

基于DETR的道路环境下双目测量系统

康轶 刘宇 王亚伟 周立君 郭城 王怡恬

DETR-based binocular measurement system in road environment

KANG Yixuan, LIU Yu, WANG Yawei, ZHOU Lijun, GUO Cheng, WANG Yitian

引用本文:

康轶, 刘宇, 王亚伟, 等. 基于DETR的道路环境下双目测量系统[J]. 应用光学, 2023, 44(4): 786–791. DOI: 10.5768/JAO202344.0402003

KANG Yixuan, LIU Yu, WANG Yawei, et al. DETR-based binocular measurement system in road environment[J]. Journal of Applied Optics, 2023, 44(4): 786–791. DOI: 10.5768/JAO202344.0402003

在线阅读 View online: <https://doi.org/10.5768/JAO202344.0402003>

您可能感兴趣的其他文章

Articles you may be interested in

基于双目视觉的大型高反光构件测量系统

Measurement system of large-scale high reflective component based on binocular vision

应用光学. 2021, 42(4): 577–585 <https://doi.org/10.5768/JAO202142.0401002>

改进Fast-RCNN的双目视觉车辆检测方法

Binocular vision vehicle detection method based on improved Fast-RCNN

应用光学. 2018, 39(6): 832–838 <https://doi.org/10.5768/JAO201839.0602001>

基于双目视觉的车身焊点定位误差补偿研究

Research on compensation for positioning errors of carbody welding points based on binocular vision

应用光学. 2021, 42(1): 79–85 <https://doi.org/10.5768/JAO202142.0102005>

硬管式无人机AAR双目视觉导航算法研究

Binocular vision navigation algorithm for AAR of flying boom UAV

应用光学. 2017, 38(6): 910–916 <https://doi.org/10.5768/JAO201738.0602002>

基于水下双目视觉的燃料组件变形检测系统

Deformation detection system of fuel assembly based on underwater binocular vision

应用光学. 2019, 40(2): 246–252 <https://doi.org/10.5768/JAO201940.0202001>

基于传像束的小视场双目结构光三维形貌恢复

3D shape restoration of binocular structure light in small field of view based on image bundles

应用光学. 2021, 42(1): 113–118 <https://doi.org/10.5768/JAO202142.0103003>



关注微信公众号, 获得更多资讯信息

文章编号: 1002-2082 (2023) 04-0786-06

基于 DETR 的道路环境下双目测量系统

康轶譔, 刘 宇, 王亚伟, 周立君, 郭 城, 王怡恬

(西安应用光学研究所 智能感知实验室, 陕西 西安 710065)

摘 要: DETR (detection transformer) 算法是一个基于 Transformer 的目标检测算法, 具有检测速度快、检测效果好的优势。介绍了一种利用 DETR 算法及双目视觉原理对道路环境下的人、车、自行车、信号灯等目标进行构建的测量系统。分析了双目测距、相机标定、目标检测以及目标匹配的原理, 并以此为基础构建了测量系统。采用目标检测算法检测视野中的目标, 利用双目视觉原理对检测到的目标进行测距, 同时分析了测量系统中测量误差的来源, 并计算其对结果的影响。该算法在 KITTI 数据集及现实环境中进行测试, 测量系统基线为 45 cm, 对 15 m~80 m 的指定目标检出率高于 90.6%, 测距误差小于 5.8%, 在 RTX 2080Ti 平台上能够实时运行。

关键词: 双目视觉; 目标检测; 测量系统; 测量误差

中图分类号: TN29

文献标志码: A

DOI: 10.5768/JAO202344.0402003

DETR-based binocular measurement system in road environment

KANG Yixuan, LIU Yu, WANG Yawei, ZHOU Lijun, GUO Cheng, WANG Yitian

(Intelligent Perception Laboratory, Xi'an Institute of Applied Optics, Xi'an 710065, China)

Abstract: Detection transformer (DETR) is a target detection algorithm based on Transformer, which has the advantages of fast detection speed and good detection effect. A measurement system based on DETR and binocular vision principle for people, vehicles, bicycles, signal lights and other targets in road environment was introduced. The principles of binocular ranging, camera calibration, target detection and target matching were analyzed, and the measurement system was constructed on this basis. The target detection algorithm was used to detect the targets in the field of vision, and the principle of binocular vision was used to measure the distance of the detected targets. The source of measurement error in the measurement system was analyzed and the influence on the results was calculated. The algorithm was tested in KITTI data set and real environment. The system baseline is 45 cm, the detection rate of 15 m~80 m specified targets is higher than 90.6%, and the ranging error is less than 5.8%. The proposed algorithm can run in real time on RTX 2080Ti platform.

