单目交通场景下基于自标定的车辆三维信息识别算法

唐心瑶, 宋焕生*, 王伟, 张朝阳, 崔华

(长安大学信息工程学院 西安 710064)
(andy19966212@126.com)

摘 要: 获取车辆的三维信息作为车型精确分类的依据, 已成为当前越来越重要的研究方向, 但交通场景中的监控相机大多为单目相机, 由于透视因素无法直接获取车辆位姿、车辆轮廓尺寸等三维信息. 针对上述问题, 提出单目交通场景下基于自标定的车辆三维信息识别算法, 首先根据典型的交通场景, 建立单目相机的摄像机模型以及较稳定的单消失点标定模型, 完成摄像机标定; 接着使用深度学习卷积神经网络中的 YOLO 模型对交通场景中的车辆进行二维目标检测. 在此基础上, 提出对角线和消失点约束的非线性优化求解算法, 结合标定信息完成车辆的三维信息识别及最佳三维目标检测. 在公开数据集 BrnoCompSpeed 和实际高速公路场景进行了实验, 实验结果表明, 该算法在多种交通场景下均能有效识别车辆三维信息, 平均识别准确率超过 90%.

关键词: 单目相机; 摄像机标定; 深度学习; 三维信息识别

中图法分类号: TP391.41 **DOI:** 10.3724/SP.J.1089.2020.18041

3D Vehicle Information Recognition Algorithm of Monocular Camera Based on Self-Calibration in Traffic Scene

Tang Xinyao, Song Huansheng*, Wang Wei, Zhang Chaoyang, and Cui Hua

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

Abstract: Obtaining 3D information of vehicles as the basis for accurate classification of vehicles has become an increasingly important research direction. However, most of the current traffic monitoring cameras are monocular cameras, which cannot directly obtain 3D information of vehicles like pose and size due to perspective factors. According to the above problem, this paper proposes a 3D vehicle information recognition algorithm of monocular camera based on self-calibration in traffic scene. Firstly, this paper builds up a monocular camera model and a stable single vanishing point calibration model according to the typical traffic scene, and completes camera calibration. Then it uses the YOLO deep learning convolution neural network for 2D vehicle detection. Based on this, it puts forward a diagonal and vanishing point constrained non-linear optimization algorithm, combining with the calibration information to complete 3D vehicle information recognition and the best 3D vehicle detection. Finally, the experiment was carried out on the public dataset called BrnoCompSpeed and in highway traffic scenes, and the results show that the algorithm can effectively complete 3D vehicle information recognition in various traffic scenarios with an average recognition accuracy of more than 90%.

收稿日期: 2019-08-04; 修回日期: 2020-03-06. 基金项目: 教育部联合基金(6141A02022610); 中央高校基金(300102249103); 陕西省重点研发计划重点项目(2018ZDXM-GY-047); 陕西省社会发展领域项目(2019SF-258); 中央高校基本科研项目(30010224910). 唐心瑶(1996—), 女, 博士研究生, 主要研究方向为计算机视觉; 宋焕生(1964—), 男, 博士, 教授, 论文通讯作者, 主要研究方向为计算机视觉; 王伟(1984—), 男, 博士, 讲师, 主要研究方向为计算机视觉、SLAM; 张朝阳(1984—), 男, 博士, 讲师, 主要研究方向为计算机视觉; 崔华(1976—), 女, 博士, 教授, 主要研究方向为计算机视觉、人工智能.

Key words: monocular camera; camera calibration; deep learning; 3D information recognition

车辆三维信息识别是智能交通系统(intelligent transportation system, ITS)中一个重要的子课题, 识别结果可应用于自动驾驶^[1-2]和车型识别^[3-4], 对于车辆收费管理和公安部门对涉案车辆侦缉均具有重要的实际意义. 目前较为常见的三维信息识别方法主要基于立体视觉的三维重建技术^[5-6], 利用 RGB-D 深度相机直接获得图像的深度信息, 但是这种方法处理速度慢且对室外环境光照敏感. 单目相机由于其具有体积小、成本低和易于在室外部署的优点, 在交通监控系统中长期占据主导地位. 但是, 单目相机所成二维图像本身具有信息损失的限制, 无法直接获取图像的深度信息, 因此必须完成车辆三维信息识别, 而目标检测^[7-8]和摄像机标定^[9-12]是 2 个必不可少的步骤. 通过目标检测可以确定车辆在二维图像上的位置, 通过摄像机标定可以建立起二维图像平面与三维世界之间的映射关系, 在此基础上求取三维世界中的物理交通参数, 进而恢复车辆三维信息.

