

引用格式:赵斌,王江卓,李建勋.基于TLD的三维单目标跟踪算法[J].电光与控制,2020,27(2):22-28. ZHAO B, WANG J Z, LI J X. A 3D single-target tracking algorithm based on TLD[J]. Electronics Optics & Control, 2020, 27(2):22-28.

基于 TLD 的三维单目标跟踪算法

赵斌¹, 王江卓², 李建勋²

(1. 空装项目管理中心, 北京 100843; 2. 上海交通大学自动化系, 上海 200240)

摘要: 目标跟踪是计算机视觉领域研究的热门话题,在实践中有着广泛的应用。主要针对三维单目标跟踪问题进行研究,考虑到三维目标检测已有较好表现,将通过视锥点云网络得到的三维目标检测结果与 TLD 目标跟踪算法相结合,并改进 TLD 算法的检测模块,构造了基于 TLD 的三维单目标跟踪算法,输入 RGB-D 图像序列,输出目标的三维跟踪结果。在 KITTI 数据集中测试结果表明该算法在三维目标跟踪上的可行性。

关键词: 三维目标跟踪; TLD; 视锥点云网络; 目标检测; RGB-D

中图分类号: TP391 **文献标志码:** A **doi:**10.3969/j.issn.1671-637X.2020.02.006

A 3D Single-Target Tracking Algorithm Based on TLD

ZHAO Bin¹, WANG Jiangzhuo², LI Jianxun²

(1. Project Management Center of Air Force Equipment, Beijing 100843, China;

2. Department of Automation, Shanghai Jiaotong University, Shanghai 200240, China)

Abstract: Target tracking is a hot topic in the field of computer vision and has a wide range of applications in practice. The 3D single-target tracking problem is mainly studied in this paper. Considering that the 3D target detection has good performance, the 3D target detection result obtained by using the frustum point cloud network is combined with the TLD target tracking algorithm, and the detection module of the TLD algorithm is improved. A 3D single-target tracking algorithm based on the TLD algorithm is proposed. The input is the RGB-D image sequence, and the output is the 3D tracking results of the target. The test results in the KITTI data set prove the feasibility of the algorithm in 3D target tracking.

Key words: 3D target tracking; TLD; frustum point cloud network; target detection; RGB-D

0 引言

近年来,随着计算机硬件性能的提升以及计算机视觉技术的发展,人工智能也逐渐成为研究热门,而视觉目标跟踪领域也得到了许多学者的广泛关注,取得了较为显著的成果。目标跟踪技术是先进辅助驾驶系统(Advanced Driver Assistant System, ADAS)的关键技术,也是近年来备受关注的自动驾驶汽车的基础技术构成之一。

在单目标研究领域,学者的研究主要是基于二维场景进行。早期的目标跟踪工作主要是基于生成式模型的研究。LUCAS 等^[1]提出的改进光流算法,通过求解运动目标的光流来进行目标跟踪;COMANICIU 等^[2]

提出的基于 Mean-Shift 均值偏移的目标跟踪算法,利用均值偏移自适应步长的特性,提升了算法实时性。这类算法主要利用了目标本身自带的外观信息建立模型,然后在后续帧的候选集中进行相似性搜索,寻找与模型最相近的样本作为当前帧目标位置,但是由于这种方法忽略了背景信息,单一的数学模型无法很好地描述跟踪目标,因此,有学者将背景信息引入跟踪模型,建立了目标跟踪的判别式模型。AVIDAN 等^[3]首次将支持向量机(Support Vector Machine, SVM)引入单目标跟踪,对光流跟踪中的偏移量进行分类,用于区分背景与目标;而 BOLME 等^[4]则将相关滤波方法引入目标跟踪,使用相关滤波器在灰度特征上进行计算,输出目标分布热度图,将热度图中像素值高的区域作为当前帧的目标预测结果,在速度和准确率上显示出更加优越的性能;HENRIQUES 等^[5-6]则基于相关滤波算法,引入循环矩阵来生成负样本,并转移到傅里叶空间求解,提出了核相关滤波(Kernelized Correlation Filter, KCF)算法,提高了跟踪效率。

收稿日期:2019-03-18 修回日期:2019-04-22

基金项目:国家自然科学基金(61673265)

作者简介:赵斌(1978—),男,新疆沙湾人,硕士,工程师,研究方向为飞行器。

而在判别式跟踪算法中,一种表现优异的算法是 KALAL 等^[7]提出的跟踪—学习—检测(Tracking-Learning-Detection, TLD)算法,将跟踪任务分解为跟踪、学习、检测 3 个任务,使用检测器和跟踪器并行运算,并且都与学习器关联交互,而学习器得到的模型又反馈给检测器和跟踪器,这种算法可靠并且可以自适应,提高了跟踪算法的鲁棒性。许多学者在 TLD 算法基础上进行了改进,CHEN 等^[8]在 TLD 算法中加入改进的卡尔曼滤波器提高了算法可靠性;JIA 等^[9]在 TLD 算法中使用 LBP 算子,使得目标具有明显的纹理属性,提高了分类器效果。

