

分类号: TP391

密 级: 公开

学校代码: 10363

学 号: 2220140103



安徽工程大学

Anhui Polytechnic University

硕士学位论文

题目 基于视觉的机器人高精度运动测量
方法研究

论 文 作 者 徐晶晶

指 导 教 师 吴路路 副教授

学 科 (专 业) 智能制造工程

研 究 方 向 机器视觉

论文提交日期: 2025 年 5 月 20 日

分类号：TP391

单位代码：10363

密 级：公开

学 号：2220140103

基于视觉的机器人高精度运动测量方法研究
**Research on High-Precision Motion Measurement
Method for Robots Based on Vision**

学生姓名：_____徐晶晶_____

指导教师：_____吴路路 副教授_____

专 业：_____智能制造工程_____

研究方向：_____机器视觉_____

论文答辩日期：2025 年 5 月 9 日

安徽工程大学学位论文原创性声明

本人郑重声明：我恪守学术道德，崇尚严谨学风。所呈交的学位论文，是本人在导师的指导下，独立进行研究工作所取得的成果。除文中已明确注明和引用的内容外，本论文不包含任何其他个人或集体已经发表或撰写过的作品及成果的内容。论文为本人亲自撰写，我对所写的内容负责，并完全意识到本声明的法律后果由本人承担。

学位论文作者签名：徐晶晶

日期：2025年5月9日

安徽工程大学学位论文版权使用授权书

学位论文作者完全了解学校有关保留、使用学位论文的规定，同意学校保留并向国家有关部门或机构送交论文的复印件和电子版，允许论文被查阅或借阅。本人授权安徽工程大学可以将本学位论文的全部或部分内容编入有关数据库进行检索，可以采用影印、缩印或扫描等复制手段保存和汇编本学位论文。

保密 ☐，在_____年解密后适用本版权书。

本学位论文属于

不保密 ☒。

学位论文作者签名：徐晶晶

指导教师签名：吴路路

日期：2025年5月9日

日期：2025年5月9日

基于视觉的机器人高精度运动测量方法研究

摘 要

随着智能制造技术的迅猛发展，工业机器人作为现代制造业的核心装备，在生产制造中的应用日益广泛。如今全球工业机器人数量持续增长，其应用场景更加多元化，这对机器人运行的可靠性、精准度和灵活性提出了更高的要求。研究其高精度运动测量方法对于提升机器人运动性能具有重要意义。传统测量方法在适用性、全局性和测量精度方面仍存在诸多局限。近年来，机器视觉技术与人工智能技术取得突破性进展，为解决这些问题提供了新的思路 and 手段。基于视觉测量的非接触性、全局观测性和高灵活性的优势，本文研究了一种基于视觉的机器人高精度运动测量方法，以期实现工业机器人运动状态的精准解析。

为实现机器人运动观测，提出一种基于改进 YOLOv8 的目标检测算法，解决传统目标检测方法在小目标检测识别方面的精度和效率受限的问题。通过引入轻量化卷积模块 GSConv，减少特征提取计算复杂度，提升模型推理速度；采用加权双向特征金字塔（BiFPN）网络结构替代原有 PAN-FPN 结构，利用可学习的加权机制自适应调整特征层的贡献度，增强多尺度特征融合，提高对小目标的检测性能；引入归一化 Wasserstein 距离（NWD）与完全交并比（CIoU）相结合的损失函数，减少模型对不同尺度特征的敏感性。实验结果表明，该方法能够在小目标检测方面取得更优的检测效果。

在目标追踪基础上，为进一步实现特征点的高精度定位，研究了一种基于改进 Zernike 矩的亚像素特征点定位方法。首先，引入各向异性扩散滤波对图像进行预处理，以抑制噪声并保留边缘细节，提高后续边缘检测的稳定性；其次，针对传统 Zernike 矩精度不足的问题，采用渐变边缘模型来描述图像边缘的灰度变化，使其更符合实际采集图像的边缘特性，基于此来获取亚像素边缘点；最后，对边缘点进行最小二乘直线拟合，实现特征点的高精度像素坐标提取。在此基础上，建立了机器人空间三维重建模型，推导了视觉系统坐标系与机器人坐标系之间的变换关系，最终实现机器人空间运动的高精度量测。

为验证所提方法的效果，搭建了双目视觉测量实验平台，针对本文所提方法展开机器人运动测量综合实验，实验结果表明本文方法可以实现较高精度的机器人运动测量。

关键词：工业机器人，视觉测量，目标检测，特征点定位，三维重建

Research on High-Precision Motion Measurement Method for Robots Based on Vision

ABSTRACT

With the rapid development of intelligent manufacturing technology, industrial robots, as the core equipment of modern manufacturing, are increasingly widely used in manufacturing. Nowadays, the number of industrial robots around the world continues to grow, and their application scenarios are more diversified, which puts forward higher requirements on the reliability, accuracy and flexibility of robot operation. It is of great significance to study its high-precision motion measurement methods to improve the robot motion performance. Traditional measurement methods still have many limitations in terms of applicability, globalization and measurement accuracy. In recent years, machine vision technology and artificial intelligence technology have made breakthroughs, providing new ideas and means to solve these problems. Based on the advantages of non-contact, global observation and high flexibility of vision measurement, this paper investigates a vision-based robot high-precision motion measurement method, with a view to realizing the accurate resolution of the motion state of industrial robots.

In order to realize robot motion observation, a target detection algorithm based on improved YOLOv8 is proposed to solve the problem of limited accuracy and efficiency of traditional target detection methods in small target detection and recognition. By introducing a lightweight convolutional module GSConv, the computational complexity of feature extraction is reduced and the model inference speed is improved; a weighted bidirectional feature pyramid (BiFPN) network structure is used to replace the original PAN-FPN structure, and the contribution of the feature layer is adaptively adjusted using a learnable weighting mechanism to enhance the multiscale feature fusion and to improve the performance of detecting small targets; and the normalized Wasserstein distance (NWD) combined with complete intersection and merger ratio (CIoU) loss function to reduce the sensitivity of the model to different scale features. The experimental results show that the method can achieve better detection results in small target detection.

On the basis of target tracking, a sub-pixel feature point localization method based on improved Zernike moments is investigated for further realizing high-precision

localization of feature points. First, an anisotropic diffusion filter is introduced to preprocess the image to suppress the noise and retain the edge details, so as to improve the stability of the subsequent edge detection; second, in view of the problem of insufficient accuracy of the traditional Zernike moments, a gradient edge model is adopted to describe the gray scale change of the image edges, so as to make it more in line with the edge characteristics of the actual captured image, and based on which the sub-pixel edge points are obtained; finally, a least-squares linear fitting is performed on the edge points, and the edge points are localized. Finally, the least squares linear fitting is performed on the edge points to realize the high-precision pixel coordinate extraction of the feature points. On this basis, the spatial 3D reconstruction model of the robot is established, and the transformation relationship between the coordinate system of the vision system and the coordinate system of the robot is deduced, so as to finally realize the high-precision measurement of the robot's spatial motion.

In order to verify the effectiveness of the proposed method, a binocular vision measurement experimental platform is set up to carry out comprehensive experiments on robot motion measurement for the method proposed in this paper, and the experimental results show that the method in this paper can realize high-precision robot motion measurement.

Xu Jingjing

Supervised by Associate Professor Wu Lulu

KEY WORDS: Industrial robot, Vision measurement, Target detection, Feature point localization, Three-dimensional reconstruction

目 录

摘 要	I
ABSTRACT	II
第 1 章 绪 论	- 1 -
1.1 研究背景与意义	- 1 -
1.2 机器人运动测量技术研究现状	- 2 -
1.2.1 机器人运动测量技术概述	- 2 -
1.2.2 视觉测量技术概述	- 4 -
1.3 视觉高精度测量方法研究现状	- 5 -
1.4 研究内容与组织结构	- 9 -
第 2 章 基于视觉的机器人运动测量基础	- 12 -
2.1 机器人运动测量基础	- 12 -
2.1.1 机器人坐标系变换	- 12 -
2.1.2 视觉测量基础	- 14 -
2.2 双目视觉机器人运动测量	- 15 -
2.2.1 机器人运动视觉测量方案	- 15 -
2.2.2 机器人运动视觉测量过程	- 17 -
2.3 本章小结	- 19 -
第 3 章 基于改进 YOLOv8 的机器人运动追踪	- 20 -
3.1 YOLOv8 模型	- 20 -
3.2 改进 YOLOv8 的机器人运动追踪	- 23 -
3.3 本章小结	- 28 -
第 4 章 特征点精确定位与机器人运动测量	- 30 -
4.1 基于改进的 Zernike 矩的特征点精确定位	- 30 -
4.1.1 各向异性扩散滤波算法	- 30 -
4.1.2 基于改进 Zernike 矩的特征点精确定位方法	- 31 -
4.1.3 特征点定位策略	- 36 -
4.2 机器人空间运动测量	- 38 -
4.2.1 双目视觉系统三维重建模型	- 38 -
4.2.2 机器人坐标系变换	- 40 -
4.3 机器人末端姿态测量	- 41 -
4.4 本章小结	- 42 -
第 5 章 工业机器人运动视觉测量实验	- 43 -
5.1 视觉测量实验平台	- 43 -
5.1.1 视觉系统硬件组成	- 43 -

5.1.2 测量系统软件环境.....	- 44 -
5.2 相机标定实验.....	- 45 -
5.3 机器人运动追踪实验.....	- 46 -
5.3.1 实验环境.....	- 46 -
5.3.2 实验数据集构建.....	- 47 -
5.3.3 评价指标.....	- 48 -
5.3.4 实验结果分析.....	- 48 -
5.4 特征点定位与机器人空间结构运动测量实验.....	- 51 -
5.4.1 特征点定位实验.....	- 51 -
5.4.2 机器人空间运动测量.....	- 52 -
5.4.3 机器人空间姿态测量.....	- 56 -
5.5 本章小结.....	- 58 -
第 6 章 总结与展望	- 59 -
6.1 总 结.....	- 59 -
6.2 展 望.....	- 60 -
参考文献	- 61 -
攻读硕士期间的主要成果	- 66 -
致 谢	- 67 -

第1章 绪论

1.1 研究背景与意义

工业机器人是一种面向工业领域的智能化装备，具备可编程性、自主作业能力以及多功能特性^[1]。随着数字化制造、工业互联网及人工智能技术的快速发展，智能化生产与自动化装配日益成为现代工业生产制造的主流趋势。工业机器人作为先进制造业中具有代表性的机电一体化数字化装备，其技术附加值高，应用领域广泛，是先进制造业的重要支撑技术和信息化社会的重要生产装备，是衡量一个国家制造业竞争力和科技创新水平的重要标志。国外的工业机器人产业起源于20世纪60年代，早期主要应用于汽车制造业的焊接和喷涂等作业任务，后期逐步拓展至机械加工、电子装配、质量检测等工业领域，为制造业的自动化转型提供了强劲动力^[2]。我国工业机器人产业的发展始于20世纪70年代，经历了70年代的萌芽期、80年代的开发期和90年代的实用化阶段，经过多年的发展已经初具规模。近年来，随着政策扶持和产业升级的加速推进，我国工业机器人市场规模持续增长，技术水平不断提升，在智能制造、航空航天、精密加工等多个领域得到了广泛应用，保有量亦呈逐年上升态势^[3]。《“十四五”机器人产业发展规划》指出，“十三五”期间，我国机器人产业取得了长足进步。然而，与国际先进水平相比，我国在自主核心技术、产业基础、高端供给能力等方面仍存在一定差距。为进一步推动机器人产业向中高端迈进，《规划》提出，到2025年，我国将成为全球机器人技术创新策源地、高端制造集聚地和集成应用新高地，制造业机器人密度实现翻番^[4]。《中华人民共和国国民经济和社会发展第十四个五年规划和2035年远景目标纲要》亦将智能制造与机器人技术列为提升制造业核心竞争力的重点方向，机器人的智能制造发展符合国家发展计划和发展远景目标。



图 1-1 工业机器人应用场景

工业机器人在工业领域的广泛应用使其对运动精度、稳定性和测量可靠性提出了更为严格的要求，推动工业机器人向智能化、高精度方向发展是机器人技术演进的关键点。在机器人的设计、生产、运行及维护等各个阶段，其动态和静态

性能以及运行状态评估均非常重要^[5]。一般地，机器人性能评估通过对其运动轨迹测量，通过数据分析将其生成评估指标，如位姿精度、重复定位精度、动态稳定性和振动特性等，这些指标能够直观反映机器人运动性能，并用于结构健康监测与误差补偿^[6]。现已有多种接触和非接触测量方法用于机器人的运动测量，以验证机器人是否符合预设的性能标准^[7]。接触式测量方法采集的数据直观稳定，适用于常见的测量场景，但由于其需要直接安装在机器人结构上，可能会引入附加的负载效应。此外，接触式测量获取的数据点通常相对离散且数据量较少，难以反映机器人的整体运动特性。与之相比，非接触式测量方法能够在不与被测对象直接接触的情况下进行测量，避免了潜在的物理影响，同时可以实现全局的测量。其中，视觉测量技术因其非接触、高精度和全局观测的特点，近年来逐渐成为机器人运动测量领域的重要手段。然而，现有视觉测量方法在实际应用场景中易受环境（如光照等）等因素制约，导致测量精度不足。因此，有必要在传统视觉测量的基础上，探索更为先进的技术手段，以满足机器人高精度、高效率的运动检测需求。

1.2 机器人运动测量技术研究现状

1.2.1 机器人运动测量技术概述

现有的机器人运动测量方法主要分为接触式测量方法和非接触式测量方法。

接触式测量是一种与被测目标直接进行物理接触的测量方法，这类方法通常将传感器等测量设备安装在机器人结构上，以采集所需的测量参数。常见的接触式测量装置有编码器、加速度传感器、应变仪和压电陶瓷传感器等。

（1）编码器主要用于测量机器人关节角度，是工业领域中常见的测量装置。其基本原理是将编码器安装在关节电机或运动结构上，以获得关节旋转角度，并结合结构参数，实现对速度、位置等运动信息的推算^[8]。根据信号输出方式，编码器分为绝对型和增量型两类。前者可以直接测出物体的绝对位置，常使用多个独立的编码器来测量物体的位置，每个编码器对应一个唯一的位置值。后者仅记录物体运动的相对位移，需依赖参考点或初始位置来计算物体的位移变化。尽管编码器具有实时响应的优势，但其测量精度容易受到迟滞效应、摩擦力矩、结构柔性等因素的影响，仅依靠编码器提供的测量数据难以实现高精度的运动测量。

（2）加速度传感器是一种用于测量物体加速度的仪器，通过监测物体的加速度变化，推算目标的运动状态、速度及位移等信息。根据敏感元件类型，加速度传感器可分为电容式、电感式和应变式等类型。这类传感器轻便且测量频带范围宽，无需参考点即可实现测量，但测量噪声大，信号容易产生测量偏差，导致

积分计算速度和位置时产生累积误差^[9]。

(3) 应变仪和压电陶瓷传感器可用于测量结构的局部形变，分别基于电阻应变原理与压电效应来感知位移和加速度信号。应变仪通常与动态信号调理器配合使用，能够提供高灵敏度、低成本的测量方案。压电陶瓷传感器则通过电荷放大器转换信号，可实现高频动态响应，对微小振动测量具有较高的灵敏度。然而，这类传感器的主要局限在于测量区域受限，仅能监测局部运动状态，同时测量信号易受环境噪声干扰，影响测量精度^[10]。



图 1-2 接触式测量装置

非接触式测量方法无需物理接触即可完成运动信息的采集，可分为基于电场的技术，如电容式传感器和电感式传感器等^[11]；基于激光的技术，如 CCD 激光位移传感技术等^[12]，以及基于视觉检测的技术。

(1) 电场技术通过检测目标与传感器间的电场变化来计算位移信息。该方法适用于短距离、高灵敏度测量，但这种方式仅适用于单自由度的测量，在多自由度任务过程中容易引入累积误差。

(2) 激光测量技术是一种基于激光物理特性的高精度非接触式测量方法，该技术主要依赖于激光的稳定性、方向性和高相干性，通过分析光的反射、折射、干涉等特性，实现对目标物体的位移、厚度、振动、速度和形变等物理参数的高精度测量。相较于传统的测量方法，激光测量技术具有测量速度快、分辨率高、非接触测量、抗干扰能力强等显著优势，使其在高精度测量需求的场景下得到了广泛应用。在众多激光测量方法中，三角测量法因其测量精度高、响应速度快、结构相对简单，能够平衡成本与测量需求，成为应用最广泛的技术之一。三角测量法的基本原理是利用已知的激光发射角度和探测器接收到的反射光位置变化，基于几何三角关系计算出目标物体的空间位置信息。这一方法特别适用于短距离和中等距离的高精度测量。与其他激光测量方法相比，三角测量法无需依赖复杂的干涉或相位调制技术，硬件成本较低，测量系统的稳定性较强，因此在工业中得到了广泛应用。

然而，尽管三角测量法在精度、响应速度和应用灵活性方面具有显著优势，但其仍然存在一定的局限性。首先，测量范围受限于探测器的视场角，随着测量

距离的增加,系统的分辨率可能会下降。其次,由于三角测量依赖于光学成像的精度,因此在远距离测量中可能会受到光学畸变、环境光干扰以及目标表面特性的影响,导致测量精度降低。此外,三角测量法在大范围测量时,其系统架构可能变得复杂,并引入额外的标定和校准需求,以确保测量精度。



图 1-3 激光跟踪仪

(3) 视觉检测技术是一种利用计算机视觉技术进行检测和判断的方法,通过视觉系统获取物体表面的图像或影像数据,并利用图像处理和分析算法来提取和分析图像中的信息,从而实现对物体特征、尺寸、位置等参数的检测和判断。相比其他测量方法,具有灵活性强、全局检测等优势。

1.2.2 视觉测量技术概述

视觉测量技术最初用于工业领域,其早期主要应用于工业自动化中的质量控制和检测环节。随着计算机技术的发展,视觉测量也逐渐拓宽至其他领域,如在医疗影像、安防监控、无人驾驶等领域。

常用的视觉检测方法分为被动视觉测量和主动视觉测量两大类。主动视觉测量利用结构光投射在被测物体表面,形成明显的特征,适用于光滑、纹理缺乏、无显著灰度与形状变化的表面区域,可避免同名点匹配困难的问题,改善三维信息的获取精度。根据投射光源模式,主动视觉测量可分为点扫描式、线扫描式和面结构光式三种类型。被动视觉测量无需依赖特殊的光源投射设备,仅利用相机拍摄被测物体的影像数据,建立被测物与相机之间的相对位置关系,从而获取被测物表面的三维信息,所需硬件系统相对简单、适用性较强。

基于视觉的测量研究尚处于发展的初期阶段。Fang 等^[13]采用单目视觉系统,实现了对六自由度工业机器人的运动测量。邱志成等^[14]针对柔性臂振动问题,使用双目视觉系统,通过特征点的三维重建,获取柔性臂的振动特性,实现了对机械臂振动的测量。祝琨等^[15]设计了一个基于双目视觉的运动测量系统,无需重新标定即可实现对三维空间中运动目标的实时跟踪与距离测量。彭聪等^[16]结合双目立体视觉和基于视频相位的分析技术,实现了轻型梁的三维振动测量。

视觉测量技术在工业自动化、振动分析、运动跟踪和健康监测等领域展现出巨大应用潜力。然而,现有视觉测量方法的测量精度仍存在提升空间。因此,进

一步优化视觉测量系统的硬件结构和软件算法,实现更高效、更精准的运动数据采集,仍是该领域未来研究的重要方向。

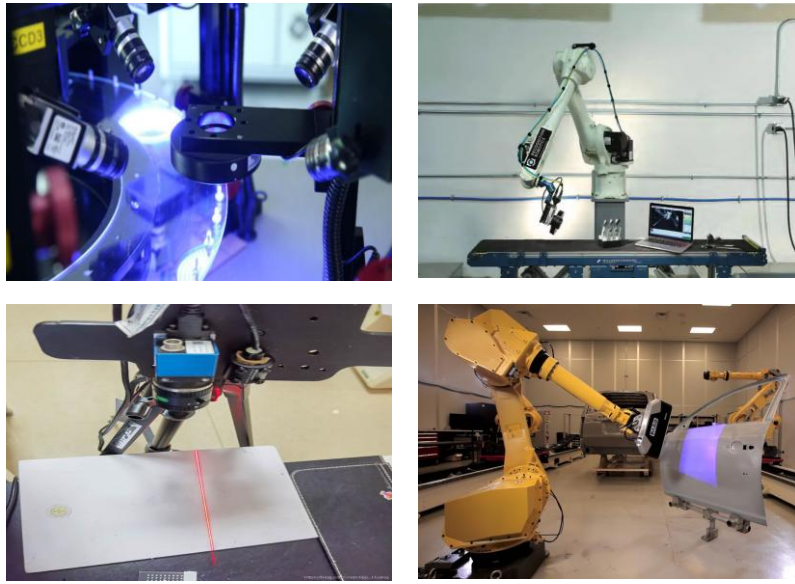


图 1-4 视觉测量技术

1.3 视觉高精度测量方法研究现状

视觉运动测量涉及相机标定、目标检测、特征点提取、三维重建等环节,这些核心技术环节直接关系到测量的精度和效率,是实现高精度运动测量的技术基础。近年来,随着计算机视觉和深度学习技术的飞速进步,视觉测量技术在理论与应用层面均取得了重要突破。以下将从视觉高精度测量方法的主要研究方向出发,探讨其在相关领域的研究进展。

在视觉运动测量过程中,视觉系统通过视觉传感器求解目标运动信息。为实现图像与空间信息的映射,需先进行相机标定。其目的在于获取相机的成像参数以及相机与世界坐标系之间的位姿关系,建立准确的成像几何模型。相机标定的精度将直接影响测量结果的准确性。传统的相机标定方法通过预设成像场景来建立模型,利用场景中的几何约束,选取合适的优化算法来求解相机参数。这类方法通常依赖特定的参照物,早期多采用高精度标定物,通过建立标定物中已知三维空间坐标与二维图像坐标间的映射关系,再利用投影几何变换解出相机的内外参数。Hallert^[17]提出了一种基于最小二乘的方法,对相机参数进行计算。Faig^[18]在研究中引入约束参数,用以对相机成像模型进行表述,通过对参数的非线性优化,显著提高了标定精度。然而,由于非线性优化计算复杂且耗时,实用性受到一定限制。随后,Tsai^[19]提出了经典的“两步法”,其核心思想是通过线性变换初步估计相机参数,并以此为初值,进一步使用非线性优化算法修正模型参数和径向畸变因子。尽管此方法显著提升了计算速度,但未充分考虑镜头切向畸变,因

此在切向畸变严重的场景中效果不够理想。1998 年,张正友教授提出了一种基于二维棋盘格的标定方法^[20],该方法仅需采用黑白格相交的棋盘格作为标定物,与传统标定方式相比,该方法精度高,操作简单,无需依赖高精度的三维标定物,成为了标定方法中的经典代表。

然而,张正友标定法依赖于棋盘格的角点检测,对光照条件、视角变化和图像分辨率较为敏感。此外,在特征点数量较多时,面临精度下降、耗时较长的问题。针对这些问题,王靖等^[21]设计了一种基于圆形特征模板的标定方法,模板特征点通过高斯误差函数建模,以处理圆形特征边缘的模糊问题。结合改进的 Zernike 矩实现特征点精确定位,并通过迭代补偿偏心误差,从而提升圆心拟合和标定的精度。黄臻臻等^[22]提出了基于同心圆模板的标定方法,利用同心圆透视投影的几何模型,建立内外圆之间的偏心误差补偿模型。通过几何约束和优化迭代算法,精准定位圆心,提高了标定精度和稳定性。杨霏等^[23]采用方向性靶标进行标定,靶标设计中加入唯一编码和方向性信息,用以解决遮挡情况下的同名点匹配问题。结合多约束优化算法,提升了标定结果的鲁棒性与适用性。