在当前主流应用中, 单目交通场景下车辆三维信息识别方法主要有 3 类: (1) 基于车辆模型的方法^[13-16]. 该类方法通过先验的车辆模型库完成车辆三维信息识别. Prokaj 等^[15]提出一种三维 CAD 模型, 结合 SIFT 特征点匹配对车辆进行建模, 但其建模过程计算量过大, 并不适用于对实时性要求较高的应用场景. 为了提高模型的求解效率, Zhang 等^[16]提出一种仅使用 12 个模型参数的可变三维 CAD 模型, 通过三维模型与二维图像之间的投影误差优化车辆三维信息. 该类方法虽然识别结果较为准确, 但是计算过程复杂, 需要制作大量的车辆模型, 而且不能适应所有车辆, 因此在实际应用中使用较少. (2) 基于深度学习的方法^[17-20]. 该类方法需要大量数据, 并且通常都需要对先验信息进行估算. Mousavian 等^[19]提出一种 MultiBin 的方法估计物体的朝向, 并且结合深度学习网络来推断物体的三维中心点和尺寸. Chabot 等^[20]提出一种 Deep MANTA 的深度学习网络, 用于推断车辆的二维和三维描述信息, 但是描述信息中也需要先验的三维方向角, 且求解过程较为复杂. (3) 基于自标定的方法^[21-23]. 该类方法不需要任何先验的车辆模型, 具有较大的灵活性, 但是其对自标定结果具有一定的依赖性, 通常需要先对交通场景路侧监控

相机进行标定, 求取二维图像平面与三维世界之间的映射关系作为车辆三维信息识别的先决条件. Yang 等^[22]提出一种 CubeSLAM 的三维目标检测方法, 它不需要预先知道物体的朝向, 只需要提取物体边缘并且已知相机高度, 但其对于场景的视角有一定限制. Dubska 等^[23]提出一种完全基于自标定的车辆三维信息识别方法, 从车辆的运动轨迹出发, 分别提取 3 个相互正交的消失点(vanishing point, VP)完成相机标定, 获取二维与三维映射尺度信息, 并在此基础上构建车辆的三维包围框, 最后利用尺度信息恢复车辆三维信息. 该方法在标定时不需要额外的人工输入, 但是 VP 的获取和标定结果的计算过程较为复杂.

综上所述, 本文在分析了国内外各种车辆三维信息识别算法的优劣后, 在关于相机标定方法^[24-25]的基础上, 提出一种单目交通场景下基于自标定及对角线和 VP 约束的车辆三维信息识别算法. 首先根据常见交通场景的几何结构, 建立简明的相机模型及稳定的单 VP 标定模型, 完成摄像机标定; 然后通过深度学习 YOLO 网络对交通场景中的车辆进行二维目标检测, 根据标定信息、对角线和 VP 约束获取车辆三维信息; 最后对车辆完成较为精确的三维目标检测. 通过在公开数据集 BrnoCompSpeed 和实际高速公路场景的实验分析, 验证了本文算法的有效性和鲁棒性.

1 摄像机标定

1.1 相机模型及坐标系的建立

本文参考 Kanhere 等^[24]提出的相机标定模型及标定参数求解方法(VWL), 并采用文献[25]方法对标定结果进行优化. 在文献[24]模型中, 将摄像机模型简化为针孔模型, 并且主点与图像中心重合, 成像平面与光轴垂直, 内参数只有焦距是未确定的, 观测路面为平面. 图 1 所示为文献[24]交通场景下相机空间模型示意图. 为方便后续分析, 设相机焦距为 f , 相机距离地面高度为 h , 相机俯仰角为 ϕ , 相机偏转角(相机光轴在道路平面投影与道路延伸方向的夹角)为 θ , 由于相机自旋角可以通过简单的图像旋转进行表示, 并且对于标定结果无影响, 因此不予考虑.

图1 交通场景下相机空间模型^[24]示意图

文献[24]模型定义3个坐标系: 世界坐标系、相机坐标系和图像坐标系. 设所建坐标系均为右手系, 世界坐标系包含 x, y, z 轴, 坐标系原点位于相机在路面的投影点, z 轴垂直于地面方向向上. 图 1a 所示为相机模型的左视图, 从中可以看出, x 轴方向垂直纸面向内, 用 $\otimes$ 表示, y 轴垂直于 xoz 平面; 图 1b 所示为相机模型的俯视图, 从中可以看出, z 轴方向垂直纸面向外, 用 $\odot$ 表示. 相机坐标系包含 x_c, y_c, z_c 轴, 坐标系原点位于相机所处位置, x_c 轴与世界坐标系下的 x 轴平行, z_c 轴正向沿着相机的光轴指向地面, y_c 轴垂直于 $x_c oz_c$ 平面指向地面. 图像坐标系包含 u, v 轴, 坐标系原点位于图 1b 中的 r 点, r 点为图像主点, 为 z_c 轴延伸至与地面的交点在图像坐标系中的投影, 水平向右为 u 轴, 竖直向下为 v 轴. 根据图 1a 中的角度关系, r 点在世界坐标系下的坐标为 $(0, h \cot \phi, 0)$.

由文献[24]推导可得到展开的投影模型为

$$\begin{bmatrix} \alpha u \\ \alpha v \\ \alpha \end{bmatrix} = \begin{bmatrix} f & 0 & 0 & 0 \\ 0 & -f \sin \phi & -f \cos \phi & fh \cos \phi \\ 0 & \cos \phi & -\sin \phi & h \sin \phi \end{bmatrix} \begin{bmatrix} x \\ y \\ z \\ 1 \end{bmatrix} \quad (1)$$

由式(1)可以求解出世界坐标系与图像坐标系之间的相互转换关系. 设图像上任意一点对应的世界坐标为 (x, y, z) , 求解出 2 个坐标系之间的相互关系为

$$u = \frac{\alpha u}{\alpha} = \frac{fx}{y \cos \phi - z \sin \phi + h \sin \phi} \quad (2)$$

$$v = \frac{\alpha v}{\alpha} = \frac{fh \cos \phi - fy \sin \phi - fz \cos \phi}{y \cos \phi - z \sin \phi + h \sin \phi} \quad (3)$$

$$y = \frac{fh \cos \phi - fz \cos \phi - hv \sin \phi + vz \sin \phi}{f \sin \phi + v \cos \phi} \quad (4)$$

$$x = \frac{u(y \cos \phi - z \sin \phi + h \sin \phi)}{f} \quad (5)$$

1.2 基于 VWL 的标定参数计算

由式(1)可知, 标定所需求解参数为 f, ϕ, θ, h .