上述算法都是针对二维场景进行的,而在三维目标跟踪领域,CHOI 等^[10]提出了一种基于模型的三维跟踪方法,在粒子滤波框架内使用边缘和关键点特征,将 2D 特征与 3D 特征相关联,可以较为准确地进行三维跟踪;王莹等^[11]则提出了一种基于 RGB-D 图像的 Tri-tracking 跟踪框架,将多种模型融合,考虑了目标的多种信息,提高了目标跟踪效果。但是这些算法大部分是使用 RGB-D 的信息在场景中框选目标所在位置的二维框,而无法得到跟踪过程中目标的三维框位置,并且大部分没有形成闭环系统,使得系统的稳定性和鲁棒性并不是太好。考虑到 TLD 算法在二维场景中的优越表现,本文将 TLD 算法引入三维目标跟踪中,使用深度学习方法得到的三维目标检测结果与 TLD 算法检测模块相结合,进行三维单目标跟踪。

1 相关理论

图 1 所示为 TLD 算法的基本原理框架图,其主要分为 3 个模块,分别是检测模块、跟踪模块和学习模块。

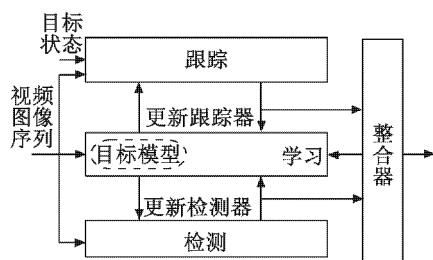


图 1 TLD 算法原理

Fig. 1 Principle of TLD algorithm

下面将详细介绍 TLD 算法的这 3 个模块。

首先,针对要跟踪的对象建立目标模型 M ,它是由正负样本图像块组成的集合,包含了跟踪目标本身及其周围的背景信息,定义 $M = \{r_1^+, r_2^+, \dots, r_m^+, r_1^-, r_2^-, \dots, r_n^-\}$,其中, r^+ 表示目标图像块的正样本, r^- 表示背景图像块的负样本。

如图 2 所示, TLD 算法的检测模块是一个级联检测器,由方差分类器、集合分类器和最近邻分类器(NCC)3

个分类器组成。

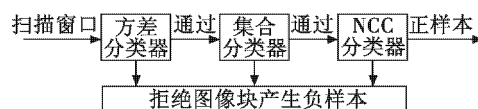


图 2 级联检测器原理

Fig. 2 Principle of cascade detector

方差分类器计算各区域灰度值的方差,将其中低于目标图像块方差的区域视作背景区域,剩下的区域认为含有前景目标,输入到集合分类器中。在集合分类器中,从图像块中任意选取两个像素点,比较这两点的亮度值,输出特征值分别为 0,1,每选取一对位置,输出一个二进制值,一对像素点进行比较形成一个蕨的节点。不同种类的样本通过同一个蕨时,输出不同的先验概率分布,当有新的未分类样本出现时,从一致的分布中寻找后验概率最大的一个,判断其属于哪一类,而最近邻分类器使用相似度度量来描述样本更新,并区分正负样本。正负样本的相似度分别为

$$S^+(r, M) = \max_{r_i^+ \in M} S(r, r_i^+) \quad (1)$$

$$S^-(r, M) = \max_{r_i^- \in M} S(r, r_i^-) \quad (2)$$

定义保守相似度为

$$S^c = \frac{S_{50\%}^+}{S_{50\%}^+ + S^-} \quad (3)$$

式中, $S_{50\%}^+(r, M)$ 称为前 50% 正样本集合的最近邻相似度,其定义为

$$S_{50\%}^+(r, M) = \max_{r_i^+ \in M \wedge i < m/2} S(r, r_i^+) \quad (4)$$

相关相似度定义为

$$S^r = \frac{S^+}{S^+ + S^-} \quad (5)$$

S^r 的变化区间为 $[0, 1]$,随着 S^r 的增大,图像块成为目标的概率也随之增大。同样的 S^c 的变化区间也是 $[0, 1]$,随着 S^c 的增大,目标存在正样本前半部分集合中的概率越大。使用 (S^r, S^c) 描述图像块及目标的相似度, S^r 为更新最近邻分类器的指标,当 $S^r(r, M) > \theta_{NN}$ 时,图像块 r 为正样本, θ_{NN} 便于使最近邻分类器趋向于收敛;使用相关相似度 S^r 进行正负样本分类,而用保守相似度 S^c 判断以检测器和跟踪器中的何者作为最终目标框位置。

