quad (8)$$

其中,  $N$  为图片被点云覆盖区域的面积,如图 4 所示。  $K$  为图像上点云个数,也是超像素的个数。最大颜色距离  $N_c$  随图片不同而不同,也随聚类不同而不同,取一个固定常数  $m$ ,取值范围  $[1, 40]$ ,一般取 10。



图 4 点云覆盖区域

Fig. 4 Point cloud coverage area

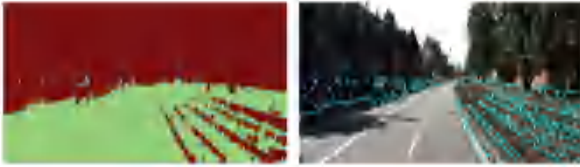
由于每个像素点都会被多个种子点搜索到,所以每个像素点都会有一个与附近不同种子点的距离



$D$ ,取距离度量  $D$  最小值对应的种子点作为该像素点的聚类中心。

(3)分配类标签。在每个种子点周围的  $2S \times 2S$  邻域内为每个像素点分配类标签。

(4)后处理。使用中值滤波以及平滑滤波对聚类结果进行处理,将障碍物边界提取出来,覆盖于原图中,得到图5。图5(a)中绿色区域代表非障碍物区域,红色代表了障碍物区域,图5(b)中蓝线为障碍物边界。



(a) 障碍物区域 (b) 障碍物边界

图5 障碍区域及其边界

Fig. 5 Obstacle area and boundary

### 1.5 可行驶区域提取

以视觉中心位置,即图片底部的中间位置的像素点为起始点,向两边延伸,若遇到障碍物边界,则停止延伸,并将沿途经过的像素点标记为0;此后,再以底部所有标记为0的像素点为起始点,向上延伸遇到障碍物边界,则停止;最终得到可行驶区域边界,如图6所示。



图6 可行驶区域边界

Fig. 6 Boundary of drivable region

## 2 实验结果与分析

### 2.1 Kitti 数据集

Kitti 数据集是目前国际上最大的自动驾驶场景下的计算机视觉算法评测数据集。该数据集用于评测立体图像 (stereo)、光流 (optical flow)、视觉测距 (visual odometry)、道路 (road)、3D 物体检测 (object detection) 和 3D 跟踪 (tracking) 等计算机视觉技术在车载环境下的性能。KITTI 包含市区、乡村和高速公路等场景采集的真实图像数据,每张图像中最多达 15 辆车和 30 个行人,还有各种程度的遮挡与

截断。KITTI 数据集中道路又分为 urban marked (um)城市有标线、urban unmarked (uu)城市无标线和 urban multiple marked (umm)城市多标线场景,共包含 289 张训练图片和 290 张测试图片,图像分辨率为  $1\ 240 \times 375$ 。定量评价时,需要将道路分割结果以二值形式的  $800 \times 400$  像素鸟瞰图 (bev) 上传至官网,进行评估。

### 2.2 评价指标

主要评价指标由 KITTI 数据集官方指定  $F_1 - Measure$ , 即加权调和平均作为最终的评价指标,计算方式见式(8):

$$F_1 - Measure = \frac{2 * P * R}{P + R} \quad (9)$$

其中,  $P$  为准确度 (PRE), 即正确预测为正的像素点, 占有所有预测为正的像素点的比例;  $R$  为召回率 (REC), 即正确预测为正的像素点占有所有正样本的比例。

此外该数据集也提供平均准确率 (AP)、假阳性率 (FPR)、假阴性率 (FNR) 指标的评估。

### 2.3 性能评估及比较

#### 2.3.1 典型场景的可视化结果

从 Kitti 公共数据集 road 中选择 uu 场景作为实验数据,得到可行驶区域的边界结果如图7所示,绿线为可行驶区域边界。将实验结果上传至 Kitti 官网评估,得到的部分评估结果如图8所示。图8中,绿色区域为正确分割像素,蓝色为假阳性像素,红色为假阴性像素。不难看出,本文的方法可以将可行驶区域与明显高出地面的障碍物,如草坪、汽车、路沿石等很好地分割出来。本文的算法不会因为道路中阴影部分与正常道路之间的纹理区别,而造成误分割。对于一些高度不明显的路缘石,以及禁止行驶的标志区域,本文不能很好地将其识别出来。

