

激光与光电子学进展

位姿视觉测量方法及应用综述

魏振忠*, 冯广堃, 周丹雅, 马岳鸣, 刘明坤, 罗启峰, 黄腾达

北京航空航天大学仪器科学与光电工程学院, 精密光机电一体化技术教育部重点实验室, 北京 100191

摘要 位置和姿态参数是表征目标空间朝向和位置的基本几何量参数, 比如飞行器的飞行位姿、高端装备大部件的对接位姿、自由运动中的人体位姿等。因此, 位姿参数的准确快速测量一直是航空航天、国防军事、工业生产、体育竞技等诸多领域的重要研究内容, 并发展出惯导、卫星、雷达、星敏感器和视觉测量等多种不同测量技术途径。本文重点针对位姿视觉测量方法及其应用进行了综述, 主要内容包括位姿视觉测量基本原理、位姿视觉测量典型方法、多源数据融合的位姿视觉测量方法和位姿视觉测量典型应用领域。

关键词 测量; 位姿参数; 视觉测量; 图像处理; PnP; 深度学习

中图分类号 TP391.4 文献标志码 A

DOI: 10.3788/LOP223420

A Review of Position and Orientation Visual Measurement Methods and Applications

Wei Zhenzhong*, Feng Guangkun, Zhou Danya, Ma Yueming, Liu Mingkun,
Luo Qifeng, Huang Tengda

Key Laboratory of Precision Opto-Mechatronics Technology, Ministry of Education, School of Instrumentation and Optoelectronic Engineering, Beihang University, Beijing 100191, China

Abstract Position and orientation are the basic geometric parameters that represent the spatial orientation and position of the target, such as the flight attitude of an aircraft, the docking position while assembling large products, and the body position of athletes in sports. Therefore, accurate and rapid measurement of position and orientation is an essential task in aerospace, national defense, military, industrial production, sports, and many other fields. To this end, various techniques have been applied to position and orientation measurements, such as inertial guidance, satellite, radar, star-sensitive device, and visual measurement. Hence, this paper summarizes the position and orientation visual measurement methods and their applications, including the basic principles and typical methods, multi-source data fusion measurement method, and the main application areas.

Key words measurement; position and orientation; visual measurement; image processing; PnP; deep learning

1 引言

在三维空间中, 位置是指目标相对于三维参考坐标系所在的空间点位, 通常用三维坐标 (x, y, z) 表示。姿态是指目标本体坐标系相对于三维参考坐标系三轴的旋转角度, 通常用方位角、俯仰角、滚转角 (α, β, γ) 表示。因此, 位置和姿态是描述目标相对于参考坐标系在哪里以及方向是什么的两项基本参数, 对于飞机、卫星等典型航空航天飞行器而言, 它们是建立飞行运动方程分析飞行运动规律, 开展飞行动力学研究的基

础数据^[1]。准确测量飞行过程中的位置和姿态参数, 既可以测试飞行性能、验证设计理论、复核实验结果, 从而提升飞行器的研制水平, 同时也可以改进控制规律、完善飞行指挥、优化飞行策略, 从而保障飞行器的飞行安全。

在惯导、卫星、雷达、星敏感器和视觉测量等众多位姿测量技术途径中, 视觉测量具有全自主、可视化、抗电磁干扰、无累积误差、近场精度高等突出特点, 尤其针对外场复杂环境下高速运动目标位姿测量任务时, 已经逐步成为优选技术途径。位姿视觉测量是将

收稿日期: 2022-12-28; 修回日期: 2023-01-21; 录用日期: 2023-01-25; 网络首发日期: 2023-02-06

基金项目: 国家自然科学基金(52127809, 51625501)

通信作者: *zhenzhongwei@buaa.edu.cn

计算机视觉技术与几何量测量原理深度交叉融合, 从而实现运动目标位置与姿态的动态实时测量。具体而言, 位姿视觉测量首先需要对运动目标进行光学成像, 进而建立起目标二维图像与三维模型的中心投影几何模型之间的关系。最终利用所得二维、三维同名特征匹配关系解算目标位置与姿态参数。位姿视觉测量可有效克服惯导、卫星、雷达和星敏感器等位姿测量技术各自存在的原理和使用上的局限性, 如惯导的误差累积、卫星的拒止环境受限、雷达的无线电静默限制、星敏感器的云层遮挡等^[2]。本文主要对使用可见光视觉系统实现位姿视觉测量的方法、系统和相关应用进行综述。随着光学成像原理和光学探测系统的进步, 这些方法有可能进一步推广并应用到其他波段光学系统或使用新的探测方式获取图像^[3-5]。

2 位姿视觉测量基本原理

2.1 摄像机成像模型

小孔成像模型是最常用的摄像机成像模型, 适用于大部分的 CCD/CMOS 摄像机, 通常基于透视投影原理对成像过程进行几何建模。如图 1 所示, 三维空间点 P 发出的主光线通过摄像机光学中心 O_c , 与摄像机像平面交于图像点 p 。摄像机坐标系定义为原点 O_c 在光学中心, Z_c 轴沿光轴指向镜头外部且垂直于像平面的方向, X_c 轴、 Y_c 轴分别与像平面的水平和垂直方向平行, 形成右手坐标系。

