

多类复杂水果目标的通用三维感知框架*

程佳兵¹ 邹湘军² 陈明猷² 谭康裕¹ 吴烽云²

(1.广州南方学院电气与计算机工程学院, 广东 广州 510970

2.华南农业大学工程学院, 广东 广州 510642)

摘要: 相比于理想的实验室环境, 在户外果园采摘的水果图像受更多复杂因素的干扰, 如背景混杂、枝叶遮挡、光照不均和机械振动等, 是制约视觉水果采摘机器人进一步发展和应用的关键因素。为此, 提出一种适应于复杂采摘环境的通用三维感知框架。首先, 构建并标定双目视觉系统; 然后, 借助深度目标检测网络去除复杂背景并获得水果所在的图像区域; 最后, 对水果区域进行立体匹配和三角测量, 得到水果的三维点云与空间位置。该框架不依赖任何人工设计, 可提高机器人在户外感知水果的准确度和稳定性, 也可在不同的采摘任务中集成, 为提高视觉水果采摘机器人感知的实用性提供有效的理论依据与技术支持。

关键词: 采摘机器人; 机器视觉; 目标检测; 立体匹配; 三角测量

中图分类号: S225; TP391.41

文献标识码: A

文章编号: 1674-2605(2021)03-0003-06

DOI: 10.3969/j.issn.1674-2605.2021.03.003

0 引言

视觉系统可帮助水果采摘机器人识别和定位果实, 进而控制末端执行器自动采摘水果, 是智能采摘机器人感知外界的重要模块之一。

目前, 大多数基于视觉的水果采摘系统仅实现目标果实的识别、检测等单个任务^[1-6]。虽取得较好效果, 但主要聚焦于对单一功能模块的研究(如单独针对图像识别、三维重构或采摘路径规划等), 未能给出一整套综合性的视觉感知流程。要实现水果采摘机器人的落地应用, 不但要考虑每一个功能模块如何实现, 同时也要从整体层面考虑技术框架的实现。

目前, 已有学者对视觉感知流程进行初步研究, 如基于双目立体视觉获得目标果实图像中的位置, 并计算果实的三维空间坐标^[7-8]、基于深度相机和颜色几何特征, 快速定位和重构水果结构^[9-11]、通过构建机器人视觉系统实现完整的采摘流程^[12-14]等。

随着计算机性能的提升, 依赖较大计算资源的深度学习被运用于水果三维感知领域, 提高图像识别准确率, 使采摘任务中的视觉感知流程更加稳定和成熟。如结合卷积神经网络和支持向量机的图像识别与三维定位方案^[15]、基于改进的目标检测网络对细小

目标果实三维定位方法^[16]、基于语义分割和立体视觉的番石榴快速识别与机械臂避障方法^[17-18]、基于深度学习方法与多机械臂协同的奇异果收获机器人等^[19]。

上述研究展示了视觉系统在水果采摘领域的成功应用, 具有重要的参考价值。然而, 其中大部分视觉系统尚处于起步阶段, 所涉及的传统方法依赖人工选取特征以及设定阈值, 限制其在复杂采样环境下的稳定性和可靠性; 而且上述研究大多针对单一水果目标, 缺乏通用性和可移植性。为此, 本文提出一种针对多类复杂水果目标的通用三维感知框架, 借助深度目标检测网络的鲁棒性提升立体视觉模块的性能, 为视觉水果采摘机器人的视觉感知提供一种稳定、自适应和实用的技术方案。

1 原理和方法

1.1 目标检测

目标检测属于二维图像处理范畴, 其作用是获取水果在图像中的区域。目标检测可去除非结构化果园中的复杂背景, 突出水果所在位置, 为后续的立体视觉算法提供良好输入, 是立体视觉预处理环节。

本框架利用 EfficientDet 网络实现水果目标区域

* 基金项目: 广东省科技计划项目(2019A050510035)