在相机标定中, VP 包含了丰富的信息. 常用的 VP 包括道路延伸方向和道路垂直方向 VP 等, 文献[24]模型中只考虑道路延伸方向 VP. 本文使用文献[26]方法对交通场景下的 VP 进行提取.

如图 1b 所示, y 轴与路面延伸方向的夹角为 θ , 可知世界坐标系中路面延伸方向的无穷远点坐标为 $\mathbf{x}_0 = [-\tan \theta \ 1 \ 0 \ 0]^T$. 由 VP 原理可知, (u_0, v_0) 为 $\mathbf{x}_0$ 在图像空间中的投影, 解出标定参数 ϕ, θ 分别为

$$\phi = \arctan(-v_0/f) \quad (6)$$

$$\theta = \arctan(-u_0 \cos \phi/f) \quad (7)$$

除了 VP, 道路虚线标志也是常见的道路几何标志. 图 2 所示为世界坐标系和图像坐标系下的几何标志关系示意图, 设道路虚线标志物理长度为 l (道路虚线标志具有国家标准, 可以从相关文件中直接获取), 虚线 2 个端点的世界坐标系及图像坐标系纵坐标分别为 y_b 和 y_f, v_b 和 v_f , 道路物理宽度为 w , 对应于图像坐标系中的一段截距, 其像素长度为 δ .

图2 交通场景下2种坐标系几何标志关系示意图

由文献[24]推导可知, h 可由 w 或 l 间接表示为

$$h = \frac{fw \sin \phi}{\delta \cos \theta} \quad (8)$$

$$h = \frac{f \tau l \cos \phi}{f^2 + v_0^2} \quad (9)$$

其中, $\tau = (v_f - v_0)(v_b - v_0)/(v_f - v_b)$.

从式(6)(7)中解出 $\sin \phi, \cos \phi, \cos \theta$ 的值代入式(8)(9), 将式(8)(9)平方后联立消去 h , 可得关于 f 的四次方程为

$$f^4 + [2(u_0^2 + v_0^2) - k_v^2]f^2 + (u_0^2 + v_0^2)^2 - k_v^2 v_0^2 = 0 \quad (10)$$

其中, 为计算方便引入中间变量 $k_v = \delta \tau l / (w v_0)$.

式(10)是关于 f 的四次方程, 据此可求解出 f . 当 f 唯一确定后, 根据式(6)(7)可以求解 ϕ 和 θ , 根据式(8)(9)可以求解 h , 所有未知标定参数都已求解, 因此完成标定过程; 再根据式(1)即可

建立起二维图像坐标与三维世界坐标之间的投影模型.

2 车辆三维信息识别

2.1 二维/三维目标检测

本文结合实际应用需求,利用深度学习目标检测中实时性较好的 YOLO 算法^[7]对交通场景下车辆进行二维目标检测,获取车辆二维包围框中心点的图像坐标、包围框的像素宽高以及车辆类型(car, truck, bus 等).

在车辆二维目标检测的基础上,结合标定参数,可以进一步完成车辆三维目标检测.图 3 所示为车辆在左视角和右视角下的二维/三维包围框模型.二维包围框模型顶点序号为 0~3,记 4 个顶点的图像坐标为 $P_i^{2D} = (u_i^{2D}, v_i^{2D})$, $i = 0, 1, 2, 3$; 三维包围框模型顶点序号为 0~7,记 8 个顶点的图像坐标为 $P_i^{3D} = (x_i^{3D}, y_i^{3D}, z_i^{3D})$, $i = 0, 1, \dots, 7$, 对应的图像坐标为 $P_j^{3Di} = (u_j^{3Di}, v_j^{3Di})$, $j = 0, 1, \dots, 7$. 表 1 列出了不同视角下车辆三维包围框模型 8 个顶点的图像坐标.

