

文章编号:1671-6833(2025)02-0011-08

基于单目视觉前馈的机械臂目标快速三维瞄准方法

陈晓鹏¹, 徐 鹏¹, 王占涛², 邢 程¹, 邱钰涵¹

(1. 北京理工大学 机电学院, 北京 100081; 2. 北京航天计量测试技术研究所, 北京 100076)

摘 要:针对基于单目视觉的机械臂目标瞄准缺少三维深度信息和收敛速度慢等问题,提出了一种单目视觉前馈的六自由度机械臂快速瞄准方法。首先,对 YOLOv4 的 CSPDarknet 网络进行裁剪,减少 Conv2d_BN_Mish 单元数量,简化特征提取网络复杂度,得到 Lite YOLOv4 以此提高目标检测速度;其次,根据从单目图像中检测得到的目标像素坐标,求取目标像素坐标的反投影归一化坐标,得到激光瞄准器期望的相对三维瞄准射线方程;再次,提出多点标定法实现单目相机、激光瞄准器以及机械臂坐标系的外参数标定,并根据外参数标定结果与激光瞄准器的期望三维射线方程计算机械臂各个关节的期望姿态;最后,基于机械臂的直接位置控制代替图像视觉伺服,加快收敛速度,实现基于单目视觉前馈的机械臂目标快速三维瞄准。经实验验证,所提方法单目平均瞄准响应时间仅为 0.611 s,同时目标瞄准成功率达 95.238%,与传统基于图像的视觉伺服的方法相比提高了 4.762 百分点。

关键词:单目视觉;视觉前馈;自主瞄准;深度学习;机械臂控制

中图分类号: TP242.6; TP249

文献标志码: A

doi: 10.13705/j.issn.1671-6833.2025.02.001

机械臂作为一种作业机器人,不仅广泛用于工业领域,在服务领域也逐渐大显身手。机器人操作一般分为两类:一类为基于位置、速度或力的接触式静态操作;另一类为非接触式动态操作^[1]。接触式操作有众多的研究,而非接触式动态操作由于机器人与目标不发生接触或者只是很少的时间发生接触,对目标的运动特性要求较高。非接触式动态操作包括机器人射箭、机器人打乒乓球、机器人打高尔夫球等^[2-3]。

机器人瞄准也是一类非接触式动态操作。与机器人打乒乓球、机器人射箭不同,机器人瞄准的操作任务是通过移动激光光源使得激光光束瞄准靶标的中心,而激光本身并没有动力学特性,轨迹也不受重力的影响,在传输过程中时间可以忽略不计,但准确性要求更高。

机器人瞄准大致分为两步:一是对靶标进行精确的视觉检测;二是基于视觉反馈伺服完成对目标的瞄准。目前涌现了大量视觉检测方法,Tan 等^[4]提出了名为 EfficientDet 的目标检测器,在小目标的识别上取得了较好的效果。Zhu 等^[5]提出了一种学

习已标记和未标记数据的自监督自集成 RGB-D 显著性检测网络,在更少的标注信息下完成 RGB-D 显著性目标检测。YOLO 系列作为经典的目标检测算法,也被广泛改进并应用于各个领域。张震等^[6]在深度分离卷积中使用 Ghost 卷积构建了新的主干模块,在保持模型精度的基础上提高了检测速度。该团队还尝试引入卷积注意力模块来进一步提高该模型的精度^[7]。Li 等^[8]提出了一种基于内部伪标签迭代更新的端到端方法,有效解决了弱监督 RGB-D 显著目标检测问题。Carion 等^[9]实现了一个以 Detection Transformer (DETR) 为主要组成部分的新框架,其检测精度和 Faster R-CNN 几乎相当。Meng 等^[10]在 DETR 的基础上提出了 Conditional DETR,解决了 DETR 训练收敛缓慢这个关键问题。但是,上述算法在实时性方面在推理时经常有所欠缺。

在视觉伺服方面,Dong 等^[11]提出了一种新的基于图像的视觉伺服控制器,通过优化一个多目标损失函数将机械臂末端执行器的运动方向限制在预定的范围内,因此,优化后的运动轨迹冗余运动明显减少。Ranganathan^[12]设计了由鱼眼相机引导的机

收稿日期:2024-10-20;修订日期:2024-12-28

基金项目:国家自然科学基金资助项目(61973031);北京理工大学科技创新计划(2024CX06013)

作者简介:陈晓鹏(1982—),男,江西赣州人,北京理工大学副教授,博士,博士生导师,主要从事机器人主动感知研究,E-mail: xpchen@bit.edu.cn。

引用本文:陈晓鹏,徐鹏,王占涛,等. 基于单目视觉前馈的机械臂目标快速三维瞄准方法[J]. 郑州大学学报(工学版), 2025,46(2):11-18. (CHEN X P, XU P, WANG Z T, et al. Fast 3D aiming method of manipulators based on monocular visual feedforward[J]. Journal of Zhengzhou University(Engineering Science), 2025,46(2):11-18.)