Key words: binocular vision; target detection; measurement system; measurement error

引言

双目视觉测量系统是基于同一目标在双目相机中存在视差而构建的非接触式被动测量系统。在三维空间中, 同一目标在不同位置相机上的投影点存在视差, 且视差大小与目标距离成反比。通过测量双目相机对同一目标的视差, 便可计算得到目标与相机之间的距离。双目视觉测量系统可以应用于无人车导航、三维场景重建、机器人引

导、体感游戏、空间跟踪等领域。

双目视觉系统以输出待检测区域的三维信息为目标, 其结果通常展示为深度图。在双目视觉系统中, 深度图的构建包含四个过程: 一是双目图像的像素级匹配; 二是匹配后的特征点聚合; 三是根据匹配点的视差计算深度; 四是对计算的深度图优化。近些年来, 相继提出许多双目立体视觉匹配算法。这种深度图生成算法是基于像素点或

收稿日期: 2022-05-19; 修回日期: 2022-06-20

基金项目: 十四五陆军预研基金 (627010402)

作者简介: 康轶譔 (1995—), 男, 硕士, 助理工程师, 主要从事深度学习方面的研究。E-mail: kangyixuan13@163.com

极小的像素块进行计算, 所以其结果能够覆盖全图, 从而获得被测空间中的全部三维信息^[1-2]。结合深度神经网络的双目视觉立体匹配算法在近年被提出, 其将深度图构建中的四个过程采用深度学习的方法实现^[3-7]。无论是传统算法或是基于深度学习的算法, 在面对一些特殊情况时都会出现由于匹配错误导致的视差计算错误, 例如无纹理区域、重复纹理区域、光滑表面的镜面反射区域、光学失真区域等。并且基于深度学习的立体视觉匹配模型在处理新环境时, 准确性会显著下降^[8-9]。

在观察图像时, 通常只有一些特定的区域或目标能够引起观察者的兴趣。在道路环境下的测量系统中, 只期望测量特定目标的距离信息, 如人、车辆、信号灯、斑马线等。无需在整幅图像的所有像素点上计算视差信息, 因此能够在很大程度上避免因匹配错误而导致的计算错误。在夜间低照度环境中, 由于驾驶员视线不佳以及疲劳等问题, 使用低照度 CMOS 摄像机结合双目视觉测量系统能够起到驾驶辅助作用。

道路环境下的测量系统主要有三类: 一是基于双目和激光点云的测量系统, 其特点是精度高但造价贵; 二是基于双目视觉的测量系统, 相对第一种方法成本较低, 但会在精度上有一定损失; 三是基于单目视觉的测量系统, 主要依靠深度学习对目标距离进行推测, 成本更低, 但对算力要求较高。本文将描述一种基于目标的双目视觉测量系统, 能够实时计算特定目标的距离信息。

在双目视觉特征匹配中常用的是点特征, 如 Harris 角点、SIFT 特征点、SURF 特征点、ORB 特征点等。在传统的目标检测方法中, 特征点信息可以被用来确定目标的分类信息。近年来, 基于深度学习的目标检测展现了巨大实力, 能很好地应对目标的光照变化、形变、遮挡等。基于深度学习的算法可以分成两大类: 一种是基于区域提案的方法; 另一种是端到端的回归检测方法。其中端到端的回归检测方法如 Faster RCNN 算法、YOLO 算法、Focus 算法等均可在常见的计算机平台上达到实时目标检测。近年来, 提出了一些基于 Transformer 的端到端目标检测算法, 省去划分锚框 (Anchor) 和非极大化抑制 (NMS) 处理步骤, 如 DeiT、Swin Transformer 算法等, 并且这些算法在 COCO 数据集上能够取得与 Faster RCNN 相当的结果^[10-12]。依托于 TensorFlow、Pytorch、TensorRT 等深度学习推理框架, 可以在常见计算机平台上

部署目标检测算法, 这得以基于目标的双目视觉测量系统拥有很大的潜力^[13]。相对于基于像素点的双目测量系统, 基于目标的双目测量系统的测量更有针对性, 并能避免大量的匹配错误, 同时也能完成目标距离的实时计算。