图 3 不同视角下车辆二维/三维包围框模型示意图

表 1 不同视角下车辆三维模型 8 个顶点世界坐标

顶点序号	左视角	右视角
0	$(x_1^{3D} - w_v, y_1^{3D}, z_1^{3D})$	$(x_1^{3D} + w_v, y_1^{3D}, z_1^{3D})$
1	$(x_1^{3D}, y_1^{3D}, z_1^{3D})$	$(x_1^{3D}, y_1^{3D}, z_1^{3D})$
2	$(x_1^{3D}, y_1^{3D} + l_v, z_1^{3D})$	$(x_1^{3D}, y_1^{3D} + l_v, z_1^{3D})$
3	$(x_1^{3D} - w_v, y_1^{3D} + l_v, z_1^{3D})$	$(x_1^{3D} + w_v, y_1^{3D} + l_v, z_1^{3D})$
4	$(x_1^{3D} - w_v, y_1^{3D}, z_1^{3D} + h_v)$	$(x_1^{3D} + w_v, y_1^{3D}, z_1^{3D} + h_v)$
5	$(x_1^{3D}, y_1^{3D}, z_1^{3D} + h_v)$	$(x_1^{3D}, y_1^{3D}, z_1^{3D} + h_v)$
6	$(x_1^{3D}, y_1^{3D} + l_v, z_1^{3D} + h_v)$	$(x_1^{3D}, y_1^{3D} + l_v, z_1^{3D} + h_v)$
7	$(x_1^{3D} - w_v, y_1^{3D} + l_v, z_1^{3D} + h_v)$	$(x_1^{3D} + w_v, y_1^{3D} + l_v, z_1^{3D} + h_v)$

本文车辆三维目标检测法步骤如下:

输入. 车辆二维目标检测.

输出. 车辆三维目标检测.

Step1. 利用深度学习 YOLO 网络对交通场景下的车辆进行二维目标检测,获得车辆二维包围框顶点图像坐标 $P_i^{2D} = (u_i^{2D}, v_i^{2D})$, $i = 0, 1, 2, 3$ 以及车辆类型. 设车辆包围框基准点为二维/三维模型中的 1 号顶点,记其在图像坐标系下的坐标为 $P_1^{2D} = (u_1^{2D}, v_1^{2D})$, 通过式(4)(5)可以将其转换为世界坐标系下的坐标 $P_1^{3D} = (x_1^{3D}, y_1^{3D}, z_1^{3D})$, 其中, $z_1^{3D} = 0$.

Step2. 设车辆实际三维物理信息为 (l_v, w_v, h_v) , l_v 表示车辆实际物理长度, w_v 表示车辆实际物理宽度, h_v 表示车辆实际物理高度,单位均为 mm. 按实际场景视角对应表 1 分别计算车辆三维包围框模型的 8 个顶点的图像坐标 $P_i^{3Di} = (u_i^{3Di}, v_i^{3Di}, z_i^{3Di})$, $i = 0, 1, \dots, 7$.

Step3. 将 Step2 中的计算结果通过式(2)(3)转换至图像坐标 $P_j^{3Di} = (u_j^{3Di}, v_j^{3Di})$, $j = 0, 1, \dots, 7$, 完成车辆三维目标检测.

图 4 所示为使用上述算法进行车辆二维/三维目标检测的示意图.

图 4 车辆二维/三维目标检测示意图

2.2 对角线约束

由于二维图像的透视特性,无法直接获取图像的深度信息,如果直接通过车辆三维包围框模型 8 个顶点世界坐标求取车辆三维信息,易受场景视角限制造成计算结果不准确.因此考虑构造一种对角线约束来精确求取车辆三维信息.

车辆三维信息识别等价于获取车辆三维信息 $X = (l_v, w_v, h_v)$, 定义车辆二维包围框对角线像素长度为

$$l_{2D} = \|P_1^{2D} - P_3^{2D}\|_2 \quad (11)$$

其中, $\|\cdot\|_2$ 表示求取 2 点之间欧氏距离的二范数.

根据车辆三维包围框模型,选取 1, 7 号点,同样可以定义车辆三维包围框对角线像素长度为

$$l_{3D} = \|P_1^{3Di} - P_7^{3Di}\|_2 \quad (12)$$

式(11)(12)中 2 个对角线像素长度之差构成一组对角线约束. 图 5 所示为车辆对角线约束的示意图,

红色/黄色线框分别为车辆二维/三维包围框, 当二维包围框与三维包围框完全贴合时, 蓝色线段表示二维/三维对角线在图像坐标系下完全重合, 此时对角线约束值为 0, 求取的车辆三维信息较为准确。

图 5 车辆对角线约束示意图

本文求取对角线约束的步骤如下:

输入. 车辆二维模型基准点图像坐标.

输出. 对角线约束.

Step1. 利用深度学习 YOLO 网络对交通场景下的车辆进行二维目标检测, 获得车辆二维包围框基准点图像坐标 $P_1^{2D} = (u_1^{2D}, v_1^{2D})$, 其余 3 个顶点图像坐标 $P_i^{2D} = (u_i^{2D}, v_i^{2D})$, $i = 0, 2, 3$, 以及车辆类型.

Step2. 将 (l_v, w_v, h_v) 看做一组未知参数, 利用式 (4)(5) 计算出基准点世界坐标 $P_1^{3D} = (x_1^{3D}, y_1^{3D}, z_1^{3D})$, 其中, $z_1^{3D} = 0$. 根据实际场景视角对应表 1 中的公式分别计算车辆三维包围框模型中 0 号顶点和 2~7 号顶点世界坐标 P_0^{3D} 和 P_2^{3D} 至 P_7^{3D} .