TLD 算法的目标跟踪模块是基于中值流跟踪器(Median-Flow)进行设计的,并且使用前后向误差进行跟踪,图 3 为跟踪过程中前后向误差示意图。如果一个目标在连续时间段内能在第 t 帧被检测到,那么它大概率也会被第 $t+1$ 帧检测到,中值流跟踪器利用这个原则,先在第 t 帧中提取特征点,并在第 $t+1$ 帧中跟踪这些特征点,之后利用反向跟踪的思路,计算回到第 t 帧的特征点的误差,筛选出误差较小的一定比例的像

素点作为最佳跟踪点,由这些点在第 t 帧和第 $t+1$ 帧间的坐标变化判断物体的实际运动轨迹,并以此选取目标框的位置和大小。

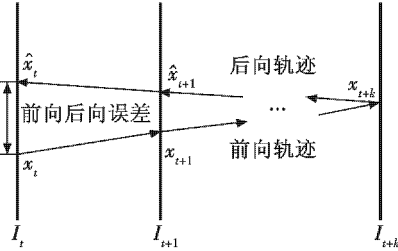


图 3 追踪器前后向误差示意图

Fig. 3 Forward and backward error of the tracer

TLD 算法的学习模块主要是使用 PN 学习,其流程如图 4 所示,其主要思想是将假阳性估计和假阴性估计分开,估计分类器的误差,是一种半监督的机器学习算法。P 专家的作用是检测出漏检的正样本,判断目标在时间上的连续性,由追踪器预测第 $t+1$ 帧目标位置,若检测器将其判断为负样本,则修正为正样本;而 N 专家的作用是改正误检的正样本,主要利用目标在空间上的连续性,如果产生了多个正样本,通过比较,选择最可信的位置。

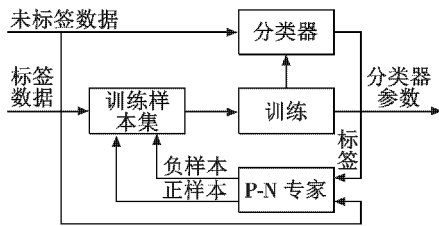


图 4 PN 学习工作流程

Fig. 4 Work flow of PN learning

2 基于 TLD 的三维单目标跟踪算法

从前文的介绍可看出,TLD 算法是一种鲁棒的单目标跟踪算法,它能够在跟踪过程中不断进行学习,优化检测器和跟踪器的效果,实时跟踪目标的特征,提高

跟踪效果。因此设计了基于 TLD 的单目标跟踪算法,将其扩展到三维空间中,使得结果更加鲁棒。

图 5 所示为本文设计的基于 TLD 的三维单目标跟踪算法框架,首先在每一帧中使用三维目标检测网络得到场景中所有目标的三维位置和大小,然后与 TLD 目标跟踪算法相结合并改进检测器,最终输出跟踪目标的三维位置和大小。

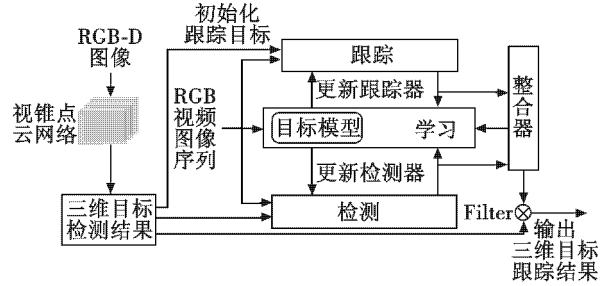


图 5 基于 TLD 的三维目标跟踪系统框架

Fig. 5 Framework of TLD-based 3D target tracking system

下面分别介绍三维目标检测网络结构和基于 TLD 的三维目标跟踪系统。

2.1 视锥点云三维目标检测网络

为了能够进行三维目标跟踪,需要根据输入的 RGB-D 图像得到目标所在的三维框位置,但是由 TLD 算法得到的是目标的二维包围框位置,而如果直接在三维空间中进行搜索建模,将大大增加时间成本,并且搜索范围的增大也会使得跟踪准确性降低。而神经网络在三维场景中有着较好的表现,因此可首先使用神经网络检测 RGB-D 场景中的目标,然后将检测结果与 TLD 算法相结合,进行三维目标跟踪,本文使用的是在三维目标检测领域表现较好的视锥点云三维目标检测网络 (FPointNet)^[12]。

视锥点云网络的结构如图 6 所示,该网络输入为 RGB-D 图像以及相机内外参数,输出为场景中目标 (本文使用的 KITTI 数据集测试,目标为汽车、行人、自行车等) 的三维框位置和大小。