1 双目视觉测量系统

图 1 是一个双目视觉测量系统。两个相机的心为 O_L 和 O_R , 摆放到理想位置, 即光轴 P_L 和 P_R 相互平行。两个相机焦距均为 f , 基线长度为 B , 同时观测目标 T , 则视差计算公式为

$$d_{\text{isparty}} = x - x' = Bf/Z \quad (1)$$

式中: x 和 x' 表示目标在两个相机焦平面上投影点坐标, 对应图 1 中 X_L 和 X_R 表示目标在两个相机焦平面上投影点距焦平面中心的距离, 满足 $x - x' = X_L + X_R$; Z 表示目标与双目相机基线之间的距离; 视差表示目标 T 在双目相机焦平面上投影点的距离差。根据式 (1), 目标到相机中心的距离与该目标投影点的视差成反比。

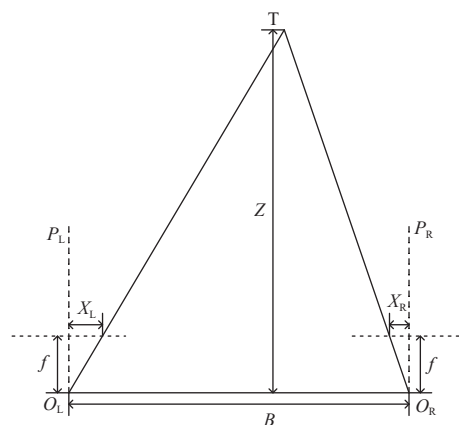


图 1 双目视觉测量系统图

Fig. 1 Diagram of binocular vision measurement system

在实际使用过程中, 如果无法将相机摆放到理想的位置关系, 则需要进行相机标定, 即求解双目相机系统的内参和外参。准确计算双目相机系统内参和外参的大小, 对于准确计算视差值具有重要意义, 图 2 表示目标 T 通过摄像系统投影为像点 Q 的过程。通过相机坐标系 c_{xyz} , 将坐标从三维世界坐标系 w_{xyz} 转化到二维像面坐标系 i_{oxy} , 计算公式为

$$x = PX = K[R|T]X \quad (2)$$

式中: x 表示目标在焦平面内的坐标 (x_i, y_i) ; X 表示目标在三维空间内的坐标 (x_w, y_w, z_w) ; P 表示相机

参数, 相机参数由内参和外参组成; K 表示相机系统内参; $[R|T]$ 表示相机系统外参。

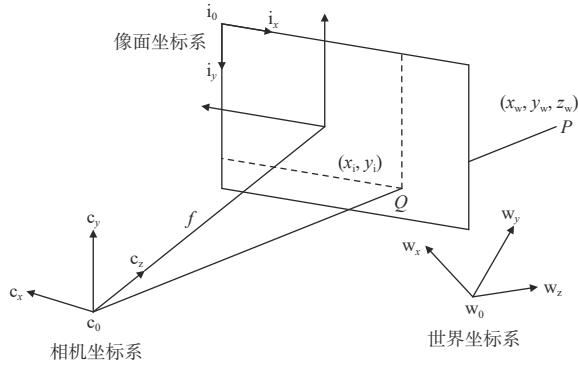


图 2 像面坐标系与真实坐标系的转化图

Fig. 2 Transformation diagram of image plane coordinate system and real coordinate system

像面坐标系到相机坐标系的转化, 可利用相机内参计算, 内参矩阵可表示为

$$K = \begin{bmatrix} f_x & 0 & c_x \\ 0 & f_y & c_y \\ 0 & 0 & 1 \end{bmatrix} \quad (3)$$

式中: f_x 和 f_y 分别表示相机在 x 方向和 y 方向上的尺度因子; c_x 和 c_y 表示相机主点。

相机坐标系到世界坐标系的转化, 可利用相机外参计算, 包括正交旋转矩阵和平移矩阵:

$$R = \begin{bmatrix} r_{11} & r_{12} & r_{13} \\ r_{21} & r_{22} & r_{23} \\ r_{31} & r_{32} & r_{33} \end{bmatrix} \quad (4)$$