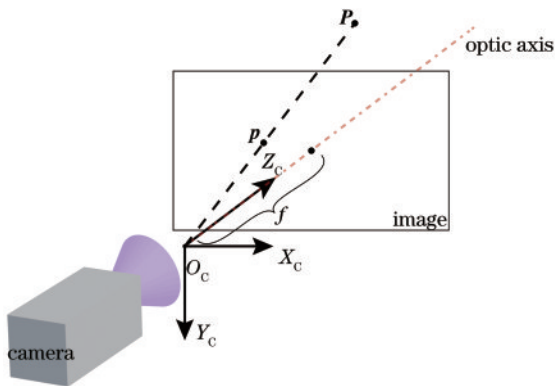


图 1 透视投影成像模型

Fig. 1 Perspective projection imaging model of a camera

根据三角相似原理, 利用齐次坐标建立三维空间点和二维图像投影点之间的映射关系, 则摄像机透视投影成像模型可表示为

$$\rho \begin{bmatrix} X \\ Y \\ 1 \end{bmatrix} = \mathbf{K} [\mathbf{R}, \mathbf{t}] \begin{bmatrix} x \\ y \\ z \\ 1 \end{bmatrix} = \begin{bmatrix} f_x & s & u_0 \\ 0 & f_y & v_0 \\ 0 & 0 & 1 \end{bmatrix} [\mathbf{R}_{3 \times 3} \quad \mathbf{t}_{3 \times 1}] \begin{bmatrix} x \\ y \\ z \\ 1 \end{bmatrix}, \quad (1)$$

式中: $(x, y, z, 1)^T$ 和 $(X, Y, 1)^T$ 分别为三维空间点和二维图像点的齐次坐标; ρ 为不为零的比例系数; \mathbf{K} 为摄

像机内参矩阵, 包含摄像机焦距 f_x 和 f_y , 主点坐标 u_0 和 v_0 , 靶面倾斜因子 s ; $[\mathbf{R}, \mathbf{t}]$ 为摄像机外参矩阵, 表示摄像机坐标系相对于世界坐标系的旋转和平移变换, 也称其为摄像机的位姿参数矩阵。

式(1)所示的透视投影成像模型将三维空间点 $P = (x, y, z)^T$ 映射为二维图像点 $p = (X, Y)^T$, 在映射过程中, 空间点距离摄像机中心的深度值丢失。因此, 已知图像点坐标可确定一条三维空间射线, 三维空间点一定位于该射线上但无法确定其具体位置, 进而导致了利用单幅图像恢复三维空间信息问题的病态性, 是单目位姿视觉测量问题的主要难点之一。

2.2 位姿参数表示

目标坐标系相对参考坐标系的位姿参数, 即空间六自由度刚体变换关系包含三自由度旋转变换和三自由度平移变换。在摄像机成像模型下, 成像目标坐标系相对于摄像机坐标系的旋转和平移关系即为待求位姿参数, 通常还需将其转换到一个统一的世界坐标系中, 如大地坐标系、场地用户坐标系等。其中, 平移关系表示为欧氏空间中的一个三维矢量 t , 旋转关系的表示有多种不同方式: 旋转矩阵、三维欧拉角、欧拉轴角、四元数、李代数等。

旋转矩阵一般表示为 $\mathbf{R}_{3 \times 3} \in \{ \mathbf{R} | \mathbf{R} \in \mathbb{R}^{3 \times 3}, \mathbf{R}^T \mathbf{R} = \mathbf{I}, |\mathbf{R}| = 1 \}$, 其与坐标系三轴的旋转变换过程一一对应, 转换过程表达清晰是其优势。欧拉角一般表示为 (ϕ, θ, γ) , 分别代表绕三个坐标轴依次进行三次旋转的角度。欧拉轴角一般表示为 (a, φ) , 代表坐标系的旋转可以表示为绕三维空间轴向 a 旋转角度 φ 。欧拉角和欧拉轴角具有直观的物理意义, 但是欧拉角存在万向锁问题, 而欧拉轴角存在旋转角度和旋转轴量纲不统一的缺点。四元数和欧拉轴角等价, 其一般表示为 $q = q_0 + q_1 i + q_2 j + q_3 k$, 其中 $q_0^2 + q_1^2 + q_2^2 + q_3^2 = 1$, i, j, k 为三个虚数单位。用四元数表示姿态的优势是参数归一且连续。由于所有三维旋转矩阵组成三自由度特殊正交群 $SO(3)$, 且该群满足李群性质, 其李代数为 $so(3)$, 是旋转的一种三自由度参数表示, 也称为旋转矢量表示法, 具体表达为 $\hat{\omega} = \log(\mathbf{R}_{3 \times 3})$, 其中 $\hat{\omega}$ 是由 ω 的三个元素组成的斜对称矩阵。