械臂与真人棋局对弈系统,采用 Kanade-Lucas-Tomasi^[13]方法实现了机械臂的视觉伺服。Paradis 等^[14]提出了深度间歇视觉伺服(deep intermittent visual servoing, IVS)框架,IVS 在自动化外科手术的执行器转移中获得了较高的成功率。Al-shanoon 等^[15]提出了深度视觉伺服(deep visual servoing, DVS)框架,DVS 是基于 RGB 相机的深度卷积神经网络与视觉伺服相结合的可靠框架。Al-shanoon 等^[16]还提出了一种用于目标姿态估计的卷积神经网络算法,并将其集成到 3D 视觉伺服系统中,使用单个摄像机实现了移动操作任务。Lee 等^[17]将基于深度学习的视觉伺服与基于力的操作基元动态序列相结合,完成了机器人装配任务。Zhong 等^[18]提出了一种自学习视觉伺服方法,能够使机械臂在非结构化环境中完成操作任务。Puang 等^[19]提出了一种新的基于学习和关键点的视觉伺服框架,只需在仿真环境中训练深度神经网络,训练后的模型可直接用于现实世界的视觉伺服任务。Ribeiro 等^[20]训练了一个卷积神经网络以执行视觉伺服控制,确保被抓取的物体始终保持在机械臂的视野范围内。

然而,目前的算法存在两个问题:一是基于深度学习的目标检测算法存在检测速度较慢等问题,影响了视觉伺服的控制周期和响应速度;二是基于图像的视觉伺服方法,当期望特征位置 and 实际特征位置很近时,收敛将变得缓慢,收敛时间较长。

本文主要针对上述问题提出改进的瞄准方法。首先,针对目标检测速度慢的问题,对经典的深度学习目标检测算法 YOLOv4 的 Backbone 中用于平衡速度和精度的 CSPDarknet53 部分进行裁剪以增加检测速度;其次,受图像目标点位于连接光心和图像点的三维射线方向上的启发,所使用的方法不通过驱动相机的移动使得图像目标实际坐标移动到期望位置,而是直接依据目标点图像坐标信息,通过反投影得到空间三维射线,并基于此射线构造期望的激光束坐标系,把这部分信息作为前馈量驱动机械臂的运动;最后,对机械臂进行微调,驱动激光发射器精确瞄准。

1 系统组成

本文基于协作机械臂 UR3e,通过单目视觉前馈实现对三维动态目标的检测和快速瞄准。

如图 1 所示,系统主要包括机械臂和可独立移动的靶标。此外,机械臂末端还配备了单目相机、光源和激光发射器。其中单目相机用于拍摄目标图

像,为控制系统提供反馈信息输入。当环境太暗无法清晰成像时,光源用于提供补充光。激光发射器可以通过蓝牙远程控制,由软件触发,触发后可以发射激光束,完成瞄准的任务。

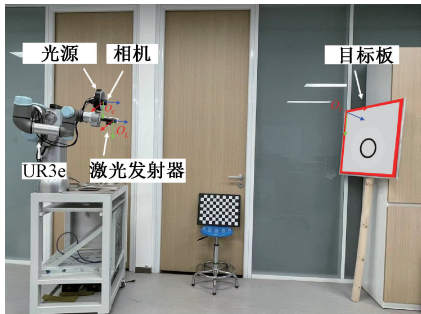


图 1 系统硬件结构
Figure 1 System hardware structure

2 瞄准方法

瞄准方法主要包括坐标系标定、目标检测、单目瞄准、前馈伺服 4 大模块,如图 2 所示。其中坐标系标定用于得到激光发射器与相机、相机与机械臂的相对位姿;目标检测用于检测瞄准目标,得到目标点相对于相机坐标系的位姿;单目瞄准用于根据期望的激光位姿和实际位姿之差计算当前机械臂末端的期望运动;前馈伺服将机械臂末端的期望运动转换为机械臂 6 个关节电机的控制指令,最终驱动机械臂控制激光发射器实现对目标的瞄准任务。

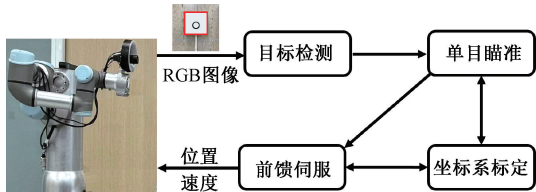


图 2 3D 目标瞄准系统的模块组成
Figure 2 Module composition of the 3D target aiming system

2.1 单目相机与激光发射器坐标系标定

实现精确瞄准需要得到两个坐标系变换矩阵,首先是相机坐标系相对于机械臂末端坐标系的变换矩阵即手眼参数;其次是激光发射器坐标系相对相机坐标系的变换矩阵。前者可通过 Daniilidis 的手眼标定算法^[21]来获取;针对后者,本文在张氏相机标定算法^[22]基础上提出如下新的外部参数标定方法。

具体方法如图 3 所示,保持棋盘格不动,手动移动机械臂在不同的位姿,使得在每个位姿下激光器都瞄准棋盘格的某一个具体的点。通过建立这些点同时在棋盘格坐标系、相机坐标系及激光坐标系下

