

王伟,唐心瑶,赵春辉,等. 单目视觉下道路场景车辆空间形态估算[J]. 长安大学学报(自然科学版),2023,43(2):100-110.
WANG Wei, TANG Xin-yao, ZHAO Chun-hui, et al. Vehicle spatial morphology estimation in road scene under monocular camera[J]. Journal of Chang'an University(Natural Science Edition), 2023, 43(2): 100-110.

DOI:10.19721/j.cnki.1671-8879.2023.02.010

单目视觉下道路场景车辆空间形态估算

王伟,唐心瑶,赵春辉,李颖,崔华

(长安大学信息工程学院,陕西西安710064)

摘要:针对受射影几何学的限制,单目相机无法直接获得准确的三维点云数据及目标尺度信息,难以获取目标的三维结构问题,提出了一种基于单目交通相机的车辆空间形态估算方法。首先建立道路场景的自动标定模型以获取 3D-2D 的投影映射及尺度信息,并基于“钻石空间”方法,利用统计轨迹直线及车辆边缘精确求取场景中的地平线,根据标定信息及灭点约束共同构建车辆空间形态的几何约束模型;然后在图像中提取车辆的实际投影约束,包括基于获得的车辆序列轮廓约束,及车辆自身体缘约束,并据此构建误差约束函数,估计车辆空间形态的投影误差;最后根据车辆的初始识别结果及先验信息,优化参数约束空间,并利用误差约束函数在约束空间中迭代求最优,得到精确的车辆空间形态信息。利用公开数据集 BrnoCompSpeed 及实际道路采集的视频数据共同验证该算法,并与现有类似算法进行比较。结果表明:该算法对于道路场景的适应性强,所需先验条件少,对于多种类型车辆在三维尺寸的估计精度高达 94% 以上。同时,该算法还可估算车辆实时的空间位置及相对于路面的偏转角度,综合空间形态估算的精度达到 92% 以上,且实时性较好,单帧多车的估算时间小于 0.5 s。与现有算法相比,该方法更适合在道路场景中利用固有的监控相机识别车辆空间形态。

关键词:交通工程;车辆空间形态估算;单目三维;道路场景车辆三维信息;自动标定;3D-2D 投影约束
中图分类号:U495 **文献标志码:**A **文章编号:**1671-8879(2023)02-0100-11

Vehicle spatial morphology estimation in road scene under monocular camera

WANG Wei, TANG Xin-yao, ZHAO Chun-hui, LI Ying, CUI Hua

(School of Information Engineering, Chang'an University, Xi'an 710064, Shaanxi, China)

Abstract: Due to the limitation of projective geometry, it was difficult for monocular camera to obtain accurate 3D point cloud and scale for the three-dimensional structure of the object. To solve this problem, a method of vehicle spatial morphology estimation based on monocular traffic camera was proposed. Firstly, an automatic calibration model of the road scene was established to obtain the 3D-2D projection mapping and scale information. Based on the “diamond space” method, the horizon line of the scene could be accurately obtained by vehicle trajectories and

收稿日期:2022-09-29

基金项目:国家自然科学基金项目(62006026);陕西省科技发展计划项目(2023-JC-YB-600,2023-JC-QN-0703);

陕西省重点研发计划项目(2020GY-027)

作者简介:王伟(1984-),男,江苏徐州人,讲师,工学博士,E-mail:wangwei_211@chd.edu.cn。

edges. Then, the geometric model of the vehicle spatial morphology could be jointly constructed with calibration information and vanishing point constraints. Secondly, the projection constraints of the vehicle were extracted from the image, including sequences of vehicle contour constraints and vehicle edge constraints. Based on these constraints, the error constraint function could be derived to estimate the projection errors of the vehicle spatial morphology. Finally, according to the initial vehicle recognition results and prior information, the parameter constraint space could be iteratively optimized according to the error constraint function, and the accurate vehicle spatial morphology information could be obtained. The experiments were validated on the public dataset BrnoCompSpeed and videos were collected from actual roads. The proposed method was also compared with similar methods. The results show that the proposed method is strongly adaptive to various road scenes with an accuracy of more than 94% for 3D vehicle size estimation, which requires few prior conditions. In the meanwhile, real-time vehicle spatial position and deflection angle relative to the road plane can be estimated with a comprehensive accuracy of more than 92% for vehicle spatial morphology estimation and a process speed of less than 0.5 seconds for a single frame with several vehicles. Moreover, compared with existing methods, the proposed method is more suitable for vehicle spatial morphology estimation by using surveillance cameras in road scenes. 8 tabs, 10 figs, 35 refs.

Key words: traffic engineering; vehicle spatial morphology estimation; monocular 3D; vehicle 3D information in road scene; automatic calibration; 3D-2D projection constraint

0 引 言

近年来,随着智能交通和自动驾驶技术的发展,车辆空间形态信息(包括车辆三维尺寸、空间位置、运动方向及趋势等信息)的准确及时获取具有重要的意义^[1]。通过估计周围车辆目标的空间形态,自动驾驶车辆可以准确预判和规划自己的行为路径^[2-4]。而在智能交通系统中,可以做车辆细粒度识别、车速测量及车流量趋势分析等进一步的应用^[5-6]。然而,相比于当前深度发展的二维(2D)目标检测,三维(3D)检测面临更多的挑战,主要体现为需要在深度信息损失的二维图像中,对目标进行物理尺寸的估计。

在当前主流应用中,3D目标检测可分为单目、双(多)目、激光雷达和深度相机的检测方式。激光雷达和深度相机获取的是点云数据^[7-10],由于点的深度数据可以直接测量,3D检测问题本质上是3D点的划分问题,而由于点云固有的稀疏性,使得目标对象边界较难估计,同时该类型设备成本较高且较为精密。近年来多应用于车载周围3D目标检测^[11-13],并没有在交通监控设备中大规模普及应用。双(多)目3D检测,是从不同的视图中得到的视差来获得深度图^[14-15],该技术理论基础成熟,但是从多视图获取视差深度本身是一种奢侈计算的方

法,同时该方法会减小视场范围,因此在车辆3D检测中并不常用。虽然由于射影几何学的限制,单视图无法直接获取每个点的深度信息,而随着深度学习的发展及同步定位与建图(simultaneous localization and mapping, SLAM)等技术在交通领域的深入应用,基于单目相机估计目标空间形态的方法得到了快速的发展,并提出了不同类型的解决方案。