$$T = [t_1 \quad t_2 \quad t_3]^T \quad (5)$$

式中: R 为正交旋转矩阵, 代表了相机坐标系相对于世界坐标系的旋转关系; T 为平移矩阵, 代表了相机坐标系相对于世界坐标系的平移关系。

相机成像系统中的透镜会在成像系统中引入一定的误差。一是由于透镜自身曲面形状造成的成像径向畸变, 对于一些广角镜头来说, 光线在远离透镜中心的地方比在靠近中心的地方更加弯曲。由透镜形状引起的径向畸变可表示为

$$x_{\text{correct}} = x(1 + k_1 r^2 + k_2 r^4 + k_3 r^6) \quad (6)$$

$$y_{\text{correct}} = y(1 + k_1 r^2 + k_2 r^4 + k_3 r^6) \quad (7)$$

式中: k_1 、 k_2 、 k_3 为径向畸变参数。

二是由于相机感光平面与焦平面的不重合导致的成像切向畸变。由相机感光平面安装误差引起的切向畸变可表示为

$$x_{\text{correct}} = x + 2p_1 xy + p_2(r^2 + 2x^2) \quad (8)$$

$$y_{\text{correct}} = y + 2p_2 xy + p_1(r^2 + 2y^2) \quad (9)$$

式中: p_1 、 p_2 为切向畸变参数。

相机的标定方法包括传统的 ZHANG Z Y 标定法和 Tsai 标定法^[14]。张正友标定法使用一张打印出的棋盘格从不同角度拍摄几组照片, 然后通过公式求解出相机的内参、外参、径向畸变和切向畸变参数。本文采用张正友标定法, 棋盘格数 20×20 , 尺寸 $200 \text{ mm} \times 200 \text{ mm}$ 。对于单个相机而言, 根据 (3) 式~(9) 式, 需标定相机内参矩阵 (包含 4 个参数)、正交旋转矩阵和平移矩阵 (包含 12 个参数), 以及畸变系数 (包含 5 个参数)。

2 双目图像中的目标定位

目标检测是在图片中精确地找到目标位置、并标注目标类别的过程, 以输出目标边界框的坐标信息和目标类别概率信息为目标。2012 年以来, 卷积神经网络开始进入高速发展状态, 区别于传统的以模板为基础的特征提取方法, 卷积神经网络中提取特征参数是经过数据集训练后生成的。正是这种卷积核参数由训练所生成的方式, 使得卷积核参数可以不受特定卷积模板的限制, 从而获得了更好的效果。相对于传统的目标检测手段, 使用深度学习的目标检测算法在准确度上得到了大幅的提升, 如 DETR 算法在 COCO 数据集的大、中、小目标上达到了 61.1、45.8、20.5 的 AP 值^[15]。这使得单幅图像上的目标检测精度有了大幅的提升, 从而为基于目标的双目测量系统提供了基础。

DETR 算法是第一个成功将 Transformer 整合到目标检测框架中的算法, 如图 3 所示。DETR 算法取消了原有目标检测框架中的先验知识, 如 Anchor、NMS 等, 在推理精度相当的情况下, DETR 相较于 Faster RCNN 拥有更快的推理速度。DETR 使用 CNN 网络提取目标的特征信息, 接着将目标特征信息输入一个 Transformer 模块 (包含编码器-解码器层) 中, 最后通过前馈网络预测目标的标准中心坐标以及宽高信息。



图 3 DETR 算法结构图

Fig. 3 Structure diagram of DETR algorithm

对于双目摄像系统而言, 在每个时刻都会获取到两张待处理图片。如果对获得的两路图像分别进行处理, 会对目标检测的实时性带来极大挑战。由于双目成像系统的两路图像获取的是同一

区域、同一时刻在不同方向上的成像信息, 在两组相机光轴夹角较小且目标距离显著大于基线距离的情况下, 同一目标由于角度不同, 在两路图像中产生的目标形态变化较小, 即两幅图像中的同一目标具有较高的相关性。本文提出一种方法, 将第一路图像输入目标检测网络中获得目标位置信息, 并以第一路图像中目标的位置信息为参考, 计算得到第二路图像中目标的位置信息。

将左侧摄像机作为第一路图像输入, 通过 DETR 算法获得目标 K 的位置信息 P_{KL} 。该目标在第二路图像中位置信息为 P_{KR} , 其中 P_{KL} 和 P_{KR} 均为标定后理想双目成像状态下的位置坐标值。由于两个相机的位置关系, P_{KR} 一定在 P_{KL} 正左侧的一定范围内。即在第二路图像上以 P_{KL} 作为初始位置信息, 向左侧寻找便可获得第二路图像中的位置信息 P_{KR} 。两路图像中的位置差 $P_{KL}-P_{KR}$ 即为像素视差的大小。