随着计算机技术的发展,基于智能化技术的相机标定方法被提出。一些智能优化算法与深度学习模型被用于相机标定工作,为提升标定精度和效率提供新的思路。刘屹东等^[24]提出了一种基于动态权重自适应变异粒子群优化的相机标定方法。通过动态调整惯性权重和引入自适应变异策略,提升了粒子群算法的全局搜索能力和收敛精度。实验表明,该方法在标定精度、鲁棒性和收敛速度上均取得了较好的效果。李春明等^[25]提出了一种基于改进灰狼遗传算法的相机标定方法。通过改进灰狼算法的种群初始化、收敛因子和位置更新策略,并融合遗传算法的优化机制,显著提升了标定精度、鲁棒性和可重复性,改进算法在平均重投影误差和计算效率上优于其他优化算法。崔岸等^[26]将 BP 神经网络引入相机标定工作中,但相比传统标定方法,其精度和收敛速度仍存在优化空间。胡志新等^[27]提出基于改进遗传算法优化 BP 神经网络的双目相机标定方法。通过优化遗传算法的交叉和变异概率,提高 BP 神经网络的全局寻优能力,从而获得更高的标定精度。罗亚波等^[28]提出了基于阻滞增长神经网络的双目相机标定方法。利用模拟生物种群增长的机制优化 BP 神经网络的初始权重和阈值,解决了传统 BP 神经网络对初始值敏感的问题,提高了标定精度和收敛性。

目标检测作为视觉运动测量的环节之一,其任务是准确识别被测目标并提取其位置信息。传统的目标检测方法包括模板匹配、滑动窗口和基于手工设计特征的分类器等^[29,30]。其中,模板匹配方法通过与预定义模板的逐一比对来识别目标对象,实现方式简单且无需数据训练,能够满足特定场景的目标检测需求。然而,这种方式对模板的依赖性较强,而且在面对目标尺度、角度和光照条件变化时,

检测精度和速度显著下降。传统目标检测方法通常需要手动设计模板或特征，在简单场景中能够取得一定的效果，但由于对特征表达的依赖性强，在面对多样化的目标形态或背景复杂的场景时检测性能受限。

随着深度学习技术的快速发展，目标检测领域迎来了新的突破。基于深度学习的目标检测方法通过引入卷积神经网络（CNN），依靠其强大的特征学习和表达能力，显著提升了检测精度。与传统方法不同，CNN 不再依赖手动设计特征，而是通过深层网络结构提取具有更高表达能力的复杂特征^[31-33]。近年来，YOLO（You Only Look Once）系列目标检测算法因其高效的检测速度和较高的检测精度，成为当前目标检测领域的主流方法之一^[34]。然而，在小目标检测任务中，YOLO 仍存在不足。小目标的特征在深层网络中容易被丢失或弱化，导致检测精度下降。此外，小目标通常在图像中占据较小区域，受背景干扰较大，加剧了检测难度。因此，如何提升小目标检测性能已成为目标检测算法研究的重要方向之一。GONG 等人^[35]提出了一种面向微小目标特征的多尺度卷积模块特征融合方法，通过设计融合因子来调节浅层与深层特征的融合比例。Sun 等^[36]针对飞行器红外小目标检测中的特征丢失和复杂背景干扰问题，提出了一种基于 YOLOv8 的红外小目标检测模型，设计 α^* -WIoU v3 损失函数优化网络，引入 BiFormer 模块和 GSConv 轻量化卷积，增强对目标信息关注能力的同时降低计算成本，此外通过增加小目标检测层，提升对小目标的检测精度。蒋伟等^[37]针对无人机航拍图像中的小目标检测问题，提出了结合可变形核卷积（AKConv）和高效多尺度注意力机制的 C2f-BE 模块，并在 Neck 中引入小目标检测层和 BiFPN 结构，提升了跨尺度特征融合能力。Yi 等^[38]针对遥感图像中小目标检测的复杂背景和低分辨率问题，设计注意力引导的双向特征金字塔网络，通过动态稀疏注意机制高效提取浅层特征，提高小目标的检测精度。吴明杰等^[39]通过引入双向路由注意力（BRA）提升对小目标区域特征的提取能力，并采用动态目标检测头（DyHead）强化特征表示能力。岳明凯等^[40]针对复杂背景下红外小目标检测易受干扰、回归损失大的问题，提出了一种改进的 YOLOv8_SG 检测算法。在原 YOLOv8 模型的网络结构上，添加了小目标检测层，并引入 SA 注意力机制与 WIoU_v3 损失函数，以增强特征提取能力，提高了对小目标的检测精度。

目标检测技术从传统特征提取方法发展到基于深度学习的端到端检测框架，在检测精度、鲁棒性和实时性方面取得了显著进步。但在实际的应用中，仍面临如小目标检测精度、资源受限环境下的轻量化设计等挑战。未来的研究应继续围绕提升检测速度、检测精度以及增强复杂环境适应能力展开。

目标检测完成后，需要从图像中提取稳定且具有辨识度的特征点。特征点提取方法主要分为传统基于手工特征的方法和基于深度学习的方法。传统的特征点

提取方法依赖于手工设计的图像特征，如角点、边缘、纹理等。Harris 角点检测是一种经典的基于梯度变化的角点检测算法^[41]，它通过计算图像梯度矩阵的特征值来确定角点位置，在旋转变化的场景下表现较为稳定，但在尺度变化较大的情况下鲁棒性较差。随后，Lowe^[42]提出了尺度不变特征变换（SIFT）算法，该方法通过构建高斯金字塔实现多尺度特征提取，并结合梯度方向直方图生成特征描述子，使其具备较强的旋转、尺度和光照不变性。然而，由于 SIFT 的计算复杂度较高，难以满足实时应用需求，Bay 等^[43]提出了加速鲁棒特征（SURF）算法，采用积分图优化卷积计算，并基于 Haar 小波响应加速特征匹配，提高了算法的计算效率，但在复杂光照变化场景下稳定性仍有限。针对计算复杂度和鲁棒性之间的权衡，Rublee 等^[44]提出了 ORB（Oriented FAST and Rotated BRIEF）算法，结合 FAST 角点检测和 BRIEF 描述子，在保证计算效率的同时具备较强的旋转不变性。然而，该方法在尺度变化较大的场景中匹配性能仍存在不足。

基于深度学习的方法通过端到端的网络结构自动学习特征表示，能够更好地适应复杂环境变化，提高特征点的稳定性和匹配精度。SuperPoint^[45]基于自监督学习框架，利用卷积神经网络自动学习特征点检测和描述子生成，使得特征点的稳定性和匹配精度显著提升。Dusmanu 等^[46]提出 D2-Net（Depth and Descriptor Network），直接从深度卷积特征图中提取关键点和描述子，在大尺度场景匹配任务中表现优异。此外，Luo 等^[47]在 D2-Net 的基础上进行了改进，提出了 ASLFeat（Adaptive Supervised Learning Feature）。采用可变形卷积（DCN）提高特征点的局部描述能力，通过特征的层级结构重建分辨率与低层信息，在维持计算量不变的前提下提升特征点精度，并引入峰值度量方法以强化特征点检测能力。

在计算机视觉领域，三维重建技术占据着极为关键的地位。根据传感器工作模式，三维重建方法可分为主动式和被动式。主动式方法基于 RGBD 深度图和点云进行三维重建，主要分为结构光法和飞行时间法。基于结构光的方法通过向目标物体投射已知结构的光斑或条纹，分析其在物体表面的变形情况来计算三维坐标，进一步可分为点云投影法和条纹投影法。点云投影法投射离散的点云图案，通过分析点云在物体表面的分布和位置变化，获取物体的三维信息，具有较高的灵活性和适应性，可用于对复杂形状物体的三维重建。条纹投影法向物体表面投射已知模式的条纹图案，根据条纹在物体表面的变形情况，计算出物体各点的高度信息，进而实现三维重建，但对于纹理缺乏的表面重建效果会降低^[48]。飞行时间法（ToF）是一种基于光波或声波的传播时间差来测量目标物体深度信息的技术^[49]。被动式方法依赖于多幅图像的几何约束恢复三维结构，按照相机数量可分为单目视觉法和立体视觉法。单目视觉法^[50]仅使用单个摄像头的二维图像来恢复场景的三维信息，所需硬件成本较低。但由于单目图像缺乏直接的深度信息，

与双目或多目视觉相比,单目三维重建在数学建模、几何约束以及学习方法方面具有更大的挑战。立体视觉法通过多台相机的几何约束恢复场景的深度信息,主要包括双目立体视觉和多目立体视觉方法。双目立体视觉法模拟人眼视觉原理,使用两个相机从不同位置拍摄同一物体,通过计算左右图像中对应点的视差,结合相机的内外参数,利用三角测量原理计算出物体的三维坐标^[51]。赵珊珊等^[52]提出一种基于双目视觉的三维重建方法,针对移动机器人在室内环境中的感知问题,构建了一个能够同时进行二维动态目标检测和三维场景地图重建的系统,有效提升了三维重建精度。张小国等^[53]提出基于双目视觉惯性的重建方法,推导双目相机提取三维路标点的偏置信息方程,结合 CREStereo 网络和截断符号距离函数,实现高动态和弱纹理条件下的高精度定位和建图。多目立体视觉法采用多个相机组成的系统进行拍摄,可提供更丰富的视角信息,增加冗余信息,提高三维重建的精度和鲁棒性^[54]。马涛等^[55]将多目相机采集到的图像进行高维特征提取,获取三维体素模型,通过采用贝叶斯规则的多目视觉组合模块融合,实现路面宏观纹理的三维重建。Wu 等^[56]提出基于视觉的多视图重建方法,利用多个相机捕获工件多视图图像,重建 3D 点云并与 CAD 模型匹配,实现工件的高精度定位。此外,结合深度学习的方法可利用卷积神经网络对大量的二维图像数据进行学习,能够直接从图像中提取特征并预测物体的三维结构,如 Pix2Vox++模型^[57]通过编码-解码架构生成粗略的 3D 信息,结合多尺度上下文感知融合模块自适应选取高质量结果,可从单张或多张图像中重建出物体的三维结构,在一些特定领域取得了较好的效果。ARAI-MVSNet^[58]采用自适应深度范围预测(ADRP)和自适应深度间隔调整(ADIA)模块,结合空洞空间金字塔特征提取网络(ASPFNet),提升了深度估计精度和模型性能。基于点云的深度学习,如 PointNet^[59]、PointNet++^[60]等模型,能够高效处理大规模点云数据,提升复杂场景的三维建模能力。

1.4 研究内容与组织结构

本文针对机器人高精度运动测量问题,探索通过视觉方式实现高精度测量的方法,以期为实现机器人运动精确调控提供数据基础。文章系统阐述了研究的理论框架、技术实现、实验设计及研究成果,并对未来研究的重点方向进行了展望。具体章节安排如图 1-5 所示:

第一章:阐述了本文的研究背景与意义,调研了机器人运动测量和视觉高精度运动测量方法的研究现状,系统分析了现有测量技术的优势与局限。

第二章:围绕机器人运动测量基础理论展开讨论,系统阐述了机器人运动测

量基础以及视觉实现机器人运动测量的方案及具体过程。

第三章：为实现机器人运动观测，对人工标识进行运动追踪。针对现有方法在小目标检测上精度和效率不足的问题，提出了一种基于改进 YOLOv8 的目标检测算法，详细阐述其原理和实现过程，并通过实验对所提算法的有效性进行了验证。

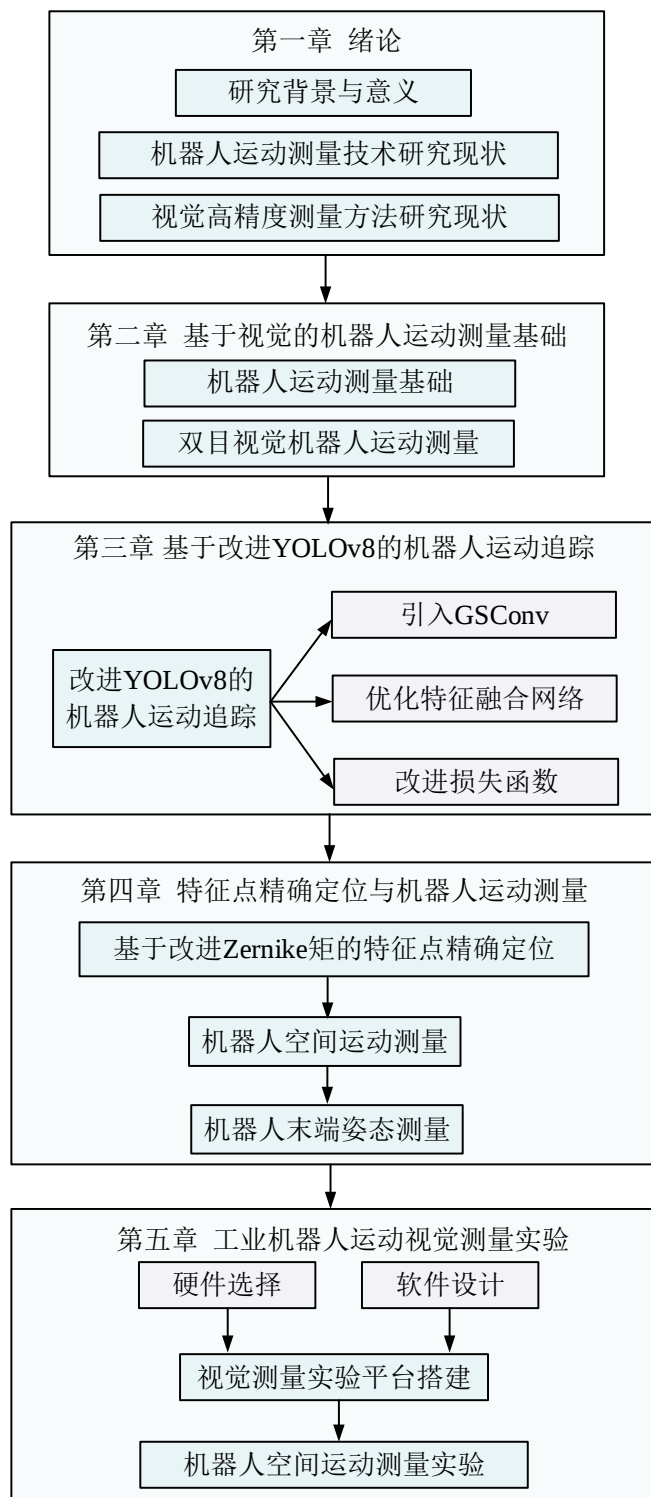


图 1-5 全文组织结构安排

第四章：在机器人运动追踪基础上，进一步开展对合作标识特征点的坐标定位。针对标识的几何特征，设计了特征点检测方案。考虑到实际边缘检测过程中标识边缘定位精度不足的问题，采用了一种基于改进 Zernike 矩的亚像素高精度边缘定位方法，完成了特征点的高精度坐标定位。最后，推导了机器人空间运动测量求解方法，给出视觉系统坐标系与机器人坐标系之间的变换方法以及机器人末端姿态求解方法。

第五章：围绕所提方法，开展机器人视觉运动测量综合实验，搭建了双目视觉测量实验平台，基于 Pytorch+CUDA 框架进行了机器人运动追踪实验。在此基础上，基于 MATLAB 软件进行特征点定位实验以及空间运动测量实验，验证了所提方法的有效性。

第六章：对本文研究工作进行了总结，分析了当前工作的优势和不足，并对未来的研究方向进行展望。

第2章 基于视觉的机器人运动测量基础

机器人运动的本质可归结为其在三维空间中的位置与姿态的动态变化，而这一变化的精准捕获与量化依赖于高精度的测量技术。在众多测量方法中，基于视觉的测量技术因其非接触性、高灵活性和多维信息获取能力，逐渐成为研究的热点。作为视觉测量的核心技术之一，双目视觉利用两个摄像头获取的图像信息进行空间几何重建，具有高精度、丰富信息获取和适应复杂场景的显著优势。本章节从数理基础出发，系统阐述机器人运动测量基础理论，并给出了机器人运动视觉测量方案及过程，为后续研究奠定基础。

2.1 机器人运动测量基础

2.1.1 机器人坐标系变换

空间坐标变换是指在三维空间中，当一个坐标系相对于另一个参考坐标系发生运动时，通过数学方法来描述坐标系间关系的过程，在数学上可理解为一组矩阵运算的组合，涉及平移、旋转或二者的组合变换。

平移变换描述了坐标系间的位置变化，而不影响其方向。若 $P = (x, y, z)^T$ 为参考坐标系中的某一点，一个坐标系相对于参考坐标系在 x 、 y 、 z 方向上的位移表示为平移向量 $t = (t_x, t_y, t_z)^T$ ，那么经过平移变换得到的新坐标以齐次坐标形式表示为：

$$P'_T = (x + t_x, y + t_y, z + t_z, 1)^T \quad (2-1)$$

可以用矩阵表示为：

$$P'_T = TP \quad (2-2)$$

其中，平移矩阵 T 为：

$$T = \begin{bmatrix} 1 & 0 & 0 & t_x \\ 0 & 1 & 0 & t_y \\ 0 & 0 & 1 & t_z \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad (2-3)$$

即平移后的新坐标系 P_{new} 可由原坐标系 P_{old} 左乘平移矩阵得到。因此，平移变换也可表述为：

$$P_{new} = Trans(t_x, t_y, t_z) \times P_{old} \quad (2-4)$$

对于旋转变换，可先理解为参考坐标系和新坐标系的原点与坐标轴均重合，新坐标系绕参考坐标系的某一轴旋转了一个角度。假设新坐标系绕参考坐标系的

x 轴旋转了角度 θ ，旋转后的坐标为 $P'_R = (x', y', z')$ ，该变换过程可表示为：

$$\begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos \theta & -\sin \theta \\ 0 & \sin \theta & \cos \theta \end{bmatrix} \begin{bmatrix} x' \\ y' \\ z' \end{bmatrix} \quad (2-5)$$

旋转变换也可表述为 $P = Rot(x, \theta) \times P'_R$ 。

机器人具有多个本体坐标系，其运动变化通常需要以坐标系之间的变换来描述。如机器人末端在运动过程中，其坐标系相对于基坐标系会经历一系列平移和旋转的变换，该变换过程可分解为依次执行的平移和旋转变换。机器人在三维空间中的运动通常可视为刚体运动，那么其运动过程可被表示为一个 4×4 的齐次变换矩阵：

$$T' = \begin{bmatrix} R & t \\ 0_{1 \times 3} & 1 \end{bmatrix} \quad (2-6)$$

其中， R 为 3×3 的旋转矩阵， t 为 3×1 的平移向量。则对三维点的位姿坐标变换可写为：

$$\tilde{p}' = T \tilde{p} = \begin{bmatrix} R & t \\ 0_{1 \times 3} & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \\ 1 \end{bmatrix} = \begin{bmatrix} Rp + t \\ 1 \end{bmatrix} \quad (2-7)$$

在不同坐标系下的变化中，该位姿的变化也可视为坐标系之间的变化关系，对于机器人常见坐标系中，涉及到空间坐标系 T_w ，机器人基坐标系 T_b 和末端坐标系 T_m 。令 P_w 为空间坐标系下的坐标， P_m 为机器人末端坐标系下的坐标。这种多坐标系变换的方式可以基于齐次矩阵的链式变换表达如下：

$$P_w = T_w^b T_b^m P_m \quad (2-8)$$

使用相机拍摄机器人的运动时，获取的点坐标数据在相机坐标系下表示，而机器人运动数据通常在机器人本体坐标系下表示。因此，以机器人基坐标系为参考坐标系，将机器人的运动数据与测量数据基于坐标系变换的方式统一到参考坐标系下，令 P_c 为相机坐标系下的坐标。此时，公式(2-8)可改写为：

$$P_c = T_c^b T_b^m P_m \quad (2-9)$$

其中， $T_c^b T_b^m$ 的结果即可表示机器人运动视觉测量的过渡矩阵 T_c^r ，令 P_r 为机器人坐标系下的坐标，将其以组合变换的形式展开可得到如下方程：

$$P_r = R_c^r P_c + T_c^r \quad (2-10)$$

2.1.2 视觉测量基础

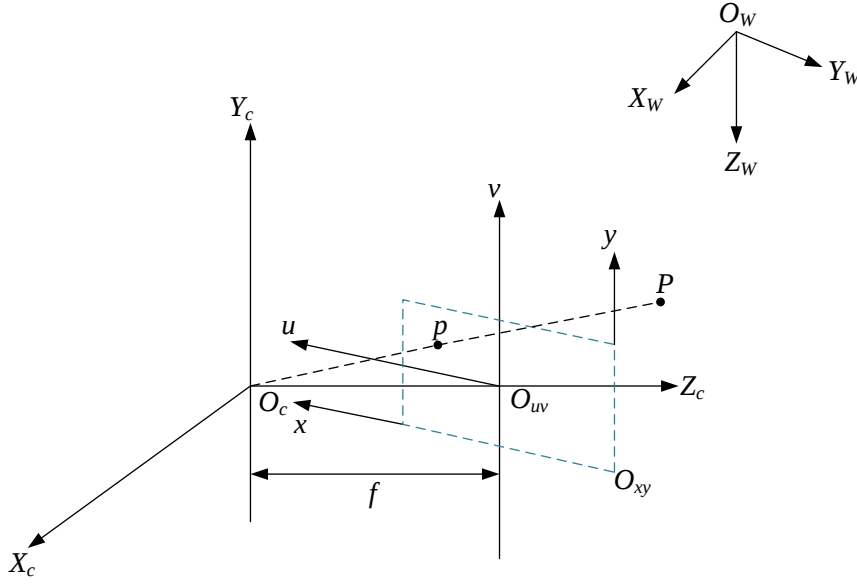


图 2-1 成像模型

视觉测量本质是精确获取三维空间的测量点到二维平面的投影映射，所涉及的相机成像模型的空间几何单元有世界坐标系、相机坐标系、图像坐标系和像素坐标系，其之间的数学转换模型如图 2-1 所示，图中点 P 代表空间中某一点在世界坐标系中的坐标值， p 代表像素在图像坐标系中的坐标值， O_c 代表光心， P 、 p 和 O_c 三点呈共线关系。

设图像坐标系原点到像素坐标系原点的位移为 (u_0, v_0) ， dx 和 dy 分别为每个像素在 x 轴方向和 y 轴方向的物理尺寸，则图像坐标和像素坐标之间的转换关系为：

$$\begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \begin{bmatrix} 1/dx & 0 & u_0 \\ 0 & 1/dy & v_0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} \quad (2-11)$$

点 P 通过透视变换投影到像素坐标系形成像点，点 P 的相机坐标与图像坐标的关系为：

$$\begin{cases} x = f_x \frac{X_c}{Z_c} \\ y = f_y \frac{Y_c}{Z_c} \end{cases} \quad (2-12)$$

则点 P 成像的相机坐标到像素坐标的变换过程可表示为：

$$\begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \frac{1}{Z_c} \begin{bmatrix} f_x/dx & 0 & u_0 \\ 0 & f_y/dy & v_0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} X_c \\ Y_c \\ Z_c \end{bmatrix} \quad (2-13)$$