基于几何约束的目标3D估计^[16-19]。早期的方法主要通过车辆固定/可变计算机辅助设计(computer aided design, CAD)模型投影到图像中,与车辆的几何特征进行贴合程度对比,获取最佳匹配结果,但由于车辆型号繁多, CAD模型数量较大,搜索效率低下,不能适用于当前复杂的场景应用中。而在实际应用中,主要关心的是车辆的3D包络盒,因此基于灭点约束的目标3D包络方法得到发展。Sochor等通过计算目标的3个垂直方向灭点得到3D包络盒,首先基于2D检测结果确定目标基准点和尺度信息,然后根据灭点和2D检测结果构建几何约束,获取最终的目标空间形态^[20-21]。该类方法只利用几何约束,处理速度快,但是由于多灭点估计的稳定性在某些视角较低,因此此类方法具有一定的局限性。

基于深度学习的目标3D估计。近年来,随着卷积神经网络(CNN)的不断成熟,利用深度学习的

方法估计车辆 3D 信息得到快速发展^[22-24]。Xu 等提出一种二阶段目标 3D 检测模型,首先提出一种深度估计子网络对图像进行深度预测,将深度图与图像特征融合,同时,将此融合特征输入区域建议网络(region proposal network, RPN)得到 2D 检测结果,然后将 2D 建议框在 3D 点云图中进行裁切,与原始建议框特征进行融合,得到最终的 3D 检测结果^[25]。此类方法使用了多种数据,同时预测深度导致模型处理效率较低。为了兼顾效率,一些学者将几何约束融入深度网络。Deep3DBox 使用 CNN 回归车辆 2D 包围盒、3D 朝向、尺寸和中心,提出将朝向的角度划分为多个子块进行预测,同时,使用几何约束将预测所得的 3D 包围盒投影到 2D 空间,与 2D 包围盒构成几何约束,用于调整车辆的位姿^[26]。Li 等提出结合深度神经网络回归学习和几何约束的 3D 目标估计的算法,并对车辆的 3D 姿态和位置进行建模,回归计算方位角和物体尺寸^[27]。然而,此类方法依赖于大量 3D 标注的精度。此外,大量研究也表明,深度学习在 2D 检测上具有较高的精度,而在 3D 检测上并没有超越传统的 SLAM 方法,主要是由于相比于点云类与几何类方法,深度学习网络在 2D 图像上很难回归出精确的尺度信息。另外当前车辆 3D 检测大都以车载相机视角为主,鲜有研究基于路侧交通相机获取车辆空间形态,而车路协同及智能交通系统却需要与路侧相机进行交互分析。

综上所述,研究单目相机下的车辆 3D 空间形态的估算有着实际工程意义及理论价值。本文在分析国内外相关车辆 3D 检测方法优缺点的基础上,提出基于路侧单目交通相机,结合实例分割与单灭点几何约束的车辆空间形态估计与优化方法。为了解决单目图像与物理空间的尺度问题,本文首先建立道路场景的空间标定模型,构建基于单灭点的 2D-3D 映射及地平线几何约束,获取图像至物理空间的尺度信息;为了能在图像中从 2D 恢复出 3D 信息,需要从序列车辆的 2D 目标中提取分割结果、初始车型识别尺寸及车辆边缘等几何信息,作为图像空间中 3D 投影的匹配约束;接着,根据标定模型及几何约束构建关于车辆空间形态参数的等价表示,并根据图像空间中的投影信息建立约束误差函数,对车辆空间形态参数在参数空间进行优化,获得精确的车辆空间形态信息;最后通过在公开数据集上和实际道路采集的视频数据试验分析,验证本文算法的有效性。该研究通过路侧单目相机即可获得车

辆空间形态,可为车路协同及智能交通等后端应用提供有效的数据支撑。

1 道路空间场景的几何约束

1.1 相机标定空间模型建立

笔者对于道路场景中的相机自动标定进行了研究^[28]。图 1 为道路场景下的相机空间模型示意。其中世界坐标系 3 个轴 $x_0y_0z_0$,且 x_0y_0 平面与路面重合, z_0 轴与地面垂直,相机坐标系 3 个轴 $x_cy_cz_c$ 。同时,假设相机无自旋角(相机自旋角可通过简单的图像旋转进行矫正,因此不予考虑)。

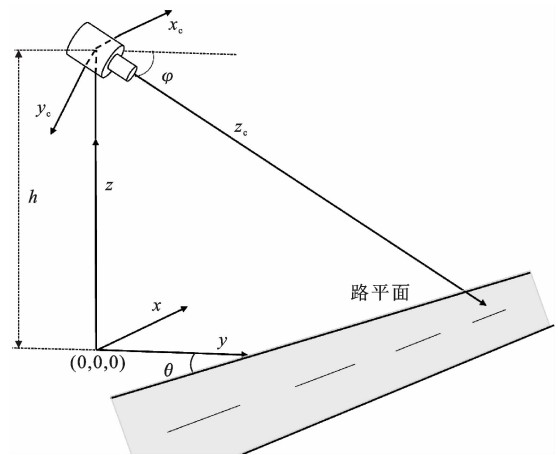


图 1 道路场景中相机空间模型示意

Fig. 1 Schematic of camera spatial model in road scene

在此模型中,设相机焦距为 f ,相机距离地面高度为 h ,相机俯仰角为 φ ,偏转角为 θ ,共同组成相机标定参数。考虑空间坐标点的齐次表示为 $\mathbf{x}=[x_w, y_w, z_w, 1]^T$,在图像坐标系齐次表示为 $\mathbf{p}=[au, av, \alpha]^T$, $\alpha \neq 0$ 为比例系数, u, v 分别为图像坐标的 2 个值。根据相机投影模型^[28],可得世界坐标投影至图像坐标的方程为

$$\begin{bmatrix} au & av & \alpha \end{bmatrix}^T = \begin{bmatrix} x_w & y_w & z_w & 1 \end{bmatrix}^T \cdot \begin{bmatrix} f & 0 & 0 & 0 \\ 0 & -f\sin(\varphi) & -f\cos(\varphi) & fh\cos(\varphi) \\ 0 & \cos(\varphi) & -\sin(\varphi) & h\sin(\varphi) \end{bmatrix} \quad (1)$$

其中,标定参数可由场景灭点、道路标志线和边缘线等几何信息等价表示^[29],由文献^[29]中的推导可得,世界坐标与图像坐标的变换公式为

$$\begin{cases} u = \frac{au}{\alpha} = \frac{fx_w}{y_w\cos(\varphi) - z_w\sin(\varphi) + h\sin(\varphi)} \\ v = \frac{av}{\alpha} = \frac{fh\cos(\varphi) - fy_w\sin(\varphi) - fz_w\cos(\varphi)}{y_w\cos(\varphi) - z_w\sin(\varphi) + h\sin(\varphi)} \end{cases} \quad (2)$$