的精确检测^[20]。EfficientDet 是目标检测领域新一代的 高性能深度网络，其结构如图 1 所示。

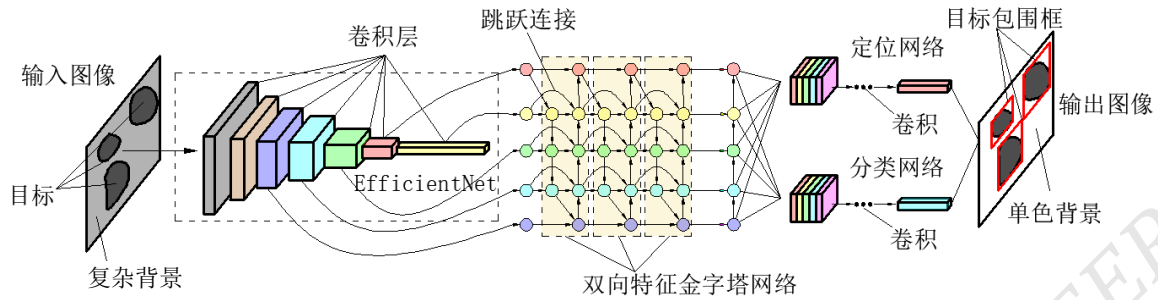


图 1 EfficientDet 网络结构示意图

EfficientNet 作为基础网络，后续连接一个双向特征金字塔网络以提取多尺度特征^[21]。EfficientDet 网络精简了双向特征金字塔网络，并采用跳跃链接和可学习权重，提高多特征融合的效率^[22]。该网络通过混合缩放方法自动确定自身的最优规模，在低资源供给的条件下同时达到高速度和高精度，这在环境复杂且计算资源受限的果园采摘任务中尤为重要。

1.2 立体视觉

1.2.1 双目标定与采样

构建双目视觉系统，对复杂非结构化环境下的多种水果进行采样，进而计算其空间形态和位置，为机器人采摘提供决策支持。

双目标定是立体视觉的前置工作，其目的是获取三角测量所需的重投影矩阵以及实现外极线校准。标定过程使用 C++ 接口的 OpenCV4.1.1 计算机视觉开源库。相机标定是机器视觉领域的常规操作，具体实现在此不赘述。标定所得的双目相机参数如表 1 所示。

表 1 双目视觉系统标定参数

	左相机	右相机
径向畸变参数	$(-0.111461 \ 0.139874)^T$	$(-0.119176 \ 0.169546)^T$
内参数矩阵	$\begin{pmatrix} 1100.45 & 0 & 976.968 \\ 0 & 1099.39 & 536.932 \\ 0 & 0 & 1 \end{pmatrix}$	$\begin{pmatrix} 1100.45 & 0 & 970.301 \\ 0 & 1099.39 & 538.069 \\ 0 & 0 & 1 \end{pmatrix}$
重投影矩阵	$\begin{pmatrix} 1 & 0 & 0 & -981.415 \\ 0 & 1 & 0 & -537.231 \\ 0 & 0 & 0 & 1025.05 \\ 0 & 0 & 0.00830239 & 0 \end{pmatrix}$	

根据标定参数进行外极线校准，可保证同一目标在左右图像上的投影满足行对齐，这可降低后续双目立体匹配的难度。校准后的双目图像如图 2 所示。任

意从中选取 A, B, C 3 个特征点，可见行对齐的效果明显。

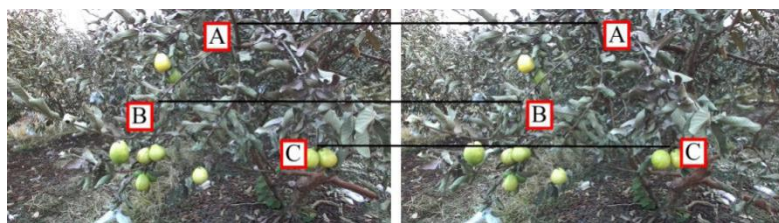


图 2 外极线校准后的双目图像

1.2.2 立体匹配和三角测量

立体匹配的目的是获取同一空间点在左右图像中的视差。遍历整张图像并计算每一个像素对应的视差后，可得到稠密视差图。稠密视差图中的每一个像素可通过三角测量原理转化成现实世界的三维点：

$$\begin{bmatrix} X \\ Y \\ Z \\ W \end{bmatrix} = \mathbf{Q} \begin{bmatrix} x \\ y \\ d \\ 1 \end{bmatrix} \quad (1)$$