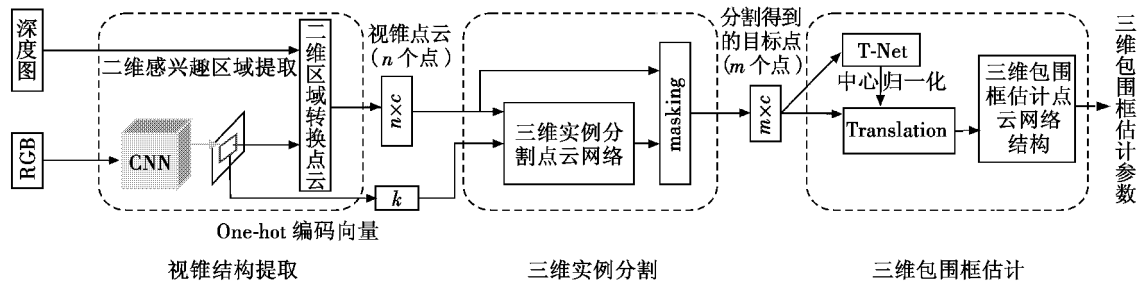


图 6 视锥点云网络结构示意图

Fig. 6 Structural diagram of frustum point cloud network

用于三维目标检测的视锥点云网络主要包含 3 个模块。第一个模块是视锥提取模块,在该模块中,首先

使用 FPN^[13] 的编码-解码器结构通过焦点损失 (focal loss)^[14] 从多层特征图中生成可能存在目标的感兴趣

区域,然后使用 Fast R-CNN^[15]网络结构从感兴趣区域中预测目标的二维包围框,最后使用已知的相机投影矩阵将深度图中二维目标检测框中的目标投影到三维空间,这样就可以得到视锥体内部所有点形成的视锥体点云图,通过将视锥进行旋转归一化使得视锥朝向中心视图,其中心轴与像平面正交(如图7(a)所示)。

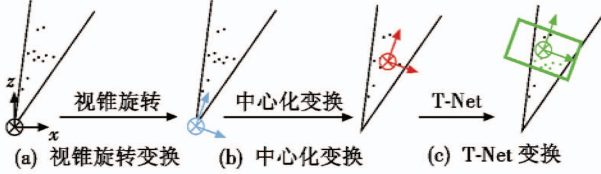


图7 视锥点云系统的坐标变换

Fig. 7 Coordinate transformation of frustum point cloud system

第二个模块是三维实例分割模块,该部分采用了文献[16]中所述的 PointNet 点云网络结构,其网络结构如图8所示。输入到网络中的是第一个模块输出的三维点云,预测视锥中每个点的概率分数,该概率分数表示三维点属于感兴趣目标的可能性,每个视锥体只包含一个感兴趣目标,目标外的三维点集合代表不相

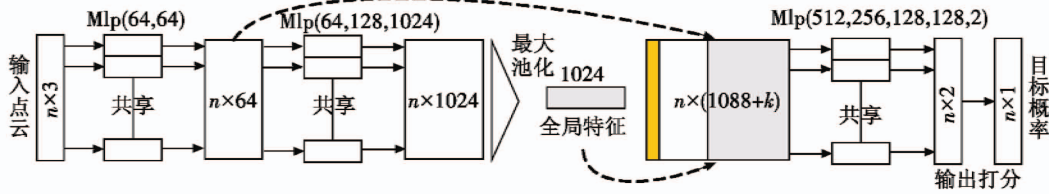


图8 三维实例分割点云网络结构

Fig. 8 Point cloud network structure after 3D instance segmentation

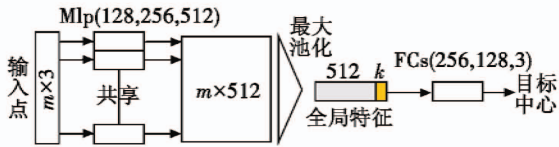


图9 估计目标中心的 T-Net 网络结构

Fig. 9 The T-Net network structure for target center estimating

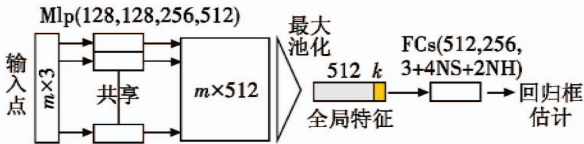


图10 三维包围框估计网络结构

Fig. 10 Network structure for 3D bounding box estimation

在训练过程中,同时优化三维实例分割网络,T-Net 以及三维框估计点云网络这3个网络,最终的损失函数表示为

$$L_{\text{multi-task}} = L_{\text{seg}} + \lambda (L_{\text{cl-reg}} + L_{\text{c2-reg}} + L_{\text{h-cls}} + L_{\text{h-reg}} + L_{\text{s-cls}} + L_{\text{s-reg}} + \gamma L_{\text{corner}}) \quad (6)$$