由式(2)可知,当给定目标高度为 z_0 时,也可将图像点坐标逆投影至世界坐标。因此,通过此相机

空间模型,可得道路场景下车辆的投影与逆投影变换,即获得了 2D 图像空间与 3D 世界空间的尺度信息。

1.2 道路场景下的地平线获取

由射影几何学的理论可知,地平线对于地平面上的刚体可形成较强的几何约束。理论上只需 2 组斜率不同且平行于地面的平行线,投影到图像获得 2 个灭点,即可得到地平线,如图 2 所示,道路场景中存在着大量的平行线组,但实际场景中的平行线存在一定的偏差,灭点也可能存在“震荡”效应^[30],因此本文基于文献[31]思路对灭点进行鲁棒的提取。该方法把图像空间中无穷的直线压缩至有限的“钻石空间”,利用累加极大值的方式求取最佳灭点。具体推导过程此处省略,直接给出钻石空间中点 $[x, y, w]_d$ 与图像空间中点 $[x, y, w]_o$ 变换公式为

$$\begin{aligned} [x, y, w]_d \rightarrow [Dy, \operatorname{sgn}(x)dx + \\ \operatorname{sgn}(y)Dy - dDw, x]_o \end{aligned} \quad (3)$$

式中: w 为比例因子; D 和 d 分别为钻石空间对角线长度的一半; $\operatorname{sgn}(\cdot)$ 代表符号函数。

因此,假设图像空间中存在直线,可推导出图像直线-钻石空间折线组映射关系为

$$\begin{aligned} \alpha = \operatorname{sgn}(ab), \beta = \operatorname{sgn}(bc), \gamma = \operatorname{sgn}(ac) \\ [a, b, c]_o \rightarrow \begin{cases} \left[\frac{\alpha a}{c + \gamma a}, \frac{-ac}{c + \gamma a}, \left[\frac{b}{c + \beta b}, 0 \right] \right] \\ \left[\frac{b}{c + \beta b}, 0 \right], \left[0, \frac{b}{a + \alpha a} \right] \\ \left[0, \frac{b}{a + \alpha a} \right], \left[\frac{-\alpha a}{c + \gamma a}, \frac{ac}{c + \gamma a} \right] \end{cases} \end{aligned} \quad (4)$$

如式(4)所示,图像空间中的无穷直线 $[a, b, c]_o$ 可以转换为钻石空间中的有限折线,并且折线由上述 3 对端点组成。

图 2(c)为一组同斜率平行线在钻石空间中的累加结果,红色点为最佳灭点位置。同理,获取多组平行线的灭点,再利用最小二乘直线拟合的方法,即可得到图像中的地平线,如图 2(d)所示。

2 车辆空间形态的构建与优化

2.1 道路场景中车辆空间集合形态约束构建

图像空间中得到只是 3D 目标在图像平面的投影,故对道路场景中的车辆空间形态进行约束,并获得其 3D-2D 的投影关系,是后续车辆空间形态估计的基础,因此将采用上文构建的标定模型及地平线对车辆进行几何约束构建。本文中,用向量 $\mathbf{V} = [l_v, w_v, h_v, \mathbf{P}_1, \theta_v]^T$ 表示车辆的空间形态,其中 (l_v, w_v, h_v) 表示其 3D 包络长宽高, \mathbf{P}_1 表示其当前

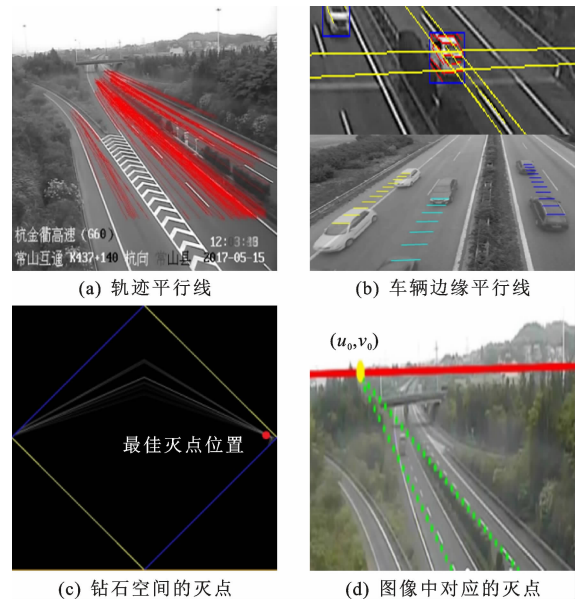


图 2 道路场景中地平线获取实例

Instance of acquisition of horizon lines in road scenes
物理空间位置,由于车辆位于路平面,因此可设 $\mathbf{P}_1 = (x_1, y_1, 0)$,并且俯仰角与自旋角均为 0,只存在偏转角,此处用 θ_v 表示车辆的偏转角,结合已知的路面偏转角 θ ,则车辆相对于路面的偏转角可表示为 $\theta_v - \theta$ 。

车辆在行驶过程中,其行驶轨迹直线与地平线的交点即可形成该车辆当前空间形态的约束灭点,如图 3 所示,即车辆刚体的 3D 包络在该灭点形成的四棱锥上,可用齐次坐标 $\mathbf{P}_0 = (-\tan(\theta_v), 1, 0, 0)$ 表示灭点物理点,并设其投影点为 (u_0, v_0) ,将 \mathbf{P}_0 代入式(1),可得

$$\begin{cases} u_0 = \frac{\alpha u_0}{\alpha} = \frac{-f \tan(\theta_v)}{\cos(\varphi)} \\ v_0 = \frac{\alpha v_0}{\alpha} = -f \tan(\varphi) \end{cases} \quad (5)$$

由此可得到 θ_v 的等价表示

$$\theta_v = \arctan\left(\frac{u_0 \cos(\varphi)}{-f}\right) \quad (6)$$

由灭点的约束,易知 $\mathbf{P}_1 \mathbf{P}_0$ 的方向向量 $\mathbf{d}_1 = (-\sin(\theta_v), \cos(\theta_v), 0)$,而 $\mathbf{P}_1 \mathbf{P}_2$ 的方向与 $\mathbf{P}_1 \mathbf{P}_0$ 垂直,因此其方向向量 $\mathbf{d}_w = (\cos(\theta_v), \sin(\theta_v), 0)$ 。同时,垂直路平面方向的方向向量可表示为 $\mathbf{d}_h = (0, 0, 1)$,根据车辆局部坐标系中 3D 包络的相对关系,基于 \mathbf{P}_1 点可推导出 3D 包络的 8 点坐标 $\mathbf{P}_1 \sim \mathbf{P}_8$ 为

$$\begin{bmatrix} \mathbf{P}_1 & \mathbf{P}_2 \\ \mathbf{P}_3 & \mathbf{P}_4 \\ \mathbf{P}_5 & \mathbf{P}_6 \\ \mathbf{P}_7 & \mathbf{P}_8 \end{bmatrix} = \begin{bmatrix} (x_1, y_1, 0) & \mathbf{P}_1 + w_v \mathbf{d}_w \\ \mathbf{P}_2 + l_v \mathbf{d}_1 & \mathbf{P}_1 + l_v \mathbf{d}_1 \\ \mathbf{P}_1 + h_v \mathbf{d}_h & \mathbf{P}_2 + h_v \mathbf{d}_h \\ \mathbf{P}_3 + h_v \mathbf{d}_h & \mathbf{P}_4 + h_v \mathbf{d}_h \end{bmatrix} \quad (7)$$