式中， x 和 y 分别为该点在左图像中的像素坐标； d 为稠密视差图中该点对应的视差； $\mathbf{Q}$ 为相机标定时获取的重投影矩阵； $[X \ Y \ Z \ W]^T$ 为三维点的齐次坐标。

通过双目视觉系统实现三维感知的关键是得到准确的视差图。实际上，立体匹配的复杂性是制约双目视觉系统性能的最重要因素之一。果园环境比结构化环境存在更多的遮挡、噪声和空间多义性，因此果园环境下的立体匹配面临更大困难，通用性和稳定性难以得到保证^[23]。

本文借助深度目标检测网络的鲁棒性解决上述问题。水果稠密视差图获取过程如图3所示。首先，利用 EfficientDet 网络检测左右图像中的水果区域，根据区域信息去除复杂背景；然后，将只含水果的图像作为立体匹配输入，可得到准确而简洁的稠密视差图，从而降低立体匹配的复杂度。该过程使用的立体匹配算法为经典的半全局立体匹配算法^[24]。

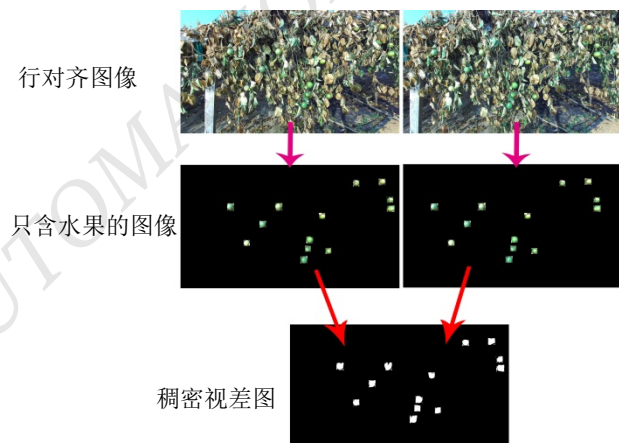


图3 水果稠密视差图获取过程

除了水果本身，枝叶障碍物的空间信息也是采摘时需要考虑的因素。与匹配水果的过程相反，对只包含枝叶的背景部分进行立体匹配。相比于水果，采摘任务对枝叶障碍物的精度需求相对较低。为降低立体匹配计算资源的消耗，将该类图像的长和宽下采样为原来分辨率的一半（960×540）后，再进行立体匹配，最后将所得的视差图通过二次线性插值复原到原分辨率（1920×1080），具体过程如图4所示。

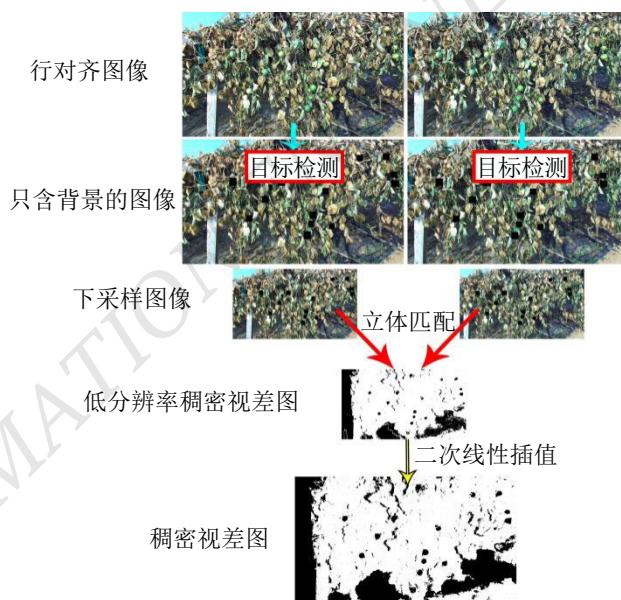


图4 枝叶障碍稠密视差图获取过程

需要特别指出的是：图3和图4所展示的视差图已被二值化，用以达到最佳的可视化效果。实际上，真正用于三角测量的原视差图是渐变的灰色图像，上面的每一个像素对应不同的灰度值。