Step3. 按式(11)计算车辆二维包围框对角线像素长度 l_{2D} , 然后按式(6)(7)将 1, 7 号顶点的坐标转换至图像坐标 P_1^{3Di}, P_7^{3Di} , 最后按式(12)计算车辆三维包围框对角线像素长度 l_{3D} . l_{2D} 与 l_{3D} 构成一组对角线约束。

2.3 VP 约束

由 VP 性质可知, 车辆三维包围框模型的 0-3, 1-2, 4-7, 5-6 点对在图像坐标系下组成的直线一定经过沿道路方向 VP, $VP = (u_0, v_0)$. 由此性质可以将其作为另外一组约束条件, 用于精确求取车辆三维信息。

上述 4 组点对在图像坐标系下构成的直线与点对中, 其中一点与道路方向 VP 构成的直线之间的夹角分别记为 $\theta_1, \theta_2, \theta_3, \theta_4$.

由余弦定理可知

$$\cos \theta_1 = \frac{\|P_0^{3Di} - P_3^{3Di}\|_2^2 + \|P_0^{3Di} - VP\|_2^2 - \|P_3^{3Di} - VP\|_2^2}{2 \cdot \|P_0^{3Di} - P_3^{3Di}\|_2^2 \cdot \|P_0^{3Di} - VP\|_2^2} \quad (13)$$

$$\cos \theta_2 = \frac{\|P_1^{3Di} - P_2^{3Di}\|_2^2 + \|P_1^{3Di} - VP\|_2^2 - \|P_2^{3Di} - VP\|_2^2}{2 \cdot \|P_1^{3Di} - P_2^{3Di}\|_2^2 \cdot \|P_1^{3Di} - VP\|_2^2} \quad (14)$$

$$\cos \theta_3 = \frac{\|P_4^{3Di} - P_7^{3Di}\|_2^2 + \|P_4^{3Di} - VP\|_2^2 - \|P_7^{3Di} - VP\|_2^2}{2 \cdot \|P_4^{3Di} - P_7^{3Di}\|_2^2 \cdot \|P_4^{3Di} - VP\|_2^2} \quad (15)$$

$$\cos \theta_4 = \frac{\|P_5^{3Di} - P_6^{3Di}\|_2^2 + \|P_5^{3Di} - VP\|_2^2 - \|P_6^{3Di} - VP\|_2^2}{2 \cdot \|P_5^{3Di} - P_6^{3Di}\|_2^2 \cdot \|P_5^{3Di} - VP\|_2^2} \quad (16)$$

以上 4 个余弦值之和构成一组 VP 约束. 图 6 所示为车辆 VP 约束的示意图, 其中红色线段组用于提取 VP; 当二维包围框与三维包围框完全贴合时, 深蓝色线段表示点对与 VP 的 2 条连线完全重合, 夹角为 0, 此时 VP 约束值为 4, 求取的车辆三维信息较为准确。

图 6 VP 约束示意图

本文求取 VP 约束的步骤如下:

输入. 车辆三维包围框模型 8 个顶点世界坐标.

输出. VP 约束.

Step1~Step2 同求取对角线约束.

Step3. 按式(2)(3)将 0~7 号顶点的坐标转换至图像坐标 P_0^{3Di} 至 P_7^{3Di} , 按式(13)~式(16)分别计算 $\cos \theta_1$ 至 $\cos \theta_4$ 的值, 构成一组 VP 约束。

根据以上算法求取的对角线约束和 VP 约束, 构建约束误差, 形如 $l_{cal} - l_{truth}$. 其中, l_{cal} 为计算得到的约束实际值, l_{truth} 为车辆二维包围框与三维包围框完全贴合时的约束理想值. 分析上述算法容易看出, 约束误差 $l_{cal} - l_{truth}$ 中的变量由参数 l_v, w_v, h_v 构成, 它们可以构成参数向量的约束空间且为非线性。

综上所述, 构建参数 $X = (l_v, w_v, h_v)$ 的非线性约束函数为

$$\arg \min_X \frac{1}{2} \left(\sum_{i=1}^N \lambda_d (l_{2D} - l_{3D})^2 + \sum_{j=1}^4 \lambda_v (\cos \theta_j - 4)^2 \right).$$

其中, 设同一车辆在视频帧中出现次数为 N , λ_d 和 λ_v 分别表示对角线约束和 VP 约束的误差系数, 一般情况下均取 1, 可根据需要进行调整, $\min_X$ 表示使约束函数值达到最小时参数 X 的值。

约束函数是非线性的, 本文采用 LM (Levenberg-Marquardt) 法对约束函数进行求解, 具有较好的收敛性. 初始值 X_0 可根据 YOLO 网络识别出的车辆类型, 查阅国家道路车辆外廓尺寸标准获得.

3 实验结果与分析

本文的实验环境和参数为 Windows 10 系统, 8 GB 内存, Intel i5-7300HQ, 2.50 GHz CPU, GTX 1050 Ti GPU, 使用开源深度学习平台 Darknet 结合 CUDA9.0 和 CUDNN7.4.

在捷克布尔诺科技大学的公开数据集 BrnoCompSpeed^[27]和实际高速公路场景, 利用本文算法进行了实验, 2 组实验均采用第 1 和第 2 节的算法, 首先利用 VWL 标定方法对交通场景路侧监控相机进行标定, 获得标定参数, 接着使用 YOLO 网络进行车辆二维目标检测, 最后构建对角线和 VP 约束求解车辆三维信息, 完成车辆三维目标检测.

实验主要可分为以下 2 个方面: (1) 验证本文算法在车辆三维信息识别的精度; (2) 比较基于自标定的方法^[21-23]与本文算法在车辆三维信息识别的效果.