假设 P_1 点在图像中的投影点 (u_1, v_1) 已知,可代入式(2)求出 P_1 点坐标,继而代入式(7)求出车辆的 3D 包围,并结合式(6)构建的 θ_v 与 u_0 等价关系,车辆空间形态向量 \mathbf{V} 也可代换为 $\mathbf{V} = [l_v, w_v, h_v, u_1, v_1, u_0]^T$,即可由 2D 信息约束表示。

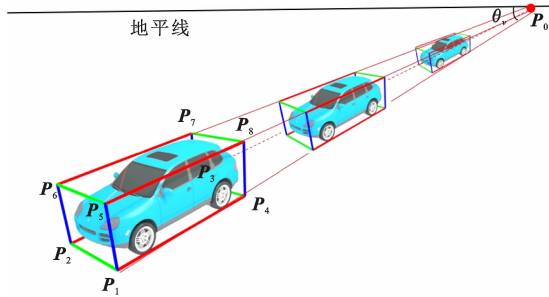


图3 道路场景中车辆空间形态几何约束示意

Fig. 3 Schematic of geometric constraints of vehicle spatial morphology in road scenes

2.2 图像空间车辆 2D 约束信息提取

上文构建了车辆空间形态的 3D-2D 投影与逆投影关系及几何约束,可知在图像中提取合适的 2D 约束是车辆空间形态恢复的基础,本文采用的车辆 2D 信息为车辆轮廓和车辆边缘直线。

实例分割网络 YOLACT^[32] 可实时获得目标的类别、在图像中的位置和像素级的分割结果,相比于二阶段分割网络 Mask RCNN^[33],其具有速度快、精度高的优点,在 Titan X 上 FPS 可达 33.5。图 4(a) 为利用该网络对场景中车辆分割的效果,为了减小视角的影响及增加约束的力度,本文提取了车辆在场景中序列的 n 个分割轮廓,设为 $\{C_i | 1 \leq i \leq n\}$ 。同时,由于车辆刚体中的纵向和横向边界多为平行线,可与灭点产生约束,因此本文首先通过 LSD^[34] 方法获取车身所有的边缘直线段,为了消除与灭点约束无关直线段的干扰,同时构造关于车辆边缘直线的约束误差值,将该直线段与灭点的夹角正切值作为依据进行距离判断(理论上边缘直线过灭点,即夹角为 0)。设边缘直线段 l'_1 斜率为 k_1 ,灭点与该边缘直线段中点形成的直线 l'_2 斜率为 k_2 ,则夹角距离判断公式 $d'(l'_1 | l'_2)$ 可表示为

$$d'(l'_1 | l'_2) = \left| \frac{k_1 - k_2}{1 + k_1 k_2} \right| \quad (8)$$

设阈值 σ ,当时 $d'(l'_1 | l'_2) > \sigma$,去除该干扰边缘,最后设车辆边缘直线约束集合为 $\{D_i | 1 \leq i \leq n\}$,其中 D_i 为序列中第 i 个车辆边缘所有夹角距离之和,图 4(b) 为提取的车辆边缘直线约束。

2.3 参数空间迭代优化车辆空间形态

由第 2.1 节的推导, $\mathbf{V} \Leftrightarrow \mathbf{V}' = [l_v, w_v, h_v, u_1, v_1,$

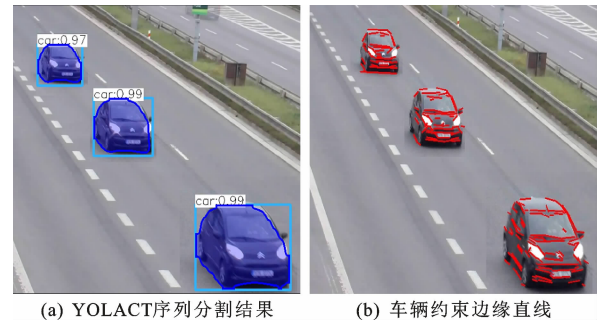


图4 车辆 2D 约束信息示例

Fig. 4 Instance of 2D vehicle constraint information

$u_0]^T$,即车辆空间形态可由车辆尺寸标量 l_v, w_v, h_v 和 2D 信息间接表示,且可关于该车辆目标提取多种 2D 冗余信息,因此可以构建 2D 冗余信息与 \mathbf{V}' 的约束函数,继而在约束空间中进行优化获取更为精确的 \mathbf{V}' 值。

由第 2.1 节可知, $\{D_i | 1 \leq i \leq n\}$ 本身为只与灭点参数 u_0 相关的误差约束。下文构造分割轮廓 C_i 与 \mathbf{V}' 的误差约束函数,易知某组已知 \mathbf{V}' 可得到唯一的 3D 包围投影凸包,理论上,当 \mathbf{V}' 构造的越准确,则与车辆的 2D 外轮廓越接近,因此,本文将投影凸包与车辆外轮廓的差值构造成约束函数,具体算法步骤如下:

算法输入,车辆空间形态 \mathbf{V}' 。

算法输出,约束误差。

步骤 1,假设 \mathbf{V}' 已知,因此可将 $p_1 = (u_1, v_1)$ 代入式(2)即可得到 $P_1 = (x_1, y_1, 0)$,并将 l_v, w_v, h_v, u_0 联合 P_1 代入式(6)、式(7),可得到 3D 包围的 8 点物理坐标。

步骤 2,代入式(2)可反求出 8 个投影点并求凸包,最终获得某组已知 \mathbf{V}' 对应的 3D 包围投影凸包,设为 $\{p_i | 1 \leq i \leq m\}$, m 为凸包的顶点数量; p_i 为凸包顶点。

步骤 3,为了加快求取速度,本文利用插值的方式求取轮廓差,具体为在相邻的 2 个投影凸包顶点间等间隔插入 w' 个顶点,则可获得投影凸包的稠密顶点 $\{p_j^i | 1 \leq j \leq m(w'+1)\}$ 。

步骤 4,求轮廓 C_i 的重心 O_i ,连接 $O_i p_j^i$ 获得与 C_i 的交点 q_j^i ,则约束误差为 $\sum_{j=1}^{m(w'+1)} p_j^i q_j^i$, $w' = 4$ 时, C_i 与 \mathbf{V}' 的误差约束示意如图 5 所示。