2.3 单目位姿视觉测量原理

单目位姿视觉测量基于目标二维图像特征与其三维计算机辅助设计(CAD)模型特征形成的2D-3D匹配关系, 利用摄像机成像模型建立优化目标函数实现目标位姿参数的测量, 如图 2 所示。测量过程包括以下关键步骤: 1) 图像预处理; 2) 图像特征提取; 3) 2D-3D 特征匹配; 4) 位姿参数解算; 5) 坐标系转换。

图像预处理主要是对输入图像进行图像质量提升、目标区域分割、颜色变换等处理。预处理的目的是降低图像噪声、减小背景干扰、提升特征分辨率等。图

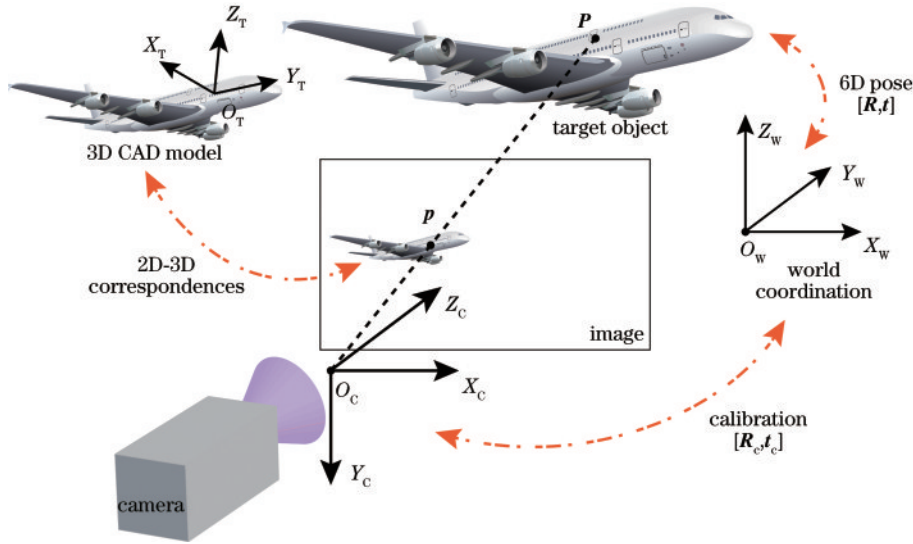


图 2 单目位姿视觉测量模型

Fig. 2 Position and orientation monocular visual measurement model

像特征提取是单目位姿视觉测量的重要步骤,其鲁棒性和精度显著影响位姿测量结果。图像特征通常是指具有鲜明几何属性的特征,如:角点、边缘线、轮廓线等,对于深度学习图像特征提取来讲也可以是神经网络提取的语义特征。2D-3D 特征匹配是针对几何图像特征与对应的三维 CAD 模型三维特征进行配对。由于透视投影过程深度信息损失,匹配问题具有病态性,因此需要结合目标三维 CAD 模型先验信息以及透视投影几何约束,才能获得准确的 2D-3D 匹配结果。位姿参数解算是利用图像特征及其 2D-3D 匹配关系求解目标位姿参数的过程,其中基于几何特征理论推导的透视投影约束方程,可直接解析求解或优化求解位姿参数,而基于深度学习提取语义特征可利用神经网络直接回归位姿参数,并通过深度学习构建特征空间到位姿参数空间的映射。坐标系转化通常是指,在测量系统坐标系下求得的位姿参数,通常需按用户要求由场地坐标系转换关系转换到用户所需的参考坐标系下,以便符合现场应用需求。

2.4 双目位姿视觉测量原理

单目位姿视觉测量的不足是由于透视投影变换过程深度信息丢失,导致难以准确获取目标空间三维坐标和尺度信息。因此,需基于目标已知的三维 CAD 模型和透视几何约束完成位姿测量。而双目视觉系统基于立体匹配和三维重构模型,可通过双目立体成像实现对目标关键特征的三维测量或者稠密三维点云的重构,从而实现目标位姿的测量。虽然系统复杂性相对较高,计算量相对大,但可有效弥补单目视觉位姿测量在目标深度恢复上的不足。

图 3 为常规双目视觉系统的三维重构模型,由三维空间处于不同位置且具有一定基线距离的两个摄像机,从两个不同视角对同一空间点成像,得到匹配像点,也称为同名图像特征点。基于三角几何约束,由同

名图像特征点的两条视线进行空间交会可唯一确定空间点 P,从而实现三维重构。图 3 中两摄像机之间的基线以及同名图像特征点的两条视线确定了双目视觉系统的极平面,极平面与两摄像机的像平面相交得到左右两条对极线,这是双目立体视觉系统进行三维重构可以使用的基本几何约束,称为极线几何约束。

基于双目视觉系统中两摄像机的成像模型以及两摄像机坐标系之间的空间转换关系,可以建立三维空间点 $P = (x, y, z)^T$ 的重构模型表示为

$$\begin{cases} x = \frac{zX_l}{f_l} \\ y = \frac{zY_l}{f_l} \\ z = \frac{f_l(f_r t_x - X_r t_z)}{X_r(r_{31} X_l + r_{32} Y_l + f_l r_{33}) - f_r(r_{11} X_l + r_{12} Y_l + f_l r_{13})} \\ = \frac{f_l(f_r t_y - X_r t_z)}{Y_r(r_{31} X_l + r_{32} Y_l + f_l r_{33}) - f_r(r_{21} X_l + r_{22} Y_l + f_l r_{23})} \end{cases}, \quad (2)$$