将图3和图4所述的2种稠密视差图进行叠加，通过式(1)可计算包含枝叶障碍物在内的高精度水果点云，如图5所示。

由图5可观察到：点云中的水果结构紧凑，表面厚实、光滑；背景点云虽可大致表征枝叶的空间位置，但表面相对粗糙，这是因为背景部分对应的视差图经过下采样后丢失了一定的细节信息。将水果与背景分开进行立体匹配的方法在保证计算效率的同时，实现视觉系统对采摘环境的整体和全方位感知。



图5 三维点云示意图

2 实验与分析

2.1 方案与设备

实验包括目标检测网络性能评估、三维点云精度评估 2 部分。通过计算所训练网络的平均精度均值 (mean Average Precision, mAP), 定量分析所训练网络在处理不同种类水果时的目标检测网络性能; 通过提取水果点云的特定几何尺寸并将其与人工测量值对比, 计算双目视觉系统的测量误差, 以此反映视觉系统的三维感知性能。

实验主要设备包括一体式双目相机、三脚架、高精度棋盘格标定板 (精度为 0.3 mm)、笔记本电脑、数显游标卡尺。相机型号为 ZED2, 分辨率为 1920×1080, 帧率为 30 帧/s, 曝光时间和增益均设为自动。笔记本电脑的 CPU 型号为 Intel core i5-9400, GPU 型号为 Nvidia GTX 1660Ti 6 GB, 内存容量为 16 GB ddr4, 操作系统为 windows 10 1803 专业版。

2.2 目标检测网络性能评估

样本库含有香蕉、菠萝、木瓜和荔枝 4 类水果图像, 每类水果图像的数量均为 250 张。样本被随机打乱后按照 8: 1: 1 的比例被划分为训练集、验证集和测试集, 分别有 800 张、100 张、100 张。训练集用于训练网络, 验证集用于监测模型的训练状态, 测试集用于验证模型的性能。模型通过 Pytorch 训练。训练完成后, 对测试集进行前向传播运算, 得到水果目标包围框; 然后, 绘制每一类水果对应的精度-召回率(P-R)曲线, 并利用式(2)计算每类水果对应的平均精度(average precision, AP)。

$$AP = \int_0^1 f_j(r)dr \quad (2)$$

采用式(3)计算 4 类水果的 mAP:

$$mAP = \frac{\sum_{j=1}^{N_k} \int_0^1 f_j(r)dr}{N_k} \quad (3)$$

式中, $N_k=4$ 为类别数; $f_j(r)$ 为第 j 类对应的 P-R 曲线。各类水果对应的 AP 及总体样本的 mAP 如图 6 所示。

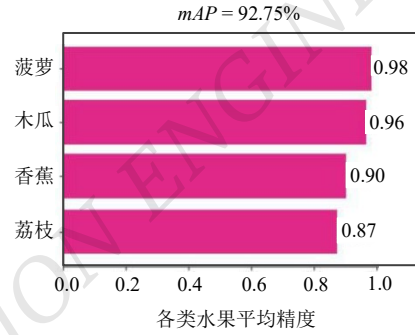


图6 4 类水果的 AP 统计值

由图 6 可知, 4 类水果的 mAP 为 92.75%。该数据表明, EfficentDet 网络能够适应复杂的非结构化果园场景, 并快速准确地完成水果目标检测任务。

2.3 三维点云精度评估

利用本文提出的三维感知框架对果实及其附近的枝叶障碍物进行点云提取, 并计算果实直径的视觉测量误差, 从而评估视觉系统所生成点云的精度。视觉测量误差计算公式为

$$\Delta \bar{D} = \frac{1}{N_F} \sum_{i=1}^{N_F} |D_i - D_{i0}| \quad (4)$$

式中, $\Delta \bar{D}$ 为水果直径的视觉测量误差; N_F 为当前帧的水果数量; D_i 为第 i 个水果的视觉测量值; D_{i0} 为第 i 个水果的人工测量值。

