

文章编号: 2095-2163(2023)04-0147-05

中图分类号: TP389.1

文献标志码: A

基于双目视觉的车辆检测跟踪与测距

郭鹏宇

(上海工程技术大学 机械与汽车工程学院, 上海 201620)

摘要: 由于道路上存在各种不安全因素, 车辆检测跟踪并测距是自动驾驶技术的重要组成部分。本文将 YOLOv4-tiny 作为检测器使之加快模型检测速度且更适合在车辆嵌入式设备中使用。考虑到目标检测失败的情况, 本文在历史缓冲区中存储以前的跟踪细节(唯一 ID)和每个对象的相应标签, 提出了一个基于中值的标签估计方案(MLP), 使用存储在前一帧的历史标签的中值来预测当前帧中对象的检测标签, 使得跟踪错误最小化, 并用双目摄像头获取图像检测车辆距离。测试新网络结构后, 检测精度(Mean Average Precision, mAP)为 80.14%, 检测速度较 YOLOv4 相比提高了 184%, 检测到的距离误差平均在 0.5% 左右。

关键词: YOLOv4-tiny; 目标跟踪; 中值算法; 双目测距

Vehicle detection, tracking and ranging based on binocular vision

GUO Pengyu

(School of Mechanical and Automotive Engineering, Shanghai University of Engineering Science, Shanghai 201620, China)

[Abstract] Due to various unsafe factors on the road, vehicle detection, tracking and ranging are the important part of automatic driving technology. In this paper, YOLOv4-tiny is used as a detector to speed up model detection and is more suitable for vehicle embedded devices. Considering the failure of object detection, this paper stores the previous tracking details (unique ID) and the corresponding label of each object in the history buffer, and proposes a median-based label estimation scheme (MLP), which uses the median value of the history label stored in the previous frame to predict the detection label of the object in the current frame, so that tracking errors are minimized. The images obtained by binocular camera are used to detect vehicle distance. After testing the new network structure, the detection accuracy (Mean Average Precision, mAP) is 80.14%, the detection speed is 184% higher than that of YOLOv4, and the detected distance error is about 0.5% on average.

[Key words] YOLOv4-tiny; target tracking; median algorithm; binocular distance measurement

0 引言

在自动驾驶辅助系统中, 基于传感器, 采用车辆检测、跟踪、测距等一系列计算机视觉算法进行环境感知, 辅助系统就能得到车辆周围信息, 以保证驾驶员安全驾驶。

基于视觉的车辆检测及测距系统主要应用在道路交通场景下, 用于辅助检测前方目标以及进行距离预警, 其性能好坏主要依赖于采用的车辆检测算法。

目前, 在使用相机进行目标检测时, 多采用传统的机器视觉检测方法。对于前方车辆目标, 该方法首先根据车辆的局部特征, 如阴影、边缘纹理、颜色分布等特征生成感兴趣区域; 然后利用对称特征等整体特征对感兴趣区域进行验证。在从产生感兴趣

区域到验证感兴趣区域的过程中, 为了达到实时检测的要求, 一般需要对图像进行灰度化, 并对灰度化后的图像进行阴影分割和边缘分析。因此, 对于相机获得的图像, 传统的机器视觉的车辆检测方法通常找到感兴趣区域的车辆的特点和梯度直方图特征(HOG^[1]), SIFT^[2]特征或 Haar-like^[3]特征通常用于获得前面的假设检验区域车辆, 即 ROI 区域; 此后用这些特征训练 SVM^[4]或 Adaboost^[5]车辆检测分类器, 计算车辆图像的特征值, 并根据车辆特征值的大小与前方车辆进行判断, 得到前车的假设测试区域验证, 完成对前车的检测。

而上述传统的机器视觉检测方法本质上是通过人工选择特征进行识别和分类。在复杂场景中, 人工特征的数量会呈几何级数增长, 这对前面车辆的识别率也有很大的影响。这种方法更适合在某种特

作者简介: 郭鹏宇(1995-), 女, 硕士研究生, 主要研究方向: 智能网联汽车。

收稿日期: 2022-05-24

定场景下的车辆识别,因为其数据规模并不大,泛化能力则较差,很难实现快速和准确的复杂应用场景的检测。近年来,随着卷积神经网络(CNN)的应用,出现了许多算法。一阶段方法包括 SSD^[6]、YOLO 系列^[7-8]、RetinaNet^[9]。两阶段方法包括 Fast R-CNN^[10]和 Faster R-CNN^[11]。最近提出的最先进的 YOLO-v4^[12]具有很高的检测精度和检测速度。