式中: L_{seg} 为三维分割损失函数; $L_{\text{cl-reg}}$ 为 T-Net 网络损失;

相关区域的点或者其他遮挡感兴趣目标的实例,将不同语义的感兴趣目标编码为 One-hot 编码(k 个类别使用 k 维的向量),并将其与点云特征进行关联。在三维实例分割之后,被分类的感兴趣目标点被提取,在获得这些分割的对象点之后,进一步对其坐标进行归一化以提高算法的平移不变性,通过将质心减去偏移量,将点云转换为局部坐标(如图7(b)所示)。

第三个模块是三维包围框估计模块,该部分使用轻量级的回归点云网络 T-Net 估计目标真正的中心,其网络结构如图9所示。网络输入为 m 个三维点坐标,输出为目标中心坐标,然后对该坐标进行转换使预测中心为原点,最后通过三维包围框预测网络预测三维包围框的位置和大小(如图7(c)所示),三维包围框预测网络如图10所示。网络输入为 m 个三维坐标点,输出为目标中心坐标(c_x, c_y, c_z),目标三维包围框的长宽高(h, w, l) (其中将三维包围框的大小分为 N_s 个类别)以及三维包围框的朝向角 θ (其中将三维包围框朝向角分为 N_H 个类别),以此来定位三维目标的位置和大小。

$L_{\text{c2-reg}}$ 为三维框中心回归损失函数; $L_{\text{h-cls}}$ 和 $L_{\text{h-reg}}$ 为朝向角估计损失函数, $L_{\text{s-cls}}$ 和 $L_{\text{s-reg}}$ 为三维框大小估计损失函数; L_{corner} 为边缘损失函数;所有的分类网络任务都使用 Softmax,所有的回归网络任务都使用 S_{L_1} 损失。

2.2 基于 TLD 的三维目标跟踪算法

假设根据视锥点云网络在第 t 帧中检测到 n 个目标,得到的目标集合为 $T_t = \{T_t^1, T_t^2, \dots, T_t^m\}$,其中, T_t^i 表示第 i 个目标的三维包围框,包含了三维包围框的中心坐标,包围框大小、朝向角等信息,将所有的 n 个目标映射到二维平面,得到目标在二维平面内的二维包围框 $B_t = \{B_t^1, B_t^2, \dots, B_t^n\}$ 。

在 TLD 算法的检测模块中,样本是由扫描窗口以一定的步长和缩放比例在目标扫描时形成的,在第 t 帧中,假设通过扫描窗口可以得到 m 个样本 $S_t = \{S_t^1, S_t^2, \dots, S_t^m\}$,而由于基于视锥点云网络的目标检测结果在数据集有着较好的表现,目标检测结果的置信度较高,因此在改进算法中,将 B_t 加入到 S_t 中得到 $\hat{S}_t = \{S_t^1, S_t^2, \dots, S_t^m, B_t^1, B_t^2, \dots, B_t^n\}$,形成新的检测样本集合,然后根据检测器中的级联分类器进行正负样本的分类。

通过改进的 TLD 检测器,在第 t 帧运算之后,可以得到目标在当前帧的位置以及大小,记录 TLD 跟踪结果中目标包围框的中心像素坐标 $(c_x^{\text{td}}, c_y^{\text{td}})$ 以及目标包围框的面积 s^{td} ,这样就得到在二维场景中的检测结果。为了得到三维跟踪结果,分别标记通过视锥点云网络得到的二维包围框集合 B_t 中每一个目标的中心像素坐标 $C = \{(c_x^1, c_y^1), (c_x^2, c_y^2), \dots, (c_x^n, c_y^n)\}$ 以及包围框面积 $S = \{s^1, s^2, \dots, s^n\}$,计算距离损失,即

$$L_{\text{dis}}^i = \|\vec{c}^i - c^{\text{td}}\|_2 + \lambda |s^i - s^{\text{td}}| \quad (7)$$

并根据

$$T_{\text{index}}^i = \arg \min_i \|\vec{c}^i - c^{\text{td}}\|_2 + \lambda |s^i - s^{\text{td}}| \quad (8)$$

寻找使得距离损失最小的那个目标标号 T_{index} ,从而选取 $T_{\text{index}}^{\text{index}}$ 作为在第 t 帧的目标跟踪结果。

根据以上叙述,可以得到目标的三维跟踪结果,算法的伪代码描述如下所示。

算法:基于 TLD 的三维单目标跟踪算法。

输入: N 帧 RGB-D 图像以及相应的相机内外参数。

输出:目标的三维跟踪结果。

1) 初始化视锥点云网络,并使用 KITTI 数据集^[17]对网络进行训练;