其中, (u, v) 表示像素坐标, f_x 表示图像坐标的 X 轴等效焦距, f_y 表示图像坐标的 Y 轴等效焦距, $\frac{1}{Z_c}$ 是透视投影的尺度因子。

相机坐标系与世界坐标系之间呈刚体变换关系, 该变换关系可表示为:

$$\begin{bmatrix} X_c \\ Y_c \\ Z_c \\ 1 \end{bmatrix} = \begin{bmatrix} R & t \\ 0 & 1 \end{bmatrix} \begin{bmatrix} X_w \\ Y_w \\ Z_w \\ 1 \end{bmatrix} \quad (2-14)$$

其中, $[X_w, Y_w, Z_w, 1]^T$ 和 $[X_c, Y_c, Z_c, 1]^T$ 分别为空间某点 P 在世界坐标系与相机坐标系中的坐标。令 $\lambda = \frac{1}{Z_c}$, 则完整的相机成像模型可表示为

$$\begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \underbrace{\frac{1}{Z_c} \begin{bmatrix} f_x/dx & 0 & u_0 \\ 0 & f_y/dy & v_0 \\ 0 & 0 & 1 \end{bmatrix}}_K \underbrace{\begin{bmatrix} R_{3 \times 3} & t_{3 \times 1} \\ 0_{1 \times 3} & 1 \end{bmatrix}}_M \begin{bmatrix} X_w \\ Y_w \\ Z_w \\ 1 \end{bmatrix} = \lambda KMP_w \quad (2-15)$$

使用相机观测机器人三维运动时, 可基于以上成像模型, 得到机器人结构任意物点在图像中的二维坐标映射, 进而, 通过三维重建的数学模型, 可将对应的二维坐标还原至相机坐标系的三维坐标系中, 得到视觉系统的测量值。

2.2 双目视觉机器人运动测量

2.2.1 机器人运动视觉测量方案

根据相机数目, 视觉测量方案可以划分为单目视觉、双目立体视觉和多目立体视觉方案。单目视觉仅需一个摄像头获取图像信息, 其成本较低, 安装和调试较为简便, 但由于缺乏视差信息, 仅适用于平面测量任务, 对于空间运动宜采用立体视觉系统。

立体视觉分为双目立体视觉和多目立体视觉。双目视觉与多目视觉相比, 具有系统结构简单、成本较低、计算量较小等优势。双目视觉仅需两台相机即可实

现深度信息的计算，硬件配置要求较低。此外，双目系统的标定和校正相对简单，且算法相对成熟。因此，本文采用双目立体视觉技术对机器人进行运动测量。如图 2-2 所示，双目视觉系统可根据相机的排布方式进一步划分为平行式双目视觉和汇聚式双目视觉。其中，平行式双目系统的两台相机的光轴平行，并且成像平面平行于基线方向，该方式在远距离测量时具有较好的适应性，且计算模型相对简单，适用于大范围场景的三维测量。但在近距离测量场景中，由于目标区域的视差信息较小，深度分辨率可能受到限制。相较之下，汇聚式双目视觉系统通过调整相机的光轴，使其在特定位置交汇，从而在目标区域形成更大的视差范围，能够有效缓解远距离目标的视差饱和问题。

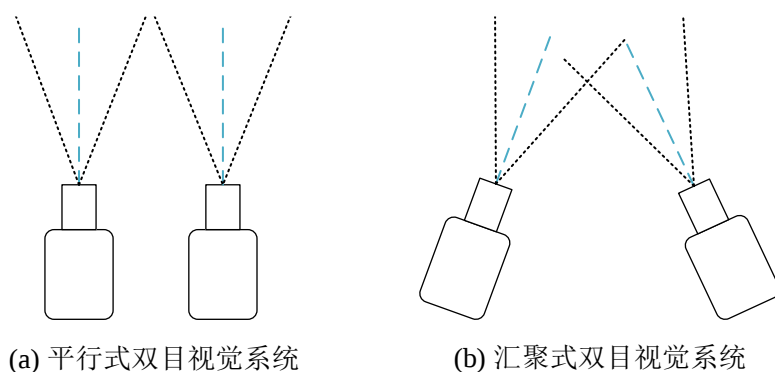


图 2-2 双目视觉系统类型

当使用视觉系统对机器人进行测量时，机器人表面易于识别的自然特征往往较为有限，难以为视觉系统提供稳定、可识别的几何特征信息。为了解决这些问题，需要以人为方式引入可控的几何特征，即采用人工标识固定于机器人表面。人工标识是一种具有明显的可识别特征的标记，例如特定的形状、颜色或图案。图 2-3 所示为机器人运动测量所设计的合作标识，其中，图 2-3(a)为黑白相间的十字交叉标识，用于机器人空间位置检测；图 2-3(b)为直角三角形标识，用于六自由度机器人的姿态检测。



图 2-3 合作标识

如图 2-4 所示，本文搭建的视觉检测平台包括被测对象机器人、辅助单元、高速相机采集模块、计算模块、集成分析模块。各模块的硬件体现如下：机器人包括三自由度直角坐标机器人和六自由度机器人。其中，三自由度机器人用于运动测量，六自由度机器人用于姿态测量。辅助单元包括补光灯、三脚架、人工标识。图像采集模块为两台高速相机与对应数据线、接口、触发器。计算模块与集

成分析模块为上位机与对应的数据处理软件。

当机器人在空间运动时,粘贴于机器人结构面上的标识可被图像采集系统拍摄,得到机器人运动图像数据流;数据流传递给处理模块进行图像处理与目标识别,提取数据流中标识所在的图像的子区域,并通过特征点定位方法精确定位标识特征点在二维图像中的确切坐标,将坐标传递给计算模块;继而,计算模块基于预先双目标定所得的标定参数与三维重建模型,计算出特征点于主相机坐标系下的空间三维坐标,集成分析模块将所有图像数据流所得出的特征点三维坐标进行整合,分析得出视觉系统所测得的机器人空间运动轨迹和位姿变化,从而获取机器人运动数据,用于对机器人运动性能进行分析。

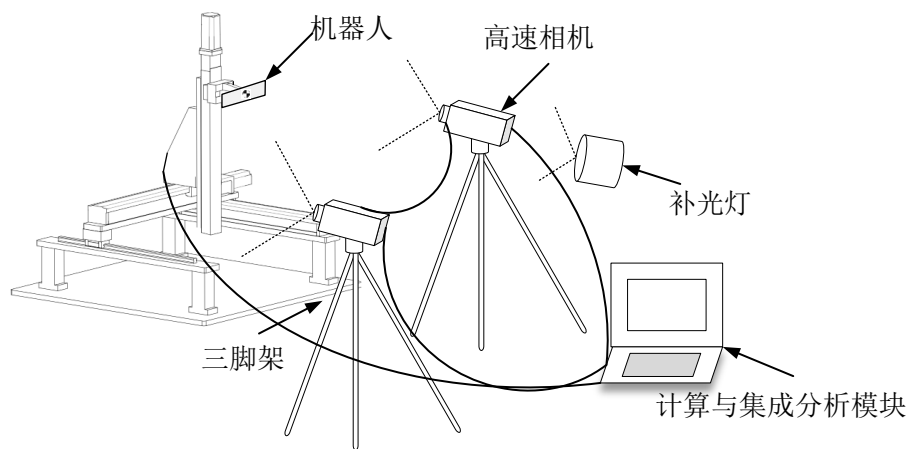


图 2-4 视觉测量平台

2.2.2 机器人运动视觉测量过程

运动测量流程如图 2-5 所示,具体步骤如下:

(1) 实验平台搭建与设备调试: 为确保视觉测量系统的高精度与稳定性,在正式测量前需对视觉测量平台进行调试。实验开始前,需根据机器人的运动范围和测量要求,合理布置测量环境。选择合适的位置安装高速相机并调整相机高度,保证相机视野能够完整覆盖机器人运动区域。同时,合理架设补光设备,采用正向打光方式,并调整补光灯的角度与光强,为实验提供光照均匀稳定的测量环境。细致调整高速相机的镜头焦距,保证相机成像清晰,避免因聚焦不当产生图像模糊或变形,确保测量数据质量稳定可靠。

(2) 相机标定: 成像模型由一系列几何参数构成,由于透镜畸变等非理想因素,成像过程中的几何变换通常是非线性的。因此,首先需要进行相机标定,矫正相机原有的非线性误差,并标定出成像模型与三维重建所需的相机几何参数。标定过程所需的标定物为棋盘格标定板,在图像采集软件中把拍摄模式调整为抓拍模式,拍摄若干张不同角度和位置的标定物图像,通过 MATLAB 标定工

具箱计算相机的内外参数，所得参数用于后续的三维重建过程。

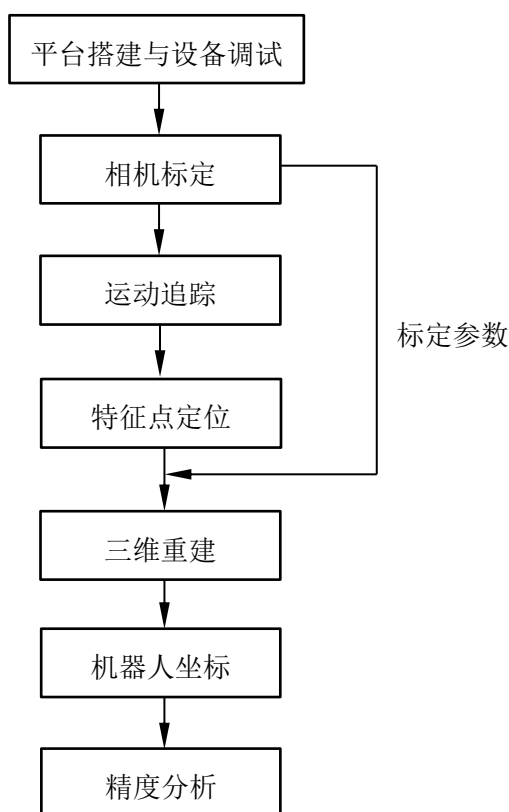


图 2-5 机器人测量流程

(3) 运动追踪：在完成相机标定后，将人工标识粘贴于机器人末端。将相机拍摄模式调整为高速采集模式，设置机器人运动轨迹、相机帧率和拍摄时间。发出机器人运动指令的同时启动高速采集，获取包含人工标识的机器人连续运动图像序列。为实现对人工标识的精确、高效追踪，采用改进的 YOLOv8 目标检测算法追踪其运动，同时与其他目标检测算法进行对比，以验证其有效性。

(3) 特征点定位：对追踪到的标识进行特征点坐标提取，标识中两两直线的交点即为特征点。通过滤波平滑、全局阈值灰度分割和 Canny 边缘检测等图像预处理操作，获得边缘的像素级信息。进而采用基于改进 Zernike 矩的边缘检测算法提取亚像素边缘，获得标识的直线边缘信息，联立直线方程求交点，得到标识特征点的亚像素级坐标。

(4) 三维重建：基于获得的特征点像素坐标，结合相机标定获得的相机内外参数，利用双目视觉三维重建模型求解特征点的空间坐标，获得特征点在相机坐标系下的三维坐标信息。基于直角三角形合作标识，进一步实现机器人末端姿态测量。利用三角标识的三个不共线特征点构建机器人末端运动坐标系，计算运动坐标系与相机坐标系之间的变换关系，并采用 SVD 分解方法对旋转矩阵进行优化，降低非正交误差，提高末端姿态测量精度。

(5) 坐标转换：采用基于点对匹配的坐标转换方法将测量得到的三维坐标

从相机坐标系转换到机器人坐标系，进而对机器人运动测量精度进行分析。

2.3 本章小结

通过视觉系统对机器人进行运动测量研究的过程需要对机器人的运动状态和视觉系统观测数据进行数学建模。本章节围绕机器人运动测量的基础理论展开，阐述了机器人运动测量理论与过程。在机器人运动测量基础部分，阐述了机器人坐标变换过程与视觉测量基础；在双目视觉机器人运动测量部分，给出了机器人运动视觉测量方案并叙述其测量过程，为后续章节运动测量研究奠定了理论基础。

第3章 基于改进 YOLOv8 的机器人运动追踪

上一章介绍了机器人视觉运动测量基础理论，机器人在物理空间的运动可以建模为几何空间中的坐标变化。对于机器人的运动测量，需要通过视觉系统捕获机器人的结构运动信息，将其转化为几何空间中的基元进行解析。在实际测量过程中，通常设计特定的合作标识标记于机器人表面，通过标识的空间运动来间接表征机器人的结构运动轨迹，将机器人的运动转化为标识的空间运动。双目视觉系统通过对合作标识的运动进行跟踪和特征提取，将标识的空间信息转化为可量化的几何运动轨迹。基于此，本章详细介绍了 YOLOv8 基础模型，并提出一种基于改进 YOLOv8 的目标检测方法，对机器人表面标识进行识别与追踪，为后续空间点定位和三维重构提供基础。

3.1 YOLOv8 模型

在目标检测的发展过程中，研究者们致力于在检测速度和检测精度之间寻求平衡。其中，YOLO（You Only Look Once）系列目标检测算法作为单阶段检测算法的代表，凭借其高效的检测速度和出色的精度，在目标检测领域得到了广泛应用。YOLO 是一种实时目标检测算法，自首次提出以来以其高效的实时检测能力迅速成为计算机视觉领域的热点。YOLO 算法的核心思想是将目标检测任务视为一个端到端的回归问题，通过单次网络推理直接预测图像目标的位置和类别，而非传统两阶段算法（如 R-CNN 系列）的候选区域方法。因此，YOLO 系列算法显著提高了目标检测的速度，并在实时应用中取得了优异表现。随后，研究人员在 YOLO 算法的基础上不断优化，先后提出了 YOLOv2、YOLOv3、YOLOv4、YOLOv5 以及 YOLOv8 等多个升级版本。YOLOv2 在原版 YOLO 的基础上，引入了锚框（Anchor Boxes）机制和多尺度训练策略，有效提升了小目标的检测性能。通过使用 Darknet-19 骨干网络，提升了特征提取能力，相比原版 YOLO，实现了检测速度与准确率之间更好的平衡。YOLOv3 引入了更深的骨干网络 Darknet-53 和特征金字塔结构（FPN），从不同尺度融合图像特征，有效地提升了多尺度目标检测能力。此外，引入了逻辑回归（Logistic Regression）方式实现类别预测，更加适应类别丰富的检测任务。YOLOv4 在 YOLOv3 的基础上，采用了基于 CSPDarknet53 的网络主干结构与双向特征融合结构，并结合了 Mish 激活函数和 Mosaic 数据增强等多种高效的数据增强策略，使模型在保持精度的同时降低了计算量。YOLOv5 着重于模型的轻量化和灵活部署，采用了 Focus 层进行特征提取，减少了计算开销，同时保留了更多空间特征。主干网络在 CSPNet

（Cross Stage Partial Network）的基础上进一步优化，减少了计算冗余，使网络在保持高效特征提取能力的同时降低了计算复杂度。YOLOv8 在前述版本的基础上，进一步在网络结构、训练策略等方面进行了优化，检测精度和计算效率相较于前代版本均有所提升。

YOLOv8 模型结构如图 3-1 所示。YOLOv8 主要由输入端（Input）、特征提取网络（Backbone）、特征融合网络（Neck）和预测头（Head）四部分组成。

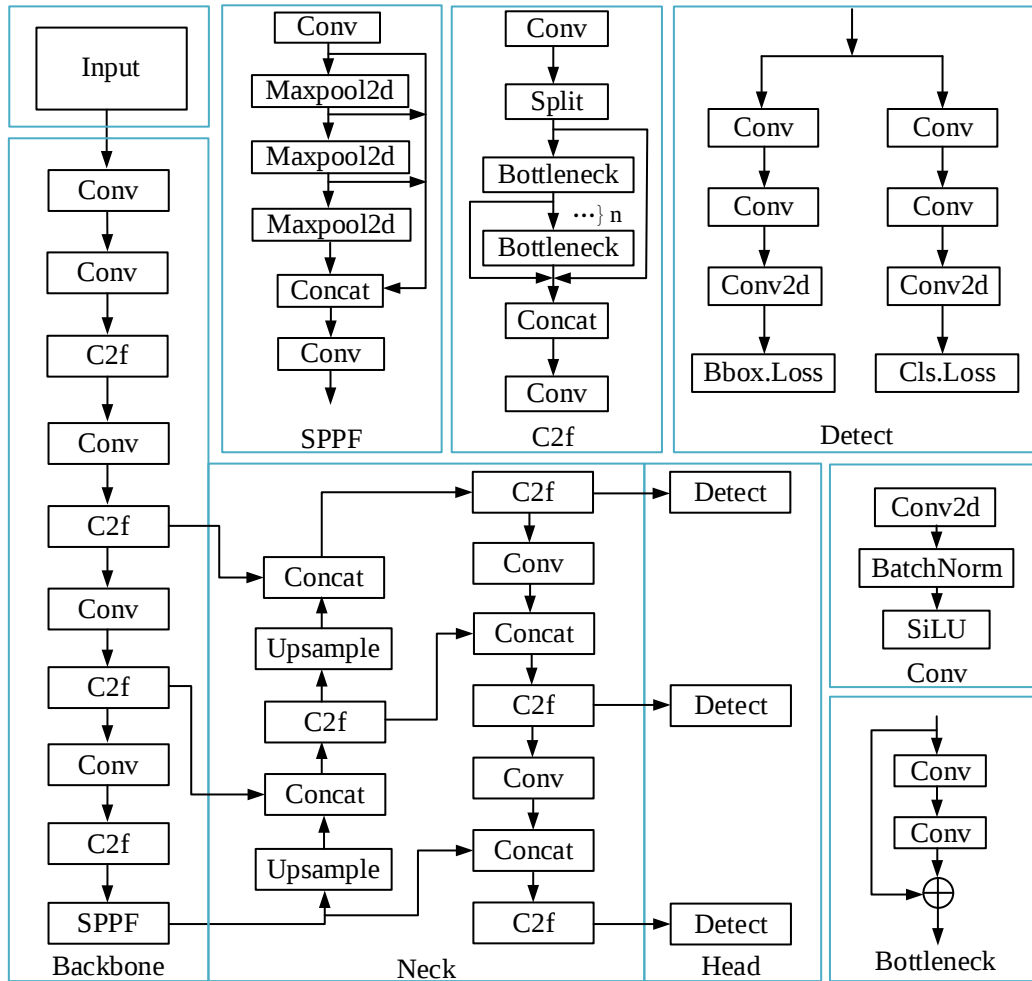


图 3-1 YOLOv8 模型结构

使用 YOLOv8 实现目标检测的过程如下：

（1）通过缩放、填充和归一化等预处理操作，将输入的原始图像调整为 640×640 固定分辨率的图像；

（2）将固定分辨率的图像送入 Backbone 模块提取多尺度特征。Backbone 采用 C2f 模块，如图 3-2 所示，C2f 模块是 YOLOv8 中引入的一种高效特征提取模块，该模块融合了 C3 模块的残差连接设计与高效聚合网络（ELAN）的跨层分支连接策略，增强了特征提取的效率和信息的流动性。此外，通过改进的空间金字塔池化模块（SPPF）对特征图进行不同尺度的池化操作，进一步提高特征图的

表达能力。SPPF 模块是在传统的空间金字塔池化（SPP）基础上改进而来的，如图 3-3(b)所示。与图 3-3(a)所示的传统 SPP 模块采用的并行多个不同尺寸的最大池化操作不同，SPPF 模块采用串行的最大池化设计，通过多个连续的最大池化层来近似获得更大感受野的特征。这种串行设计在功能上与传统 SPP 模块基本等效，但显著减少了模型的计算量与参数量，进一步提高了网络的推理效率。通过这种池化策略，SPPF 模块可以高效地捕获多尺度上下文信息，增强特征表达的能力，同时保留重要的空间位置信息，增强网络对不同感受野特征的捕获能力；

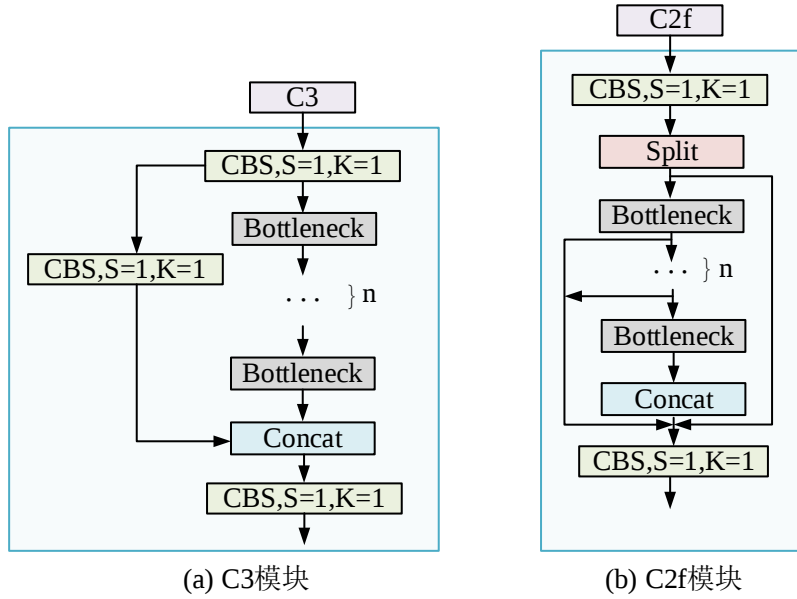


图 3-2 C3 模块和 C2f 模块结构图

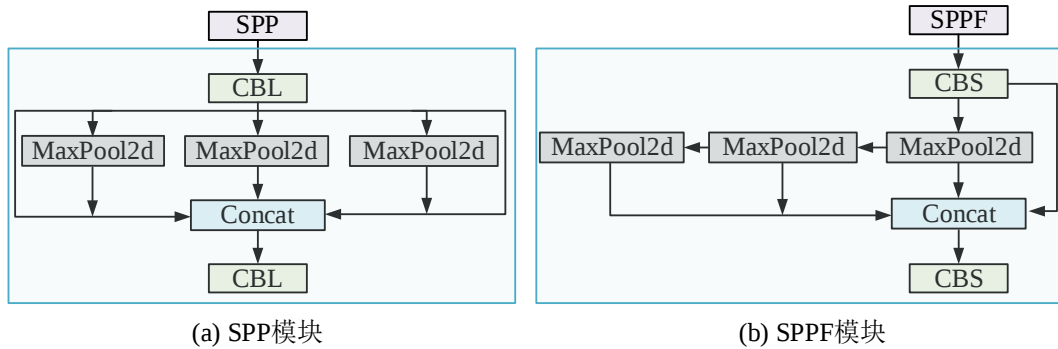


图 3-3 SPP 模块和 SPPF 模块结构图

(3) 将 Backbone 生成的多尺度特征送入 Neck 进行特征融合。Neck 部分采用 PAN-FPN 结构的双向融合策略进行特征处理，能够改善复杂背景下目标检测的能力；

(4) 通过 Head 部分输出目标检测结果。Head 采用解耦头（Decoupled Head）与无锚框（Anchor-Free）设计，结构图如图 3-4 所示。该结构将目标检测任务拆分为分类、目标置信度和边界框回归两个分支，通过变焦加权损失（VFL Loss）