因此, 每 1 帧仅有 1 个视觉测量误差数据。

以百香果、番石榴、柑橘 3 类水果为研究对象, 每类水果采样 10 帧, 共获得 30 个视觉测量误差数据。获取视觉测量值时, 通过 CloudCompareStereo 软件打开点云, 直接测量点云中的水果直径; 获取人工测量值时, 通过精度为 0.01 mm 的数显游标卡尺现场测量

水果直径。用于测量的部分点云如图7的右半侧所示。

3 结论

针对复杂的户外果园采摘环境, 本文提出一种通用三维感知框架, 可得到水果及其附近的枝叶障碍物空间位置。该框架充分结合了深度网络和立体视觉系统的优势, 通过对多类水果的目标检测与点云提取实验, 验证其稳定性和有效性, 有望在各类水果采摘机器人中实现灵活集成。

本文还有一定的提升空间。首先, 可通过部署更高分辨率的相机来提高背景点云的稠密程度; 其次, 可对已有的点云采用滤波算法, 获得更加紧凑和精确的目标空间结构。

参考文献

- [1] 王红君, 牟其松, 岳有军, 等. 基于YOLOv3的水果采摘通用检测模型研究[J]. 中国科技论文, 2021, 16(3): 336-342.
- [2] 彭红星, 黄博, 邵园园, 等. 自然环境下多类水果采摘目标识别的通用改进SSD模型[J]. 农业工程学报, 2018, 34(16): 155-162.
- [3] 伍锡如, 雪刚刚, 刘英璇. 基于深度学习的水果采摘机器人视觉识别系统设计[J]. 农机化研究, 2020, 42(2): 177-182, 188.
- [4] 熊俊涛, 刘振, 林睿, 等. 自然环境下树上绿色芒果的无人机视觉检测技术[J]. 农业机械学报, 2018, 49(11): 23-29.
- [5] 熊俊涛, 刘振, 汤林越, 等. 自然环境下绿色柑橘视觉检测技术研究[J]. 农业机械学报, 2018, 49(4): 45-52.
- [6] 张帆. 基于深度卷积神经网络的水果图像识别算法研究[D]. 北京: 中央民族大学, 2020.
- [7] 杜泉雄, 张鹏超, 姚晋晋, 等. 基于双目立体视觉的采摘机器人设计[J]. 机电信息, 2018(36): 145, 147.
- [8] XIONG J, HE Z, LIN R, et al. Visual positioning technology of picking robots for dynamic litchi clusters with disturbance[J]. Computers and Electronics in Agriculture, 2018, 151: 226-237.
- [9] 黄博. 基于深度学习和RGB-D深度信息的水果识别与定位研究[D]. 广州: 华南农业大学, 2018.
- [10] 吴刚. 基于RGB-D相机的树上水果识别方法研究[D]. 无锡: 江南大学, 2020.
- [11] 丁宇祺. 果实自动化采摘的识别与定位方法研究[D]. 沈阳: 沈阳工业大学, 2020.
- [12] GE Y, XIONG Y, FROM P J. Symmetry-based 3D shape completion for fruit localisation for harvesting robots[J]. Biosystems Engineering, 2020, 197: 188-202.



图7 部分水果采样图像及其对应的三维点云

30 帧水果样本图像对应的视觉测量误差样本统计图如图8所示。其中, 误差样本最大值为 3.648 mm, 平均值 μ 为 2.037 mm, 标准差 σ 为 0.824 mm。视觉系统达到了水果采摘机器人在复杂环境下的三维感知要求^[25]。