2) 在输入的第一帧图像中使用视锥点云网络得到其中的所有目标,并选取其中一个目标作为跟踪对象,将其投影到二维平面作为 TLD 算法的初始输入,令 $t = 1$;

3) $t = t + 1$,在第 t 帧图像中,利用视锥点云网络检测所有的三维目标 $T_t = \{T_t^1, T_t^2, \dots, T_t^n\}$,并将其映射到二维像平面,得到 $B_t = \{B_t^1, B_t^2, \dots, B_t^n\}$;

4) 使用 TLD 算法对第 t 帧图像进行单目标跟踪,在检测器模块,将 B_t 与原有的检测样本 S_t 相结合形成新的样本,最后得到二维目标的中心像素坐标 $(c_x^{\text{td}}, c_y^{\text{td}})$ 以及目标包围框的面积 s^{td} ;

5) 计算集合 B_t 与 TLD 目标跟踪结果之间的距离度量损失 $L_{\text{dis}}^i = \|\vec{c}^i - c^{\text{td}}\|_2 + \lambda |s^i - s^{\text{td}}|$;

6) 选择使得度量损失最小的目标 $T_{\text{index}}^{\text{index}}$ 作为第 t 帧图像的三维目标跟踪结果;

7) 如果 $t = N$,则终止,输出每一帧的三维目标跟踪结果,否则转到 3)。

3 实验结果及分析

本文在 KITTI 数据集上进行测试分析,KITTI 数据集是利用自动驾驶平台 Annieway 开发出的具有挑战性的真实计算机视觉基准,KITTI 数据集支持很多视觉任务,包括三维立体结构、光流、视觉测距、三维物体

检测、三维目标跟踪等。

首先,使用 KITTI 数据集进行三维目标检测网络的训练。KITTI 的官方训练集包含了 7481 帧图像,测试集包含了 7581 帧图像,在训练过程中,将官方训练集分别划分为 3717 帧训练集(training)和 3769 帧验证集(validation),从而使得训练集和验证集属于不同的视频序列。使用 Adam 优化器进行训练,设置初始的学习率为 0.001,并且每六万次迭代学习率减半,对于除了最后一个分类或回归之外的所有可训练层,使用批量归一化(batch normalization)进行处理。在第一模块进行二维目标提取的时候,首先在 ImageNet 和 COCO 目标检测数据集上与训练模型权重,并在 KITTI 数据集的二维目标检测数据集上进一步微调用来分类和预测二维目标检测框。

训练完成后,输入到网络中一张图片,就可以得到图片中所有目标(包括行人、汽车、自行车等)的三维包围框的中心位置以及大小朝向信息。在 KITTI 数据集中选取几帧图像用来测试,图 11 展示了三维目标检测网络得到的检测结果。从图 11 中可以看出,场景中对包括行人目标、车辆目标在内的所有目标都进行了标注。

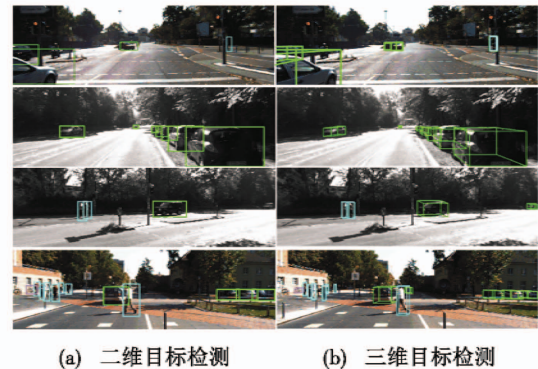


图 11 目标检测结果

Fig. 11 Target detection result

以图 11 中的第一幅图片为例,三维目标检测网络的输出如表 1 所示。其中, B_{box} 表示以像素为单位的二维包围框的左上角和右下角坐标, D 表示目标在相机坐标系中的长宽高, L 表示目标在相机坐标系中的位置, R_y 表示目标在相机坐标系中绕 y 轴的旋转角, S_{sq} 表示目标检测结果的置信度(概率分数)。

图 12 为使用基于 TLD 的三维目标跟踪算法进行目标跟踪的示例(设置参数 $\lambda = 0.01$)。图 12 中展示了 5 帧(第 2, 4, 8, 13, 18 帧)的跟踪结果,将图 11 中最后一幅图像作为视频序列的第一帧,并对该视频场景下的一辆汽车进行跟踪。