的约束,完成激光坐标系和相机坐标系之间的标定。

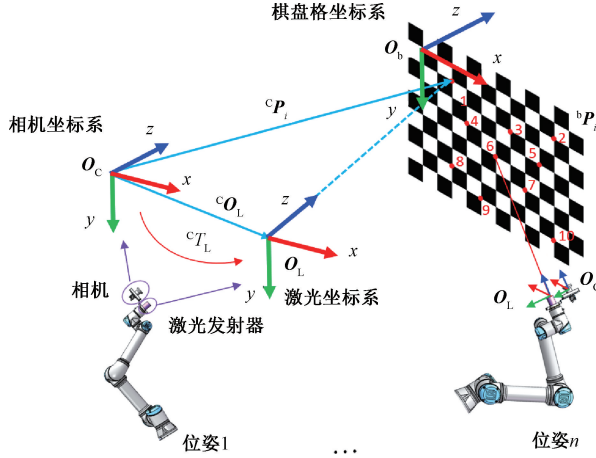


图 3 不同坐标系之间的约束关系

Figure 3 Constraint relationships among different coordinate systems

cT_L 表示相机坐标系与激光发射器坐标系之间待求解的变换矩阵,如式(1)所示:

$${}^cT_L = \begin{bmatrix} {}^cR_L & {}^cO_L \\ \mathbf{0} & \mathbf{1} \end{bmatrix} = \begin{bmatrix} {}^cX_L & {}^cY_L & {}^cZ_L & {}^cO_L \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{1} \end{bmatrix}. \quad (1)$$

式中: cR_L 为旋转分量; cO_L 为平移分量。作为单位向量, cX_L 、 cY_L 、 cZ_L 相互正交:

$$\begin{cases} {}^cX_L^2 + {}^cY_L^2 + {}^cZ_L^2 = 1; \\ {}^cY_L^2 + {}^cY_{Ly}^2 + {}^cY_{Lz}^2 = 1; \\ {}^cZ_L^2 + {}^cZ_{Ly}^2 + {}^cZ_{Lz}^2 = 1. \end{cases} \quad (2)$$

在图 3 中,棋盘格的单个方格的边长为 30 mm。当激光发射器发射的激光束刚好照射到棋盘角点位置时,有:

$${}^cO_L - {}^{c_i}P_i = \lambda_i {}^cZ_L. \quad (3)$$

在每次观测中,相机的中心点坐标可能会发生变化,因此增加了下标 i 。同时,即使下标发生变化, cO_L 、 cZ_L 分别为未知待标定的平移量和 z 轴方向矢量,它们不会随着相机位姿变化而变化。这里 ${}^{c_i}P_i$ 为棋盘角点 P_i 在相机坐标系中的坐标,满足:

$${}^{c_i}P_i = {}^{c_i}T_b {}^bP_i. \quad (4)$$

式中: ${}^{c_i}T_b$ 表示相机坐标系与棋盘格坐标系之间的变换矩阵,可以通过张氏相机标定算法^[22]来获得; bP_i 为棋盘角点在自身棋盘格坐标系中的坐标,可以根据棋盘格的尺寸得到,每个点的 z 轴方向坐标值为零。

由于同方向的向量叉乘结果为 0,式(3)可以改写为

$$\begin{pmatrix} {}^cO_{Lx} \\ {}^cO_{Ly} \\ {}^cO_{Lz} \end{pmatrix} - {}^{c_i}T_b {}^bP_i \times \begin{pmatrix} {}^cZ_{Lx} \\ {}^cZ_{Ly} \\ {}^cZ_{Lz} \end{pmatrix} = \mathbf{0}. \quad (5)$$

式(5)可转换为 ${}^cZ_{Lx} {}^cO_{Ly}$ 、 ${}^cZ_{Ly} {}^cO_{Lx}$ 、 ${}^cZ_{Ly} {}^cO_{Lz}$ 、 ${}^cZ_{Lz} {}^cO_{Lx}$ 、 ${}^cZ_{Lz} {}^cO_{Ly}$ 、 ${}^cZ_{Lz} {}^cO_{Lz}$ 等变量的线性方程组:

$$AX = \mathbf{0}. \quad (6)$$

式(6)中有 3 个方程,5 个独立未知数。实际操作中选取 $i=1,2,\dots,10$,共 10 个点,如图 3 所示,这些点尽可能覆盖棋盘格的横坐标和纵坐标,并采用 DLT^[23]联立式(6)和式(2)进行求解。

2.2 Lite YOLOv4 目标检测

由于光照、距离等不可控因素,实际环境往往非常复杂,传统算法的泛化能力较弱。并且对于移动目标,检测算法耗时更少,因此,本文结合了深度学习算法来检测圆心。

为了提高算法的准确性和速度,以 YOLOv4^[24]为基础,对其网络结构进行了一定的修改。YOLOv4 的网络结构主要包括 3 个部分:Backbone、Neck 和 Head。作为图像特征提取的关键模块,Backbone 的设计需要根据任务要求权衡精度和速度。在 YOLOv4 中,这部分使用了 CSPDarknet53^[24]模型。作为融合不同尺寸特征图特征信息的重要模块,Neck 利用了 SPP (spatial pyramid pooling)^[25]和 PAN (path aggregation network)^[26]。作为图像特征预测、生成边界框及预测类别的模块,Head 融合先验值并根据离散特征尺度的推理输出预测最终结果。图 4 为修改后的 YOLOv4 算法的网络结构。首先,对于每个 CSPDarknet 网络,Conv2d_BN_Mish 单元数从 5 个减少到 2 个,省略了特征拼接操作。因此,对于每个下采样模块,除了残差单元外,还添加了两个卷积操作;其次,卷积核通道数从 64 个减少到 32 个。这些操作大大减少了需要优化的参数数量和计算量。