和分布焦点损失（DFL Loss）+完全交并比损失（CIoU Loss）对分类与回归任务分别进行优化，输出三个不同大小的特征图，实现多尺度目标的检测。

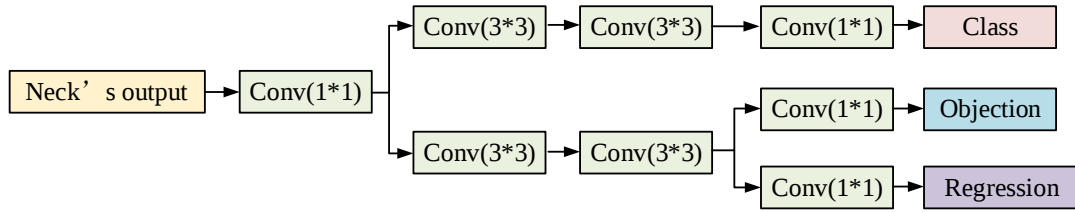


图 3-4 解耦头结构图

3.2 改进 YOLOv8 的机器人运动追踪

虽然 YOLOv8 在目标检测任务中展现出了较高的检测精度和推理速度，但在应对小目标检测时依然存在一些局限性。小目标的特征信息在深层卷积网络的下采样过程中容易被削弱或丢失，从而导致检测精度下降。此外，小目标由于在图像中占据的像素比例较小，容易受到背景噪声和干扰因素的影响，进一步增加了检测的复杂性。针对现有小目标检测算法存在的漏检率高、检测效率不足问题，本文探索以 YOLOv8 为基础的高效、高精度目标检测算法。YOLOv8 作为 YOLO 系列的前沿版本，不仅继承了 YOLO 系列的高效检测特性，同时在网络结构设计和特征提取能力方面进行了优化，使其在目标检测任务中展现更加优异的性能。为了更好地适应机器人运动追踪这一特定任务，本研究对 YOLOv8 网络进行了针对性改进，以增强其对小目标特征的提取能力，提高检测精度，并在复杂背景下保持较强的鲁棒性。

（1）引入 GSConv 模块

GSConv 是一种新型的轻量化卷积模块，其结构如图 3-5(a)所示。GSConv 通过结合普通卷积和深度可分离卷积，能够有效减少计算量和参数量，实现特征的高效提取。其中，深度可分离卷积可进一步分解为深度卷积与逐点卷积。如图 3-6 所示，深度卷积采用单通道的卷积核，分别作用于输入特征图的各个通道，有效捕捉空间特征信息；逐点卷积则采用尺寸为 1×1 的卷积核，用于跨通道融合深度卷积的输出特征，从而降低网络的计算负担。

GSConv 模块通过将输入特征图分为两个平行的分支分别进行卷积操作，其中一个分支采用普通卷积，另一个分支则将普通卷积生成的特征图进一步采用深度可分离卷积进行特征提取。随后，将这两个卷积分支的输出特征在通道维度进行拼接融合，实现普通卷积与深度可分离卷积之间的信息互补。为解决深度可分离卷积固有的通道交互问题，借助混洗（Shuffle）操作通过对融合特征的通道随机重排，将普通卷积与深度可分离卷积的生成信息进行交融，交换不同通道上的

局部特征，有效融合通道信息和空间信息，增强模型的非线性特征表示能力。

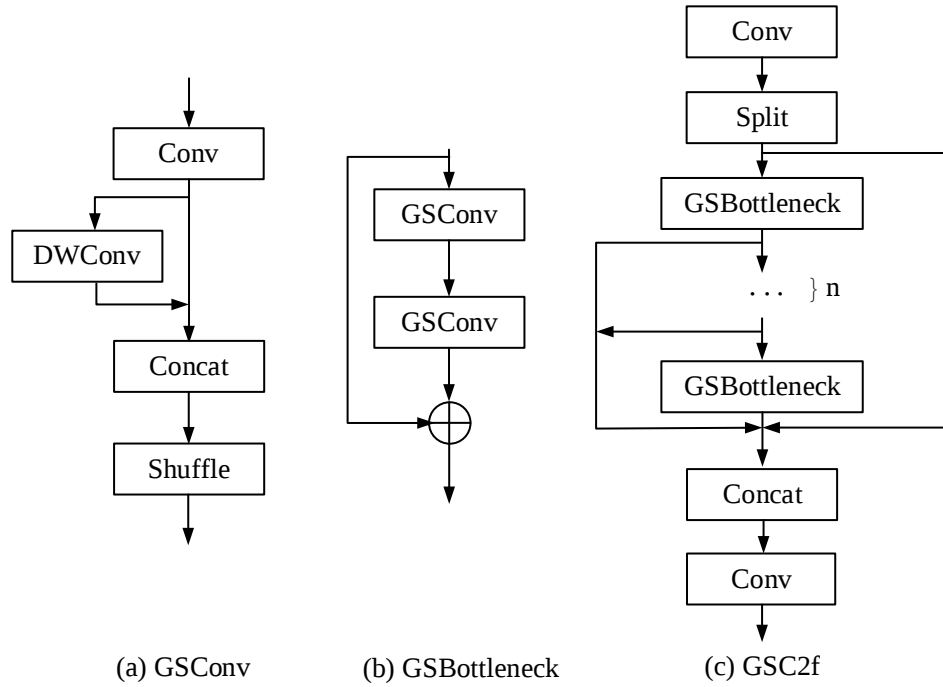


图 3-5 GSConv、GSBottleneck、GSC2f 模块结构示意图

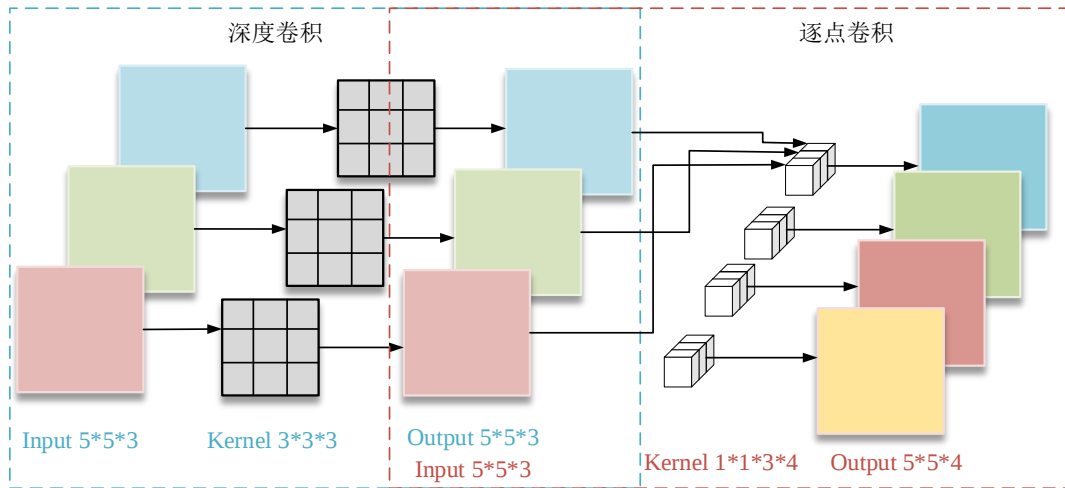


图 3-6 深度可分离卷积

本文在 YOLOv8 中引入 GSConv 模块，重新构建特征融合模块（GSC2f），以优化多尺度特征的处理效率。为降低计算开销，兼顾推理速度和检测精度的平衡，仅在 Neck 部分使用 GSC2f 模块替代 C2f 模块，使用 GSConv 替换 Neck 部分的 Conv 模块。这种局部替换的设计是基于网络不同部分特征作用的差异做出的合理选择。一般地，经过 Backbone 阶段后，输入图像的空间信息往往随着网络深度增加而逐步压缩，其空间维度（即高度与宽度）逐渐降低，而通道数量逐渐增加，这一过程不可避免地造成了部分空间细节信息的丢失。高密度的卷积连接能够尽量保持通道信息之间的相关性，但会导致计算复杂度的增加；相对应地，过于稀疏的卷积结构可能会过度阻断这种跨通道连接，影响特征的表达能力。

GSCnv 模块则通过高效、合理的结构设计，在降低计算负担的同时尽可能地保持通道之间关键的特征交互。

然而，如果在整个网络中均采用 GSCnv 模块，则可能会造成网络结构深度的增加，降低训练效率。当特征图传输至 Neck 部分时，其空间维度已压缩至较低水平，且通道维度较高，进一步的特征压缩已无必要。因此，在 Neck 部分引入 GSC2f 模块是一种更为合理且高效的选择。通过这种设计，可以在不额外损失空间特征表达能力的基础上，实现通道维度上的轻量化特征融合，确保网络在较低计算负担下仍然保持出色的检测精度。基于 GSCnv 设计的 GSBottleneck 和 GSC2f 模块分别如图 3-5(b)、3-5(c)所示。

(2) 优化特征融合网络

YOLOv8 的 Neck 部分采用特征金字塔网络和路径聚合网络(PAN)组成 PAN-FPN 结构，融合由 Backbone 提取的多尺度特征。YOLOv8 的 Backbone 将特征图划分为五个不同尺度的层级，深层特征图具有较大的感受野，分辨率较低，包含丰富的语义信息，适合检测大目标；浅层特征图具有较高的分辨率，能够提供更多的图像细节信息，是小目标检测的重要信息来源。如图 3-7(a)所示，FPN 通过自顶向下的路径将高层特征逐级传递到低层，并通过横向连接与低层特征进行融合，从而增强多尺度特征的语义信息。但这种自顶向下的单向信息流模式对低层细节特征整合的能力有限。为解决这一问题，PANet 在 FPN 的基础上增加了一条自底向上的路径，使低层的细节特征信息能够向上传递至高层，实现双向的信息流动，如图 3-7(b)所示。这种双向特征融合机制不仅保留了高层特征的语义信息，还能有效融合低层的定位信息，增强了网络对不同尺度目标的感知能力。然而，PANet 在特征融合过程中仍存在局限性。PANet 虽然实现了双向的信息流动，但并未考虑它们在融合特征时对各个层级特征的重要性，其融合特征时采用固定权重的简单相加策略，缺乏对不同特征层的贡献度进行动态调整的能力，导致在应对复杂背景和具有干扰因素的小目标检测场景时表现出一定的局限性。

针对上述问题，本文采用加权双向特征金字塔网络(BiFPN)结构对 YOLOv8 中的 PAN-FPN 结构进行替换，以进一步优化特征融合机制。如图 3-7(c)所示。BiFPN 在融合过程中采用了可学习的加权机制，通过为每个特征层分配可学习的权重，使得网络可以自适应地调整每个特征层的贡献度，从而更加灵活地处理小目标检测任务。同时，BiFPN 通过跨尺度连接，保留必要的相邻特征层连接，删除仅有一个输入边的节点，减少了计算开销，提高融合效率，能够在保持较低计算复杂度的同时，显著提升检测精度和鲁棒性。

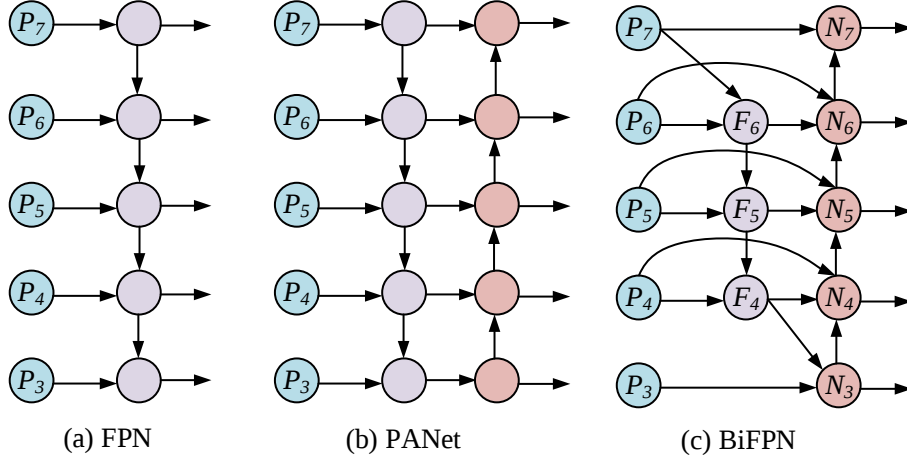


图 3-7 特征金字塔结构

以 P_5 为例，BiFPN 特征融合过程如下：

$$F_5 = \text{Conv}\left(\frac{\omega_1 P_5 + \omega_2 \text{Resize}(P_6)}{\omega_1 + \omega_2 + \varepsilon}\right) \quad (3-1)$$

$$N_5 = \text{Conv}\left(\frac{\omega'_1 P_5 + \omega'_2 F_5 + \omega'_3 \text{Resize}(N_4)}{\omega'_1 + \omega'_2 + \omega'_3 + \varepsilon}\right) \quad (3-2)$$

其中， P_5 为第 5 层的输入特征； F_5 是自上而下的路径中第 5 级的中间特征； N_5 是自下而上的路径中第 5 级的输出特征； Conv 表示卷积运算操作； ω 是各层特征图的权重； Resize 是对高低特征图的上下采样运算。

(3) 改进损失函数

YOLOv8 将 CIoU Loss 和 DFL Loss 相结合，用于优化边界框预测的回归过程。这种组合方法分为两个步骤：第一步，DFL 通过对边界框分布概率进行建模，通过交叉熵的方式，计算预测边界框与真实标注框之间的概率分布损失。第二步，通过 CIoU 进一步评估预测框与真实框之间的整体误差，综合考虑预测框与真实框的重叠程度、中心点的距离和宽高比的相似性，对预测框进行全局优化。CIoU 损失函数的计算公式如下：

$$\text{Loss}_{\text{CIoU}} = 1 - \text{IoU} + \frac{\rho^2(b, b^*)}{c^2} + \frac{4}{\pi^2} \left(\arctan \frac{w^*}{h^*} - \arctan \frac{w}{h} \right)^2 \quad (3-3)$$

其中， IoU 为交并比，衡量预测框和真实框的重叠程度； $\rho^2(b, b^*)$ 为预测框和真实框中心点之间的欧氏距离； c 为预测框和真实框的最小闭合框对角线长度； h 为高度； w 为宽度； w^* 和 h^* 分别为真实框的宽度和高度。

CIoU 虽然在 IoU 的基础上进行了改进，通过引入中心点距离和宽高比相似性优化了边界框的定位能力，但仍存在一定的局限性。首先，CIoU 未对难易样本进行平衡处理，对于小目标检测，因小目标通常占据图像中的较小区域，微小的偏移就会显著减少预测框与真实框的重叠区域，从而导致 IoU 值大幅下降，即

小目标的 IoU 值对偏移非常敏感，这种敏感性会使模型在训练过程中对小目标的梯度更新变得剧烈，优化过程不稳定，增加了优化难度；而大目标对位置偏移的鲁棒性较强，导致模型在优化过程中更倾向于关注大目标的梯度更新。此外，CIoU 在计算宽高比相似性时引入了反三角函数，虽然提升了边界框拟合能力，但增加了模型的计算复杂度，在处理大规模数据集时可能增加额外的开销。最后，CIoU 在优化过程中仍然依赖于 IoU 的重叠区域计算，当预测框与真实框无交集时为 0，导致梯度消失，优化能力受到限制。

针对 CIoU 在小目标检测中的不足，引入一种基于归一化 Wasserstein 距离（NWD）的边界框回归损失函数。其核心思想是将预测框和标注框建模为二维高斯分布，并根据其分布参数计算它们之间的归一化 Wasserstein 距离，量化两者的相似性，从而优化边界框的位置分布。假设预测框和标注框的高斯分布参数分别为 $N_p(\mu_p, \Sigma_p)$ 和 $N_t(\mu_t, \Sigma_t)$ ，其中 $\mu_p = (x_p, y_p)$ 和 $\mu_t = (x_t, y_t)$ 分别为预测框和标注框的中心点， $\Sigma_p = \text{diag}(w_p^2, h_p^2)$ 和 $\Sigma_t = \text{diag}(w_t^2, h_t^2)$ 分别为预测框和标注框的协方差矩阵。二者之间的 Wasserstein 距离在二维高斯分布下的计算公式为：

$$W(\mu_p, \Sigma_p, \mu_t, \Sigma_t) = \|\mu_p - \mu_t\|_2^2 + \text{Tr}(\Sigma_p + \Sigma_t - 2(\Sigma_p^{1/2} \Sigma_t \Sigma_p^{1/2})^{1/2}) \quad (3-4)$$

其中， $\|\mu_p - \mu_t\|_2^2$ 表示预测框和标注框中心点之间的欧式距离。上式可简化为：

$$W(\mu_p, \Sigma_p, \mu_t, \Sigma_t) = \|\mu_p - \mu_t\|_2^2 + \|\Sigma_p^{1/2} - \Sigma_t^{1/2}\|_F^2 \quad (3-5)$$

其中， $\|\cdot\|_F^2$ 表示 Frobenius 范数。带入边界框的具体几何参数，进一步简化得：

$$W(N_p, N_t) = \left\| \begin{bmatrix} cx_p, cy_p, \frac{w_p}{2}, \frac{h_p}{2} \end{bmatrix}^T, \begin{bmatrix} cx_t, cy_t, \frac{w_t}{2}, \frac{h_t}{2} \end{bmatrix}^T \right\|_2^2 \quad (3-6)$$

$W(N_p, N_t)$ 是一种距离度量，其值能反映预测框和标注框之间的几何差异，但无法直接用于模型优化中的相似度评价，如 IoU 通常采用的 0 到 1 之间的相似度值。因此，为了使其更适用于目标检测任务，采用指数归一化的方式，将距离度量转化为新的归一化 Wasserstein 距离。其计算公式为：

$$NWD(N_p, N_t) = \exp\left(-\frac{\sqrt{W(N_p, N_t)}}{C}\right) \quad (3-7)$$

其中， C 为常数，取数据集中目标框的平均绝对大小。

与传统的 IoU 类损失函数不同，NWD 不依赖于预测框与真实框重叠区域的计算，因此，在二者无交集时，NWD 仍然能够有效提供梯度信息。该方法通过

全局性的 Wasserstein 距离，统一度量目标分布间的差异，避免了单纯依赖重叠区域而带来的不足。此外，NWD 对不同尺度的目标不敏感，能够平衡不同大小目标的优化过程，适合用于小目标检测任务。本文采用 CIoU 和 NWD 结合的策略，从几何特性和分布特性两个维度优化边界框，既保留了 CIoU 对边界框几何关系的精确建模，又利用 NWD 对目标分布的全局度量弥补了传统 IoU 类方法的不足。由 CIoU 和 NWD 重构成的损失函数如下：

$$L_{CIoU+NWD} = \alpha \cdot CIoU + (1 - \alpha) \cdot NWD \quad (3-8)$$

其中， α 为权重占比。改进后的 GSBi-YOLOv8 模型结构如图 3-8 所示。

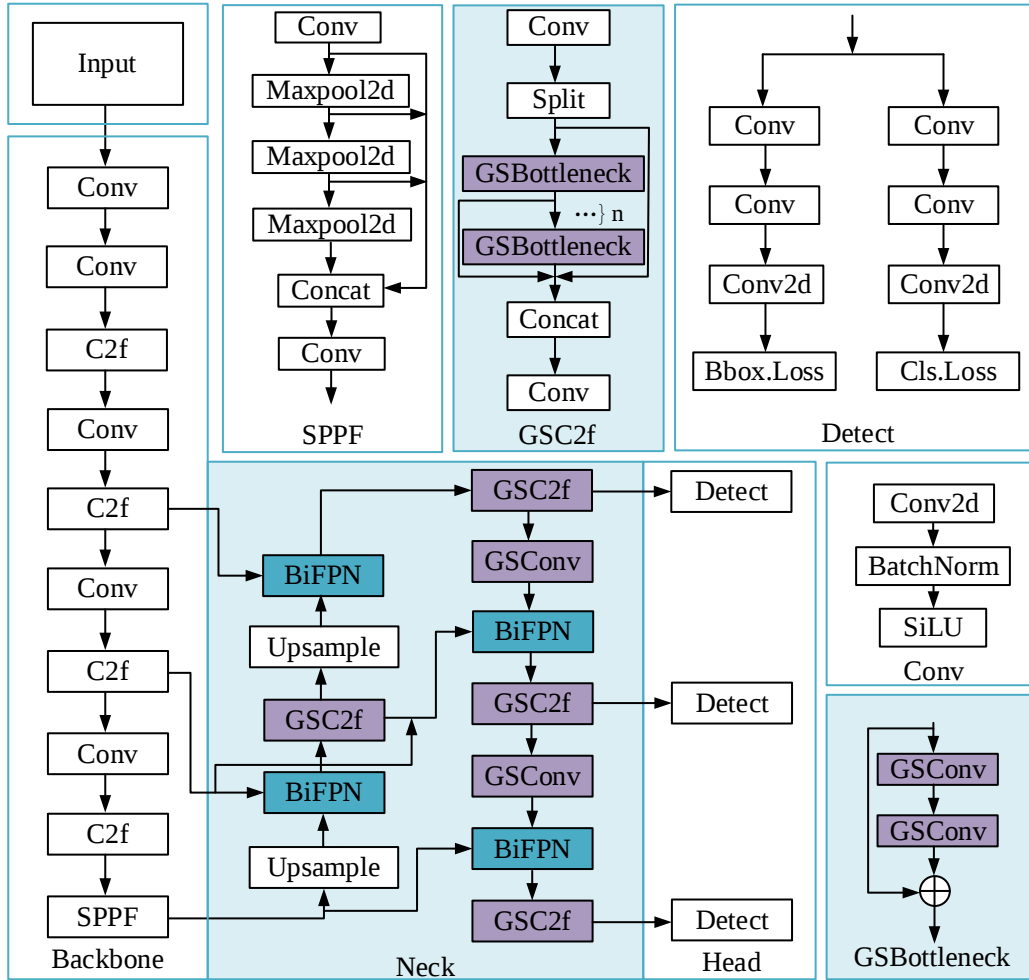


图 3-8 GSBi-YOLOv8 模型结构

3.3 本章小结

本章旨在对机器人表面合作标识进行运动追踪，为后续运动测量奠定基础。针对现有小目标检测算法存在的漏检率高、检测效率不足的问题，探索以 YOLOv8 为基础的高效、高精度目标检测算法。通过在 Neck 部分引入轻量级卷积 GSCConv，在降低计算复杂度的同时保持较高的特征表达能力，在精度与效率

之间实现平衡。同时采用 BiFPN 结构替代原有的 PAN-FPN 结构，通过跨尺度连接方式和可学习的加权机制自适应调整特征层的贡献度，保留更多浅层特征，增强多尺度特征的融合效果，提升小目标检测性能。此外，采用 NWD 与 CIoU 相结合的损失函数，通过 NWD 捕捉目标间的分布差异，结合 CIoU 优化边界框的定位精度，减少对不同尺度特征信息的敏感性，提升对小目标检测的适应能力。

第 4 章 特征点精确定位与机器人运动测量

在上一章节中，通过 GSBi-YOLOv8 算法完成对粘贴在机器人结构表面的标识的识别，获取了标识在整幅图像中的区域，完成对标识的初步定位。实现机器人运动的高精度测量的核心在于准确获取每一帧图像中被测量标识的像素坐标信息。本章在初步定位的基础上，进一步进行合作标识的特征点定位，获取特征点的像素坐标。在此基础上，推导从特征点像素坐标转换为机器人空间坐标系下的三维坐标的转换过程，进而完成对机器人运动的解析。

4.1 基于改进的 Zernike 矩的特征点精确定位

4.1.1 各向异性扩散滤波算法

在图像处理研究中，噪声抑制与边缘保持之间的平衡是一个值得长期关注的问题。传统的图像平滑方法（如高斯滤波）通过空间卷积实现去噪，但往往会导致边缘模糊和细节丢失，从而降低测量精度。为此，本文采用各向异性扩散滤波方法，该方法基于偏微分方程，通过构造梯度驱动的扩散系数，使得扩散过程在平坦区域较强，而在边缘区域受到抑制，保持结构细节，从而在去噪的同时保留重要的结构信息。