表1 目标检测网络输出信息

Table 1 Output information of the target detection network

目标	$B_{\text{box}}/\text{pixel}$	D/m	L/m	R_y/rad	S_{tg}
行人	[1001,147,1025,215]	[1.784,0.642,0.752]	[10.335,0.821,18.161]	3.586	0.293
车辆	[505,171,587,209]	[1.496,1.595,3.826]	[-2.550,1.102,29.174]	0.833	0.993
车辆	[1,203,291,366]	[1.517,1.596,3.879]	[-3.181,1.684,3.358]	-1.542	0.975
车辆	[0,183,124,275]	[1.550,1.681,4.236]	[-12.529,1.746,15.241]	1.452	0.786

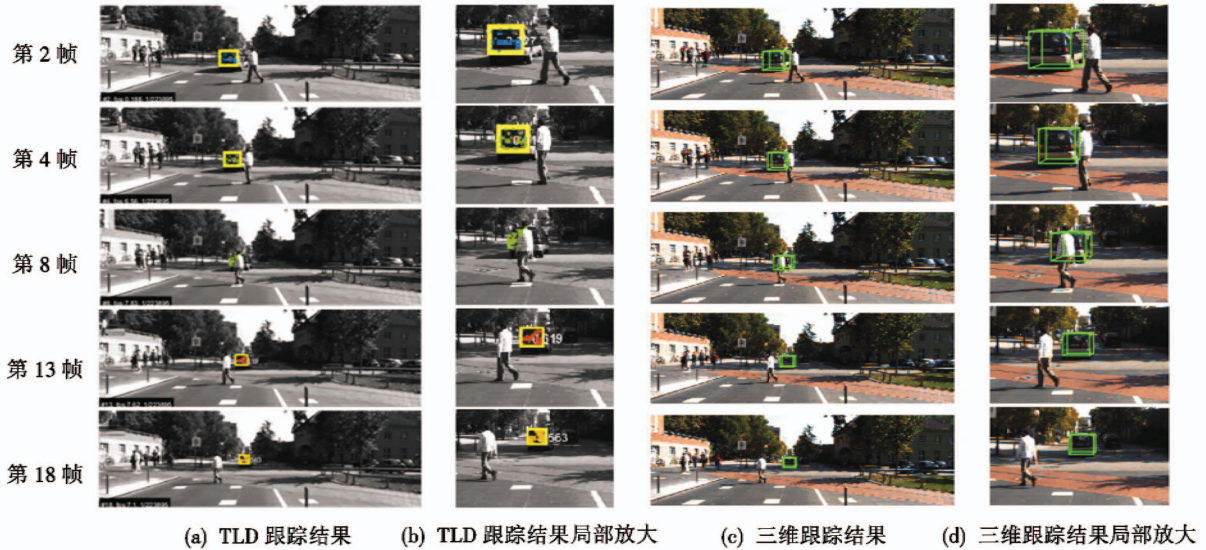


图12 基于TLD的三维单目标跟踪结果示例

Fig. 12 Example of 3D single-target tracking result based on TLD

从图12中结果可以看出,基于TLD的三维单目标跟踪算法能够在存在多目标的复杂场景中准确跟踪到标定目标,并忽略其他次要目标的影响。并且从第8帧的跟踪结果可以看出,当场景中要跟踪的目标被其他目标遮挡时(该场景中行人遮挡了要跟踪的汽车目标),TLD二维目标跟踪算法对于跟踪的目标出现跟丢的现象,而在三维场景中,虽然汽车被部分遮挡,但是根据未遮挡部分的三维结构也可以得到目标的三维检测结果,而本文提出的三维单目标跟踪算法就具备了这个特性,在这种情况下不会跟丢目标。此外,本文提出的算法能够更好地体现目标在场景中的三维结构信息,并可根据目标包围框的大小以及朝向变化来估计目标在三维场景中的运动趋势,最终的实验结果表明该算法在KITTI数据集中表现良好。

4 总结

本文提出了一种基于TLD的三维单目标跟踪算法,利用了TLD算法闭环鲁棒并且跟踪准确性优良的特点,将其扩展到三维空间,并与视锥点云网络的三维目标检测结果相结合,实现了在三维空间中的目标跟踪,在KITTI数据集下的测试结果表明了算法对于三维目标跟踪的可行性,相比于传统的TLD算法,本文算法能够较

好地处理场景中目标存在部分遮挡的情况。

参考文献

- [1] LUCAS B D, KANADE T. An iterative image registration technique with an application to stereo vision[C]//International Joint Conference on Artificial Intelligence, 1981: 674-679.
- [2] COMANICIU D, RAMESH V, MEER P. Real-time tracking of non-rigid objects using mean shift[C]//IEEE Conference on Computer Vision and Pattern Recognition, IEEE, 2000:142-149.
- [3] AVIDAN S. Support vector tracking[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2004, 26(8): 1064-1072.
- [4] BOLME D S, BEVERIDGE J R, DRAPER B A, et al. Visual object tracking using adaptive correlation filters[C]//IEEE Conference on Computer Vision and Pattern Recognition (CVPR), IEEE, 2010:2544-2550.
- [5] HENRIQUES J F, CASEIRO R, MARTINS P, et al. Exploiting the circulant structure of tracking-by-detection with kernels[C]//European Conference on Computer Vision, 2012:702-715.