2.3 目标反投影视觉前馈瞄准算法

图 5 为目标点相对于激光坐标系的俯仰方位。如图 5 所示,一般来说,在瞄准目标点 M 之前,首先需要获取其 3D 坐标。这样就可以很容易地求解出相对激光坐标系的俯仰方位角,然后控制激光发射器的位姿,使点 M 位于新的激光坐标系的 z 轴方向上。

然而,三维目标需要采用双目立体相机,单目相机难于获得三维坐标。由于所观测的目标为 2D 目标,采用基于图像的视觉伺服存在精度不高的问题。同时,当期望图像位置与实际图像位置距离较近时,

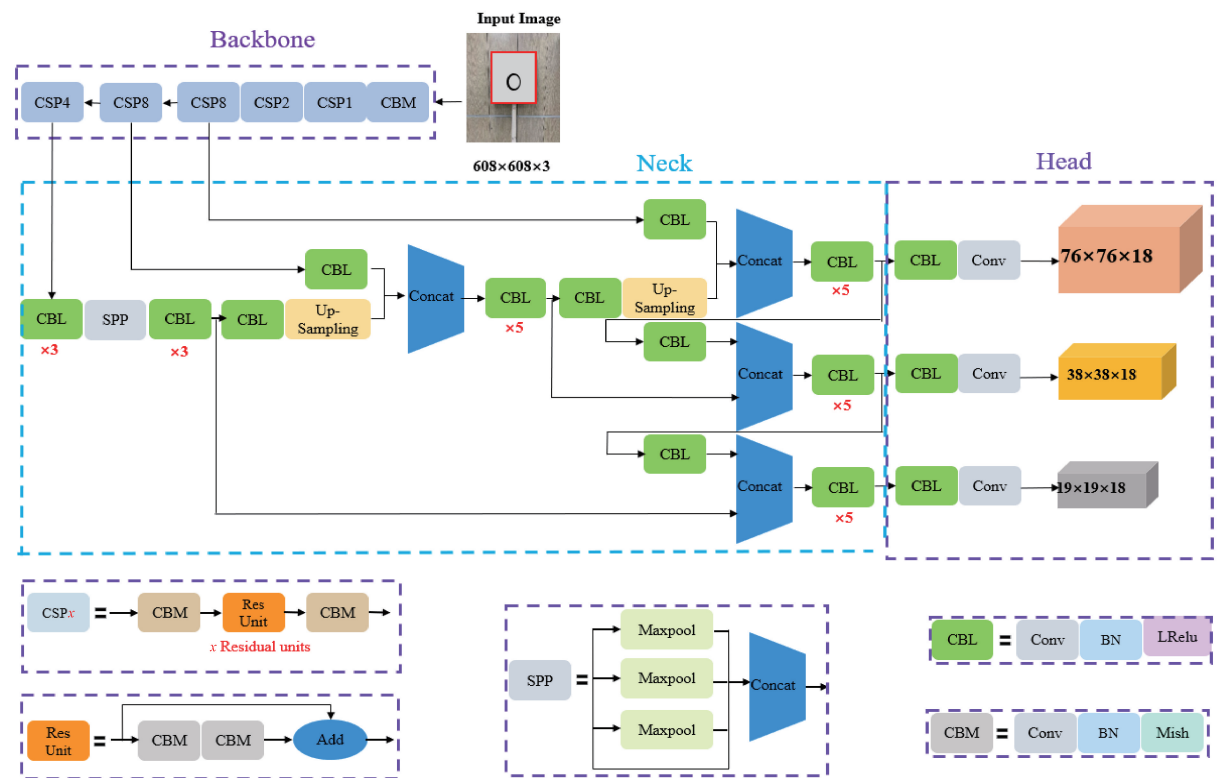


图 4 Lite YOLOv4 网络结构
Figure 4 Lite YOLOv4 network structure

收敛速度会变慢,而为了保证精度不得不花较长时间等待以得到较小的伺服误差。

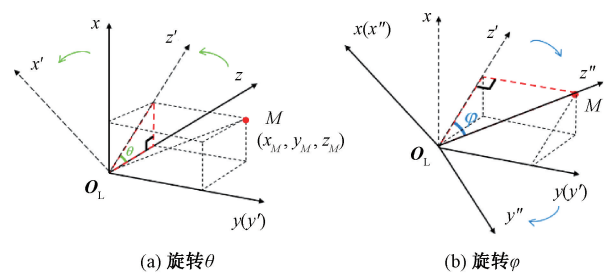


图 5 目标点相对于激光坐标系的俯仰方位
Figure 5 The elevation angle of the target point relative to the laser coordinate system