为提高搜索效率, 仅在第二路图像的候选框 R_K 中进行搜索, 候选框的尺寸满足以下条件:

$$(x_{P_K}, y_{P_K}) - (x_{R_K}, y_{R_K}) = (r_K, 0) \quad (10)$$

$$(w_{P_K}, h_{P_K}) - (w_{R_K}, h_{R_K}) = (2r_K, h_K) \quad (11)$$

式中: x 、 y 表示候选框的中心点坐标; w 、 h 表示候选框的宽高值; 其中 r_K 为目标在双目相机中沿基线方向的最大像素视差值; h_K 为目标在双目相机垂直基线方向上的最大像素视差值。如果相机标定结果足够好, h_K 的值应为 0。考虑实际的运行环境, 可以适当增加 h_K 的值以确保候选框 R_K 能完整地包含目标。

在候选框中, 采用图像相关的方法寻找对应目标。考虑到较近的目标尺度较大, 进行卷积运算时耗时较多, 可以采用图像金字塔方式, 从小尺度到大尺度不断优化查找。经过测试, 可使小尺度缩放图的像素面积在 16×16 像素到 32×32 像素之间进行第一轮寻找, 接着以 2 的倍数不断增大缩放图面积, 以达到最优搜索结果。

3 测量精度分析

目标 T 的视差计算公式如(1)式所示, 则目标距离 Z 的计算公式为

$$Z = \frac{Bf}{(x-x')} = \frac{B}{2 \tan \frac{\theta}{2} x_{\text{pixel}}} P_{\text{ixelx}} \quad (12)$$

式中: B 表示基线长度; θ 表示相机视场角; x_{pixel} 表示目标在双目图像水平方向上的像素视差; P_{ixelx}

表示相机的水平分辨率。将 Z 对 x_{pixel} 求导, 得到:

$$\frac{\partial Z}{\partial x_{\text{pixel}}} = -\frac{B}{2 \tan \frac{\theta}{2} x_{\text{pixel}}^2} P_{\text{ixelx}} \quad (13)$$

由双目图像上像素视差的误差引起的目标测距误差可表示为

$$\frac{dZ}{Z} = -\frac{2 \tan \frac{\theta}{2} Z}{B P_{\text{ixelx}}} d(x_{\text{pixel}}) \quad (14)$$

式中: dx_{pixel} 表示目标在双目图像上像素视差的误差, 从(14)式可得, 由双目图像上像素视差的误差引起的目标测距误差与目标实际距离成正比。

同理可得, 由基线测量误差和视场角测量误差引起的目标测距误差可分别表示为

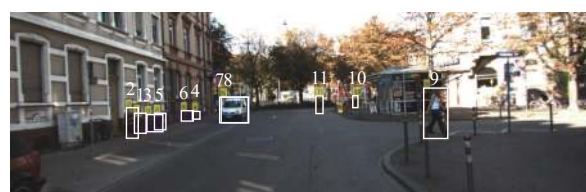
$$\frac{dZ}{Z} = \frac{1}{B} dB \quad (15)$$

$$\frac{dZ}{Z} = \frac{1}{2 \tan \frac{\theta}{2}} d\theta \quad (16)$$

式中: dB 和 $d\theta$ 分别表示基线和视场角的测量误差, 由基线测量误差和视场角测量误差引起的目标测距误差与目标实际距离没有关系, 仅是系统误差, 并可以通过多组实验进行消除^[16]。

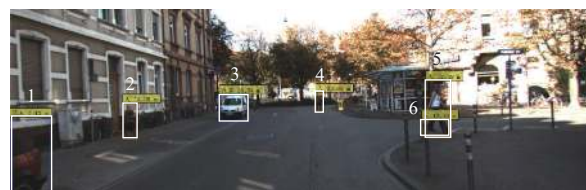
4 实验结果

图 4 表示了在 KITTI 数据集上取得的测距效果, 其中图 4(a)表示图像标签真值, 图 4(b)表示测距效果^[17]。



1: 自行车: 91% 14.78 m; 2: 人: 97% 16.24 m; 3: 自行车: 96% 18.74 m;
4: 自行车: 92% 25.82 m; 5: 自行车: 95% 20.06 m; 6: 自行车: 90% 24.91 m;
7: 卡车: 95% 24.65 m; 8: 汽车: 98% 24.74 m; 9: 人: 100% 11.49 m;
10: 人: 83% 42.17 m; 11: 人: 100% 31.83 m.