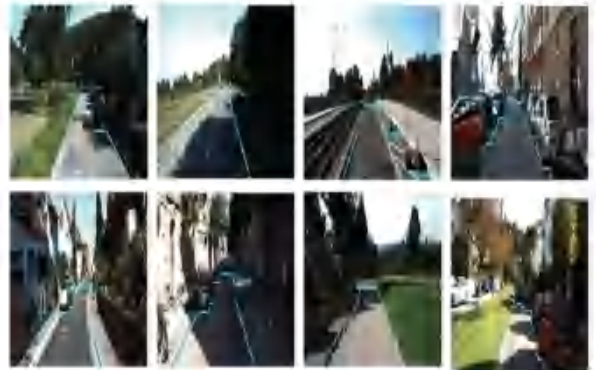


图7 实验结果

Fig. 7 Experimental results



图 8 评估结果

Fig. 8 Evaluation results

### 2.3.2 与其他方法的比较

本文选取了Shinzato等人<sup>[2]</sup>的方法以及基于条

表 1 Kitti 数据集评估与比较

Tab. 1 Kitti dataset evaluation and comparison

方法	$F - me/\%$	AVP/ $\%$	PRE/ $\%$	REC/ $\%$	FPR/ $\%$	FNR/ $\%$	Time/s
本文的方法	84.96	74.43	79.42	91.32	7.71	8.68	1.20
Shinzato	81.99	70.41	74.92	90.54	10.29	9.46	0.36
FusedCRF	84.49	72.35	77.13	93.40	9.02	6.60	2.00
MixedCRF	85.69	75.12	80.17	92.02	7.42	7.98	6.00

## 3 结束语

本文融合了单目相机图像的颜色信息和激光雷达的三维空间关系,结合了两者的优势,同时借鉴了超像素的思想,提出了一种新颖简单的可行驶区域检测方法,并取得了较好的检测效果。采用 Kitti 数据集进行验证,结果表明,该方法可以有效地检测出可行驶区域,具有运行速度和精度的综合优势。

### 参考文献

[1] CUI Yaodong, CHEN Ren, CHU Wenbo, et al. Deep learning for image and point cloud fusion in autonomous driving: A review[J].

件随机场 CRF 的方法进行检测结果对比,对比结果见表 1。本文的  $F - me$  达到了 84.96%, 平均准确率达到 74.43%, 分别超过了 Shinzato 的方法将近 3% 和 4%; 本文与一些 CRF 的方法对比, 在  $F - me$  指标上相差不大, 但是在运行速度上, 本文的速度比 FusedCRF 和 MixedCRF 这 2 个方法提升了许多。

总的来说, 本文提出的方法在检测精度和速度方面, 展现出了更加全面的优势。但本文采用的思想对于高出地面的障碍物检测效果很好, 对于一些地势平坦、但禁止行驶的区域不能很好地识别出来。

IEEE Transactions on Intelligent Transportation Systems, 2022, 23 (2): 722-739.

[2] SHINZATO P Y, WOLF D F, STILLER C. Road terrain detection: Avoiding common obstacle detection assumptions using sensor fusion [C]//2014 IEEE Intelligent Vehicles Symposium Proceedings. Dearborn, MI, USA: IEEE, 2014: 687-692.

[3] TALUKDER A, MANDUCHI R, RANKIN A, et al. Fast and reliable obstacle detection and segmentation for cross-country navigation [C]//Intelligent Vehicle Symposium, 2002. IEEE. Versailles, France: IEEE, 2002: 610-618.

[4] REN X, MALIK J. Learning a classification model for segmentation [C]//Computer Vision, IEEE International Conference on. Nice: IEEE Computer Society, 2003: 10-17.

[5] ACHANTA R, SHAJI A, SMITH K, et al. Slic superpixels[R]. Swiss: EPFL, 2010.

(上接第 64 页)

[14] 彭伟乐, 武浩, 徐立. 基于注意力机制面向短文本多分类的关键词权重优化[J]. 计算机应用, 2021, 41(S2): 19-24.

[15] 付静, 龚永罡, 廉小亲, 等. 基于 BERT-LDA 的新闻短文本分类方法[J]. 信息技术与信息化, 2021(02): 127-129.

[16] 张小川, 戴旭尧, 刘璐, 等. 融合多头自注意力机制的中文短文本分类模型[J]. 计算机应用, 2020, 40(12): 3485-3489.

[17] 邓扬, 张晨曦, 李江峰. 基于弹幕情感分析的视频片段推荐模型[J]. 计算机应用, 2017, 37(04): 1065-1070, 1134.

[18] 朱思淼, 魏世伟, 魏思恒, 等. 基于弹幕情感分析和主题模型的视频推荐算法[J]. 计算机应用, 2021, 41(10): 2813-2819.

[19] 陈建美. 中文情感词汇本体的构建及其应用[D]. 大连: 大连理工大学, 2009.

[20] WANG Shaokang, CHEN Yiyao, MING Hongjun, et al. Improved danmaku emotion analysis and its application based on Bi-LSTM model[J]. IEEE Access, 2020, 8: 114123-114134.