式中: $(X_i, Y_i, 1)^T, i = l, r$ 是同名图像特征点齐次坐标; $r_{ij}, i, j = 1, 2, 3$ 是两摄像机坐标系间的旋转矩阵 $R_{3 \times 3}$ 中的元素; $t = (t_x, t_y, t_z)^T$ 是平移矢量; f_l, f_r 为两摄像机的有效焦距。因此,基于三维重构模型式(2)即可基于双目视觉系统测量重构空间点的三维坐标。进一步,基于测得的三维数据结合目标的三维 CAD 模型,可进一步计算目标的位姿参数,如图 4 所示。

基于双目立体视觉系统的三维重构进行位姿测量的主要步骤包括:1)图像预处理;2)图像特征提取;3)立体匹配;4)三维重构;5)位姿参数计算;6)坐标系转换,其中 1、2、6 与单目位姿视觉测量的内容基本

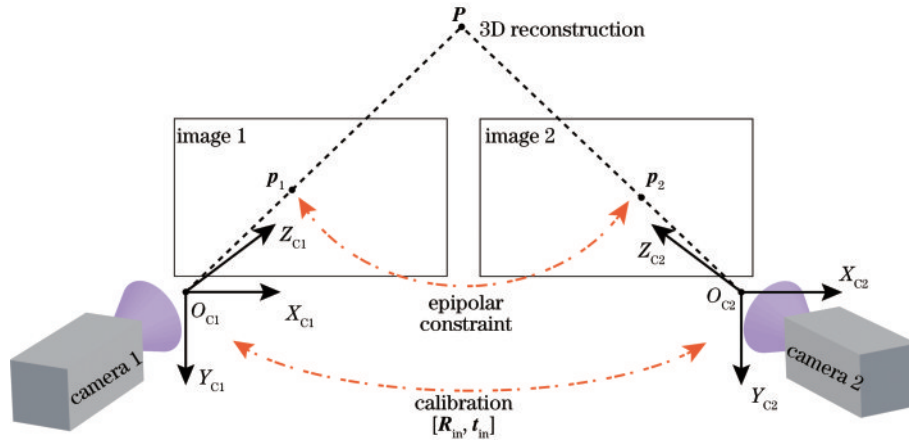


图 3 双目视觉系统三维重构模型

Fig. 3 3D reconstruction model of stereo visual measurement system

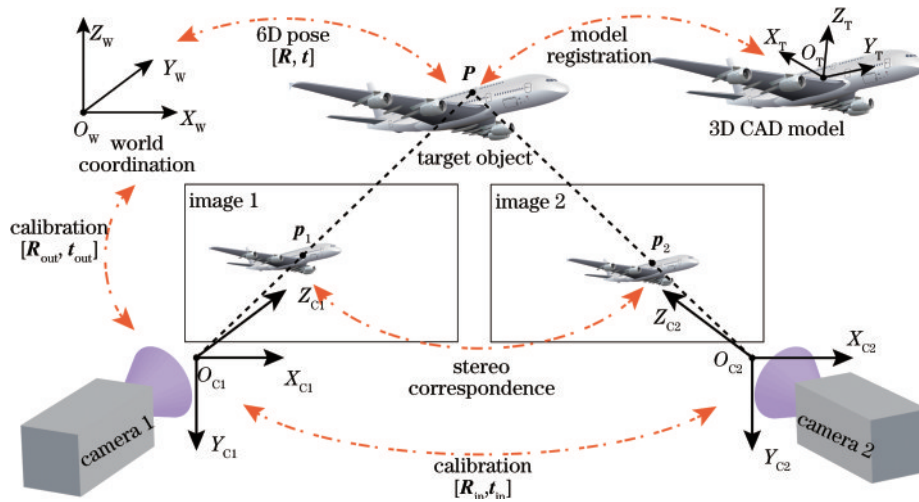


图 4 基于双目视觉系统三维重构的位姿视觉测量过程

Fig. 4 Position and orientation measurement based on 3D reconstruction of stereo visual measurement system

一致。

立体匹配是利用前述对极几何约束实现同名图像特征的匹配,其难点在于对极几何仅能将匹配点约束在对极线上,并不是一一匹配,仍然存在歧义性。因此往往需要利用灰度分布、序一致性约束等其他先验信息进一步筛选匹配对。当得到同名特征匹配结果之后,即可利用式(2)重构关键特征信息或者稠密点的三维坐标。基于传感器测量坐标系下的三维重构结果,需要与被测目标坐标系下的三维同名特征进行3D空间匹配,从而求得目标位姿参数。由于求解过程是一个3D匹配关系与位姿参数同步耦合求解的问题,难度较大,计算速度也受到限制。