本论文提出一种新的单目瞄准算法,该算法无须获得如图 5 所示的目标的三维坐标,而是受到图 5 的启发,视觉瞄准所需要获得的并不是 P 点,只要获得经过 P 点的空间三维射线,就可以操作机械臂,使激光瞄准器的 z 轴和这条空间三维射线重合即可。通过图像坐标点反投影求解,恰好可以得到一条三维空间的射线,以这条射线作为激光器瞄准的 z 轴,则可以实现视觉前馈的三维瞄准。

图 6 为单目瞄准算法中的映射关系。在图 6 中, ${}^cT_{cd}$ 为要求解的相机位姿偏移矩阵; f 为焦距; a 为 OA_2 的长度; b 为 OA_3 的长度。 $A_1(u_0, v_0)$ 为成像面中心的像素坐标, $A_3(u_3, v_3)$ 为目标板中心点的

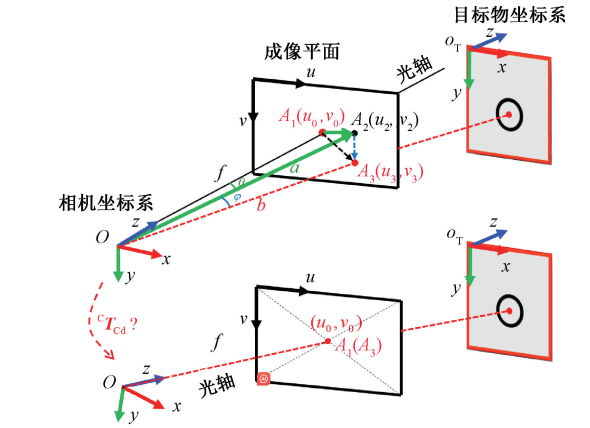


图 6 单目瞄准算法中的映射关系
Figure 6 Mapping relationship in monocular aiming algorithm

像素坐标。此外, δ_x, δ_y 表示像素的物理尺寸。所以瞄准过程可以分为以下两个步骤。

首先,将相机坐标系绕 y 轴旋转角度 θ , 由于 ΔOA_1A_2 为直角三角形,有:

$$\theta = \arctan \frac{(u - u_0) \delta_x}{f} \quad (7)$$

式中: δ_x 为单个像素 x 方向的物理尺寸。

其次,相机坐标系围绕先前更新的 x 轴旋转角度 ϕ , 在 ΔOA_2A_3 中, ϕ 可以根据余弦定律得到

$$\phi = \arccos \left(\frac{a^2 + b^2 - (v - v_0)^2 \delta_y^2}{2ab} \right) \quad (8)$$

其中,

$$\begin{cases} a = \sqrt{f^2 + (u - u_0)^2 \delta_x^2}; \\ b = \sqrt{f^2 + (u - u_0)^2 \delta_x^2 + (v - v_0)^2 \delta_y^2}. \end{cases} \quad (9)$$

${}^c T_{Cd}$ 包含两部分:旋转分量 R_{xyz} 和平移分量

t_{xyz} :

$${}^c T_{Cd} = \begin{bmatrix} R_{xyz} & t_{xyz} \\ \mathbf{0} & \mathbf{1} \end{bmatrix}. \quad (10)$$

平移分量 t_{xyz} 可以设置为零向量:

$$t_{xyz} = [0 \ 0 \ 0]^T. \quad (11)$$

结合式(9),求解得旋转分量 R_{xyz} :

$$R_{xyz}(\varphi, \theta, 0) = R_z(0) R_y(\theta) R_x(\varphi). \quad (12)$$

最后,发送期望的末端执行器位姿命令 ${}^w T_{e1d}$ 以实现激光发射器坐标系中的瞄准过程:

$${}^w T_{e1d} = {}^w T_C {}^C T_{Cd} {}^L T_C {}^C T_{e1}. \quad (13)$$

该目标反投影视觉前馈瞄准算法如下。

算法 1 目标反投影视觉前馈瞄准算法。

① 标定相机和激光器之间的相对位姿 ${}^c T_L$:

在平面标定棋盘格内选取 10 个或以上的点;

for 每一个待标定点 do

根据棋盘格尺寸确定其在棋盘格坐标系下的空间坐标;

手动移动机械臂,让激光瞄准待标定角点;

根据棋盘格计算相机相对棋盘格的外参数矩阵 ${}^c T_b$;

根据 10 个点所得到的参数获得式(6)中 A 的表达式,并采用 DLT 方法求解出外参数 ${}^c T_L$;

② 采用 Lite YOLOv4 获得目标的中心点坐标 (u, v) ;

③ 根据式(7)、式(8)计算俯仰角 θ 和方位角 φ ;

④ 根据式(11)、式(12)计算相机位姿偏移矩阵 ${}^c T_{Cd}$;

⑤ 根据式(13)计算机机械臂末端执行器期望位姿 ${}^w T_{e1d}$;