3.1 BrnoCompSpeed 数据集实验

公开数据集 BrnoCompSpeed 共包含 6 个场景的视频和背景图像, 每个场景均可分为左、中、右 3 个视角, 共约 200 GB 数据量. 如图 7 所示, 选取其中一个场景的 3 个视角对实验进行详细阐述. 图 7a~图 7c 分别为数据集中同一场景 3 个视角的场景背景图像, 道路均不含弯道部分. 图 8a 所示为基于 VWL 的相机标定示意图, 表 2 所示为 3 个视角路侧监控相机的标定结果. 图 8b 所示为采用深度学习 YOLO 模型对场景中的车辆进行二维目标检测的结果, 其中颜色框为检测到的车辆目标, 颜色框上部的黑色字母为目标车辆的类型, 颜色框下部的红色数值为车辆二维包围框基准点的图像坐标, 中视角车辆基准点的选择可使用左视角或右视角中的任意一种. 表 3~表 5 分别对应 3 个视角求解车辆

图 7 数据集场景 3 个不同视角背景图

图 8 数据集 3 个不同视角车辆三维信息识别实例图

表2 数据集不同视角的摄像机标定结果

视角	f	ϕ/rad	θ/rad	h/mm	VP
1	2991.97	0.1863	0.2711	10844.10	(114, -24)
2	3944.66	0.1549	0.0311	7910.72	(836, -76)
3	3347.43	0.2515	-0.2437	7672.77	(1819, -320)

表3 数据集左视角 Car 三维信息迭代结果及误差

迭代次数	基准点图像坐标	三维信息/mm	约束误差/像素
1	(543, 864)	(4669.37, 1319.04, 1200.85)	90.3561
2	(558, 881)	(4787.78, 1385.71, 1301.25)	84.8928
3	(561, 914)	(4807.91, 1403.70, 1334.40)	81.4546
4	(576, 930)	(4971.70, 1436.75, 1420.23)	70.1292
5	(582, 948)	(5079.39, 1496.43, 1437.63)	62.7345
6	(600, 988)	(5193.25, 1507.57, 1447.45)	26.8530
7	(606, 1003)	(5208.38, 1555.94, 1453.80)	23.9319
8	(627, 1022)	(5231.67, 1645.83, 1459.42)	18.4616

三维信息过程的参数迭代结果和约束误差值,车辆的真实物理三维信息分别为(5.5 m, 1.8 m, 1.6 m), (6.5 m, 1.6 m, 1.5 m), (6.0 m, 1.8 m, 1.7 m)。如图8c所示,由彩色线段组成的立方体为利用本文算法对车辆完成三维信息识别后进行车辆三维目标检测的结果。

与表3~表5对应,图9a~图9c分别为求解车辆三维信息时迭代次数与误差的曲线图,误差分

图9 数据集3个不同视角车辆三维信息识别迭代次数与误差曲线图

3.2 实际高速公路场景实验

为了进一步验证本文算法在实际场景的应用能力,选取了3个国内实际高速公路场景进行了实验验证。表6所示为3个场景路侧监控相机的标定结果。图10所示为所选3个高速公路场景的背景图像和基于VWL的相机标定示意图,图10a场景来自安平高速,图10b和图10c场景均来自杭金衢高速(G60)的杭向—吕塘角枢纽路段,且带有部分弯道。图11所示为在3个实际高速公路场景中求解车辆三维信息的测试实例,每个场景选取了不

为2类:三维信息误差(包含车辆长度误差、车辆宽度误差、车辆高度误差)和约束误差,三维信息误差的单位为mm,约束误差的单位为像素。可以看出,通过本文算法求解车辆三维信息最终都能够收敛,并且取得比较好的结果。

表4 数据集中视角 Truck 三维信息迭代结果及误差

迭代次数	基准点图像坐标	三维信息/mm	约束误差/像素
1	(717, 910)	(6316.21, 1444.89, 1212.64)	55.8432
2	(717, 926)	(6344.64, 1457.57, 1234.22)	53.3654
3	(714, 936)	(6383.23, 1475.57, 1276.00)	49.2130
4	(714, 941)	(6403.13, 1483.67, 1296.40)	42.7201
5	(711, 954)	(6422.03, 1497.03, 1316.91)	35.9161
6	(711, 971)	(6426.35, 1519.47, 1357.53)	35.2093
7	(702, 986)	(6429.21, 1527.06, 1378.72)	25.2260

表5 数据集右视角 Truck 三维信息迭代结果及误差

迭代次数	基准点图像坐标	三维信息/mm	约束误差/像素
1	(735, 873)	(5681.48, 1633.33, 1602.00)	105.2550
2	(717, 890)	(5696.82, 1640.26, 1625.30)	66.7127
3	(678, 905)	(5731.75, 1651.74, 1637.00)	63.1999
4	(678, 922)	(5732.14, 1682.86, 1648.00)	47.9342
5	(645, 946)	(5762.36, 1693.94, 1672.50)	19.4817
6	(621, 967)	(5802.51, 1608.87, 1689.90)	8.9099
7	(612, 983)	(5856.71, 1719.53, 1705.80)	5.3432