(a) 图像标签真值



1: 汽车: 4.13 m; 2: 人: 15.95 m; 3: 汽车: 26.64 m;
4: 人: 34.08 m; 5: 人: 12.42 m; 6: 人: 13.43 m.

(b) 测距效果图

图 4 KITTI 数据集测试结果

Fig. 4 Test results of KITTI dataset

在计算视差时,由于数据集中的左右摄像头获取的图像已经完成校正,所以可以直接利用数据集给出的基线长度进行计算。

表 1 表示了双目视觉测量系统在 KITTI 数据集上的检测结果,包含检出目标数、成功测距数、测距正确数以及平均测距误差,测距误差定义为测距距离差值与实际距离之间的比值。对于近距离的目标,算法得到的距离值与真值间存在较大误差,主要原因是由于目标在三维空间内存在一定体积,所以本文的测距基准点与数据集的定义有一定差异。一般车身长度为 3 m,对于 10 m 左右的车辆,测距基准点的不同会对测距结果产生影响。

表 1 双目视觉测量系统检测结果分析

Table 1 Detection results analysis of binocular vision measurement system

真实目标 距离/m	目标数	检出 目标数	成功 测距数	测距 正确数	平均测距 误差/%
5~15	7 308	7 108	6 845	6 257	7.9
15~25	7 464	7 014	6 727	6 219	5.1
25~40	7 836	7 175	6 934	6 626	4.8
40~80	4 214	3 820	3 674	3 591	5.8
80~120	2 127	1 825	1 751	1 693	6.4

表 2 为本文算法与其他双目测量算法在 KITTI 数据集上对 15 m~80 m 目标的测量性能进行的比较。本文算法相对于其他双目深度算法,在速度上有较大优势,检出率相当;相对于单目算法,在测距误差上有较大优势,能够达到其他双目算法的测距误差,即本文算法在针对道路目标的测量方面优于其他算法。

表 2 本文算法与其它算法对 15 m~80 m 目标的测量结果比较

Table 2 Detection results comparison of 15 m~80 m targets with proposed algorithm and other algorithms

算法	检出率/%	测距误差/%	运行帧率/(帧/s)
本文算法	92.3	5.2	21
GC-Net	93.1	5.9	1.1
GANet	93.5	5.2	10.2
PSMNet	93.9	4.8	2.4
Monodepth	92.9	6.9	15

在实际环境下进行测试,相机分辨率为 1 920×1 080,相机视场角为 60°,基线长度为 18 cm。设置

最近可测距距离为 2 m,根据式(11)计算得出 r_K 最小为 200。在测试中设置 r_K 为 200, h_K 为 2。对于目标在双目图像的定位,取 1 个像素点的误差作为像素视差误差,理论上能够达到的最远测距距离为 300 m。但根据式(14),在距离较远时,双目测距精度极大降低,实验中取双目测量系统的实际测距范围为 2 m~40 m。通过式(14)计算得到目标测距误差为

$$\frac{dZ}{Z} = -0.333 Z \quad (17)$$

即对于距离为 2 m 的目标,其测距误差为 0.66%;对于距离为 40 m 的目标,其测距误差为 13.2%(对应距离误差为 5.2 m)。图 5 为双目测量系统测试效果图,表 3 为测试结果分析。



图 5 双目测量系统测试效果

Fig. 5 Detection effect of binocular measurement system

表 3 双目测量系统测试结果分析

Table 3 Detection results analysis of binocular measurement system

目标序号	目标真实距离/m	输出概率值	测距值/m	测距误差/%
1	18.94	0.99	18.25	3.64
2	19.81	0.98	18.95	4.34
3	21.18	1.00	20.82	1.70
4	22.56	0.99	21.92	2.84

目标检测算法在 RTX 2080Ti 上运行,单张图片的检测时间为 42.13 ms。整体算法对于单帧图像处理时间为 47.96 ms,检测帧速为 21 帧/s,能达到实时测量。