图像的各向同性扩散可类比于热传导方程：

$$\frac{\partial I}{\partial t} = D \nabla^2 I \quad (4-1)$$

其中， D 为扩散系数， $\nabla^2 I$ 为图像的拉普拉斯算子， $\nabla^2 I = \frac{\partial^2 I}{\partial x^2} + \frac{\partial^2 I}{\partial y^2}$ 。

该公式描述的是一个均匀介质中的扩散过程，即像素的变化仅取决于其周围像素的均值，这种方法适用于平滑图像，但会模糊边缘。为了在边缘处抑制扩散，引入空间可变扩散系数 $c(x, y, t)$ ，使扩散速率随图像梯度变化：

$$\frac{\partial I}{\partial t} = \nabla \cdot (c(x, y, t) \nabla I) \quad (4-2)$$

其中，选取反比形式的扩散系数：

$$c(x, y, t) = \frac{1}{1 + \left(\frac{|\nabla I|}{K} \right)^2} \quad (4-3)$$

其中， K 是控制边缘敏感度的参数。较小的 K 使得扩散更容易在高梯度区域停止，较大的 K 会使扩散在高梯度区域的抑制作用减弱，使边缘区域发生较强的

扩散。

根据散度定理，展开该式：

$$\frac{\partial I}{\partial t} = \frac{\partial}{\partial x} \left(c(x, y) \frac{\partial I}{\partial x} \right) + \frac{\partial}{\partial y} \left(c(x, y) \frac{\partial I}{\partial y} \right) \quad (4-4)$$

在平坦区域，梯度 $|\nabla I|$ 小， $c(x, y) \approx 1$ ，扩散接近各向同性；在边缘区域，梯度 $|\nabla I|$ 大， $c(x, y) \approx 0$ ，扩散被抑制。

令 $I_{i,j}$ 为代表图像像素值的离散表示，使用有限差分法进行数值逼近：

$$\frac{\partial I}{\partial t} \approx \frac{I_{i,j}^{t+1} - I_{i,j}^t}{\Delta t} \quad (4-5)$$

离散化的各向异性扩散方程可表示为：

$$I_{i,j}^{t+1} = I_{i,j}^t + \lambda \sum_{\eta \in (N, S, E, W)} c_{\eta} (I_{\eta}^t - I_{i,j}^t) \quad (4-6)$$

其中， N 、 S 、 E 、 W 为像素的四个邻域方向； c_{η} 为相应方向的扩散系数； λ 为步长参数，通常设定在 0~0.25 之间。

4.1.2 基于改进 Zernike 矩的特征点精确定位方法

在机器人运动测量中，合作标识的特征点提取精度将影响定位结果的可靠性和准确性。然而，在实际应用中常面临以下问题：标识边缘受光照、噪声等因素影响，表现为渐变而非陡峭变化，难以实现精确定位。传统像素级边缘检测精度有限，难以满足高精度检测的需求。角点检测算法（如 Harris 和 FAST）对噪声较为敏感，且在复杂边缘条件下可能出现误提取。此外，标识在测量过程中可能发生旋转，要求特征提取方法具有较强的旋转不变性。针对这些问题，本文引入改进的 Zernike 矩，其通过单位圆上的正交多项式对边缘区域亮度分布进行建模，既能有效描述边缘渐变特性，又具备旋转不变性和亚像素级精度优势。

如图 4-1 所示，在理想边缘灰度模型中，直线 L 在单位圆内的部分被定义为理想边缘，该直线分隔区域两侧的灰度值分别为 h 和 $h+k$ ，其中 k 表示前景与背景之间的灰度差值。 l 为原点到直线 L 的垂直距离。单位圆内的每个点通过其到边缘的垂直距离 l 和 l 与 x 轴的夹角 φ 来唯一描述。图 4-1(a)为原始的理想边缘模型，图 4-1(b)为图 4-1(a)围绕原点旋转后的理想边缘阶跃模型。

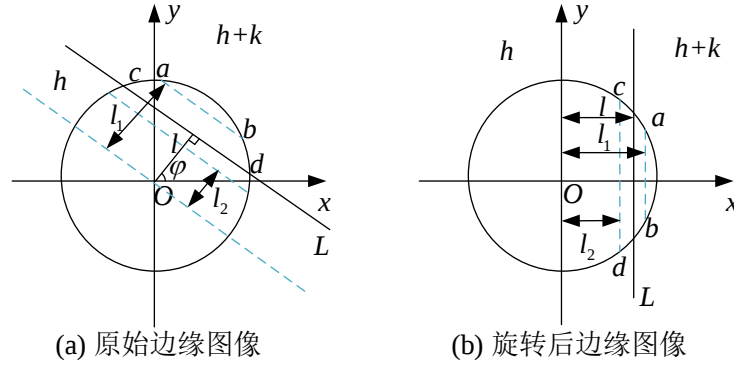


图 4-1 理想边缘阶跃模型

对于二维图像 $f(x, y)$ ，其 Zernike 矩的数学方程如下

$$Z_{mn} = \sum_x \sum_y f(x, y) V_{mn}^*(\rho, \theta), x^2 + y^2 \leq 1 \quad (4-7)$$

其中， m 和 n 为 Zernike 矩的阶数， $*$ 表示复共轭， $V_{mn}(\rho, \theta)$ 为积分核函数，其公式如下：

$$V_{mn}(\rho, \theta) = \sum_{t=0}^{(n-|m|)/2} \frac{(-1)^t (n-t)! \rho^{n-2t}}{t! (\frac{n+|m|}{2} - t)! (\frac{n-|m|}{2} - t)!} e^{-im\theta} \quad (4-8)$$

设图像旋转一个角度 φ 后的 Zernike 矩为 Z'_{mn} ，根据复数运算的规则， Z'_{mn} 与旋转前的矩 Z_{mn} 之间存在以下关系：

$$Z'_{mn} = Z_{mn} e^{-im\varphi} \quad (4-9)$$

这一公式表明，图像在旋转后，Zernike 矩的幅值保持不变，仅相位发生变化，体现了其旋转不变性。 Z'_{00} 、 Z'_{11} 和 Z'_{20} 的积分核函数为：

$$\begin{cases} V_{00} = 1 \\ V_{11} = x + jy \\ V_{20} = 2x^2 + 2y^2 - 1 \end{cases} \quad (4-10)$$

根据图 4-1(b) 所示的理想模型，旋转后的 Zernike 矩计算方法如下所示：

$$\begin{aligned} Z'_{00} &= h\pi + \frac{k\pi}{2} - k \arcsin l - kl\sqrt{1-l^2} \\ Z'_{11} &= \frac{2k(1-l^2)^{3/2}}{3} \\ Z'_{20} &= \frac{2kl(1-l^2)^{3/2}}{3} \end{aligned} \quad (4-11)$$

由式(4-9)的旋转不变性可得：

$$\begin{cases} Z'_{00} = Z_{00} \\ Z'_{11} = Z_{11}e^{-j\varphi} \\ Z'_{20} = Z_{20} \end{cases} \quad (4-12)$$

由 Z'_{11} 和 Z_{11} 的关系可得:

$$\begin{aligned} \text{Im}(Z'_{11}) &= \text{Im}(Z_{11})\cos(\phi) - \text{Re}(Z_{11})\sin(\phi) = 0 \\ Z'_{11} &= \text{Re}(Z'_{11}) = \text{Re}(Z_{11})\cos(\phi) + \text{Im}(Z_{11})\sin(\phi) \end{aligned} \quad (4-13)$$

结合(4-11)可得:

$$\begin{cases} \varphi = \arctan\left(\frac{\text{Im}(Z_{11})}{\text{Re}(Z_{11})}\right) \\ l = \frac{Z_{20}}{Z'_{11}} \\ k = \frac{3Z'_{11}}{2(1-l^2)^{3/2}} \\ h = \frac{Z_{00} - \frac{k\pi}{2} + k \arcsin l + kl\sqrt{1-l^2}}{\pi} \end{cases} \quad (4-14)$$

由此可得图像子像素边缘位置为:

$$\begin{bmatrix} x_{sp} \\ y_{sp} \end{bmatrix} = \begin{bmatrix} x \\ y \end{bmatrix} + l \begin{bmatrix} \cos \varphi \\ \sin \varphi \end{bmatrix} \quad (4-15)$$

Zernike 矩的值是通过将模板和图像进行卷积运算操作得到,若采用 $N \times N$ 大小的模板,考虑模板放大效果,积分单位圆半径变成 $N/2$,同时 l 放大为原来的 $N/2$ 。因此,边缘像素点的亚像素坐标可进一步表示为:

$$\begin{bmatrix} x_{sp} \\ y_{sp} \end{bmatrix} = \begin{bmatrix} x \\ y \end{bmatrix} + \frac{N}{2} l \begin{bmatrix} \cos \varphi \\ \sin \varphi \end{bmatrix} \quad (4-16)$$

上述的亚像素级边缘检测方法是基于理想阶跃模型下的 Zernike 矩,但其并未充分考虑图像实际采集过程中诸多复杂影响因素,如光照分布、相机成像畸变、传感器噪声等。因此,在真实场景中,图像边缘通常并非理想的阶跃变化,导致基于该模型描述难以准确刻画边缘灰度的实际变化特征。实际边缘通常更接近渐变过渡的灰度分布,因此其可用渐变变化的灰度值表示,如图 4-2 所示为边缘模型的二维展示。

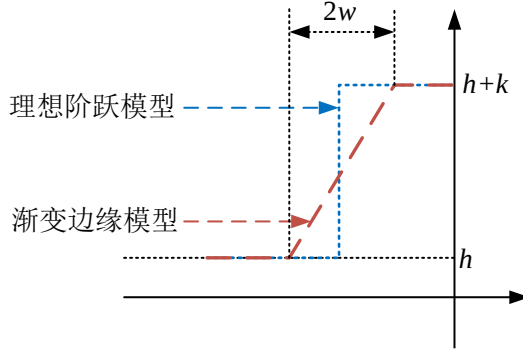


图 4-2 边缘模型

由图 4-2 可知，渐变边缘模型相较于理想阶跃模型，引入了参数 w ，采用如式(4-17)的分段函数描述渐变边缘模型：

$$f(x') = \begin{cases} h, & x' \leq l - w \\ h + k[x' - (l - w)] / (2w), & l - w < x' < l + w \\ h + k, & x' \geq l + w \end{cases} \quad (4-17)$$

则渐变边缘模型的 Z'_{11} 、 Z'_{20} 计算公式如下：

$$\begin{aligned} Z'_{11} = & \frac{k}{24w} [3\arcsin(l + w) - 3\arcsin(l - w) \\ & - (5(l - w) - 2(l - w)^3)(1 - (l - w)^2)^{1/2} \\ & + (5(l + w) - 2(l + w)^3)(1 - (l + w)^2)^{1/2}] \end{aligned} \quad (4-18)$$

$$Z'_{20} = \frac{k}{15w} [(1 - (l - w)^2)^{5/2} - (1 - (l + w)^2)^{5/2}] \quad (4-19)$$

由以上两个矩的计算公式可看出，存在以下两种情况：

(1) 当 w 趋于 0 时， Z'_{11} 和 Z'_{20} 的值与理想阶跃模型下的计算公式相同，由此可见，理想阶跃模型是渐变边缘模型在参数 w 趋于 0 时的形式。

(2) 当 w 不趋于 0 时，公式包含三个未知数，无法直接解析出 l 的精确解。关于 l 的近似表达可如下式：

$$\frac{Z_{20}}{Z'_{11}} \approx l [1 - (1 + l/2)w^2] \quad (4-20)$$

由此可求解出新的表达式如下：

$$l_m = \left[1 - w^2 - \sqrt{(w^2 - 1)^2 - 2w^2 Z_{20} / Z'_{11}} \right] / w^2 \quad (4-21)$$

计算边缘点处的法向量：

$$n = \begin{bmatrix} n_x \\ n_y \end{bmatrix} = \begin{bmatrix} \cos \varphi \\ \sin \varphi \end{bmatrix} \quad (4-22)$$

进一步，可求解出新的边缘点亚像素坐标：

$$\begin{bmatrix} x'_{sp} \\ y'_{sp} \end{bmatrix} = \begin{bmatrix} x \\ y \end{bmatrix} + \frac{N}{2} l_m \begin{bmatrix} \cos \varphi \\ \sin \varphi \end{bmatrix} \quad (4-23)$$

对于 w 值的选取，本文采用一种自适应参数选取策略。该方法的基本思想是，从边缘的初步定位点开始，沿着边缘法线方向在两边进行采样，并利用上述渐变边缘模型拟合，最终确定边缘宽度 $2w$ 。边缘的初步定位点可按照一定间距进行采样，为更精确地确定采样点的灰度值，本文使用双线性插值法计算。插值计算时，选取采样点最近的四个像素点进行估值计算。

双线性插值法的原理是在水平和垂直方向分别执行线性插值计算。如图 4-3 所示，若点 Q 为待插值的点，其周围的像素点为 A 、 B 、 C 、 D ，并已知这些点的灰度值为 $f(A)$ 、 $f(B)$ 、 $f(C)$ 、 $f(D)$ ，插值计算的具体步骤如下：

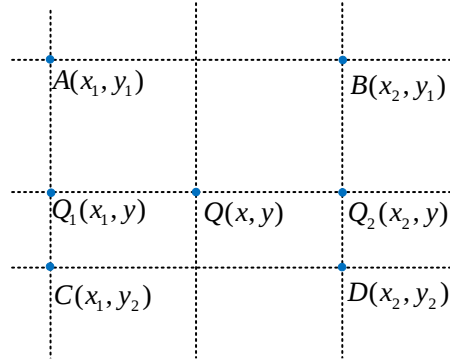


图 4-3 双线性插值示意图

首先，在 AC 和 BD 方向上进行线性插值操作，获取点 P_1 和 P_2 的灰度估算值如下：

$$f(P_1) = \frac{y - y_1}{y_2 - y_1} f(A) + \frac{y_2 - y}{y_2 - y_1} f(C) \quad (4-24)$$

$$f(P_2) = \frac{y - y_1}{y_2 - y_1} f(B) + \frac{y_2 - y}{y_2 - y_1} f(D) \quad (4-25)$$

然后，在 P_1P_2 方向执行线性插值运算，最终得到点 P 的灰度值如下：

$$f(P) = \frac{x_2 - x}{x_2 - x_1} f(P_1) + \frac{x - x_1}{x_2 - x_1} f(P_2) \quad (4-26)$$

获得初定位点周围的采样点后，采取渐变边缘模型的公式拟合，以进一步优化边缘信息。为更简洁地表达计算过程，可采用公式(4-27)表述。最终，拟合得到的参数 h 、 k 、 x_1 、 x_2 中， $x_2 - x_1$ 即表示 $2w$ 。

$$I(x) = \begin{cases} h, & x < x_1 \\ h + k \frac{x - x_1}{x_2 - x_1}, & x_1 \leq x \leq x_2 \\ h + k, & x > x_2 \end{cases} \quad (4-27)$$

基于改进 Zernike 矩的亚像素边缘检测步骤如下：

- (1) 对图像进行滤波等预处理，并提取边缘的像素级初定位点；
- (2) 求解像素级边缘点处的 Zernike 矩 Z_{11} 、 Z_{20} ；
- (3) 依据线性渐变边缘模型拟合边缘，计算出边缘宽度 $2w$ ；
- (4) 求解 l 与 x 的夹角： $\varphi = \arctan\left(\frac{\text{Im}(Z_{11})}{\text{Re}(Z_{11})}\right)$ ；
- (5) 求解子像素边缘点到像素级边缘点的距离 l ；

- (6) 求亚像素边缘点坐标： $\begin{bmatrix} x_{sp} \\ y_{sp} \end{bmatrix} = \begin{bmatrix} x \\ y \end{bmatrix} + \frac{N}{2} l \begin{bmatrix} \cos \varphi \\ \sin \varphi \end{bmatrix}$ 。

4.1.3 特征点定位策略

在图 2-3(a)中，定义标识图案中心点为特征点，如图 4-4 所示，对于该中心点，可采用基于几何特征的计算方法，通过对十字标识的两条垂直直线进行拟合，求取其交点坐标。为了实现特征点的精准提取，本文采用一种结合 4.1.2 节 Zernike 矩亚像素点提取与直线拟合的特征点提取方法。定位流程如图 4-5 所示。首先，对输入图像进行预处理，通过 4.1.1 节各向异性扩散滤波算法平滑图像以减少噪声干扰，并通过全局阈值灰度分割将图像由灰度格式转化为二值形式。然后，使用 Canny 边缘检测算子提取交叉线的像素级别的边缘信息。接着，基于改进 Zernike 矩计算边缘的矩特征，进行亚像素级边缘细化。基于此，根据式(4-28)计算边缘点相对于中心点的斜率，判断其属于横向或纵向边缘，据此对边缘点进行分类。最后，采用最小二乘法对提取的两组边缘点分别进行直线拟合，拟合直线的交点即为特征点坐标。

$$k_i = \left| \frac{x_i - x_{center}}{y_i - y_{center}} \right| \Rightarrow \begin{cases} k_1 \leq k_i \leq k_2, (x_i, y_i) \subseteq l_1 \\ \text{else}, (x_i, y_i) \subseteq l_2 \end{cases} \quad (4-28)$$

其中， (x_i, y_i) 为待分类的边缘点坐标， (x_{center}, y_{center}) 为亚像素点质心， k_1 、 k_2 为水平边缘点的判断阈值。

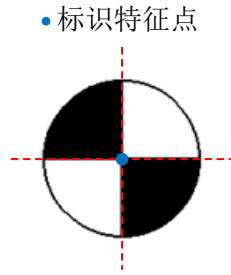


图 4-4 圆形内十字标识特征点定位示意图

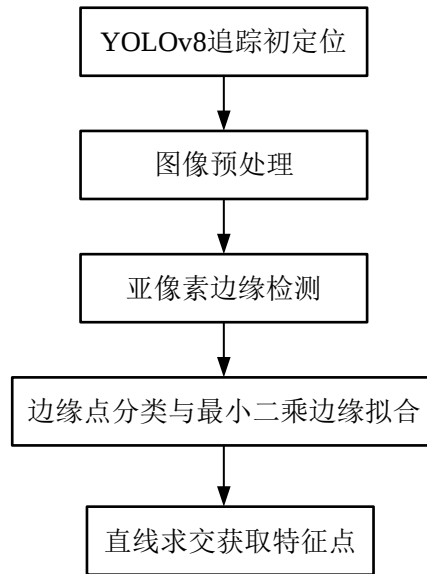


图 4-5 十字标识特征点定位流程图

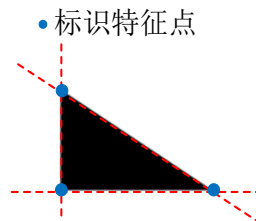


图 4-6 直角三角形标识特征点定位示意图

如图 4-6 所示，定义图 2-3(b)直角三角形标识的三个顶点为特征点，求解思路与十字标识特征点一致，每两条边的交点即为三角形顶点。图像的预处理步骤和亚像素边缘提取与十字标识特征点提取中描述的过程相同。在此基础上，为了提高边缘点分类的准确性，在 Canny 边缘检测的基础上，引入霍夫直线变换以提取直线边缘作为参考，辅助后续的定位与分类工作。然后依据点到参考直线的距离大小关系对亚像素边缘点进行分类。最后，采用最小二乘法对三条边进行拟合。依据几何结构，将三角形的三条边定义为直角短边、直角长边和斜边，相邻两条边的交点即为特征点。三角形标识特征点检测算法流程如图 4-7 所示。

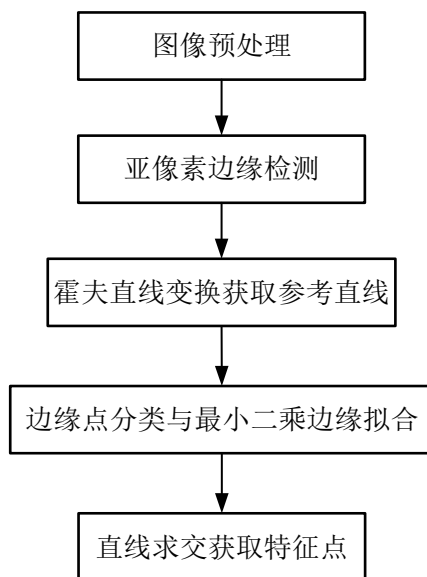


图 4-7 三角形标识特征点定位流程

4.2 机器人空间运动测量

4.2.1 双目视觉系统三维重建模型

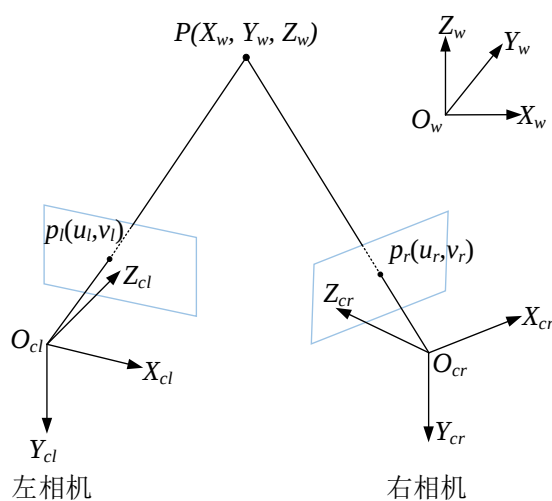


图 4-8 三维重建模型

获取特征点坐标后，利用双目视觉系统对特征点进行三维重建。重建过程基于图 4-8 所示的三维重建模型。双目视觉系统通过两台相机对三维空间同一点 P 进行成像，并结合相机标定参数，实现空间特征点的三维坐标恢复。双目测量系统的数学模型可表述如下：

$$\begin{cases} P_l = R_l P_w + t_l \\ P_r = R_r P_w + t_r \end{cases} \quad (4-29)$$

其中， P_l 和 P_r 分别为左右相机坐标系中的点坐标， R_l 、 R_r 和 t_l 、 t_r 分别为

左右相机的旋转矩阵和平移矩阵，消去式(4-29)中的 P_w ，可得右相机坐标系和左相机坐标系之间的关系式：

$$P_r = R_r R_l^{-1} (P_l - t_l) + t_r \quad (4-30)$$

以左相机为基准相机，则右相机到左相机的空间变换关系可表述为 $P_l = R P_r + t$ ，带入式(4-30)可得

$$\begin{cases} R = R_r R_l^{-1} \\ t = t_r - R t_l \end{cases} \quad (4-31)$$

结合相机成像模型，可得到双目视觉系统的三维重建模型表示为

$$\begin{cases} \lambda_l p_L = K_L P_L \\ \lambda_r p_R = K_R P_R \\ P_L = R P_R + t \end{cases} \quad (4-32)$$

式(4-32)中， K_L 和 K_R 分别是左右相机的内参矩阵， λ_l 和 λ_r 为尺度因子。 R 和 t 描述了两个相机之间的相对位姿关系， p_L 和 p_R 分别为像素坐标 p_l 和 p_r 的向量形式， P_L 和 P_R 分别表示重建点 P_l 和 P_r 的向量形式。将式(4-32)的第三方程 $P_L = R P_R + t$ 带入第一方程 $\lambda_l p_L = K_L P_L$ 中，可以得到如下形式：

$$\lambda_l p_L = K_L (R P_R + t) \quad (4-33)$$

为消除 λ_l ，将式(4-33)等号两侧均叉乘 p_L ，得到：

$$p_L \times K_L R P_R + p_L \times K_L t = 0 \quad (4-34)$$

在上式中提取 P_R ，整理为如下形式：

$$\begin{bmatrix} pKR_1 \cdot pKt_3 - pKR_3 \cdot pKt_1 \\ pKR_2 \cdot pKt_3 - pKR_3 \cdot pKt_2 \end{bmatrix} [P_R] = 0 \quad (4-35)$$