同类型的车辆作为测试对象,颜色框为检测到的车辆目标,颜色框上部的黑色字母为目标车辆的类型,颜色框下部的红色数值为二维包围框基准点的图像坐标,由彩色线段组成的立方体为利用

表6 高速公路不同场景的摄像机标定结果

场景	f	ϕ/rad	θ/rad	h/mm	VP
1	3960.24	0.0801	0.1192	12940.10	(484, 222)
2	3685.93	0.0477	-0.1055	12523.10	(1351, 364)
3	3639.65	0.0658	0.1423	12168.42	(437, 300)

图 10 高速公路 3 个不同场景背景图及标定示意图

图 11 高速公路 3 个不同场景车辆三维信息识别实例图

本文算法对车辆进行三维目标检测的结果, 每个车辆下部标注了算法求解的车辆三维信息结果. 从测试结果可以看出, 利用本文算法对不同场景不同类型的车辆均可以较为准确地求取车辆三维信息, 且绘制的三维包围框均与车辆有较高的贴合度.

表 7 展示了多场景多视角下, 利用本文算法得到的部分车辆三维信息识别结果及精度.

表 8 总结了 3 类常见方法与本文算法所需要的先验信息. 可以看出, 本文算法不受场景视角的限制, 也不需要模型库和方向角等先验信息, 具有较好的通用性.

在实际应用中, 车辆三维信息识别的精度与场景视角有密切关联, 场景视角通常使用相机偏转角 θ 表示. 在统计不同偏转角 θ 的情况下, 基于自标定的方法与本文算法整体精度的对比曲线图如图 12 所示. 从图 12 可以看出, 相机偏转角 θ 范围为 $[-45^\circ, 45^\circ]$, 当相机偏转角 θ 变换至 0° 时, 基于自标定的车辆三维信息识别方法精度会受较大影响, 而本文算法能够在较大程度上克服该缺陷. 经过实验验证, 本文算法在不同偏转角及初始车辆三维信息不准确的情况下均具有较高的识别精度, 综合识别精度超过 90%.

表 7 多场景多视角车辆三维信息识别结果及精度

序号	车辆类型	基准点图像坐标	三维信息/m	实际三维信息/m	误差/m	精度/%
1	Car	(716, 319)	(4.47, 1.65, 1.46)	(4.2, 1.8, 1.6)	(0.27, 0.15, 0.14)	92.3
2	Truck	(551, 583)	(6.43, 1.53, 1.38)	(6.5, 1.6, 1.5)	(0.07, 0.07, 0.12)	95.5
3	Car	(611, 1000)	(4.86, 1.72, 1.71)	(4.0, 1.8, 1.7)	(0.86, 0.08, 0.01)	92.4
4	Car	(455, 167)	(4.78, 1.24, 1.74)	(4.0, 1.5, 1.7)	(0.78, 0.26, 0.04)	88.0
5	Car	(237, 176)	(4.71, 1.41, 1.32)	(4.0, 1.5, 1.5)	(0.71, 0.09, 0.18)	89.0
6	Truck	(584, 229)	(6.23, 1.65, 2.56)	(6.3, 1.6, 2.5)	(0.07, 0.05, 0.06)	97.8
7	Truck	(360, 279)	(7.82, 1.43, 2.51)	(7.0, 1.5, 2.5)	(0.82, 0.07, 0.01)	94.8
8	Car	(588, 503)	(4.08, 1.63, 1.33)	(4.5, 1.8, 1.7)	(0.42, 0.17, 0.37)	86.5
9	Truck	(834, 929)	(12.83, 2.42, 3.44)	(13.0, 3.0, 3.0)	(0.17, 0.58, 0.44)	88.9
10	Car	(148, 567)	(4.15, 1.63, 1.35)	(4.3, 1.6, 1.5)	(0.15, 0.03, 0.15)	94.9
11	Truck	(949, 1070)	(5.16, 2.32, 1.87)	(5.5, 2.5, 2.0)	(0.34, 0.18, 0.13)	93.4
12	Bus	(934, 846)	(13.89, 2.33, 3.81)	(15.0, 2.5, 3.5)	(1.11, 0.17, 0.31)	92.6

表 8 不同方法需要的先验信息对比

方法	先验信息		
	模型库	方向角	场景视角
基于车辆模型 ^[13-16]	√	×	√
基于深度学习 ^[17-20]	×	√	√
基于自标定 ^[21-23]	×	×	√
本文	×	×	×

图 12 不同方法车辆三维信息识别的综合精度

4 结 语

通过实验验证, 本文提出的单目交通场景基于标定信息、对角线和 VP 约束的车辆三维信息识别算法, 在多种场景不同视角下均取得了较好的效果. 本文的主要贡献有: (1) 提出一种基于对角线和 VP 约束的车辆三维信息识别算法, 可以完成较为准确的车辆三维信息识别. (2) 将深度学习二维目标检测的方法与摄像机标定结合, 对车辆目标完成三维目标检测.

但是, 通过本文算法求取车辆三维信息时, 对于相机标定结果有较大的依赖性, 且在相机偏转

角接近 0° 时, 具有一定的误差. 在未来的工作中, 可以使用场景适应性更好的标定算法, 融合其他类型的几何约束, 进一步提高车辆三维信息识别的准确度.