⑥ 由机械臂末端执行器期望位置执行位置指令到达指定点并发射一次激光。

3 实验和结果

3.1 实验配置

在本文实验中,机械臂采用 UNIVERSAL ROBOTS 的 UR3e;相机采用 allied vision 相机,最高图像分辨率为 1936×1216 像素,最高帧率可达 220 帧/s,搭配 6 mm 定焦镜头;其他连接及辅助部件见本文第 1 节系统组成。实验运行的计算机硬件配置

GPU 为 NVIDIA TITAN V, CPU 为 Intel(R) Xeon(R) Silver 4114 CPU@2.20 GHz;软件运行 Ubuntu 18.04 操作系统、RT Preempt 实时系统内核;深度学习环境为 CUDA 11.3、darknet;软件各模块基于 ROS melodic 完成各模块的实现和集成。

3.2 目标检测实验

本文利用基于霍夫变换的传统检测算法^[27]、原始 YOLOv4 算法^[24]和本文所提修改后的 Lite YOLOv4 算法进行对比实验。实验采用的数据集包括了在不同角度、不同距离、异常光照条件下的 400 张包含目标板的实际场景图像,按照 9:1 的比例分配给训练集和验证集进行训练。输入模型的图片大小 $608 \times 608 \times 3$ 像素。在数据集中随机采集 22 张图像作为测试集,图 7 显示了 3 种检测算法的平均单帧检测时间和平均检测误差。

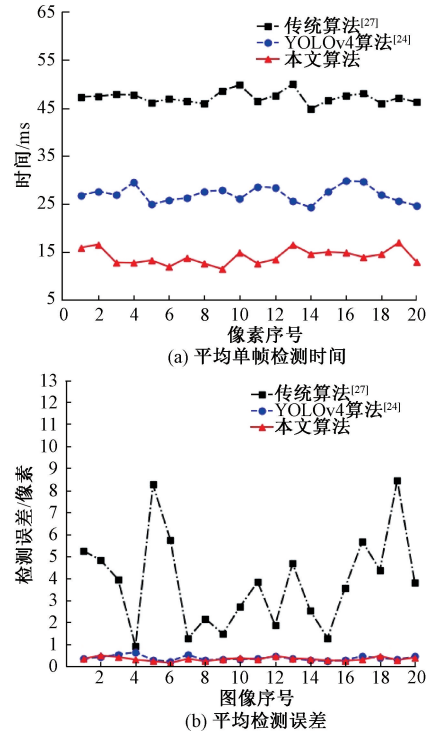


图 7 目标检测算法的时间与精度对比

Figure 7 Comparison of time and accuracy of target detection algorithms

平均检测误差中使用的圆心坐标真值通过人工标注得到。如果检测结果在目标板上的圆圈内,则认为检测过程成功。如果检测失败,则该图像将不计入平均检测误差的计算公式。

表 1 为目标检测算法结果对比。如表 1 所示,对于传统的目标检测算法^[27],单帧图像的平均检测时间为 47.218 ms,检测帧率 21.18 帧/s,检测成功率为 90.91%,平均检测误差为 3.838 个像素点。对于原始的 YOLOv4 算法^[24],单帧图像的平均检测时间为

27.055 ms, 检测帧率 36.96 帧/s, 检测成功率 100%, 平均检测误差为 0.377 个像素点。对于本文算法, 单帧图像的平均检测时间为 14.122 ms, 检测帧率提升至 70.81 帧/s, 同时检测成功率保持 100%, 平均检测误差为 0.341 个像素点。

表 1 目标检测算法结果对比

算法	单帧时间/ms	检测误差/像素	成功率/%
传统算法 ^[27]	47.218	3.838	90.91
YOLOv4 算法 ^[24]	27.055	0.377	100.00
本文算法	14.122	0.341	100.00

3.3 单目瞄准实验

为了验证单目瞄准算法的有效性, 本文统一使用 Lite YOLOv4 算法进行目标检测, 分别使用传统的基于图像的视觉伺服算法和单目瞄准算法进行瞄准实验。将目标板放置在 21 个不同位置进行对比实验, 用来比较两种算法的响应时间和成功率。图 8 显示了两种瞄准算法的平均响应时间。对于每次瞄准实验, 如果机械臂最终能够将激光束照射到目标板的圆圈内, 则认为这个瞄准过程是成功的。

对于视觉伺服算法, 其瞄准过程的平均响应时间为 4.934 s, 成功率为 90.476%; 而对于本文的单目瞄准算法, 平均响应时间仅为 0.611 s, 瞄准成功率为 95.238%。

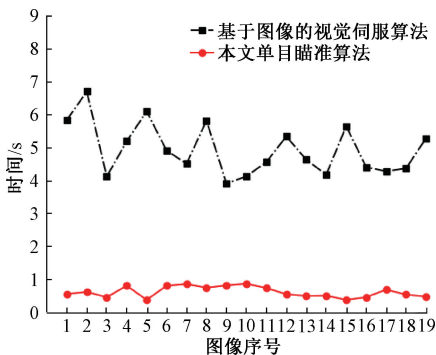


图 8 两种瞄准算法的平均响应时间

Figure 8 Average response time of two aiming algorithms

4 结论

本文分别进行了目标检测实验和实物瞄准实验。实验结果显示, 本文所提出的 Lite YOLOv4 算法在基本不降低检测精度的情况下, 将检测帧率从 36.96 帧/s 提高到 70.81 帧/s, 速度约为原 YOLOv4 的两倍。本文提出的单目瞄准算法相比于传统的基于图像的视觉伺服算法, 平均响应时间从 4.934 s 降为 0.611 s; 成功率从 90.476% 提高到 95.238%。

实验结果表明, 本文方法可以通过单目视觉前馈很好地提高 3D 目标瞄准的速度和精度。

参考文献:

[1] MATHESON E, MINTO R, ZAMPIERI E G G, et al. Human-robot collaboration in manufacturing applications: a review[J]. Robotics, 2019, 8(4): 100.