目前,对于多目标车辆轨迹跟踪技术主要可分为两大类。一类是传统方法,如利用背景差分法、帧差法、光流法等方法提取运动目标,传统方法部署方便,资源消耗低,但受先验知识限制,跟踪稳定性差,准确性不高。另一类是基于卷积神经网络的、称为深度学习的方法,深度学习可以学习更多的目标特征,能在连续的视频帧中检测出目标对象。深度学习精度方法精度高,但其计算量较大,实时性不高,因此,基于视频跟踪的车辆检测算法仍需改进。

研究可知,基于视觉相机的测距方法主要有单目测距和双目测距两种。这2种方法的共同特点是通过相机采集图像数据,随后从图像数据中得到距离信息。单目检测方法的优点是成本低,缺点是对检测精度的依赖过大。此外,从实用的角度来看,在汽车上安装单目相机时,由于汽车的颠簸,汽车的俯仰角经常发生变化,导致精度显著下降。双目测距的方法是通过计算2幅图像的视差直接测量距离。

1 车辆检测与跟踪

本文使用的目标检测算法是 YOLOv4-tiny,其中 YOLO 表示 You Only Look Once,由 Bochkovskiy 等学者开发。YOLOv4-tiny 是 YOLOv4 的压缩版本,虽在平均精度方面受到了影响,但却可以在低计算能力下高效运行。与未压缩版本的4个 YOLO 头相比,YOLOv4-tiny 只使用了2个 YOLO 头,并使用了29个预训练卷积层作为基础。YOLO 各变量参数设置见表1,卷积层各变量参数设置见表2。

上一代 YOLO 的非 maxpool 抑制(NMS)等遗留

特性和一些新特性、包括加权剩余连接(WRC)、Mosaic 数据增强在内有效提高了算法在模糊图像中识别类的能力,降低了识别类所需的处理能力。YOLOv4-tiny 提供了较高的帧率,同时具有中间地带平均精度与常用模型并列。

在本文中,使用 YOLOv4-tiny 算法作为车辆的检测器,并且使用 DeepSORT^[13]算法作为初始车辆跟踪器。

表1 YOLO 各变量参数设置
Tab. 1 YOLO parameter settings

YOLO 变量	参数
<i>Mask</i>	0, 1, 2
<i>anchors</i>	10, 14, 23, 27 37, 58 81, 82 135, 169 344, 319
<i>classes</i>	4
<i>num</i>	6
<i>jitter</i>	0.3
<i>scale_x_y</i>	1.05
<i>cls_normalizer</i>	1.0
<i>iou_loss</i>	<i>ciou</i>
<i>ignore_thresh</i>	0.7
<i>truth_thresh</i>	1
<i>random</i>	0
<i>Resize</i>	1.5
<i>nms_kind</i>	<i>greedydyncs</i>
<i>beta_nms</i>	0.6

表2 卷积层各变量参数设置
Tab. 2 The convolution layer parameter settings

卷积层变量	参数
<i>batch_normalize</i>	1
<i>filters</i>	64
<i>size</i>	3
<i>stride</i>	2
<i>pad</i>	1
<i>activation</i>	<i>leaky</i>

图1显示了2个ID及其前3个标签。对于ID#137的车辆,本文方法预测的标签用加黑来标记。

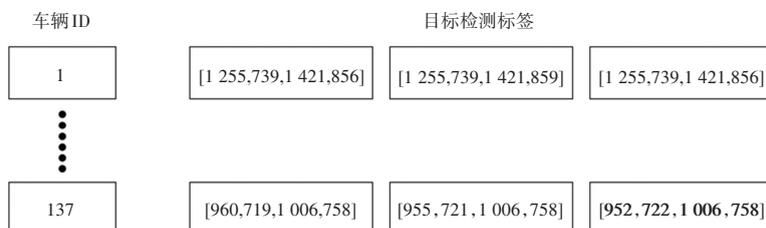


图1 应用 MLP 后的历史缓冲区示例图

Fig. 1 A history buffer example after applying MLP

本文使用历史缓冲区来调整和改进每个检测标签的视觉质量和在帧中的显示。如果有任何车辆检测标签缺失,那么本文的 MLP 为该车辆生产一个估计的检测标签。延时使用一系列的检测标签前存储在历史缓冲区来预测未检测到车辆的检测标签 ID 在给定的框架(见图 1)。条件估计为特定车辆检测标签,标签 ID 应该至少在 2 个连续帧出现。为了预测缺失的检测标签,本文对当前帧 t 应用以下公式:

$$\nabla = l(t-1)_{(i)} - l(t-2)_{(i)} \quad (1)$$

$$l(t)_{(i)} = l(t-1)_{(i)} + \nabla \quad (2)$$

这里, $l(x_{\min}, y_{\min}, x_{\max}, y_{\max})$ 表示每个车辆 ID 基于调整边界框标签的中值, ∇ 表示边界框位置的变化从时间戳 $(t-2)$ 到 $(t-1)$; i 表示每辆车唯一的 ID。