由此对于车辆序列中的第 i 个目标,可构建 6 维参数 $\mathbf{V}'_i = [l_v^i, w_v^i, h_v^i, u_1^i, v_1^i, u_0^i]^T$ 的约束函数,即

$$\min \left(\lambda_1 \left(\sum_{j=1}^{m(w'+1)} \frac{p_j^i q_j^i}{m(w'+1)} \right) + \lambda_2 D_i \right)$$

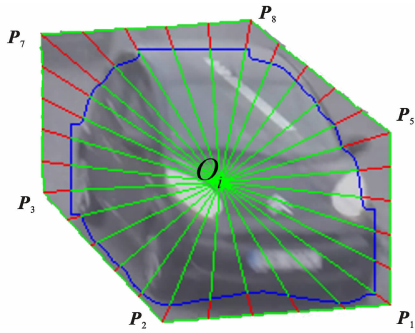


图 5 轮廓约束误差示意

Fig. 5 Schematic of contour constraint error

s. t.

$$\begin{aligned} l_v^s \leq l_v^i \leq l_v^e, \omega_v^s \leq \omega_v^i \leq \omega_v^e, h_v^s \leq h_v^i \leq h_v^e, \\ (u_1^i, v_1^i) \in R_i, u_0^i - \epsilon \leq u_0^i \leq u_0^i + \epsilon \end{aligned} \quad (9)$$

式中: l_v^i, ω_v^i, h_v^i 的约束范围可由识别出的车辆型号给出,不同的车型(如轿车,卡车等)有不同的取值范围; $l_v^s, l_v^e, \omega_v^s, \omega_v^e, h_v^s, h_v^e$ 分别表示不同车型尺寸取值范围的下限和上限; λ_1, λ_2 分别表示轮廓约束和边缘约束的权重系数。锚点 (u_1^i, v_1^i) 的取值范围 R_i 可由车辆目标的 2D 包络给出,灭点范围在 u_0^i 附近设置阈值 ϵ 。在构建出约束函数及参数空间后,本文采用收敛速度较快的最小二乘梯度下降法进行优化求解。

3 试验结果

公开的 3D 目标检测数据集大都基于车载相机,而本文算法的应用场景为基于路侧交通相机,因此采用 BrnoCompSpeed 交通相机视频数据集^[35]进行验证,该数据集可以提供每辆被测车精确的型号、物理位置及实时速度等多种空间形态信息。本文的试验环境和参数如下: Windows 10 系统, 8 GB 内存, Intel 2.50 GHz CPU, 英伟达 GTX1060Ti 显卡, 使用开源深度学习网络 YOLACT 做目标识别及轮廓分割, 检测目标带有类别标签, 为了进一步加快检测速度及减少干扰, 可以通过类别标签过滤人、自行车等无关目标。对于单帧图像中的车辆空间形态估计平均耗时小于 500 ms。对千余辆各类型的车辆目标进行空形态的估算, 平均精度超过 94%。下文通过几组典型车辆的空间形态估计结果说明本文算法的有效性。图 6 为本文算法对典型车辆的空间形态构建过程及结果。

首先根据识别出的车辆类别(不同车辆类别的 l_v, ω_v, h_v 统计数据如表 1 所示)及 2D 包络, 利用标定信息和灭点约束进行初始 3D 包络并构建

投影凸包的约束误差; 然后基于提取的车辆 2D 约束信息, 在约束空间迭代优化, 图 6(c) 为基于 2D 约束迭代优化结果; 最后将优化结果反投影至 3D 空间, 获取最终的车辆空间形态的表达, 如图 6(d) 所示。

实例选取 2 辆小轿车和 1 辆卡车, 虽然每种类型的车辆外轮廓尺寸范围不同, 但常见的深度学习目标检测网络仅能识别出粗粒度的车辆分类, 如轿车、卡车与公交车等。初始 V' 赋值时, 可参考表 1 给出的范围, 设置所识别车辆 l_v, ω_v, h_v 取值范围, 设 2D 包络的右下顶点 (r_x, r_y) , 且其长和宽为 l_r, ω_r , 给予 (u_1, v_1) 的范围则为 $\{(u_1 - \omega_r/4, u_1), (v_1 - l_r/4, v_1)\}$, 最后给予灭点 u_0 的取值范围为 $(u_0 - 10, u_0 + 10)$, 优化算法对于初始 V' 中各参数在取值范围内随机选取一个初始值, 在取值范围组成的约束空间对 V' 进行迭代优化。本文综合考虑了优化精度和优化所需时间, 设终止条件为最大迭代次数为 200 或者迭代误差值小于 10 像素。图 7 为该 3 组实例迭代优化误差下降曲线, 表 2~表 4 为各优化迭代 V' 取值, 并给出了最终对应的空间形态 V 值及与真实值的相对误差 e , 如表 5 所示。表 5 中相对误差计算方法为计算估计的空间形态 V 值与数据集中实测 V_r 值的向量距离, 即

$$e = \frac{|\mathbf{V} - \mathbf{V}_r|}{|\mathbf{V}_r|} \times 100\% \quad (10)$$

可见, 试验车辆的空间形态均在约束空间得到了较好优化, 但最终总是存在一定的约束误差, 原因在于: ①车辆非严格的长方体, 因此与 3D 包络的投影凸包本身存在一定的误差, 但不影响最终空间形态的估计结果; ②约束来自于图像中提取的 2D 信息, 如车辆轮廓及边缘提取的越精确, 原则上迭代的结果越准确; ③初始约束凸空间越小, 最终迭代优化结果越好, 如车辆 1, 2 初始识别为 car, 则 l_v, ω_v, h_v 的约束空间较小, 因此迭代优化很快收敛, 而车辆 3 识别为 truck, 则参数约束空间较大, 收敛速度较慢, 最终精确度也稍差一点。基于表 1, 可得知车辆 1 为小型轿车, 车辆 2 为中大型轿车, 车辆 3 为厢式货车。

本文车辆空间形态估算优化算法的耗时与图像分辨率大小、图像中待估车辆数量及优化迭代次数均相关。在实际处理中, 可将图像分辨率统一设置成常用的 1 200 像素 \times 800 像素(低于此分辨率不改变)。同时, 大多数情况下场景中车辆数少于 10(堵车情况下可能高于此数值)。因此, 以单帧场景中车辆数量为 10 左右, 验证优化迭代次数对于最终精度