5 结论

文中基于 DETR 的双目测量系统能够高效地完成一定区间范围内的被动式测距。该系统的测距范围与基线长度成反比,测距误差与基线长度

成反比, 通过调整基线长度能构建出适用于不同区间的测距系统。本文的双目测量系统相对于基于深度计算的方法, 能够很好地避免双目图像在匹配过程中带来的误匹配问题, 在针对道路环境下人、车、自行车、信号灯等常见目标的检测准确度方面有优势。相对于基于深度学习的双目计算方法, 本文方法在运算速度上有较大优势, 且能在主流显卡上实时运行。

参考文献:

- [1] NIE G Y, CHENG M M, LIU Y, et al. Multi-level context ultra-aggregation for stereo matching[C]//2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Long Beach: IEEE, 2020: 3278-3286.
- [2] DENG H, LIAO Q M, LU Z Q, et al. Parallax contextual representations for stereo matching[C]//2021 IEEE International Conference on Image Processing (ICIP). Anchorage: IEEE, 2021: 3193-3197.
- [3] ZBONTAR J, LECUN Y. Stereo matching by training a convolutional neural network to compare image patches[J]. *Journal of Machine Learning Research*, 2016, 17: 2287-2318.
- [4] PANG J H, SUN W X, REN J S, et al. Cascade residual learning: a two-stage convolutional neural network for stereo matching[C]//2017 IEEE International Conference on Computer Vision Workshops (ICCVW). Venice: IEEE, 2018: 878-886.
- [5] KHAMIS S, FANELLO S, RHEMANN C, et al. StereoNet: guided hierarchical refinement for real-time edge-aware depth prediction[M]//Computer Vision-ECCV 2018. Munich: Springer International Publishing, 2018: 596-613.
- [6] CHANG J R, CHEN Y S. Pyramid stereo matching network[C]//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City: IEEE, 2018: 5410-5418.
- [7] XU H F, ZHANG J Y. AANet: adaptive aggregation network for efficient stereo matching[C]//2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Seattle: IEEE, 2020: 1956-1965.
- [8] LIANG Z F, FENG Y L, GUO Y L, et al. Learning for disparity estimation through feature constancy[C]//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City: IEEE, 2018: 2811-2820.
- [9] POGGI M, PALLOTTI D, TOSI F, et al. Guided stereo matching[C]//2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Long Beach: IEEE, 2020: 979-988.
- [10] ASHISH V, NOAM S, NIKI P, et al. Attention is all you need[C]//2017 Conference on Neural Information Processing Systems. Long Beach: IEEE, 2017: 3058-3068.
- [11] TOUVRON H, CORD M, DOUZE M, et al. Training data-efficient image transformers and distillation through attention[EB/OL]. (2021-01-15) [2022-03-15]. <https://arxiv.org/abs/2012.12877>.
- [12] LIU Z, LIN Y T, CAO Y, et al. Swin transformer: hierarchical vision transformer using shifted windows[C]//2021 IEEE/CVF International Conference on Computer Vision (ICCV). Montreal: IEEE, 2022: 9992-10002.
- [13] 张琦, 胡广地, 李雨生, 等. 改进Fast-RCNN的双目视觉车辆检测方法[J]. *应用光学*, 2018, 39(6): 832-838.
ZHANG Qi, HU Guangdi, LI Yusheng, et al. Binocular vision vehicle detection method based on improved Fast-RCNN[J]. *Journal of Applied Optics*, 2018, 39(6): 832-838.
- [14] ZHANG Z Y. A flexible new technique for camera calibration[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2000, 22(11): 1330-1334.
- [15] CARION N, MASSA F, SYNNAEVE G, et al. End-to-end object detection with transformers[M]//Computer Vision-ECCV 2020. Berlin: Springer International Publishing, 2020: 213-229.
- [16] 崔恩坤, 滕艳青, 刘佳伟. 立体视觉测量系统标定误差补偿[J]. *应用光学*, 2020, 41(6): 1174-1180.
CUI Enkun, TENG Yanqing, LIU Jiawei. Calibration error compensation technique of stereoscopic vision measurement system[J]. *Journal of Applied Optics*, 2020, 41(6): 1174-1180.
- [17] GEIGER A, LENZ P, URTASUN R. Are we ready for autonomous driving? The KITTI vision benchmark suite[C]//2012 IEEE Conference on Computer Vision and Pattern Recognition. Providence: IEEE, 2012: 3354-3361.