3 位姿视觉测量典型方法

从位姿视觉测量的发展历程看,具体采用何种方法常与使用的具体测量特征紧密相关。总体而言,这些测量特征主要有特征、线特征和区域特征,其中线特征又涉及直线和曲线。因此,本文力图以测量特征为参照,对典型的位姿视觉测量方法进行详细综述,其

中融合了测量特征的典型提取与匹配方法。

3.1 基于点特征的位姿视觉测量方法

图像的点特征是实现位姿测量的最基本的图像特征,具有完备的几何描述和透视投影几何约束,因此,传统的基于严格理论推导的位姿视觉测量方法大都据此构建位姿求解方程,求解具有明确物理意义的最优解。而基于深度学习技术能够进一步提升点特征获取的鲁棒性,有利于解决旋转、尺度、光照等复杂干扰变化情况下2D-3D匹配和位姿求解的难题。

3.1.1 基于点特征的经典位姿视觉测量方法

图5给出了基于点特征的经典位姿视觉测量方法的发展过程。这类方法涉及的关键步骤包括:图像点特征提取、2D-3D匹配和位姿参数求解。经典的点特征提取方法通过判别图像中色彩或灰度的显著变化得到提取点特征,例如:Harris等^[6]提出的算子通过近似局部二阶泰勒展开的微分描述局部像素变化,尺度不变特征变换(SIFT)^[7]在多尺度下利用高斯微分算子寻找显著点特征。这些点特征需要和三维CAD模型中对应特征建立匹配关系即2D-3D匹配关系。由于

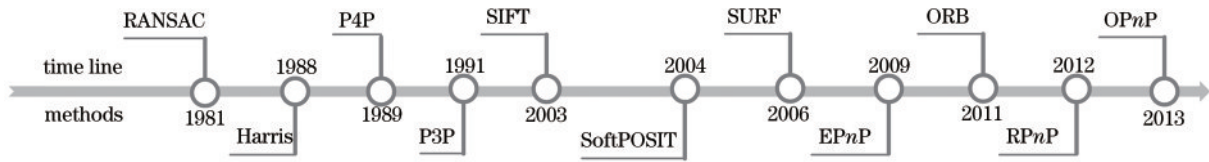


图 5 基于点特征的经典位姿视觉测量方法发展过程

Fig. 5 Development of classical object pose measurement methods based on point feature

点特征中可能包含大量非显著或干扰特征点,且透视投影过程损失深度信息导致二维与三维特征之间无法直接建立关联,因此 2D-3D 匹配一直是位姿视觉测量的难点之一。SoftPOSIT^[8]是直接同步进行 2D-3D 匹配和位姿求解的经典方法,它利用迭代思路交替更新匹配关系和目标位姿。这类方法的精度和速度取决于有效特征点的比例和匹配关系初值的质量。另一种解决方案是在线事先制作关键帧图像库,得到关键帧图像与三维 CAD 模型的匹配关系,进而在线上只需将实时图像与关键帧图像进行二维匹配即可^[9]。这类方法将 2D-3D 匹配难题转换为 2D-2D 图像间点特征匹配问题,因此可以借助 SIFT^[7]、加速稳健特征(SURF)^[10]、ORB^[11]等经典的点特征描述方法完成匹配。然而,这类方法的匹配效果取决于关键帧图像库的规模和关键帧筛选方法,为此需要在精度和速度间进行权衡。利用上述方法得到点特征 2D-3D 匹配关系后,即可在摄像机透视投影模型的约束下建立优化函数求解目标位姿参数。基于点特征的位姿测量被称为 n 点透视问题 (PnP),基于此发展了一系列经典的求解方法,称为 PnP 方法。比如,使用固定 n 个点的 P3P 法^[12]、P4P 法^[13]等,经证明 PnP 问题存在唯一解最少需要 4 个点。为了提升解的鲁棒性,通常使用更多点特征的方法通过平差,减小点特征提取误差或 2D-3D 匹配误差对位姿参数求解的影响。经典的方法有如 Lepetit 等^[14]提出的 EPnP 算法,通过构造四个控制点并以此为基底构建线性方程将求解复杂度降低至 $O(n)$ 。另外,常用的 RPnP^[15]、OPnP^[16]等非线性方法通过求解高阶非线性方程,寻求最优解并提升精度,如图 6 所示,对比了多种 PnP 算法对不同形状三维物体的位姿测量精度。针对 2D-3D 匹配中可能存在的错误,使用随机采样一致性(RANSAC)算法^[17]配合 PnP 算法可以有效筛选错误匹配提升鲁棒性。

3.1.2 基于点特征的深度学习位姿视觉测量方法

图 7 给出了基于点特征的深度学习位姿视觉测量方法的发展过程。这类方法是通过大量样本训练深度神经网络,使其习得点特征提取、匹配和位姿解算的方法。相比基于经典理论模型的方法,深度学习位姿视觉测量方法极大提升了鲁棒性,在应对遮挡、背景干扰等复杂情况时性能突出。基于点特征的深度学习位姿视觉测量方法可以进一步分为使用稀疏关键点和稠密点特征的 PnP 计算方法,以及基于点特征约束的直接