图6 车辆空间形态优化估计示例

Fig. 6 Instance of vehicles spatial morphology estimation and optimization

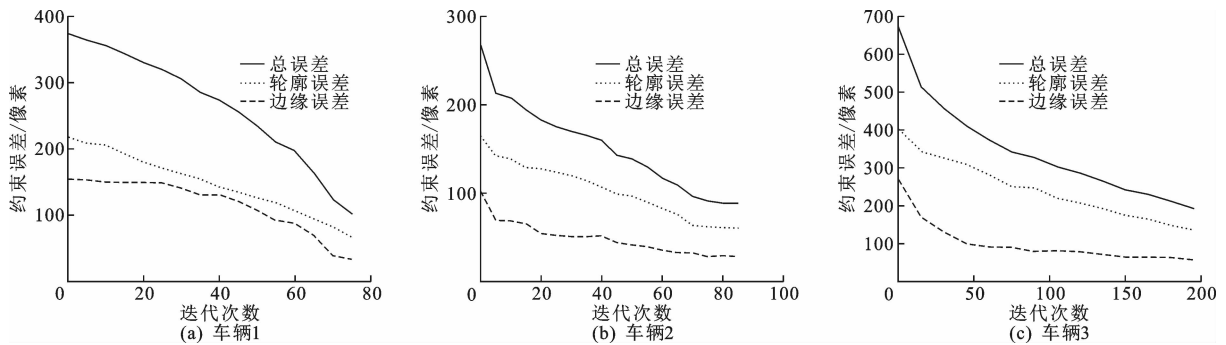


图7 约束空间迭代误差

Fig. 7 Iterative errors in constraint space

表1 各类型车辆外轮廓尺寸范围

Tab. 1 Outline size range of various types of vehicles m

车型尺寸	长	宽	高
小型轿车	3.6~4.4	1.5~1.7	1.3~1.5
中大型轿车	4.3~5.2	1.7~2.0	1.4~1.8
⋮	⋮	⋮	⋮
货车(厢式)	6.2~12.5	1.9~2.4	1.8~2.7
货车(重载)	12.5~18.5	2.4~2.8	2.4~2.7

表2 车辆1参数空间迭代优化结果及误差

Tab. 2 Iterative optimization results and errors of vehicle 1 parameter space

迭代次数	l_v/m	w_v/m	h_v/m	(u_1, v_1)	u_0	误差/像素
1	3.73	1.72	1.30	(982,815)	157	375
2	3.69	1.71	1.32	(989,820)	149	365
3	3.72	1.71	1.36	(985,820)	138	357
⋮	⋮	⋮	⋮	⋮	⋮	⋮
75	3.91	1.68	1.42	(987,820)	136	102

和耗时的影响,发现迭代150次之后精度增加接近停滞,对于小型车收敛时间更快,如图7所示,通过统计,在此条件下,单帧图像中所有车辆空间形态估算平均耗时小于500ms,平均精度超过94%。

对同一车辆在空间序列中进行空间形态求取,

获得参数序列 $\{V'_i | 1 \leq i \leq n\}$,将会加大约束力度,通过更多的图像2D约束,获取更精确的最终车辆空间形态估计结果,如结合时间参数得到速度、加速度运动趋势等信息,图8为对同一车辆序列的空间形态识别结果。

表 3 车辆 2 参数空间迭代优化结果及误差

Tab. 3 Iterative optimization results and errors of vehicle 2 parameter space

迭代次数	l_v/m	w_v/m	h_v/m	(u_1, v_1)	u_0	误差/像素
1	3.72	1.72	1.32	(1 012,846)	129	267
2	3.79	1.73	1.39	(1 009,849)	131	213
3	3.88	1.78	1.46	(1 006,849)	136	207
⋮	⋮	⋮	⋮	⋮	⋮	⋮
87	4.80	1.82	1.72	(1 011,849)	142	89

表 4 车辆 3 参数空间迭代优化结果及误差

Tab. 4 Iterative optimization results and errors of vehicle 3 parameter space

迭代次数	l_v/m	w_v/m	h_v/m	(u_1, v_1)	u_0	误差/像素
1	6.88	2.63	2.11	(1 009,818)	134	671
2	6.94	2.62	2.26	(1 012,819)	136	513
3	7.72	2.58	2.39	(1 015,819)	141	457
⋮	⋮	⋮	⋮	⋮	⋮	⋮
196	9.69	2.34	2.65	(1 039,821)	161	192

表 5 试验车辆最终空间形态值及误差

Tab. 5 Final spatial shape values and errors of experimental vehicles

车辆	l_v/m	w_v/m	h_v/m	P_1/m	$\theta_v/(\circ)$	相对误差/%
车辆 1	3.91	1.68	1.42	(3.53,37.76,0)	15.1	5.3
车辆 2	4.80	1.82	1.72	(6.45,36.44,0)	16.2	4.1
车辆 3	9.69	2.34	2.65	(10.32,37.72,0)	14.7	9.7

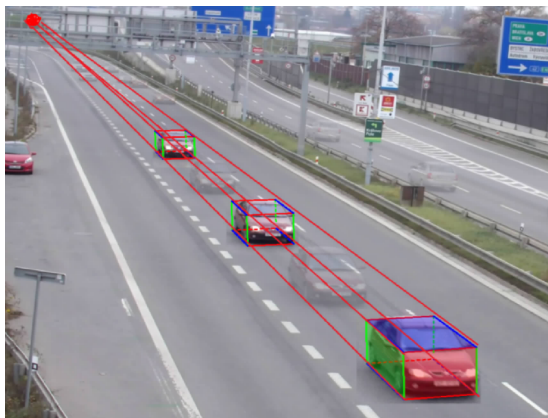


图 8 车辆序列空间形态优化估计结果

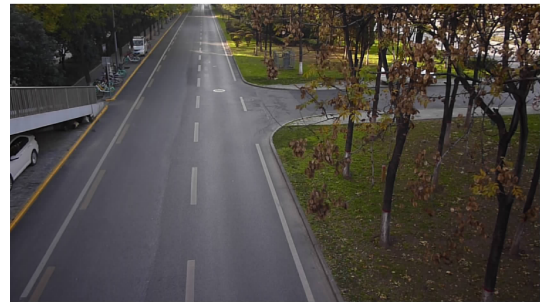
Fig. 8 Estimation and optimization results of vehicle spatial morphology in sequences

为了进一步证明本文算法的有效性,也针对实际场景进行了验证,西安市南二环某段交通摄像头采集的视频数据,如图 9 所示。利用第 1.1 节提出的相机自标定方法,获得场景的标定参数及投影矩阵,如表 6 所示。

对于实际场景中多类型数百辆车辆进行了测试,统计出平均精度也超过 94%,在 2 个实际场景



(a) 实际场景1



(b) 实际场景2

图 9 实际场景

Fig. 9 Actual scenes

表 6 实际场景自动标定结果及投影矩阵

Tab. 6 Automatic calibration results and projection matrix of actual scenes

场景	f	$\varphi/(\circ)$	h/m	投影矩阵 P
1	5 749.81	4.20	7.88	$\begin{bmatrix} 5\ 749.81 & 0 & 0 & 0 \\ 0 & -420.86 & -5.73 \times 10^3 & 4.52 \times 10^7 \\ 0 & 0.997 & -0.073 & 576.79 \end{bmatrix}$
2	1 853.22	12.24	7.95	$\begin{bmatrix} 1\ 853.22 & 0 & 0 & 0 \\ 0 & -3.92 \times 10^2 & -1.81 \times 10^3 & 1.44 \times 10^7 \\ 0 & 0.977 & -0.212 & 1.69 \times 10^3 \end{bmatrix}$