[2] 王大浩, 平雪良. 基于规定性能的六自由度机械臂视觉伺服控制[J]. 传感器与微系统, 2022, 41(3): 104-108.

WANG D H, PING X L. Six-degree-of-freedom manipulator visual servo control based on prescribed performance[J]. Transducer and Microsystem Technologies, 2022, 41(3): 104-108.

[3] LING X, ZHAO Y S, GONG L, et al. Dual-arm cooperation and implementing for robotic harvesting tomato using binocular vision[J]. Robotics and Autonomous Systems, 2019, 114: 134-143.

[4] TAN M X, PANG R M, LE Q V. EfficientDet: scalable and efficient object detection[C]//2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE, 2020: 10778-10787.

[5] ZHU L, WANG X Q, LI P, et al. S³ net: self-supervised self-ensembling network for semi-supervised RGB-D salient object detection[J]. IEEE Transactions on Multimedia, 2023, 25: 676-689.

[6] 张震, 王晓杰, 晋志华, 等. 基于轻量化 YOLOv5 的交通标志检测[J]. 郑州大学学报(工学版), 2024, 45(2): 12-19.

ZHANG Z, WANG X J, JIN Z H, et al. Traffic sign detection based on lightweight YOLOv5[J]. Journal of Zhengzhou University (Engineering Science), 2024, 45(2): 12-19.

[7] 张震, 陈可鑫, 陈云飞. 优化聚类 and 引入 CBAM 的 YOLOv5 管制刀具检测[J]. 郑州大学学报(工学版), 2023, 44(5): 40-45, 61.

ZHANG Z, CHEN K X, CHEN Y F. YOLOv5 with optimized clustering and CBAM for controlled knife detection[J]. Journal of Zhengzhou University (Engineering Science), 2023, 44(5): 40-45, 61.

[8] LI J J, JI W, BI Q, et al. Joint semantic mining for weakly supervised RGB-D salient object detection[C]//Proceedings of the 35th International Conference on Neural Information Processing Systems. New York: ACM, 2021: 11945-11959.

[9] CARION N, MASSA F, SYNNAEVE G, et al. End-to-end object detection with transformers[C]//European Conference on Computer Vision. Cham: Springer, 2020: 213-229.

- [10] MENG D P, CHEN X K, FAN Z J, et al. Conditional DETR for fast training convergence[C]//2021 IEEE/CVF International Conference on Computer Vision (ICCV). Piscataway: IEEE, 2021: 3631–3640.
- [11] DONG J X, ZHANG J. A new image-based visual servoing method with velocity direction control[J]. Journal of the Franklin Institute, 2020, 357(7): 3993–4007.
- [12] RANGANATHAN G. An economical robotic arm for playing chess using visual servoing[J]. Journal of Innovative Image Processing, 2020, 2(3): 141–146.
- [13] SHIRLEY D R A, RANJANI K, ARUNACHALAM G, et al. Automatic distributed gardening system using object recognition and visual servoing[C]//Inventive Communication and Computational Technologies. Berlin: Springer, 2021: 359–369.
- [14] PARADIS S, HWANG M, THANANJEYAN B, et al. Intermittent visual servoing: efficiently learning policies robust to instrument changes for high-precision surgical manipulation[C]//2021 IEEE International Conference on Robotics and Automation (ICRA). Piscataway: IEEE, 2021: 7166–7173.
- [15] AL-SHANOON A, LANG H X. Robotic manipulation based on 3-D visual servoing and deep neural networks [J]. Robotics and Autonomous Systems, 2022, 152: 104041.
- [16] AL-SHANOON A, WANG Y J, LANG H X. DeepNet-based 3D visual servoing robotic manipulation[J]. Journal of Sensors, 2022, 2022: 3511265.
- [17] LEE Y S, VUONG N, ADRIAN N, et al. Integrating force-based manipulation primitives with deep learning-based visual servoing for robotic assembly [C] //ICRA 2022 Workshop: Reinforcement Learning for Contact-Rich Manipulation. Piscataway: IEEE, 2022.
- [18] ZHONG X G, SHI C Q, LIN J, et al. Self-learning visual servoing for robot manipulation in unstructured environments[C]//International Conference on Intelligent Robotics and Applications. Cham: Springer, 2021: 48–57.
- [19] PUANG E Y, PENG TEE K, JING W. KOVIS: key-point-based visual servoing with zero-shot sim-to-real transfer for robotics manipulation [C] //2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS). Piscataway: IEEE, 2020: 7527–7533.
- [20] RIBEIRO E G, DE QUEIROZ MENDES R, GRASSI V J. Real-time deep learning approach to visual servo control and grasp detection for autonomous robotic manipulation[J]. Robotics and Autonomous Systems, 2021, 139: 103757.
- [21] DANILIDIS K. Hand-eye calibration using dual quaternions [J]. The International Journal of Robotics Research, 1999, 18(3): 286–298.
- [22] ZHANG Z. A flexible new technique for camera calibration[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2000, 22(11): 1330–1334.
- [23] SUTHERLAND I E. Three-dimensional data input by tablet[J]. Proceedings of the IEEE, 1974, 62(4): 453–461.
- [24] BOCHKOVSKIY A, WANG C Y, LIAO H Y M. YOLOv4: optimal speed and accuracy of object detection [EB/OL]. (2020-05-23) [2024-08-15]. <https://doi.org/10.48550/arXiv.2004.10934>.
- [25] HE K M, ZHANG X Y, REN S Q, et al. Spatial pyramid pooling in deep convolutional networks for visual recognition[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2015, 37(9): 1904–1916.
- [26] LIU S, QI L, QIN H F, et al. Path aggregation network for instance segmentation[C]//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2018: 8759–8768.
- [27] YUEN H, PRINCEN J, ILLINGWORTH J, et al. Comparative study of Hough Transform methods for circle finding[J]. Image and Vision Computing, 1990, 8(1): 71–77.