其中， pKR_i 表示 $p_L \times K_L R$ 结果的第 i 个行向量， pKt_i 表示 $p_L \times K_L t$ 结果的列向量第 i 个元素。类似地，对公式 $\lambda_r p_R = K_R P_R$ 进行同样处理，可得：

$$\begin{bmatrix} K_{R11} & K_{R12} & K_{R13} - u_R \\ K_{R21} & K_{R22} & K_{R23} - v_R \end{bmatrix} [P_R] = 0 \quad (4-36)$$

最后，合并式(4-35)和式(4-36)如下：

$$\begin{bmatrix} pKR_1 \cdot pKt_3 - pKR_3 \cdot pKt_1 \\ pKR_2 \cdot pKt_3 - pKR_3 \cdot pKt_2 \\ K_{R11} & K_{R12} & K_{R13} - u_R \\ K_{R21} & K_{R22} & K_{R23} - v_R \end{bmatrix} [P_R] = 0 \quad (4-37)$$

采用 SVD 分解法对方程进行求解，即可得到重建出的坐标 P_r 。

4.2.2 机器人坐标系变换

完成双目测量系统的三维重建后，所获取的点坐标数据位于相机坐标系下，而机器人的运动数据通常在机器人坐标系下表示。由于双目测量系统的重建结果无法直接对应于机器人坐标系，为了实现测量数据与机器人运动数据的一致性，需要将相机坐标系下的数据转换到机器人坐标系下，这一过程需要确定相机坐标系到机器人坐标系之间的刚性变换矩阵，即旋转矩阵 R_c^r 和平移向量 t_c^r 。两个坐标系间的转换关系如下：

本文 2.1.1 节已推导出了相机坐标系到机器人坐标系之间所需的过渡矩阵，现通过奇异值分解法给出过渡矩阵的求解过程，将原有公式改写为如下形式：

$$P_i^r = R_c^r P_i^c + t_c^r \quad (4-38)$$

其中， P_i^c 和 P_i^r 分别为相机坐标系与机器人坐标系下的对应点坐标。首先构建最小二乘目标函数如下：

$$f(R_c^r, t_c^r) = \arg \min \sum_{i=1}^n \|P_i^r - R_c^r P_i^c - t_c^r\|^2 \quad (4-39)$$

对该函数关于 t_c^r 求偏导并令导数为零，可得：

$$t_c^r = \frac{1}{n} \sum_{i=1}^n (P_i^r - R_c^r P_i^c) = \overline{P^r} - R_c^r \overline{P^c} \quad (4-40)$$

其中， $\overline{P^r}$ 和 $\overline{P^c}$ 分别表示机器人坐标系与相机坐标系下所有对应点的质心。令 $P_i^r - \overline{P^r} = m_i^r$ ， $P_i^c - \overline{P^c} = m_i^c$ ，结合式(4-39)和式(4-40)得：

$$f(R_c^r, t_c^r)_{\min} = \sum_{i=1}^n [(m_i^r)^2 + (m_i^c)^2 - 2m_i^r R_c^r m_i^c] \quad (4-41)$$

定义 $M^c = \{P_1^c - \overline{P^c}, P_2^c - \overline{P^c}, \dots, P_n^c - \overline{P^c}\}$ ， $M^r = \{P_1^r - \overline{P^r}, P_2^r - \overline{P^r}, \dots, P_n^r - \overline{P^r}\}$ ，采用奇异值分解(SVD)法对矩阵进行分解：

$$SVD(M^r R_c^r M^c) = U \Sigma V^T \quad (4-42)$$

那么 R_c^r 的解为：

$$R_c^r = UV^T \quad (4-43)$$

最后，将求得的 R_c^r 带入式(4-40)，即可计算出 t_c^r 。

4.3 机器人末端姿态测量

在三维空间中，三个不共线点能够唯一确定一个刚体的空间位置和朝向，因此，通过获取机器人末端上任意不共线的三个靶标点位坐标，可以建立稳定的参考基准，求解出末端空间姿态。一般地，需在机器人末端安装一个具有至少三点的合作标识，构建末端运动坐标系，进而计算末端姿态。基于这一原理，本文采用了一个如图 4-9 所示的直角三角形合作标识进行运动坐标的建系，那么三角形的三个顶点可构成不共线的三个点坐标，基于 4.2 节所描述的方法重构出特征点在相机坐标系下的空间坐标，记作 $P_1 = (x_1, y_1, z_1)$, $P_2 = (x_2, y_2, z_2)$, $P_3 = (x_3, y_3, z_3)$ 。

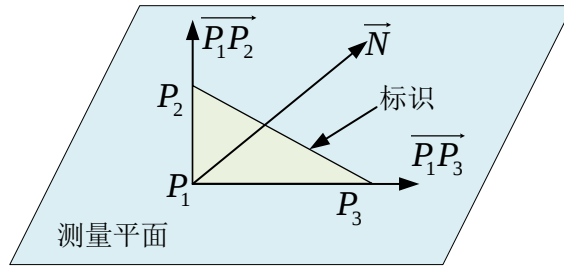


图 4-9 位姿坐标系构建示意图

根据直角三角形三条边的关系，运动坐标系的坐标轴表述如下：

$$\begin{aligned}\vec{X}_{tcp} &= \frac{[x_2 - x_1, y_2 - y_1, z_2 - z_1]}{\sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2 + (z_2 - z_1)^2}} \\ \vec{Y}_{tcp} &= \frac{[x_3 - x_1, y_3 - y_1, z_3 - z_1]}{\sqrt{(x_3 - x_1)^2 + (y_3 - y_1)^2 + (z_3 - z_1)^2}} \\ \vec{Z}_{tcp} &= \vec{X}_{tcp} \times \vec{Y}_{tcp}\end{aligned}\quad (4-44)$$

同时，定义运动坐标系的原点为 $P_0 = \frac{1}{3}(P_1 + P_2 + P_3)$ 。因此，由相机坐标系到运动坐标系的旋转矩阵可表示为 $R_{TCP}^C = [\vec{X}, \vec{Y}, \vec{Z}]$ 。

由于特征点测量过程存在误差，计算得到的旋转矩阵可能会不满足正交性质。为解决这一问题，通过数学优化方法对旋转矩阵进行修正。设理想情况下存在误差和优化后的旋转矩阵分别为 R_1 、 R_0 、 R_1 ，那么误差模型为：

$$R_0 = R + N \quad (4-45)$$

待优化目标函数定义为：

$$f(R_1) = \arg \min \|R_1 - R_0\|_F^2 \quad (s.t. R_1 R_1^T = I) \quad (4-46)$$

式中，将 Frobenius 范数展开并整理如下：

$$\|R_1 - R_0\|_F^2 = 3 + \text{trace}(R_0^T R_0) - 2\text{trace}(R_1^T R_0) \quad (4-47)$$

由于 $3 + \text{trace}(R_0^T R_0)$ 是一个定量，因此目标函数可转换为

$$f(R_1) = \arg \min(\text{trace}(R_1^T R_0)) \quad (4-48)$$

对 R_0 进行 SVD 分解，代入上式，并结合 R_1 的正交性，得到：

$$\text{trace}(V^T R_1^T U S) = \text{trace}(IS) = \sum_{i=1}^3 \sigma_{ii} s_i \leq \sum_{i=1}^3 s_i \quad (4-49)$$

因此最优解为 $R_1 = UV^T$ 。至此， R_{TCP}^C 可修正为 UV^T ，可获得相机坐标系到运动坐标系的变换矩阵：

$$T_{TCP}^C = \begin{bmatrix} R_{TCP}^C & t_{TCP}^C \\ 0^T & 1 \end{bmatrix} \quad (4-50)$$

为了让测量数据具有一致的参考基准，通常会将机器人末端的位姿描述在机器人基坐标系下。设机器人末端位姿在 i 时刻的位姿表示为 Rot_i ，在机器人基坐标的值为 Rot_i^B ，在相机坐标下的值为 Rot_i^C ，那么位姿变换公式如下：

$$Rot_i^B = T_{TCP}^C T_C^B Rot_i \quad (4-51)$$

其中， T_C^B 为相机坐标系到机器人坐标系的变换矩阵。

4.4 本章小结

本章围绕机器人运动视觉测量展开了详细分析，从特征点的高精度定位到三维重建与坐标系变换，逐步搭建了完整的视觉测量框架。首先，引入各向异性扩散滤波算法，在有效抑制噪声的同时保留图像中的重要边缘信息。然后，推导基于 Zernike 矩的亚像素边缘定位过程，针对实际测量过程中的边缘渐变问题，采用一种基于渐变边缘模型的改进 Zernike 矩的边缘定位算法，以提升特征点的亚像素级定位精度，为后续合作标识特征点的准确提取奠定基础。针对十字标和三角形合作标识的几何特点，分别设计了结合改进 Zernike 矩的特征点提取方案，实现标识特征点的像素坐标提取。在此基础上，推导了双目三维重建模型，实现了像素坐标到相机坐标的三维重建。随后，针对相机坐标系与机器人坐标系间的转换问题，采用最小二乘法与奇异值分解相结合的方法，推导出坐标系间的变换参数，实现了视觉测量数据与机器人运动数据的统一表达。最后，给出了末端姿态测量方法，构建了机器人末端运动坐标系，实现了机器人末端的姿态解析。

第 5 章 工业机器人运动视觉测量实验

本章围绕前文提出的研究思路与算法，搭建了视觉系统及实验平台，对所提方法的实际效果与适用性展开全面验证。首先，进行相机标定与机器人运动追踪实验，进而完成特征点的像素定位。基于此，进一步开展机器人空间结构运动测量实验。此外，将所提测量方法与传统方法进行对比，通过分析测量误差和精度差异，评估本文所提测量方案在工业机器人视觉测量任务中的性能。

5.1 视觉测量实验平台

5.1.1 视觉系统硬件组成

为了验证本文所提方法的有效性，搭建如图 5-1 所示的双目视觉机器人运动测量系统，开展测量性能验证实验。

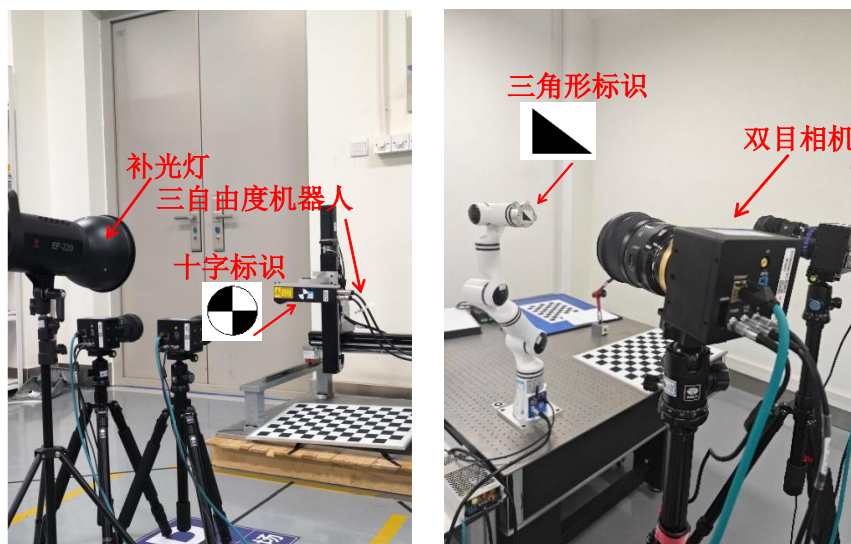


图 5-1 视觉测量系统

测量系统主要组成包括：（1）两台相同参数的高速相机和配备镜头。使用的相机与镜头参数分别如表 5-1、5-2 所示。高速相机支持内触发模式和外触发模式。在内触发模式下，采集频率和曝光时间由相机内部时钟控制。在外触发模式下，采集过程由外部触发信号控制，相机通过外部触发源接收同步信号，在指定时间点触发图像采集；（2）三自由度直角坐标机器人和六自由度机器人。其中，三自由度直角坐标机器人在 X、Y、Z 轴上可达最大范围分别为 400mm、400mm、300mm，其在 X、Y、Z 方向上运动最小分辨精度分别为 0.1um、0.1um、0.08um。六自由度机器人的重复定位精度为 0.05mm；（3）上位机控制器；（4）补光设备和三脚架。补光灯用于提供均匀光照，避免因光照问题导致特征提取误差；三脚

架用于支撑和固定高速相机，可根据实验需要调整高度。

在实验过程中，需对实验器材进行精细调整，以确保测量结果的准确性与稳定性。首先，调整相机的摆放角度和位置，以获得最佳的视场覆盖。然后，微调镜头的焦距，同时观察上位机上的图像采集系统上的画面，调整到成像清晰。同时，对机械臂的运动范围进行设置，保证其在 X、Y、Z 轴上的轨迹覆盖实验设计的测量区域。通过外部触发器实现双目相机与机械臂运动的同步采集，确保数据采集的时序一致性。为了避免光照不均对视觉测量的干扰，实验环境中需使用补光灯进行均匀照明。

表 5-1 相机参数

传感器	分辨率(H×V)	帧率(s ⁻¹)	位深	像素大小(um ²)	快门模式
COMS	2320×1720	<13000	8bit	7×7	Global

表 5-2 镜头参数

光圈范围	滤镜螺纹直径(mm)	焦距(mm)	放大倍率	视角
F/16-F/1.4	77	24	1:5.3	84.1°

5.1.2 测量系统软件环境

为了实现工业机器人运动视觉测量，本实验构建了完整的软件环境支持系统。测量系统的软件部分主要包括图像采集、机器人运动控制、相机标定、运动追踪以及空间运动测量模块。表 5-3 列出了各模块对应的软件实现平台及其开发环境。

表 5-3 视觉测量系统功能模块与软件支持

功能模块	平台	编程语言
图像采集	高速双目视觉系统	/
机器人运动控制	自主开发	C++
相机标定	MATLAB	MATLAB
运动追踪	PyTorch	Python
空间运动测量模块	MATLAB	MATLAB

图像采集模块负责从双目相机获取机器人运动过程中的图像数据，为后续的视觉处理提供原始输入。系统采用高速相机进行数据采集，以保证在机器人运动过程中获取清晰的图像。机器人运动控制模块用于对机器人进行运动轨迹规划和执行，通过发送控制指令实现目标轨迹的自动化运行。机器人运动操作界面如图 5-2 所示。该模块基于 C++开发，并与底层控制硬件通信，以确保机器人在设定

的路径上平稳运行。相机标定模块采用 MATLAB 相机标定工具箱进行双目相机的标定计算。该模块通过输入标准标定板图像和标定板尺寸，完成相机的内参和外参估计。标定结果将用于后续的三维重建。运动追踪模块用于检测机器人结构上的人工标识，为后续的特征点定位输入数据。空间运动测量模块包括特征点提取、三维重建、位姿解算及误差分析部分。利用特征点检测算法提取人工标识特征点的像素坐标。进而，基于双目三维重建模型和相机标定参数重建标识点的三维坐标，并将其转换至机器人坐标系中进行误差分析。



图 5-2 机器人运动控制软件界面

5.2 相机标定实验

对于双目视觉测量系统，为准确测量运动目标的位置，在测量前需要对相机进行标定，以获取相机的内部参数和两台相机之间的位姿关系。本实验使用 MATLAB 2022b 中的双目标定工具箱进行相机标定，采用如图 5-3 所示的棋盘格标定板为校准对象，棋盘格角点规格为 12 行 9 列，45mm×45mm。实验过程中，需确保棋盘格角点全部位于相机视野内，同时不宜距离相机过远，标定板占据相机拍摄画面的 1/4 至 1/2 为宜。同时对打光光源进行调整，防止标定板在拍摄过程中出现偏暗、过曝或反光等不良情况，确保光照均匀，标定板清晰可见。在上述前提条件下，为确保标定结果的准确性，减少系统误差，采集数据时需多次改变标定板的摆放位置和姿态，每次调整后，双目相机同步对标定板进行一次拍摄，实验共拍摄 15 组标定板图片，具体如图 5-4 所示。

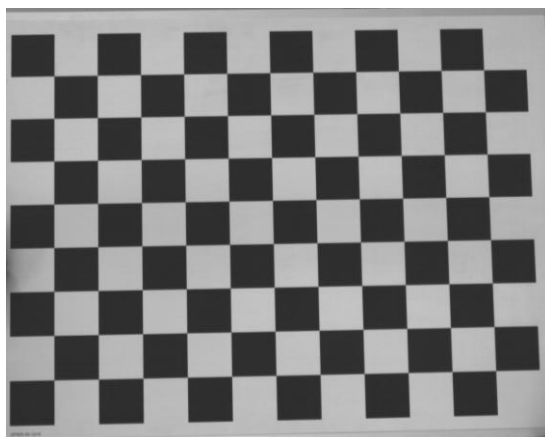


图 5-3 棋盘格标定板

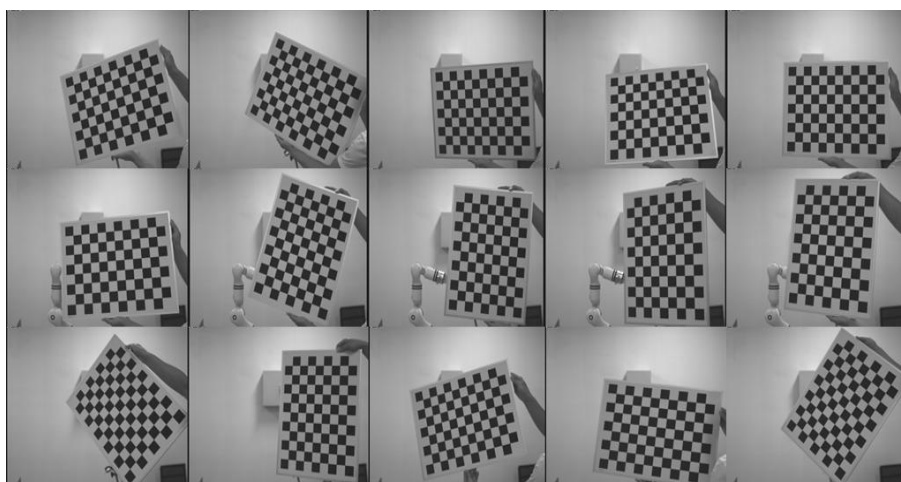


图 5-4 不同位姿的标定板图像

通过 MATLAB 双目标定工具箱输入拍摄的标定板图像和尺寸，完成两台相机的标定。在标定完成后，生成了每台相机的参数文件。本实验的标定结果为： $\mathbf{K}_L=[3612.760, 0, 1198.694; 0, 3612.910, 844.262; 0, 0, 1]$ ； $\mathbf{K}_R=[3610.236, 0, 1174.792; 0, 3600.477, 851.456; 0, 0, 1]$ ； $\mathbf{R}=[0.941, -0.017, 0.339; -0.005, 0.999, 0.037; -0.340, 0.034, 0.940]$ ； $\mathbf{t}=[-349.992; 3.675; -7.752]$ 。标定的准确性通过重投影误差评估，本实验的平均重投影误差为 0.06 像素。

5.3 机器人运动追踪实验

对相机进行标定后，将合作标识粘贴于机器人表面，对机器人的运动数据进行采集，捕获机器人运动的离散点位置和连续空间运动的图像。基于此，通过 GSBi-YOLOv8 算法对采集图像进行识别，完成对标识的初定位。

5.3.1 实验环境

实验环境配置如表 5-4 所示，训练过程中输入图像尺寸为 640×640 ，训练周期为 200 个 epochs，批量大小设置为 16。模型使用 Adam 优化器进行参数优化，

初始学习率为 0.001，动量参数为 0.937，权重衰减系数为 5×10^{-4} 。

表 5-4 实验环境配置

名称	参数
操作系统	Linux
GPU	NVIDIA RTX A6000
编程语言	Python 3.8
框架	PyTorch 1.12.1+CUDA 12.4

5.3.2 实验数据集构建

本文研究内容应用的任务场景为“机器人运动测量与分析”，目前未有公开的标准数据集，由于机器人运动测量过程中机器人位姿的改变导致检测目标在图像中的大小、形态等特征各不相同，因此，本文将被检测人工标识粘贴于机器人上，操控机器人运动并通过相机进行图像采集，自主构建数据集，拟检测的目标为图 2-3(a)十字标识。为了避免因数据采集场景的单一性和样本种类不足所导致的过拟合问题，并增强模型的鲁棒性和泛化能力，通过改变机器人运动时的位姿、相机与机器人距离远近、光照条件等，获取包含不同形态、大小的目标图像，并对图片进行数据增强处理，包括旋转、增加高斯噪声、遮挡、模糊等操作，构建数据集，共获取图像 3000 张。按照 7: 2: 1 的比例划分训练集、验证集和测试集，用于评估检测模型的性能。数据集中部分图像如图 5-5 所示。根据小目标定义，目标在图像中的尺寸相较于图像整体尺寸比例小于 0.1 时，均为小目标^[61]。本文采集的数据集图像的像素尺寸为 2368×1720 ，目标尺寸范围约为 $25 \times 25 \sim 120 \times 120$ ，目标尺寸占图像尺寸的比例小于 0.1，因此本文实验中的目标可定义为小目标。



图 5-5 数据集实例

5.3.3 评价指标

本文将从精确率(Precision)、召回率(Recall)、平均精度均值(mAP)、mAP50、mAP50-95、每秒帧数(FPS)作为模型性能的评价指标,各指标的定义如下。

$$Precision = \frac{TP}{TP + FP} \quad (5-1)$$

$$Recall = \frac{TP}{FP + FN} \quad (5-2)$$

$$AP = \int_0^1 P(R) dR \quad (5-3)$$

$$AP = \int_0^1 P(R) dR \quad (5-4)$$

$$mAP = \frac{1}{C} \sum_{i=1}^C A_{p_i} \quad (5-5)$$

其中, TP (True Positive) 表示正确检测出的标识数量, FP (False Positive) 表示标识出现误检的数量, FN (False Negative) 表示标识漏检的数量, i 表示类别, C 表示总类别数。精确率表示准确检测到目标的比例; 召回率体现了算法对正样本的识别能力; mAP 是算法整体精度的综合指标, 表示多个阈值条件下的平均精度的均值, 用于评价算法在不同重叠率条件下的检测性能; mAP50 表示当交并比 (IoU) 阈值设置为 0.5 时的平均精度; mAP50-95 表示在 IoU 阈值从 0.5 到 0.95 (步长为 0.05) 范围内所有类别的平均精度的均值; FPS 表示算法每秒可以处理的图像帧数, 衡量算法的推理效率。

5.3.4 实验结果分析

为验证 3.2.2 节所提出的 CIoU 与 NWD 相结合的损失函数对 YOLOv8 小目标检测精度的影响, 找到式(3-8)最优的结合策略, 通过调整不同的 α 值进行精度实验, 结果如表 5-5 所示。

表 5-5 选取不同 α 值时的检测结果

α	Precision /%	Recall /%	mAP50 /%	mAP50-95 /%
$\alpha = 0.2$	93.7	95.3	95.5	79.9
$\alpha = 0.4$	93.8	95.5	95.6	79.8
$\alpha = 0.6$	94.2	95.6	95.9	80.3
$\alpha = 0.8$	94.6	95.8	96.4	80.7
$\alpha = 1.0$	94.3	95.6	96.0	80.5