参考文献(References):

- [1] Geiger A, Lenz P, Urtasun R. Are we ready for autonomous driving? The KITTI vision benchmark suite[C] //Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Los Alamitos: IEEE Computer Society Press, 2012: 3354-3361
- [2] Chen X Z, Kundu K, Zhang Z Y, *et al.* Monocular 3D object detection for autonomous driving[C] //Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Los Alamitos: IEEE Computer Society Press, 2016: 2147-2156
- [3] Zhang Yan. Research on recognition method of vehicle type based on three-dimensional estimation with monocular camera[D]. Xi'an: Xi'an University of Technology. Department of Automation & Information Engineering, 2017(in Chinese) (张岩. 基于单目摄像头三维估计的车型识别方法研究[D]. 西安: 西安理工大学自动化与信息工程学院, 2017)
- [4] Zeeshan Zia M, Stark M, Schiele B, *et al.* Detailed 3D representations for object recognition and modeling[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2013, 35(11): 2608-2623
- [5] Zhang Kailin, Zhang Liang. 3D object recognition and 6DoF pose estimation in scenes with occlusions and clutter based on C-SHOT 3D descriptor[J]. Journal of Computer-Aided Design & Computer Graphics, 2017, 29(5): 846-853(in Chinese) (张凯霖, 张良. 复杂场景下基于 C-SHOT 特征的 3D 物体识别与位姿估计[J]. 计算机辅助设计与图形学学报, 2017, 29(5): 846-853)
- [6] Song S, Lichtenberg S P, Xiao J X. SUN RGB-D: a RGB-D scene understanding benchmark suite[C] //Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Los Alamitos: IEEE Computer Society Press, 2015: 567-576
- [7] Redmon J, Divvala S, Girshick R, *et al.* You only look once:

- unified, real-time object detection[C] //Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Los Alamitos: IEEE Computer Society Press, 2016: 779-788
- [8] Liu W, Anguelov D, Erhan D, *et al.* SSD: single shot multibox detector[C] //Proceedings of European Conference on Computer Vision. Heidelberg: Springer, 2016: 21-37
- [9] Zhang Z Y. A flexible new technique for camera calibration[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2000, 22(11): 1330-1334
- [10] Dawson D N, Birchfield S T. An energy minimization approach to automatic traffic camera calibration[J]. IEEE Transactions on Intelligent Transportation Systems, 2013, 14(3): 1095-1108
- [11] Zheng Y, Peng S L. Apractical road side camera calibration method based on least squares optimization[J]. IEEE Transactions on Intelligent Transportation Systems, 2014, 15(2): 831-843
- [12] Zhang Z X, Tan T N, Huang K Q, *et al.* Practical camera calibration from moving objects for traffic scene surveillance[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2013, 23(3): 518-533
- [13] Corral-Soto E R, Elder J H. SlotCars: 3D modelling for improved visual traffic analytics[C] //Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Los Alamitos: IEEE Computer Society Press, 2017: 16-24
- [14] Lin Y L, Morariu V I, Hsu W, *et al.* Jointly optimizing 3D model fitting and fine-grained classification[C] //Proceedings of European Conference on Computer Vision. Heidelberg: Springer, 2014: 466-480
- [15] Prokaj J, Medioni G. 3D model based vehicle recognition[C] // Proceedings of the IEEE Workshop on Applications of Computer Vision. Los Alamitos: IEEE Computer Society Press, 2009: 1-7
- [16] 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, 2011, 21(1): 1-13
- [17] Sochor J, Herout A, Havel J. BoxCars: 3D boxes as CNN input for improved fine-grained vehicle recognition[C] //Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Los Alamitos: IEEE Computer Society Press, 2016: 3006-3015
- [18] Sochor J, Špaňhel 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
- [19] Mousavian A, Anguelov D, Flynn J, *et al.* 3D bounding box estimation using deep learning and geometry[C] //Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Los Alamitos: IEEE Computer Society Press, 2017: 5632-5640
- [20] Chabot F, Chaouch M, Rabarisoa J, *et al.* Deep MANTA: acoarse-to-fine many-task network for joint 2D and 3D vehicle analysis from monocular image[C] //Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Los Alamitos: IEEE Computer Society Press, 2017: 1827-1836
- [21] Dubská M, Herout A, Juránek R, *et al.* Fully automatic road-side camera calibration for traffic surveillance[J]. IEEE Transactions on Intelligent Transportation Systems, 2015, 16(3): 1162-1171
- [22] Yang S C, Scherer S. CubeSLAM: monocular 3D object SLAM[J]. IEEE Transactions on Robotics, 2019, 35(4): 925-938
- [23] Dubská M, Herout A, Sochor J. Automatic camera calibration for traffic understanding[C] //Proceedings of British Machine Vision Conference. Guildford: BMVA Press, 2014: 1-12
- [24] 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
- [25] Wang Wei, Zhang Chaoyang, Tang Xinyao, *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(in Chinese)
(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)
- [26] Dubská M, Herout A. Real projective plane mapping for detection of orthogonal vanishing points[C] //Proceedings of British Machine Vision Conference. Guildford: BMVA Press, 2013: 1-10
- [27] Sochor J, Juránek R, Špaňhel J, *et al.* Comprehensive dataset for automatic single camera visual speed measurement[J]. IEEE Transactions on Intelligent Transportation Systems, 2018, 20(5): 1633-1643