2 双目测距

双目视差示意如图 2 所示。由图 2 可知,2 个摄像头的中心距是 B , 两个摄像头观察同一点 P , P_1 的坐标为 (X_1, Y_1) , P_2 的坐标为 (X_2, Y_2) , 由于 2 台相机始终是平行的,高度相同,所以在左右 2 张图像中 P 点的 Y 坐标是相同的,在 X 方向上存在视差。因此,可以通过标定相机内外参数来确定图像坐标与世界坐标之间的关系。双目视差原理如图 3 所示。

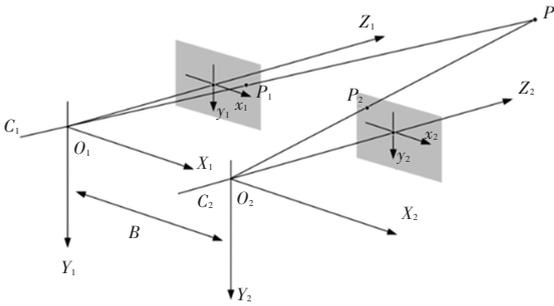


图 2 双目视差示意图

Fig. 2 Schematic diagram of binocular parallax

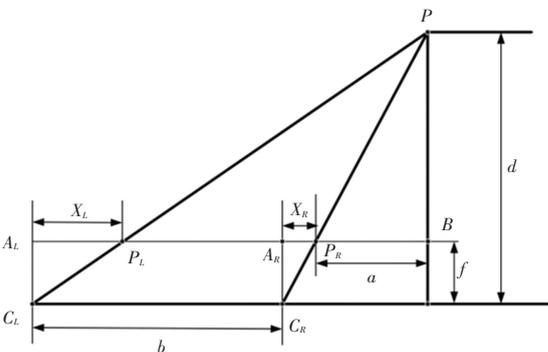


图 3 双目视差原理图

Fig. 3 Principle of binocular parallax

图 3 中, C_L 和 C_R 是 2 个摄像头的光学中心,摄像头之间的距离是 b ,相机的焦距为 f , P 点在左右图像的投影点为 $P_l(X_l, Y_l)$, $P_r(X_r, Y_r)$, $|A_L, P_L| = X_L$, $|A_R, P_R| = X_R$, $|P_R, B| = a$, 从三角形相似关系可知:

$$\frac{d-f}{d} = \frac{a}{a+x_R}$$

$$\frac{d-f}{d} = \frac{b-x_L+x_R+a}{b+x_R+a} \quad (3)$$

由式(3)可知:

$$a = \frac{b x_R}{x_L - x_R} - x_R \quad (4)$$

由此,空间中 P 点到相机的距离为:

$$d = f \frac{a+x_R}{x_R} = \frac{bf}{x_L-x_R} \quad (5)$$

P 在相机坐标系中的三维坐标可以由几何关系得到:

$$\begin{cases} X = \frac{b x_L}{x_L - x_R} \\ Y = \frac{b y}{x_L - x_R} \\ Z = \frac{b f'}{x_L - x_M} \end{cases} \quad (6)$$

对于车辆的测距,本文取检测到的边界框内每辆车的中心来表示被检测物体到双目相机中心的距离。

3 实验结果与分析

将 YOLOv4-tiny 与其他常用的目标检测算法进行比较,将其 mAP 与 FPS 进行比较,得到表 3 中的结果。

本文提出的车辆检测与跟踪方法使用了 TensorFlow 库和基于 YOLOv4-tiny 模型的 DeepSORT 算法。经综合比较,使用 YOLOv4-tiny 的精度和检测速度是可以接受的,精度比 YOLOv3-tiny 高,速度比 YOLOv4 的方法更快。

YOLOv4-tiny 模型检测车辆效果如图 4 所示。

表 3 各模型帧率和 mAP 对比

Tab. 3 Frame rate and mAP comparison

模型	mAP / %	帧率 (FPS)
YOLOv4	85.08	14.12
YOLOv4-tiny	80.14	40.11
YOLOv3	83.32	16.99
YOLOv3-tiny	69.03	52.77