实验结果表明, 不同的 α 值会影响 YOLOv8 的检测性能。当 $\alpha = 0.8$ 时, 检测性能达到最佳, 优于 $\alpha = 1$ 时的全权重 CIoU 结果。这说明采用 NWD 与 CIoU 相结合的策略, 能够更好平衡样本特征分布间的差异, 更准确地评估框的重叠程度,

从而提升小目标检测能力。因此，取 $\alpha = 0.8$ 作为后续实验的参数。

为进一步验证本文所提出的 CIoU 与 NWD 相结合的 NWD-CIoU 损失函数相较于其他损失函数的优势，进行了添加不同损失函数的对比实验，实验结果如表 5-6 所示。对比分析显示，将 NWD-CIoU 作为回归损失函数时，模型的整体检测效果最优，验证了所提方法的可行性。

表 5-6 损失函数实验结果

模型	Precision /%	Recall /%	mAP50 /%	mAP50-95 /%	FPS
SIoU	94.4	93.3	96.3	80.6	202.3
WIoU	94.1	95.2	95.6	79.7	205.1
EIoU	94.2	95.6	96.1	80.3	204.6
CIoU	94.3	95.6	96.0	80.5	202.3
NWD-CIoU	94.6	95.8	96.4	80.7	204.5

为分析各改进模块对 YOLOv8 模型检测性能的影响，开展消融实验，逐步引入改进模块，评估每个模块对模型性能的影响。以 YOLOv8 系列中的 YOLOv8s 作为基准模型，在基准模型中分别引入 GSConv 模块、BiFPN 模块、NWD-CIoU 损失函数、GSConv+BiFPN 模块以及所有模块。消融实验结果如表 5-7 所示。

表 5-7 消融实验结果

方法	Precision /%	Recall /%	mAP50 /%	mAP50-95 /%	FPS
YOLOv8s	94.3	95.6	96.0	80.5	202.3
+GSConv	93.9	95.2	95.8	80.0	216.2
+BiFPN	98.2	98.3	98.9	82.5	200.0
+NWD-CIoU	94.6	95.8	96.4	80.7	204.5
+GSConv+BiFPN	97.9	97.9	98.7	82.2	214.6
Y-ALL	98.5	98.2	98.9	82.5	214.9

由表 5-7 可以看出，引入 GSConv 模块能有效提升模型的推理速度，FPS 提高了 6.9%；采用 BiFPN 结构代替原始的 PAN-FPN 结构能够有效提升模型检测的精度，检测精度由原有的 94.3% 提升至 98.2%，提高了 3.9%，召回率提高了 2.7%，mAP50 和 mAP50-95 分别提高了 2.9% 和 2.0%；将 NWD-CIoU 作为回归损失函数，各项指标均有提升；引入 GSConv+BiFPN 模块，FPS 提升了 6.1%，精确率提高了 3.6%，召回率提高了 2.3%，mAP50 和 mAP50-95 分别提高了 2.7% 和 1.7%；将本文提出的多个改进结合后，相比基准模型，FPS 提升了 6.2%，精确率提高了 4.2%，召回率提高了 2.6%，mAP50 和 mAP50-95 分别提高了 2.9% 和 2.0%，改进算法能够在提升精度的同时实现速度优化。

为验证改进算法的效果，本文对比了改进算法与其他常用目标检测算法在自

建数据集上的检测效果，实验结果如图 5-6 和表 5-8 所示。

表 5-8 算法效果比对结果

方法	Precision /%	Recall /%	mAP50 /%	mAP50-95 /%	FPS
NCC ^[62]	78.3	/	/	/	43.5
SSD ^[63]	93.5	90.2	95.0	78.8	50.6
Faster R-CNN ^[33]	93.8	91.2	95.5	80.3	16.8
YOLOv5s	93.1	92.0	94.2	78.2	81.9
YOLOv8s	94.3	95.6	96.0	80.5	202.3
YOLOv9s ^[64]	96.7	96.5	97.6	72.5	185.2
GSBi-YOLOv8	98.5	98.2	98.9	82.5	214.9

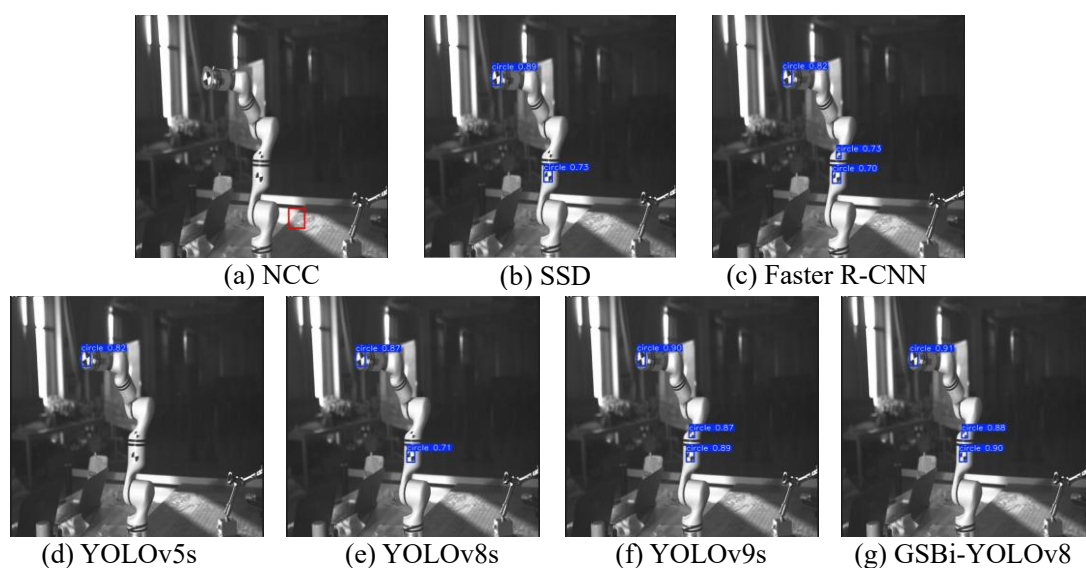


图 5-6 检测结果图

由图 5-6 和表 5-8 可以看出，由于受目标形态和尺度变化的影响，基于归一化互相关的传统模板匹配算法（NCC）检测精度和速度远低于所提改进算法。相较于 SSD、Faster R-CNN、YOLOv5s、YOLOv8s 和 YOLOv9s，本文所提算法在检测精确率、召回率、mAP 和检测效率上均有所提升。SSD、Faster R-CNN 和 YOLOv5s 的处理速度较慢，远低于本文所提算法；相较于 YOLOv8s 原模型，所提算法在主要性能指标上均有所提升，检测精确率、召回率、mAP50、mAP50-95 和检测效率分别提升 4.2%、2.6%、2.9%、2.0%、6.2%；相比 YOLOv9s，本文算法在精确率、召回率、mAP50 和 mAP50-95 上分别提升了 1.8%、1.7%、1.3% 和 10.0%，其中 mAP50-95 提升尤为显著，表明改进算法在小目标检测任务中具有更高的精度和鲁棒性。此外，检测效率提升 16.0%，在提高小目标检测精度的同时保持了较高的计算效率。实验结果表明，本文所提改进策略能够有效提升小目标检测的性能。

5.4 特征点定位与机器人空间结构运动测量实验

5.4.1 特征点定位实验

基于 5.3 节完成对粘贴在机器人结构表面的标识的初步定位后，进一步，基于 4.1 节方法获取标识特征点的像素坐标。图 5-7 展示了十字标识中心点定位过程。如图 5-7(a)所示，首先进行像素级边缘检测，然后求解亚像素边缘点，并求取边缘点的质心，获得中心点坐标，如图 5-7(b)所示。进而，根据边缘点到中心点所连成直线的斜率对直线进行分类，获取如图 5-7(c)所示的两类直线。最后，如图 5-7(d)所示，对直线进行拟合，交点即为所求特征点。特征点检测结果如图 5-7(e)所示。

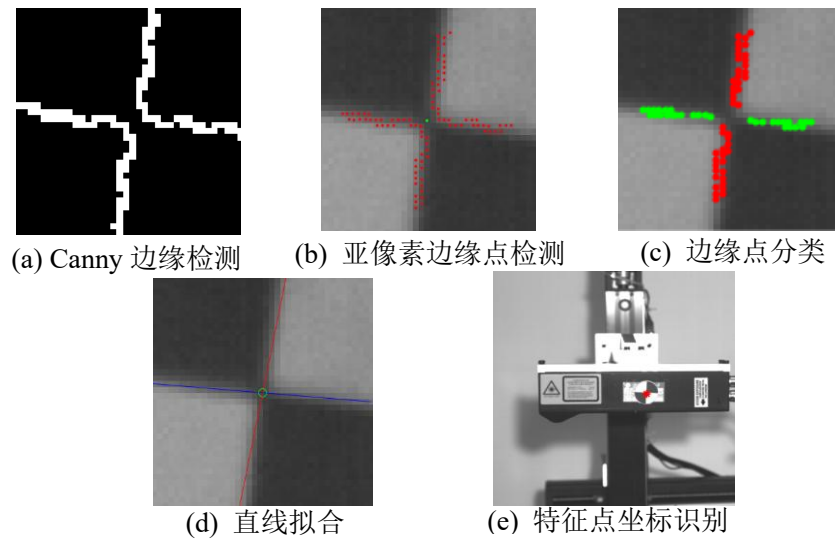


图 5-7 十字标识特征点检测过程

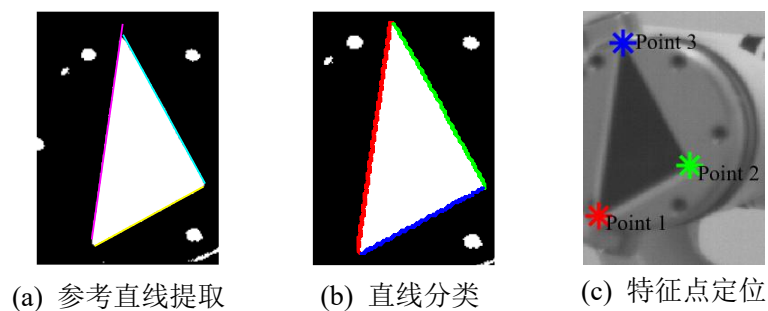


图 5-8 三角形特征点检测过程

三角形标识特征点检测过程如图 5-8 所示，采用霍夫直线变换获取三角形三条边的参考直线，如图 5-8(a)。进而，对直线进行分类，如图 5-8(b)所示，分别用绿色、蓝色和红色标记三角形的直角长边、直角短边和斜边的亚像素点。分类完成后，使用最小二乘法分别对三条边的亚像素点进行直线拟合；最后，三条直线的交点即为三角形的三个顶点，用三种不同颜色标记出了三个顶点，图 5-8(c)

展示了图像中对特征点标记的情况。

5.4.2 机器人空间运动测量

基于 5.4.1 节可获得左右相机下的标识特征点像素坐标，结合相机标定结果，采用 4.2 节所述方法进行机器人空间运动测量。实验共获取 18 组参考点与对应的观测值，重构于左相机坐标系下，如图 5-9 所示。对应的左相机坐标系坐标和机器人坐标系坐标如表 5-9 所示。

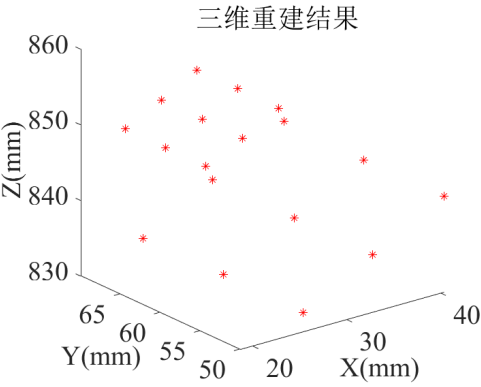


图 5-9 参考点三维重建图

表 5-9 参考点坐标数据

序号	左相机坐标系坐标(mm)			机器人坐标系坐标(mm)		
	X	Y	Z	X	Y	Z
1	40.402	49.870	842.607	40	-20	630
2	32.823	49.908	837.529	40	-30	630
3	25.318	49.833	832.617	40	-40	630
...	...	...	...	...	...	...
16	18.725	54.230	852.096	20	-35	625
17	22.596	54.194	854.471	20	-30	625
18	26.456	54.234	857.028	20	-25	625

为了验证所提方法的有效性，采用传统的角点检测方法对标识的特征点进行像素坐标提取并进行三维重建。随后，将重建结果与本文方法的结果进行对比分析。为确保对比的统一性和准确性，所有重建结果均转换至机器人坐标系下进行比较和验证，图 5-10 所示为结果的可视化图。为了进一步量化两种方法的性能，计算了误差的最大值、均值和标准差，结果如表 5-10 所示。

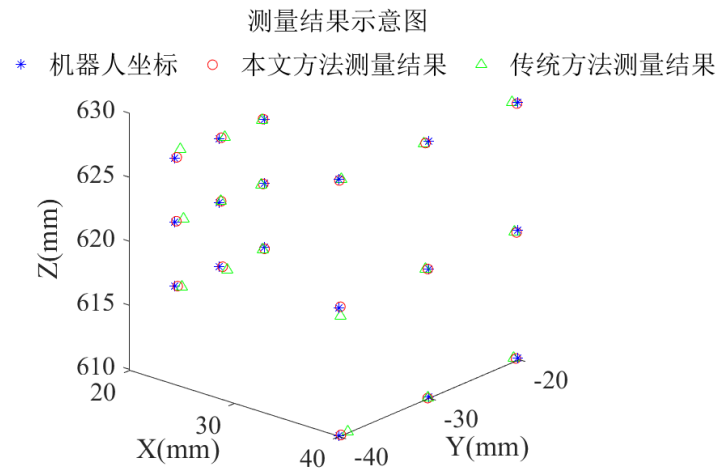


图 5-10 测量结果示意图

表 5-10 传统方法和本文方法误差比较

所用方法	误差项		
	最大值(mm)	均值(mm)	标准差(mm)
传统方法	0.891	0.392	0.316
本文所提方法	0.318	0.189	0.134

可以看出,采用传统角点检测方法与本文所提出的改进方法相比,本文所提方法测量误差结果明显优于传统方法。从误差的具体数值来看,采用本文方法后,误差最大值由 0.891 减小到 0.318mm,均值由 0.392 减小到 0.189mm,标准差由 0.316 减小到 0.134mm。无论是在最大误差还是整体误差分布上,本文方法均表现出更高的精度和稳定性。结合图 5-10 中数据可视化的对比结果,可以观察到采用本文方法的测量结果与实际值之间的偏差更小,验证了本文所提方法的有效性。

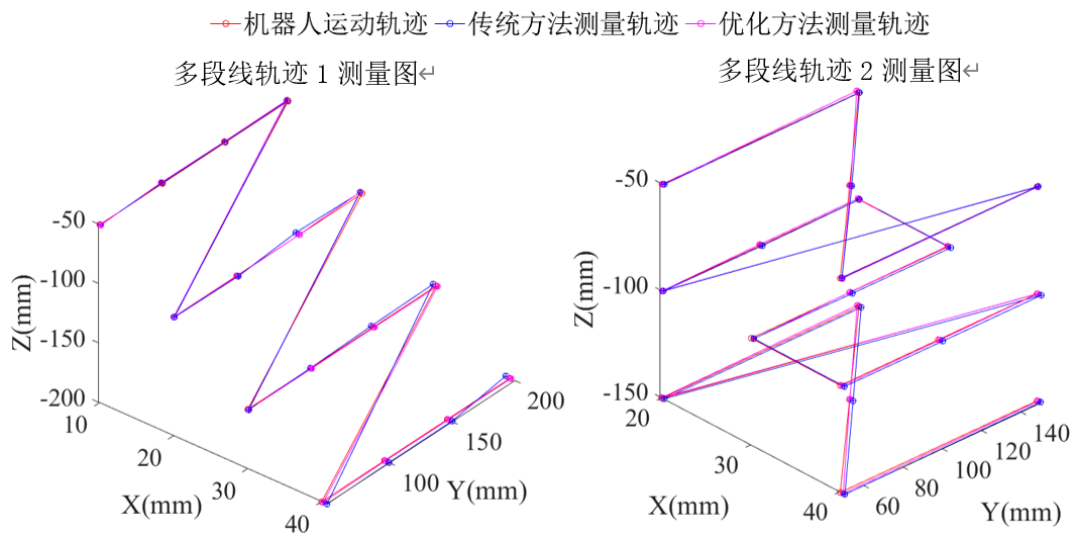


图 5-11 末端测量轨迹图

表 5-11 机器人空间运动测量轨迹偏差比较

所用方法	误差项	最大值(mm)	均值(mm)	标准差(mm)
多段线轨迹 1	传统方法	0.729	0.371	0.226
	优化方法	0.386	0.195	0.090
多段线轨迹 2	传统方法	0.571	0.395	0.121
	优化方法	0.390	0.222	0.096

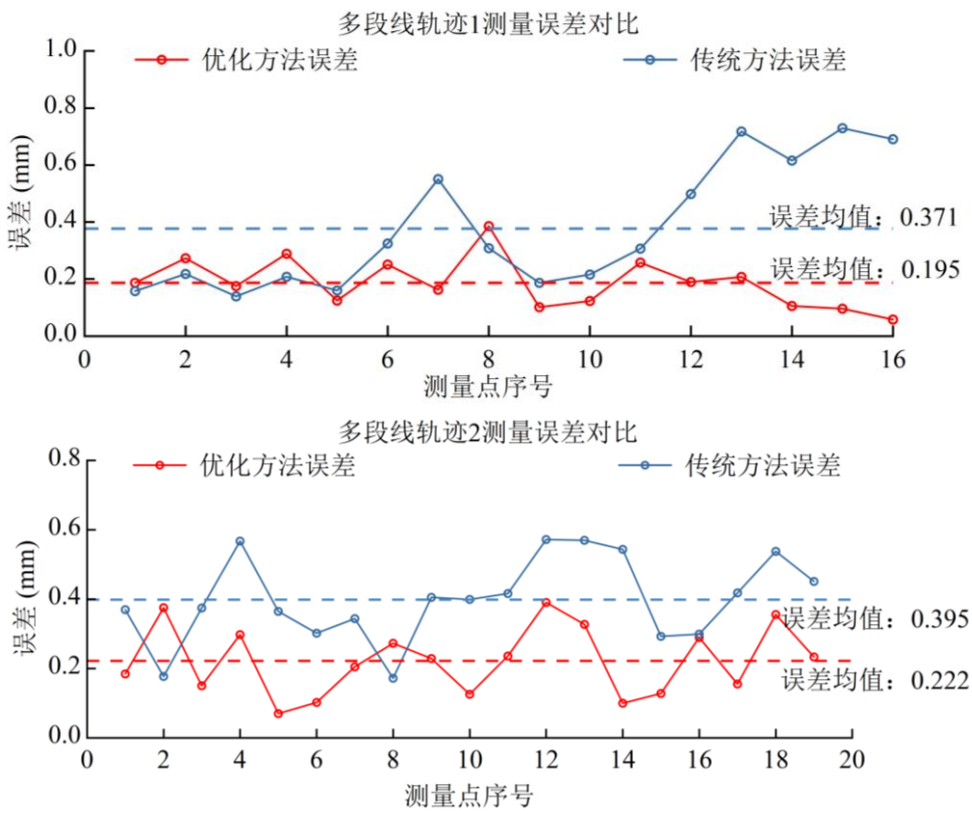


图 5-12 多段线轨迹测量误差结果图

为进一步验证本文方法的有效性，设计并开展了连续轨迹运动测量。驱动机器人沿空间多段线移动，末端测量轨迹如图 5-11 所示，轨迹测量误差结果如表 5-11 和图 5-12 所示。由结果可知，本文所提方法相比传统方法，能够显著提升重建精度并降低误差波动。在多段线轨迹 1 中，优化方法的最大误差相比传统方法减少了 47.1%，由 0.729mm 降至 0.386mm；误差均值减少了 47.4%，由 0.371mm 降至 0.195mm；误差标准差减少了 60.1%，由 0.226mm 降至 0.090mm。同样，在多段线轨迹 2 中，优化方法的最大误差减少了 46.1%，从 0.571mm 降至 0.390mm；误差均值减少了 43.8%，由 0.395mm 降至 0.222mm；误差标准差降低了 20.7%，由 0.121mm 降至 0.096mm。实验结果表明，本文方法能够有效提升运动测量的稳定性和可靠性，具有更高的求解精度。

进一步，验证机器人沿圆弧和直线两种轨迹下连续运动的精度。机器人运动

轨迹参数如表 5-12 所示，机器人运动的线速度为 500mm/min。

表 5-12 机器人运动轨迹参数

圆弧	起点坐标 (-30, 100, -60)	终点坐标 (30, 150, -100)	圆心, 半径 (0, 100, 100), $r=50$
直线	起点坐标 (-20, 150, 0)	终点坐标 (20, 50, -100)	方向向量 [40; -100; 100]

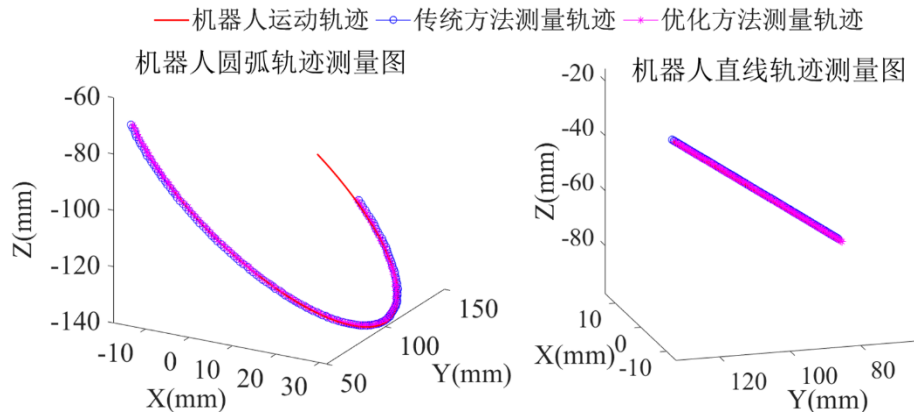


图 5-13 机器人运动轨迹测量图

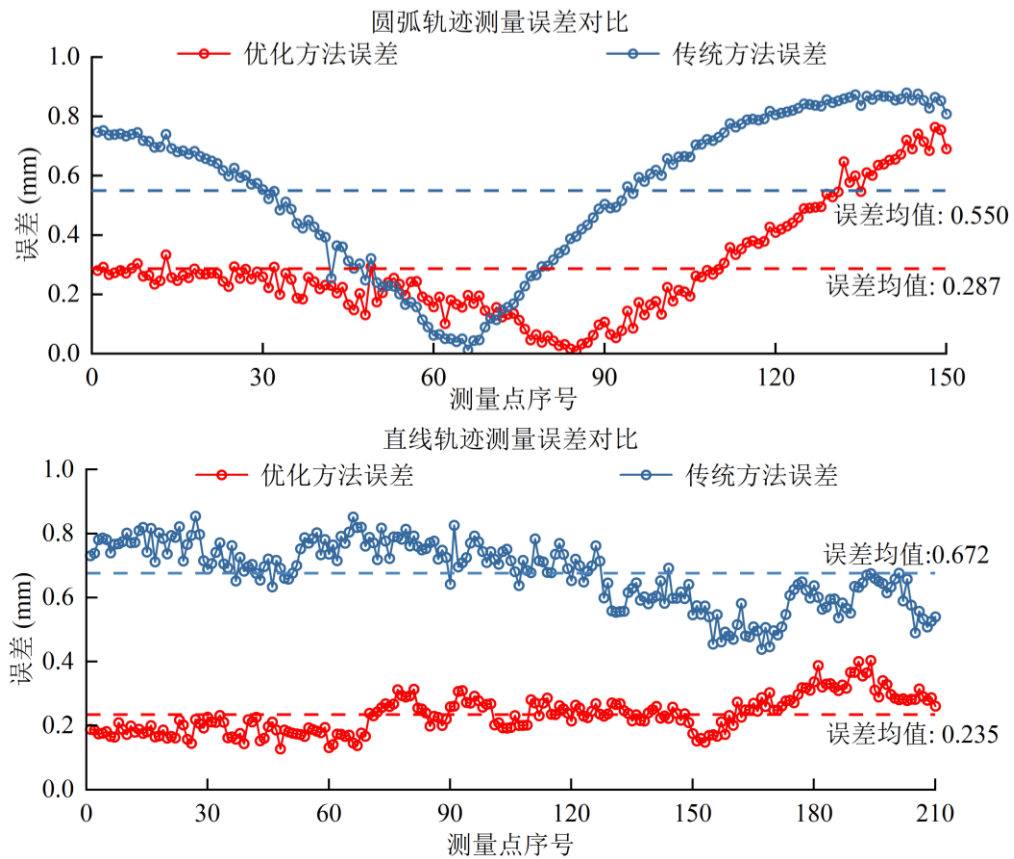


图 5-14 机器人运动轨迹测量误差

表 5-13 机器人空间运动测量轨迹偏差比较

所用方法	误差项	最大值(mm)	均值(mm)	标准差(mm)
圆弧轨迹	传统方法	0.879	0.550	0.260
	优化方法	0.753	0.287	0.182
直线轨迹	传统方法	0.854	0.672	0.107
	优化方法	0.403	0.235	0.058