中,实际采集的 3 种典型类型车辆(轿车,SUV 及公交车)空间形态估算结果实例,如图 10 所示。对于轿车和 SUV 等小型车,估算精度较高,符合前述分析和预期结果,而对于公交车类大型车,估算精度比货车要高,原因是由于公交车的尺寸范围变化没有货车大,即初始约束空间较小,因此也获得了较满意的估算结果,见表 7。

对比现有典型算法,如表 8 所示,文献[16]中方法由于只利用灭点进行约束,因此实时性最高,但其可获取的空间形态参数较少,同时存在需要先验标定及精度不高等缺点;文献[20]中方法识别精度高及可识别的空间形态参数较多,但其预处理步骤为计算奢侈的点云估计,因此实时性最差;文献[22]中方法由于利用深度网络回归目标空间参数,实时性一般,同时依赖大量的 3D 标注训练及先验标定。本文方法在获得较高精度的情形下,可做到较高

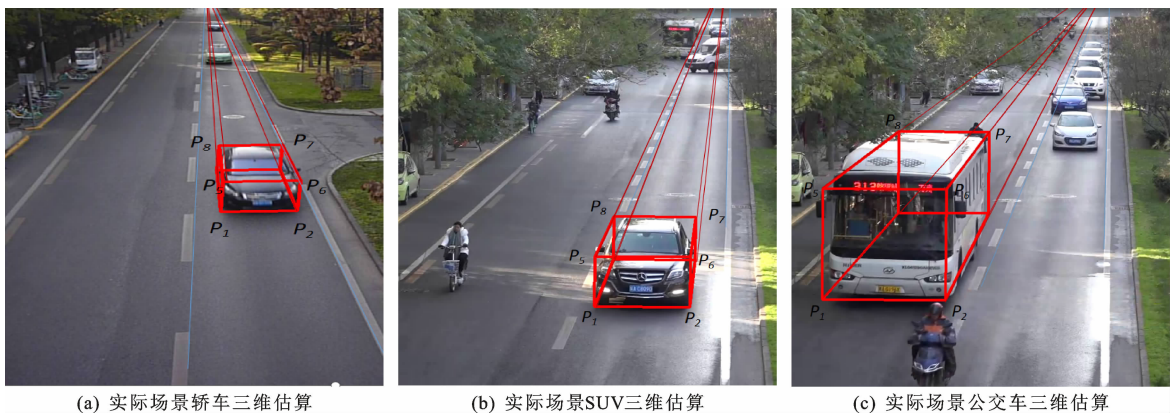


图 10 实际场景车辆 3D 估算实例

Fig. 10 Instance of 3D vehicle estimation in actual scenes

表 7 实际场景车辆实际尺寸估算结果及相对误差

Tab. 7 Estimated results and relative errors of actual vehicle size in actual scenes

车辆	3D 尺寸估计/cm	相对误差/%
轿车	(425,163,140)	2.6
SUV	(524,196,185)	3.8
公交车	(1 235,250,306)	5.1

表 8 本文算法与其他算法的对比

Tab. 8 Comparisons of algorithm in this paper with other algorithms

方法	视角	先验标定	可识别的空间形态参数			单帧平均耗时/s	3D 尺寸精度/%
			长宽高	偏转方向	空间定位		
几何灭点约束 ^[16]	路侧	✓	✓	×	×	0.2	83.7
点云估计法 ^[20]	车载	×	✓	✓	✓	4.2	94.6
深度网络与几何约束结合法 ^[22]	车载	✓	✓	✓	×	2.3	93.7
本文方法	路侧	×	✓	✓	✓	0.5	94.2

的实时性及识别出多维度的空间形态信息。

4 结 语

(1)通过试验验证,本文提出的基于单目交通相机车辆空间形态估计算法,在公开数据集及实际道路采集的视频数据下,对于多类型的车辆空间形态估计均取得了较好的结果。本算法的主要贡献有:①不需要先验手动标定,可自动获取相机的标定信息及图像与物理空间的尺度信息;②构建了基于单灭点的道路场景下车辆空间形态几何约束模型,该模型简单稳定,可有效构建出车辆 3D 包络、物理空间位置及偏转角与图像投影信息的映射关系;③针对 2D-3D 恢复的信息损失,本文通过车辆轮廓及车

辆边缘等多冗余信息的几何约束,在约束空间进行迭代优化,不仅可有效恢复出车辆 3D 信息,并可提高车辆空间形态的估计精度。

(2)车辆在图像中不仅包括轮廓及边缘方向约束,也存在对称结构及部件相对位置等约束信息,融合多类型的投影约束可增强约束力度,并提高最终识别精度,在未来工作中可以考虑加入此类约束。同时,车辆的细粒度 2D 目标识别(譬如识别出轻型卡车与重型卡车,而不是仅能识别出卡车分类),能够缩小初始参数空间,提高迭代优化效率与效果,随着深度学习的发展,后续考虑采用细粒度目标识别网络进行初始车型识别。最后,在视野中存在车辆相互遮挡的情况下,本文采用了车辆的序列投影作为约束,一定程度弥补了遮挡造成的信息损失,后续将会利用跟踪等更多手段,获取更多位置约束,提高形态识别精度。

参考文献:

References:

[1] NOVAK L. Vehicle detection and pose estimation for autonomous driving[D]. Czech;Czech Technical University in Prague,2017.

[2] GUPTA I,RANGESH A,TRIVEDI M. 3D bounding boxes for road vehicles: A one-stage, localization prioritized approach using single monocular images [C]//ECCV. Proceedings of the European Conference on Computer Vision (ECCV). Cham; Springer,2018; 1-16.

[3] CHEN X Z,KUNDU K,ZHANG Z Y,et al. Monocular 3D object detection for autonomous driving[C]// IEEE. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). New York;IEEE,2016;2147-2156..