- [6] HENRIQUES J F, CASEIRO R, MARTINS P, et al. High-speed tracking with kernelized correlation filters[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2015, 37(3):583-596.
- [7] KALAL Z, MIKOLAJCZYK K, MATAS J. Tracking-learning-detection[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2012, 34(7):1409-1426.
- [8] CHEN L L, ZHU S H, LI X X, et al. Target tracking via improved TLD algorithm[C]//The 27th Chinese Control and Decision Conference, 2015:5942-5947.
- [9] JIA C X, WANG Z L, WU X, et al. A tracking-learning-detection(TLD) method with local binary pattern improved[C]//Robotics and Biomimetics, 2015:1625-1630.
- [10] CHOI C, CHRISTENSEN H I. Robust 3D visual tracking using particle filtering on the special Euclidean group; a combined approach of keypoint and edge features[J]. *The International Journal of Robotics Research*, 2012, 31(4):498-519.
- [11] 王莹. 基于 Kinect 的 Tri-tracking 视频跟踪算法研究[D]. 大连:大连理工大学, 2012.
- [12] QI C R, LIU W, WU C, et al. Frustum pointnets for 3D object detection from RGB-D data[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018:918-927.
- [13] LIN T Y, DOLLÁR P, GIRSHICK R, et al. Feature pyramid networks for object detection[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017:2117-2125.
- [14] LIN T Y, GOYAL P, GIRSHICK R, et al. Focal loss for dense object detection[C]//Proceedings of the IEEE International Conference on Computer Vision, 2017:2980-2988.
- [15] GIRSHICK R. Fast R-CNN[C]//Proceedings of the IEEE International Conference on Computer Vision, 2015:1440-1448.
- [16] QI C R, SU H, MO K, et al. Pointnet: deep learning on point sets for 3D classification and segmentation[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017:652-660.
- [17] GEIGER A, LENZ P, URTASUN R. Are we ready for autonomous driving? The KITTI vision benchmark suite[C]//IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2012:3354-3361.

(上接第 9 页)

- [11] QIAN J H, HE Z S. Mainlobe interference suppression with eigenprojection algorithm and similarity constraints[J]. *Electronics Letters*, 2016, 52(3):228-230.
- [12] QIAN J H, HE Z S, JIA F D, et al. Mainlobe interference suppression in adaptive array[C]//IEEE International Conference on Signal Processing, 2016:470-474.
- [13] KHABBAZIBASMENJ A, VOROBYOV S A, HASSANIEN A. Robust adaptive beamforming based on steering vector estimation with as little as possible prior information[J]. *IEEE Transactions on Signal Processing*, 2012, 60(6):2974-2987.

(上接第 13 页)

- [10] 刘月平,姜秋喜,孙志勇. 网络雷达对 Swerling II 目标检测性能分析[J]. *电路与系统学报*, 2013, 18(1):255-262.
- [11] HE Q, BLUM R S, HAIMOVICH A M. Non-coherent MIMO radar for target estimation; more antennas means better performance[C]//Conference on Information Sciences & Systems, 2009:109-113. doi:10.1109/CISS.2009.5054700.
- [12] LIU W, LEI T. Multi-target detection algorithm for MIMO radar with arbitrarily distributed array-target configuration[C]//The 8th International Conference on Intelligent Human-Machine Systems and Cybernetics(IHMSC), 2016:541-544. doi:10.1109/IHMSC.2016.73.
- [13] ZHOU S, LIU H. Space-partition-based target detection for distributed MIMO radar[J]. *IEEE Transactions on Aerospace and Electronic Systems*, 2014, 49(4):2717-2729.
- [14] 谷文堃,王党卫,马晓岩,等. 分布式 MIMO 雷达覆盖范围分析[J]. *空军预警学院学报*, 2015, 29(6):391-395.
- [15] 费太勇,谭贤四,林强,等. 长基线统计 MIMO 雷达检测性能研究[J]. *现代雷达*, 2017(7):75-79.
- [16] 王盛利. 分布式 MIMO 雷达威力覆盖研究[J]. *中国电子科学研究院学报*, 2010, 5(6):551-555.
- [17] CHERNYAK V S. 双(多)基地雷达系统[M]. 周万幸,吴鸣亚,胡明春,等译. 北京:电子工业出版社, 2011.