回归方法。

稀疏关键点是事先挑选和规定目标三维 CAD 模型上的显著点特征,随后训练神经网络从图像中检测、提取这些点特征,即利用深度神经网络直接完成关键点特征提取和 2D-3D 匹配过程。这类方法的关键在于点特征的选取和表示。例如 Rad 等^[18]提出的 BB8 等方法使用目标模型三维检测框的 8 个角点,然而这些角点可能偏离目标区域,因此测量精度有限。另一种关键点选取思路是使用贪婪策略,在目标三维模型上依次选取相互间距离最大的点^[19],目前尚无理论证明被选取关键点的最优性。稀疏关键点可以表示为原始坐标形式^[20],也可以通过热度图^[21]表示,或者用如图 8(a)所示的二维矢量场^[22]表示。由于后两者更接近卷积网络善于处理的张量形式往往提取精度更高。

稠密点特征是指每个图像像素位置均对应一个目标三维点,形成一种逐像素稠密的 2D-3D 匹配关系,这类方法具有更好的鲁棒性,能够提升位姿精度并克服局部遮挡问题。如图 8(b)所示,一种直观的稠密点特征表示是将每个图像像素处的三维坐标当作三个颜色通道值,由此获得用三维张量表示稠密点特征^[23]。另一种方案是借鉴计算图形学中的 UV 贴图技术,事先制作 UV 映射,将坐标信息存储在贴图中,而神经网络只需预测 UV 贴图坐标即可索引到原始三维坐标^[24]。

无论是使用稀疏关键点还是稠密点特征,最终均需要借助 PnP 算法,利用神经网络预测的点特征 2D-3D 匹配关系求解目标位姿参数,其缺陷在于神经网络仅针对特征提取和匹配进行优化训练,未能充分发挥其针对位姿测量任务的潜力。因此,最新发展的点特征引导的直接回归方法利用神经网络替代 PnP 算法^[25],随后与点特征提取模块一并参与端到端优化训练^[26],并经实验证明相比于 PnP 算法能够得到更精确的位置结果^[23]。如图 8(c)所示,相比于一般的直接回归方法,利用点特征引导神经网络学习位姿回归任务,能够辅助神经网络理解透视投影约束,提升训练的收敛速度和网络预测精度^[27]。

表 1 对比了几种不同的基于点特征的位姿视觉测量方法的精度、速度、适用性。尽管基于点特征的经典位姿视觉测量方法在近年来的研究中得到了长足的发展,但是其特征点选取困难、计算量较大以及在复杂环境下鲁棒性不足的问题,仍然没有得到解决。而随着