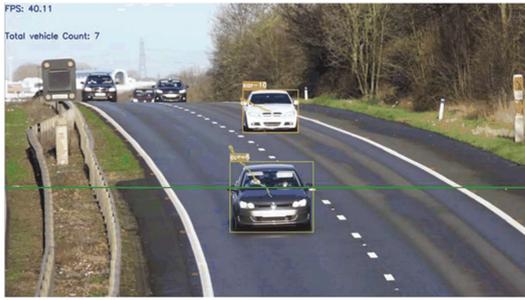
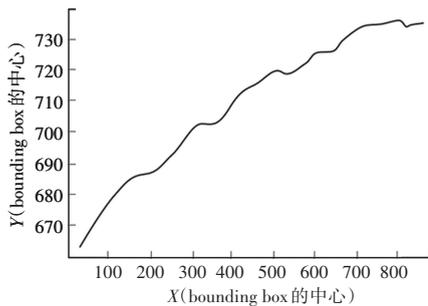


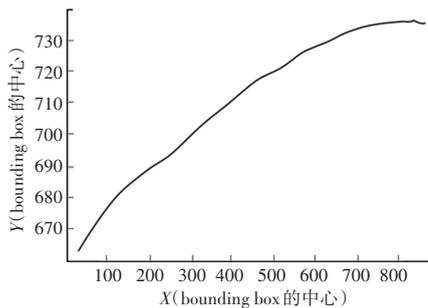
图4 YOLOv4-tiny 模型检测车辆效果

Fig. 4 Cars video detection using YOLOv4-tiny model

使用本文方法前后汽车的标签变化曲线如图5所示。对于 ID#39 的车辆,图 5(a)是使用方法前,图 5(b)是使用方法后,相同的汽车标签变得更加平滑。



(a) 使用本文方法前



(b) 使用本文方法后

图5 使用本文方法前后汽车的标签变化

Fig. 5 The label changes before and after using the method in this paper

在目标跟踪时,从历史缓冲区中预测缺失标签的方法往往会产生不好的结果,因为对象检测器的可视化结果经常显示不稳定和闪烁的边框标签。在应用本文的基于中值的算法后,可以得到高度稳定的标签。因此,本文方法提高了目标检测器的视觉性能,并为目标检测器和跟踪器提供了对缺失标签的更好估计。

利用双目相机取检测到的边界框内每辆车的中心来表示被检测物体到双目相机中心的距离。

仿真测试结果见表4。从距离测试的结果来看,测试精度相对较高,基本保持在 0.5% ~ 0.6% 之间。

表4 测量结果分析

Tab. 4 The measured results analysis

实验组数	测量距离/cm	实际距离/cm	误差/%
1	1 567.00	1 559.11	0.503 5
2	1 655.00	1 646.14	0.535 3
3	1 738.00	1 729.16	0.508 6
4	1 893.00	1 883.17	0.519 3
5	1 983.00	1 971.20	0.595 1
6	2 236.00	2 223.22	0.571 6
7	2 489.00	2 475.26	0.552 0

4 结束语

本文介绍了一种用于自动驾驶的实时检测跟踪与测距系统。通过本文提出的实时同步方法,该系统方便了车辆实时同步检测;利用双目摄像头, YOLOv4-tiny 和 DeepSORT 算法对车辆进行检测和跟踪,并提出中值标签预测方法优化跟踪效果,同时实现了对前方车辆的精确测距。整个系统在检测和测距方面取得了较高的精度和实时性。对于自动驾驶的应用,该系统可以结合许多智能技术,如目标预警、自动避障等。与此同时,该系统还有很大的改进空间。在检测方面,通过优化算法提高检测性能,通过训练更多类型的物体,如行人、非机动车等,为自动驾驶提供更多的道路信息。在这个系统中,测距是指从双目相机的中心到物体的距离。在实际情况下,车辆的具体位置到物体的距离可以根据相机的安装位置和车辆的实际长度来计算。通过优化双目测距算法,可以提高测距精度。

参考文献

- [1] TAIGMAN Y, YANG Ming, RANZATO M A, et al. DeepFace: Closing the gap to human-level performance in face verification [C]//2014 IEEE Conference on Computer Vision and Pattern Recognition. Columbus, OH, USA: IEEE, 2014:1701-1708.
- [2] MA Xiaoxu, GRIMSON W E L. Grimson. Edge-based rich representation for vehicle classification [C]//Tenth IEEE International Conference on Computer Vision. Beijing: IEEE, 2005:1185-1192.
- [3] XU Qing, GAO Feng, XU Guoyan. An algorithm for front-vehicle detection based on Haar-like feature[J]. Automotive Engineering, 2013, 35(4):381-384.