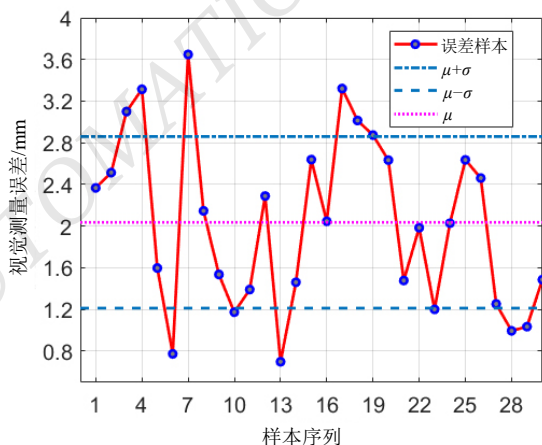


图8 视觉测量误差样本统计图

- [13] ARAD B, BALENDONCK J, BARTH R, et al. Development of a sweet pepper harvesting robot[J]. Journal of Field Robotics, 2020,37(6):1027-1039.
- [14] WIBOWO T S, SULISTIJO A, RISNUMAWAN A. End-to-end coconut harvesting robot[C]//2016 International Electronics Symposium (IES). IEEE, 2016: 444-449.
- [15] 董戈.基于深度学习和图像处理的水果收获机器人抓取系统[J].农机化研究,2021,43(3):260-264.
- [16] 杨平.基于深度学习的水果采摘定位算法的研究[D].哈尔滨:哈尔滨理工大学,2020.
- [17] LIN G, TANG Y, ZOU X, et al. Guava detection and pose estimation using a low-cost RGB-D sensor in the field[J]. Sensors, 2019,19(2):428.
- [18] LI J, TANG Y, ZOU X, et al. Detection of fruit-bearing branches and localization of litchi clusters for vision-based harvesting robots[J]. IEEE Access, 2020,8:117746-117758.
- [19] WILLIAMS H A, JONES M H, NEJATI M, et al. Robotic kiwifruit harvesting using machine vision, convolutional neural networks, and robotic arms[J]. Biosystems Engineering, 2019,181:140-156.
- [20] TAN M, PANG R, LE Q V, et al. Efficientdet: Scalable and efficient object detection[C]//Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2020: 10781-10790.
- [21] TAN M, LE Q. Efficientnet: Rethinking model scaling for convolutional neural networks[C]//International Conference on Machine Learning. PMLR, 2019: 6105-6114.
- [22] LIU S, QI L, QIN H, et al. Path aggregation network for instance segmentation[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2018: 8759-8768.
- [23] TANG Y, WANG C, LUO L, et al. Recognition and localization methods for vision-based fruit picking robots: a review[J]. Frontiers in Plant Science, 2020, 11: 510.
- [24] HIRSCHMULLER H. Accurate and efficient stereo processing by semi-global matching and mutual information[C]//2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05). IEEE, 2005, 2: 807-814.
- [25] CHEN M, TANG Y, ZOU X, et al. Three-dimensional perception of orchard banana central stock enhanced by adaptive multi-vision technology[J]. Computers and Electronics in Agriculture, 2020, 174: 105508.

General 3D Perception Framework for Multiple Types Complex Fruit Objects

Cheng Jiabing¹ Zou Xiangjun² Chen Mingyou² Tan Kangyu¹ Wu Fengyun²

(1. School of Electrical and Computer Engineering, Nanfang College, Guangzhou, 510970, China

2. College of Engineering, South China Agricultural University, Guangzhou 510642, China)

Abstract: Compared with an ideal laboratory environment, fruit images sampled in an outdoor orchard are often disturbed by more noises, such as complex backgrounds, occlusion of branches and leaves, uneven illumination and mechanical vibration, etc., which have always been the main constraints on the further development of visual fruit picking robots. In response to this problem, a general three-dimensional perception framework adapted to the complex picking environment was developed. First, the binocular vision system was constructed and calibrated; then, the complex background was removed with the help of a robust deep object detection network to obtain the image area where the fruit is located; finally, the fruit area was conducted stereo matching operation and triangulated to obtain the spatial position of the fruits. The framework has fully combined the robustness of the deep neural network and the excellent perception ability of the stereo vision system. It does not rely on any artificially designed features and can improve the accuracy and stability of the robot's perception of outdoor fruits. This research provides an theoretical basis and technical support for improving the stability and practicability of the fruit picking robot.

Key words: picking robot; machine vision; object detection; stereo matching; triangulation

作者简介:

程佳兵, 女, 1994 年生, 硕士, 主要研究方向: 图像处理、机器视觉。E-mail: chengjb@nfcu.edu.cn