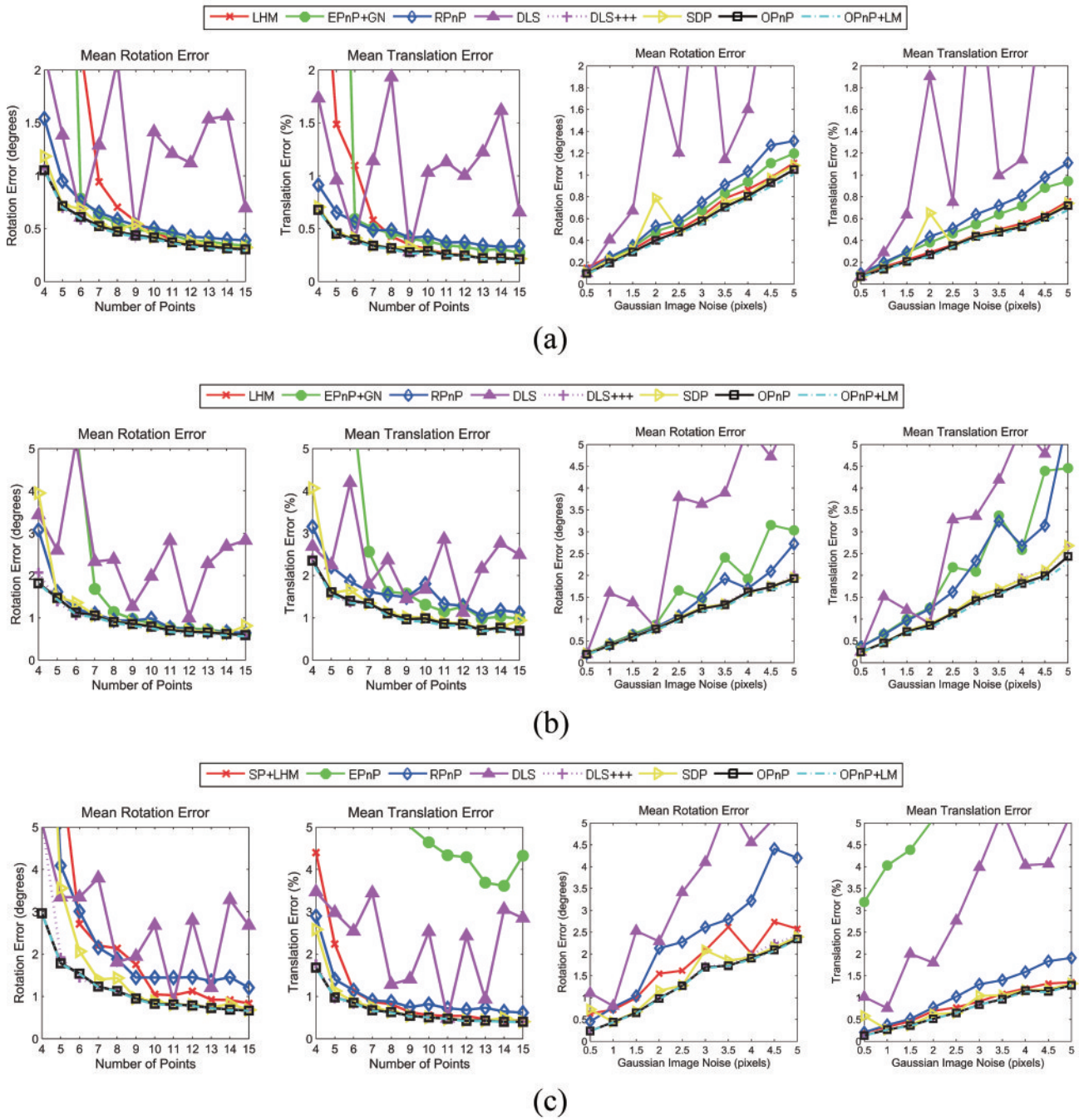


图 6 多种 PnP 算法对比^[16]。(a)一般三维点分布；(b)半奇异点分布；(c)共面点分布
 Fig. 6 Comparison of different PnP algorithms^[16]. (a) Ordinary-3D point configuration; (b) quasi-singular point configuration; (c) planar point configuration

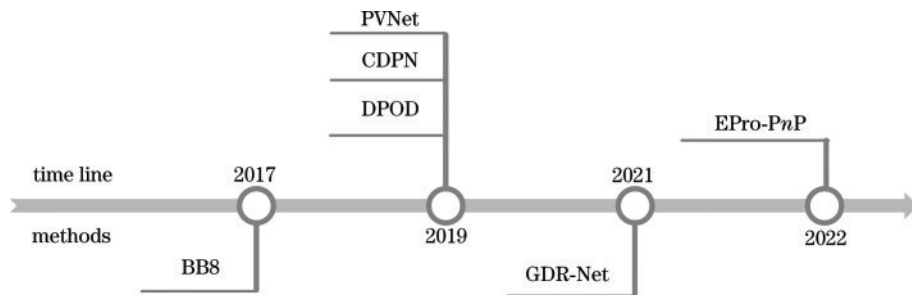


图 7 基于点特征的深度学习位姿视觉测量方法发展过程
 Fig. 7 Development of deep learning object pose measurement methods based on point feature