机器人沿圆弧和直线轨迹运动的结果如图 5-13、图 5-14 和表 5-13 所示。实验结果表明，在圆弧轨迹测量中，优化方法的最大误差由 0.879mm 降低至 0.753mm，降低了 14.3%；均值误差由 0.550mm 降至 0.287mm，降低了 47.8%；标准差从 0.260mm 下降到 0.182mm，降低了 30.0%。在直线轨迹测量中，优化方法的最大误差由 0.854mm 降至 0.403mm，降低了 52.8%；均值误差从 0.672mm 下降至 0.235mm，降低了 65.0%；标准差由 0.107mm 下降到 0.058mm，降低了 45.8%。从误差曲线可以看出，优化方法在整个轨迹范围内均能有效减少测量误差，误差值更加稳定，与理论轨迹的偏差更小。优化方法的测量结果较传统方法更加接近真实轨迹，说明该方法具有更高的精度和稳定性，能够有效降低测量误差。与已有文献^[65]中视觉测量方法的实验结果比，本文所提方法在误差指标上更低，表现出更好的测量性能，验证了本文方法在机器人运动轨迹测量中具有较高的精度。

5.4.3 机器人空间姿态测量

为验证本文提出的测量方法在机器人空间姿态测量任务中的有效性与精度表现，根据 4.3 节内容，开展机器人空间姿态测量实验。将直角三角形标识粘贴于机器人末端，设置机器人末端执行器进行姿态变换，本次实验共采集 20 组数据。在机器人基坐标系下的相对姿态观测结果如图 5-15 所示，测量结果和每个轴的误差分析分别如表 5-14、表 5-15 和图 5-16 所示。

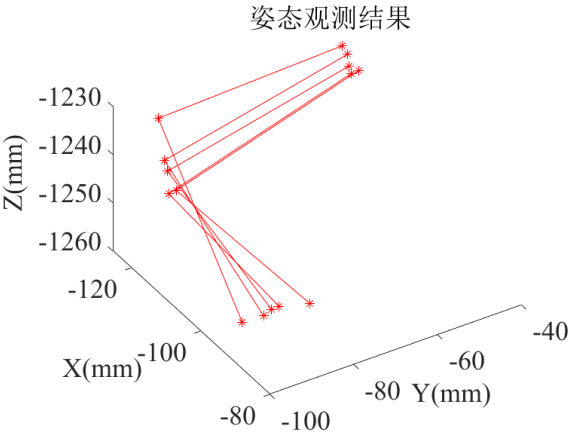


图 5-15 空间姿态观测图

表 5-14 姿态测量结果

序号	运动相对姿态			观测姿态		
	$\Delta R_x / ^\circ$	$\Delta R_y / ^\circ$	$\Delta R_z / ^\circ$	$\Delta R_x / ^\circ$	$\Delta R_y / ^\circ$	$\Delta R_z / ^\circ$
1	-5	0	0	-5.128	-0.038	-0.135
2	-5	0	0	-5.074	0.041	0.123
3	-5	0	0	-4.873	-0.169	-0.131
4	-5	0	0	-4.889	0.114	-0.142
5	10	2	0	9.665	2.501	-0.158
6	10	2	0	10.176	1.673	0.175
7	10	2	0	9.890	2.199	0.124
8	10	2	0	9.762	2.166	0.153
9	0	-5	-5	0.108	-4.875	-5.259
10	0	-5	-5	0.115	-5.367	-5.224
11	0	-5	-5	0.114	-5.211	-5.198
12	0	-5	-5	-0.288	-5.109	-5.140
13	-1	-0.5	-5	-0.945	-0.581	-5.159
14	-0.5	0	-5	-0.327	-0.297	-5.306
15	1	-3	-2	1.207	-3.315	-2.160
16	-5	3	-2	-4.903	2.806	-1.869
17	5	-2	1	4.849	-2.163	1.171
18	0	-4	1	0.271	-4.414	0.827
19	0	-4	1	0.089	-4.318	0.809
20	0	-4	1	0.107	-4.387	0.890

表 5-15 X、Y、Z 轴误差

坐标轴	误差项	最大值	均值	标准差
X		0.335°	0.154°	0.075°
Y		0.501°	0.227°	0.128°
Z		0.306°	0.168°	0.047°

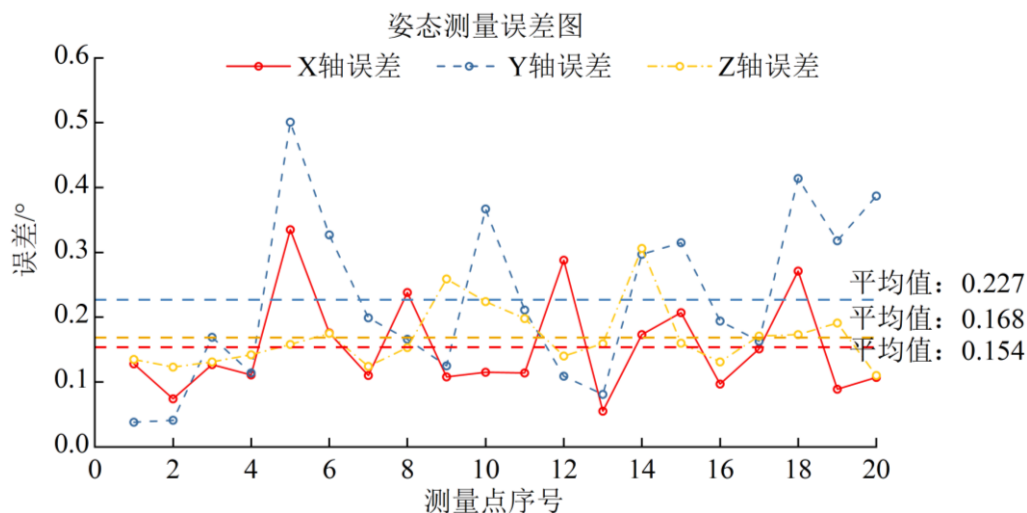


图 5-16 姿态三轴误差图

根据姿态测量结果可知，X、Y、Z 轴上，误差最大值分别为 0.335° 、 0.501° 和 0.306° ，误差均值分别为 0.154° 、 0.227° 和 0.168° 。姿态测量误差控制在 0.6° 以内，验证了本文所提姿态测量方法的可行性。

5.5 本章小结

本章围绕前述理论与算法，搭建双目视觉测量系统，开展工业机器人运动测量效果验证实验。设计并完成了包括相机标定、运动追踪、特征点定位、机器人运动测量以及机器人空间姿态测量在内的多项工作。结果表明，所提出的 GSBi-YOLOv8 算法能够准确高效完成机器人运动追踪任务，特征点定位方法能够准确提取特征点像素坐标，进而提升三维重建精度。综合实验结果，可以看出，所提视觉测量方案能够实现工业机器人的高精度运动测量。

第 6 章 总结与展望

6.1 总 结

工业机器人作为工业自动化的重要装备，广泛应用于汽车制造、金属成形、冶金等多个领域，极大提高了作业效率，节约了生产成本。研究工业机器人的运动测量方法，对于进一步提高工业机器人运动的精准度、灵活性及适应性具有重要意义。视觉测量技术因其非接触性、高灵活性以及全局观测的优势，在工业机器人运动测量领域展现出巨大潜力。本文针对工业机器人高精度运动测量中的关键技术问题，探寻基于视觉的机器人高精度运动测量方法，本文的主要研究成果如下：

（1）针对小目标检测过程中漏检率高、检测效率不足的问题，提出 GSBi-YOLOv8 算法，对设计的合作标识进行高效检测，通过优化模型的小目标检测能力，提高了模型对标识的检测精度和速度。

（2）设计了一种基于改进 Zernike 矩和最小二乘拟合的亚像素级坐标提取方法，提升特征点检测精度与鲁棒性。构建了机器人运动的三维重建模型，研究了坐标转换方法，结合最小二乘法与奇异值分解法，实现了视觉坐标系与机器人坐标系间的统一表达。给出了机器人末端运动坐标系与姿态解析方法，实现了六自由度机器人末端姿态的测量。

（3）基于所提测量方案，开展了机器人目标追踪、特征点定位和空间运动测量实验。结果表明，本文提出的视觉测量方法能够实现机器人高精度运动测量，验证了本文方法的效果。

6.2 展 望

本文提出的基于视觉的机器人高精度运动测量方法在理论与实践均取得了一定成果，但仍有进一步研究和优化的空间。在未来的研究中，可从以下几个方面进行拓展：

（1）目前基于 MATLAB 标定工具箱的方法已经能够取得较为理想的精度，但仍有一定的优化空间。在优化参数维度较高时，该方法面临陷入局部最优的风险。此外，在大规模角点提取时，存在耗时长、精度下降等问题。未来相机标定研究技术应从以下几个方向进行研究：一是进一步提升精度和稳定性，提升亚像素级特征点检测精度，降低角点检测误差；从参数数量的角度平衡标定的精准性与稳定性，适当轻量化标定模型，保证模型表达能力的同时，避免参数冗余，确保标定结果的高可靠性；二是实现标定过程的自动化、智能化与轻量化，充分利用高性能计算设备和多传感器数据融合技术，快速完成标定任务；三是建立全面的评价体系，从标定精度、速度及复杂场景适应性等多维度出发，开发满足实际需求的高效标定方案。

（2）随着工业场景对实时性要求的提高，优化视觉测量算法的计算效率与资源占用仍是未来的重点方向。可通过引入更高效的深度学习模型或加速技术，进一步提高系统的运行效率，为边缘计算应用提供支持。

参考文献

- [1] 吴锦辉, 陶友瑞. 工业机器人定位精度可靠性研究现状综述[J]. 中国机械工程, 2020, 31(18): 2180-2188.
- [2] 刘伟, 刘顺, 邓朝晖等. 工业机器人定位误差补偿技术研究进展[J]. 机械工程学报, 2023, 59(17): 1-16.
- [3] 郑岩. 基于双目立体视觉的机械臂运动位置测量[D]. 华中科技大学, 2021.
- [4] 钱海章, 张强, 李帅. “十四五”规划下中国制造供给能力及发展路径思考[J]. 数量经济技术经济研究, 2022, 39(01): 28-50.
- [5] Zhang D Q, Peng Z Y, Ning G S, et al. Positioning accuracy reliability of industrial robots through probability and evidence theories[J]. Journal of Mechanical Design, 2021, 143(1): 15.
- [6] Mousavi S, Gagnol V, Bouzgarrou B C, et al. Stability optimization in robotic milling through the control of functional redundancies[J]. Robotics and Computer-Integrated Manufacturing, 2018, 50: 181-192.
- [7] Placzek M, Piszczek L. Testing of an industrial robot's accuracy and repeatability in off and online environment[J]. Maintenance and Reliability 2018, 20(3): 455-464.
- [8] 王海鹏, 郭瑜, 尹兴超. 工业机器人工况下的增量式光栅编码器误差补偿法[J]. 振动与冲击, 2024, 43(22): 225-231.
- [9] 易杰, 赵斌, 张文涛等. 新型光纤加速度传感器在振动台试验中的应用[J]. 振动与冲击, 2024, 43(19): 43-51.
- [10] 邱志成. 柔性机械臂的振动测量和控制研究进展综述[J]. 信息与控制, 2021, 50 (02): 141-161.
- [11] 肖青伟, 段发阶, 黄婷婷等. 电容式高频振动幅值测量方法研究[J]. 仪表技术与传感器, 2020, (09): 106-111.
- [12] Abdollahpouri M, Gergely Takács, Roha-Ilkiv B. Real-time moving horizon estimation for a vibrating active cantilever[J]. Mechanical Systems and Signal Processing, 2017, 86(A): 1-15.
- [13] Fang T, Ding Y. A sampling-based motion planning method for active visual measurement with an industrial robot[J]. Robotics and Computer-Integrated Manufacturing, 2022, 76: 102322-102322.
- [14] 邱志成, 肖骏, 刘金国. 柔性臂振动的双目三维检测和主动控制[J]. 电机与控制学报, 2019, 23(10): 129-138
- [15] 祝琨, 杨唐文, 阮秋琦等. 基于双目视觉的运动物体实时跟踪与测距[J]. 机器人, 2009, 31(04): 327-334.
- [16] 彭聪, 缪卫东, 曾聪. 基于机器视觉的轻型梁三维振动测量方法[J]. 北京航空航天大学

- 学报, 2021, 47(02): 207-212.
- [17] 唐洪. 基于深度学习和双目视觉的工业机器人抓取方法研究[D]. 华南理工大学, 2018.
- [18] 苏晋. 摄像机标定方法研究[D]. 东北大学, 2010.
- [19] Tsai R Y. A versatile camera calibration technique for high-accuracy 3D machine vision metrology using off-the-shelf TV cameras and lenses[J]. IEEE Journal on Robotics and Automation, 1987, 3(4): 323-344.
- [20] Zhang Z. A flexible new technique for camera calibration[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 1998, 22(11): 1330-1334.
- [21] 王靖, 魏亮, 向文豪等. 考虑圆形特征边缘模糊和偏心误差修正的高精度相机标定方法[J]. 红外与激光工程, 2021, 50(12): 369-379.
- [22] 黄臻臻, 黄晓梅, 郑莲玉等. 同心圆偏心误差补偿迭代的相机标定方法[J]. 光学学报, 2022, 42(19): 73-81.
- [23] 杨霏, 殷玉龙, 卢荣胜等. 基于方向性靶标和多约束优化的双目相机标定[J]. 光学学报, 2022, 42(08): 195-206.
- [24] 刘屹东, 贾振堂. 基于非线性优化的改进相机内参数标定方法[J]. 激光与光电子学进展, 2022, 59(18): 347-355.
- [25] 李春明, 吕大勇, 远松灵. 基于改进灰狼遗传算法的相机标定[J]. 激光与光电子学进展, 2024, 61(18): 218-227.
- [26] 崔岸, 袁智, 王龙山. 基于自适应神经网络的双摄像机标定[J]. 计算机工程与应用, 2009, 45(21): 55-57+60.
- [27] 胡志新, 王涛. 改进遗传算法优化 BP 神经网络的双目相机标定[J]. 电光与控制, 2022, 29(1): 75-79.
- [28] 罗亚波, 罗健. 基于阻滞增长神经网络的双目视觉标定方法[J]. 华中科技大学学报(自然科学版), 2021, 49(12): 71-75+82.
- [29] 梁路宏, 艾海舟, 肖习攀等. 基于模板匹配与支持矢量机的人脸检测[J]. 计算机学报, 2002, 25(1): 22-29.
- [30] Dalal N, Triggs B. Histograms of oriented gradients for human detection[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2005: 886-893.
- [31] Girshick R, Donahue J, Darrell T, et al. Region-based convolutional networks for accurate object detection and segmentation[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2015, 38(1): 142-158.
- [32] Girshick R. Fast R-CNN[C]//Proceedings of the IEEE International Conference on Computer Vision. 2015: 1440-1448.
- [33] Ren S Q, He K M, Girshick R, et al. Faster R-CNN: Towards Real-Time Object Detection with

- Region Proposal Networks[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2016, 39(6): 1137-1149.
- [34] Redmon J, Divvala S, Girshick R, et al. You only look once: unified, real-time object detection[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2016: 779-788.
- [35] Gong Y Q, Yu X H, Ding Y, et al. Effective fusion factor in FPN for tiny object detection[C]//IEEE Winter Conference on Applications of Computer Vision. 2021: 1160-1168.
- [36] Sun S Z, Mo B, Xu J W, et al. Multi-YOLOv8: an infrared moving small object detection model based on YOLOv8 for air vehicle[J]. Neurocomputing, 2024, 588: 127685.
- [37] 蒋伟, 王万虎, 杨俊杰. AEM-YOLOv8s: 无人机航拍图像的小目标检测[J]. 计算机工程与应用, 2024, 60(17): 191-202.
- [38] Yi H, Liu B, Zhao B, et al. Small object detection algorithm based on improved YOLOv8 for remote sensing[J]. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 2024, 17: 1734-1747.
- [39] 吴明杰, 云利军, 陈载清等. 改进 YOLOv5s 的无人机视角下小目标检测算法[J]. 计算机工程与应用, 2024, 60(2): 191-199.
- [40] 岳明凯, 权康男, 张骢等. 基于改进 YOLOv8 的红外小目标检测算法研究[J]. 红外技术, 2024, 46(11): 1286-1292.
- [41] Harris C, Stephens M. A combined corner and edge detector[C]//Proceedings of The 4th Alvey Vision Conference, Manchester, 1988: 147-151.
- [42] Lowe D G. Distinctive image features from scale-invariant keypoints[J]. International Journal of Computer Vision, 2004, 60: 91-110.
- [43] Bay H, Ess A, Tuytelaars T, et al. Speeded-up robust features (SURF)[J]. Computer vision and image understanding, 2008, 110(3): 346-359.
- [44] Rublee E, Rabaud V, Konolige K, et al. ORB: An efficient alternative to SIFT or SURF[C]//2011 International conference on computer vision. 2011: 2564-2571.
- [45] Detone D, Malisiewicz T, Rabinovich A. Superpoint: Self-supervised interest point detection and description[C]//Proceedings of the IEEE conference on computer vision and pattern recognition workshops. 2018: 224-236.
- [46] Dusmanu M, Rocco I, Pajdla T, et al. D2-net: A trainable CNN for joint description and detection of local features[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2019: 8092-8101.
- [47] Luo Z, Zhou L, Bai X, et al. ASLFeat: Learning local features of accurate shape and localization[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern

- Recognition. 2020: 6589-6598.
- [48] 李宜峰, 周顺, 李庞跃等. 基于结构光的偏振主被动融合三维重建方法[J]. 光子学报, 2024, 53(12): 51-62.
- [49] Paredes A L, Conde M H, Loffeld O. Sparsity-Aware 3-D ToF Sensing[J]. IEEE Sensors Journal, 2023, 23(4): 3973-3989.
- [50] Tao Y, Douté S, Muller J-P, et al. Ultra-High-Resolution 1 m/pixel CaSSIS DTM Using Super-Resolution Restoration and Shape-from-Shading: Demonstration over Oxia Planum on Mars [J]. Remote Sensing, 2021, 13(11): 2185.
- [51] 李德玉, 肖龙飞, 魏汉迪. 基于双目立体视觉的实验室波浪场实时重建[J]. 船舶力学, 2024, 28(10): 1463-1471.
- [52] 赵珊珊, 米曾真, 陈韧. 基于双目视觉的运动目标检测与三维重建[J]. 计算机集成制造系统, 2024, 30(07): 2419-2430.
- [53] 张小国, 沈德玉, 张梓涵等. 基于双目视觉惯性的同步定位与稠密场景重建方法[J]. 中国惯性技术学报, 2024, 32(11): 1086-1094+1101.
- [54] 韩雨坤, 潘翀, 王文涛等. 基于多目立体视觉和神经网络标定的表面形貌测量方法研究[J]. 实验流体力学, 2021, 35(06): 44-51.
- [55] 马涛, 童峥, 张一鸣等. 基于多目视觉深度神经网络的路面宏观纹理三维重建方法[J]. 中国公路学报, 2023, 36(03): 70-80.
- [56] Wu J, Jiang H, Wang H, et al. Vision-based multi-view reconstruction for high-precision part positioning in industrial robot machining[J]. Measurement, 2025, 242: 116042.
- [57] Xie H, Yao H, Zhang S, et al. Pix2Vox++: Multi-scale context-aware 3D object reconstruction from single and multiple images[J]. International Journal of Computer Vision, 2020, 128(12): 2919-2935.
- [58] Zhang S, Xu W, Wei Z, et al. ARAI-MVSNet: A multi-view stereo depth estimation network with adaptive depth range and depth interval[J]. Pattern Recognition, 2023, 144: 109885.
- [59] 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.
- [60] Qi C R, Yi L, Su H, et al. Pointnet++: Deep hierarchical feature learning on point sets in a metric space[J]. Advances in neural information processing systems, 2017, 30: 5099-5108.
- [61] 董刚, 谢维成, 黄小龙等. 深度学习小目标检测算法综述[J]. 计算机工程与应用, 2023, 59(11): 16-27.
- [62] Mori M, Kashino K. Fast template matching based on normalized cross correlation using adaptive block partitioning and initial threshold estimation[C]//2010 IEEE International

Symposium on Multimedia. 2010: 196-203.

- [63] Liu W, Anguelov D, Erhan D, et al. SSD: Single shot multibox detector[C]//European Conference on Computer Vision. 2016: 21-37.
- [64] Wang C Y, Yeh I H, Mark Liao H Y. Yolov9: Learning what you want to learn using programmable gradient information[C]//European Conference on Computer Vision. 2024: 1-21.
- [65] 杨继东, 陶金池, 覃畅. 基于视觉测量的 Tripod 机器人运动标定方法研究[J]. 传感器与微系统, 2021, 40(05): 25-29.

攻读硕士期间的主要成果

- [1] 徐晶晶, 吴路路, 熊新炎. 基于改进 YOLOv8 的小目标检测算法[J]. 安徽工程大学学报 (已录用)

致 谢

在本论文的完成之际，我向所有给予我帮助和支持的师长、同学、家人以及朋友们表示最诚挚的谢意。

三年研究生期间，常常遇到问题，感谢我的导师吴路路给予的包容与耐心指导。无论是在课题选择、研究方法还是论文撰写的各个环节，都给予了耐心的帮助和宝贵的建议。感谢课题组的师兄们、同门、合作公司领导 and 同事在研究过程中对我的帮助。感谢父母和朋友在读研期间对我的理解和无条件支持。

2017 年 9 月，我带着行李箱和父母一起来到芜湖，一位学长接过我的行李箱将我领入校园，我在安徽工程大学的读书生涯由此开始。七年的本科和硕士学习生涯中，我亲眼见证了母校的发展，学院的老师和领导关心学生的生活和学习，对学生给予帮助，使我在求学道路上倍感温暖。

衷心感谢大家的帮助，希望大家越来越好！

尚德敏学 唯实惟新