Fast 3D Aiming Method of Manipulators Based on Monocular Visual Feedforward

CHEN Xiaopeng¹, XU Peng¹, WANG Zhantao², XING Cheng¹, QIU Yuhan¹

(1. School of Mechatronic Engineering, Beijing Institute of Technology, Beijing 100081, China; 2. Beijing Aerospace Institute for Metrology and Measurement Technology, Beijing 100076, China)

Abstract: To address the issues of the lack of three-dimensional depth information and slow convergence speed in monocular vision-based manipulator aiming, a fast aiming algorithm based on monocular visual feedforward approach for a 6D of robotic arm was proposed. First of all, the CSPDarknet of YOLOv4 was cropped through reducing the number of the Cov2D_BN_Mish units and simplifying the complexity of the backbone network so as to form a Lite YOLOv4 algorithm to accelerate object detection speed. Then, the pixel coordinates of the target object de-

tected from monocular images were inversely projected to a 3D ray emitted from the camera center. The 3D ray was exactly the target ray for the aiming laser. What was more, a multiple common points based calibration approach was proposed to obtain the extrinsic parameters of the monocular camera and the aiming laser, and the expected joint poses of the robotic arm were then deducted based on the expected 3D ray coordinates. Finally, direct position control was used to replace visual servoing control based on the expected joint poses to accelerate convergence speed, and realize the monocular visual feedforward based fast 3D aiming for robotic arms. Experiments verified that the average monocular aiming time was 0.611 second. The aiming success rate was 95.238%, which was 4.762 percentage points higher than traditional image based visual servoing.

Keywords: monocular vision; visual feedforward; autonomous aiming; deep learning; manipulator operation

(上接第 10 页)

Multi-electrode Coordinated Upper Limb Rehabilitation Motion Control Method Based on ADRC

HUO Benyan¹, WANG Yanan¹, ZHANG Zan¹, DONG Anqin², LIU Yanhong¹

(1. School of Electrical and Information Engineering, Zhengzhou University, Zhengzhou 450001, China; 2. The Fifth Affiliated Hospital, Zhengzhou University, Zhengzhou 450000, China)

Abstract: Functional electrical stimulation (FES) was an important rehabilitation treatment for stroke-induced disabilities. However, existing FES devices primarily used single-pair surface electrodes to stimulate muscles, making it difficult to precisely locate the optimal stimulation position during limb movements. Additionally, the accuracy of rehabilitation movement control could be affected by internal and external disturbances. To achieve precise control of upper limb movements. A multi-electrode coordinated control method based on active disturbance rejection control (ADRC) was proposed in this study. First, a multi-electrode functional electrical stimulation system for upper limb rehabilitation was designed. Next, an electrode switching strategy was determined by analyzing the dynamics model of the upper limb, dynamically adjusting the electrode combination according to the angle range during angle tracking to track the optimal stimulation position. Subsequently, ADRC was introduced as the controller to estimate and compensate for system disturbances in real-time. Finally, healthy subjects were recruited for simulations and experimental validation. The results showed that the proposed control method achieved precise control of upper limb movements. Compared with the fixed electrode strategy, the average tracking error was reduced by approximately 50.41%, and the root mean square error was reduced by approximately 43.30%. Furthermore, analysis of the subjects' electromyographic signals indicated that the electrode switching strategy reduced mean absolute value *MAV* and median frequency *MF* by approximately 44.21% and 17.97%, respectively, while mitigating muscle fatigue.

Keywords: functional electrical stimulation; multi-electrode switching strategy; upper limb rehabilitation system; electrode array; active disturbance rejection control