- [4] PENG T, LIU X, FANG R, et al. Lane-change path planning and control method for self-driving articulated trucks[J]. *Journal of Intelligent and Connected Vehicles*, 2020, 3(2):49-66.
- [5] HE H S, SHAO Z Z, TAN J D. Recognition of car makes and models from a single traffic-camera image [J]. *IEEE Transactions on Intelligent Transportation Systems*, 2015, 16(6):3182-3192.
- [6] LAN J H, LI J, HU G D, et al. Vehicle speed measurement based on gray constraint optical flow algorithm[J]. *Optik*, 2014, 125(1):289-295.
- [7] QI C R, LIU W, WU C X, et al. Frustum PointNets for 3D object detection from RGB-D data[C]//IEEE. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. New York: IEEE, 2018:918-927.
- [8] AHMED S M, CHEW C M. Density-based clustering for 3D object detection in point clouds[C]//IEEE. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. New York: IEEE, 2020:10605-10614.
- [9] ZHOU Y, TUZEL O. VoxelNet: End-to-end learning for point cloud based 3D object detection[C]//IEEE. *Proceedings of the Conference on Computer Vision and Pattern Recognition*. New York: IEEE, 2018:4490-4499.
- [10] KU J, MOZIFIAN M, LEE J, et al. Joint 3D proposal generation and object detection from view aggregation [C]//IEEE. *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. New York: IEEE, 2018:1-8.
- [11] SHI S S, GUO C X, JIANG L, et al. PV-RCNN: Point-voxel feature set abstraction for 3D object detection [C]//IEEE. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. New York: IEEE, 2020:10526-10535.
- [12] NAJIBI M, LAI G D, KUNDU A, et al. DOPS: Learning to detect 3D objects and predict their 3D shapes [C]//IEEE. *Proceedings of the Conference on Computer Vision and Pattern Recognition (CVPR)*. New York: IEEE, 2020:11910-11919.
- [13] SHI S S, WANG X G, LI H S. PointRCNN: 3D object proposal generation and detection from point cloud [C]//IEEE. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. New York: IEEE, 2019:770-779.
- [14] CHEN X Z, KUNDU K, ZHU Y K, et al. 3D object proposals using stereo imagery for accurate object class detection [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2018, 40 (5):1259-1272.
- [15] LI P L, CHEN X Z, SHEN S J. Stereo R-CNN based 3D object detection for autonomous driving [C]//IEEE. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. New York: IEEE, 2019:7636-7644.
- [16] CAETANO T S, CAELLI T, SCHUURMANS D, et al. Graphical models and point pattern matching [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2006, 28(10):1646-1663.
- [17] ROMANONI A, SORRENTI D G, MATTEUCCI M. Backward-simulation particle smoother with a hybrid state for 3D vehicle trajectory, class and dimension simultaneous estimation[J]. *Machine Vision and Applications*, 2015, 26(2):369-385.
- [18] ZHANG Z X, TAN T N, HUANG K Q, et al. Three-dimensional deformable-model-based localization and recognition of road vehicles[J]. *IEEE Transactions on Image Processing*, 2012, 21(1):1-13.
- [19] CORRAL-SOTO E R, ELDER J H. Slot cars: 3D modelling for improved visual traffic analytics[C]//IEEE. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. New York: IEEE, 2017:889-897.
- [20] SOCHOR J, SPANHEL J, HEROUT A. BoxCars: Improving fine-grained recognition of vehicles using 3D bounding boxes in traffic surveillance [J]. *IEEE Transactions on Intelligent Transportation Systems*, 2019, 20(1):97-108.
- [21] YANG S C, SCHERER S. CubeSLAM: Monocular 3D object SLAM [J]. *IEEE Transactions on Robotics*, 2019, 35(4):925-938.
- [22] HE T, SOATTO S. Mono3d++: Monocular 3d vehicle detection with two-scale 3d hypotheses and task priors[C]//AAAI. *Proceedings of the AAAI Conference on Artificial Intelligence*. Menlo Park: AAAI Press, 2019:8409-8416.
- [23] QIN Z Y, WANG J L, LU Y. MonoGRNet: A geometric reasoning network for monocular 3D object localization[C]//AAAI. *Proceedings of the AAAI Conference on Artificial Intelligence*. Menlo Park: AAAI Press, 2019:8851-8858.
- [24] DING M Y, HUO Y Q, YI H W, et al. Learning depth-guided convolutions for monocular 3D object detection[C]//IEEE. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*

- (CVPR). New York:IEEE,2020:11669-11678.
- [25] XU B, CHEN Z Z. Multi-level fusion based 3D object detection from monocular images[C]//IEEE. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). New York:IEEE, 2018:2345-2353.
- [26] MOUSAVIAN A, ANGUELOV D, FLYNN J, et al. 3D bounding box estimation using deep learning and geometry[C]//IEEE. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). New York:IEEE,2017:5632-5640.
- [27] LI B Y, OUYANG W L, SHENG L, et al. GS3D: An efficient 3D object detection framework for autonomous driving [C]//IEEE. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). New York:IEEE,2019:1019-1028.
- [28] 王伟,张朝阳,唐心瑶,等.道路场景下相机自动标定及优化算法[J].计算机辅助设计与图形学学报,2019,31(11):1955-1962.
WANG Wei, ZHANG Chao-yang, TANG Xin-yao, et al. Automatic self-calibration and optimization algorithm of traffic camera in road scene[J]. Journal of Computer-Aided Design & Computer Graphics,2019, 31(11):1955-1962.
- [29] KANHERE N K, BIRCHFIELD S T. A taxonomy and analysis of camera calibration methods for traffic monitoring applications[J]. IEEE Transactions on Intelligent Transportation Systems, 2010, 11 (2): 441-452.
- [30] 张润初,杜倩云,俞祝良,等.一种利用参考图象与路面信息的道路监控摄像机标定方法[J].公路交通科技,2014,31(11):137-141.
ZHANG Run-chu, DU Qian-yun, YU Zhu-liang, et al. A calibration method for road monitoring cameras exploiting reference images and roadway information[J]. Journal of Highway and Transportation Research and Development,2014,31(11):137-141.
- [31] DUBSKA M, HEROUT A. Real projective plane mapping for detection of orthogonal vanishing points [C]//BMVC. Proceedings of the British Machine Vision Conference. Cambridge: British Machine Vision Association,2013:1-10.
- [32] BOLYA D, ZHOU C, XIAO F Y, et al. YOLACT: Real-time instance segmentation[C]//IEEE. Proceedings of the IEEE International Conference on Computer Vision (ICCV). New York: IEEE, 2019: 9156-9165.
- [33] HE K M, GKIOXARI G, DOLLAR P, et al. Mask R-CNN[C]//IEEE. Proceedings of the IEEE International Conference on Computer Vision. New York: IEEE,2017:2980-2988.
- [34] GROMPONE VON G R, JAKUBOWICZ J, MOREL J M, et al. LSD: A fast line segment detector with a false detection control[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence,2010,32(4): 722-732.
- [35] SOCHOR J, JURANEK R, SPANHEL J, et al. Comprehensive data set for automatic single camera visual speed measurement[J]. IEEE Transactions on Intelligent Transportation Systems,2019,20(5):1633-1643.