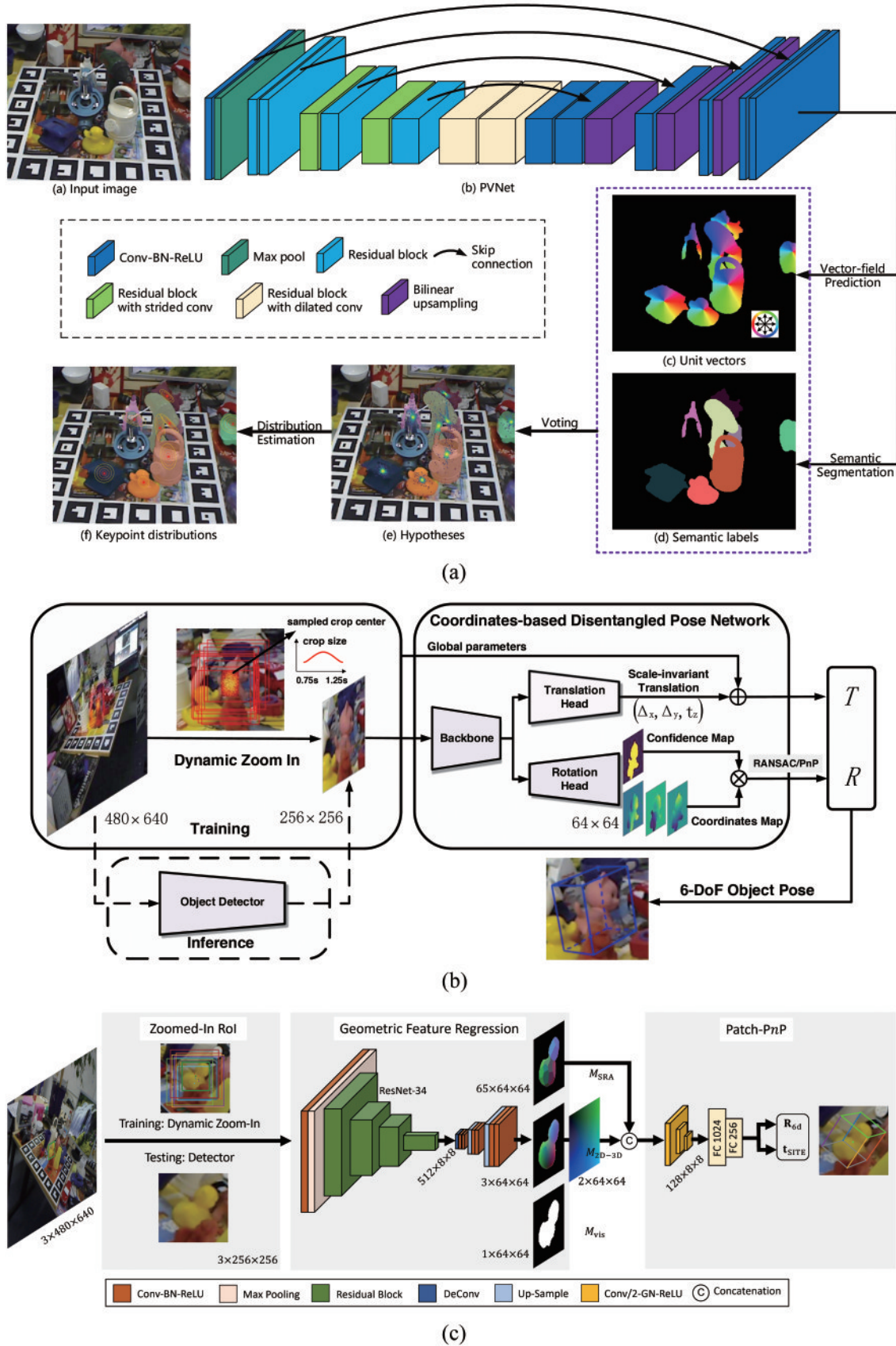


图 8 基于点特征的深度学习位姿视觉测量方法。(a)基于稀疏关键点的PVNet方法^[22];(b)基于稠密点特征的CDPN方法^[23];(c)基于点特征引导直接回归的GDR-Net方法^[27]
 Fig. 8 Deep learning based visual pose measurement using point feature. (a) PVNet^[22] based on sparse keypoints; (b) CDPN^[23] based on dense points; (c) GDR-Net^[27] based on points-guided direct regression

表 1 基于点特征的位姿视觉测量方法对比

Table 1 Comparison on object pose measurement methods based on point feature

Method	Year	Accuracy	Speed	Feasibility
SURF ^[10]	2006	Good	Common	Scale and rotation invariance
ORB ^[11]	2011	Common	Good (10 times faster than SURF)	Suitable for applications requiring speed priority
BB8 ^[18]	2017	Common	Common	The deviation between corner and target area leads to limited measurement accuracy
PVNet ^[22]	2019	Good	Good	Suitable for severe occlusion or truncation
CDPN ^[23]	2019	Better	Better	Better robustness and local occlusion resistance
GDR-Net ^[27]	2021	Better	Better	Ensuring the real-time performance, accuracy and robustness

深度学习在位姿测量领域的广泛应用,基于点特征的深度学习位姿视觉测量方法得到了进一步的发展。相较于经典的点特征位姿视觉测量方法,深度学习的方法能够有效实现光照条件恶劣以及遮挡情况下的点特征提取,同时能够极大的降低运算量,缩短系统的运行时间。同时,基于深度学习的方法在其训练过程的收敛速度以及训练模型的通用性问题中,仍需要进一步的研究,这也会是未来基于点特征的位姿视觉测量方法的研究热点。

3.2 基于线特征的位姿视觉测量方法

少纹理/无纹理物体的点特征提取数量往往受到限制,且点特征对于遮挡、光照等环境影响因素较为敏感而影响位姿测量的鲁棒性和精度。线特征用于实现三维空间位姿状态的描述,对于光照变化和图像噪声等具有不变性,且对于遮挡具有较强的鲁棒性,因此,基于线特征的位姿视觉测量方法得到广泛研究。典型的线特征包括道路、建筑物、室内陈设等场景中的直线特征,以及待测物体边缘/轮廓所呈现的曲线特征等。基于直线特征的位姿视觉测量方法可首先通过直线特征的提取和匹配,进而由 Perspective- n -Line (P n L)算法求解目标位姿。此外,基于边缘/轮廓特征的位姿测量方法通过不断投影迭代优化位姿参数,直至达到物体三维 CAD 模型投影与目标边缘/轮廓完全重合的优化目标。

3.2.1 基于直线特征的经典位姿视觉测量方法

基于直线特征的位姿测量方法与基于点特征的位姿测量方法类似,首先需进行直线特征提取,进而建立直线

2D-3D 对应匹配关系。典型的直线特征提取方法包括 Hough 变换^[28]、LSD 直线检测算法^[29]、EDLines 算法^[30]和基于深度学习的直线检测方法^[31]等,直线匹配方法包括 MSLD 算法^[32]、基于直线带状区域描述子 LBD 的匹配算法^[33]等。当已知空间 3D 直线及其对应的图像 2D 直线,便可求解世界坐标系与摄像机坐标系的相对位姿参数 R 和 t ,如图 9^[34]所示。

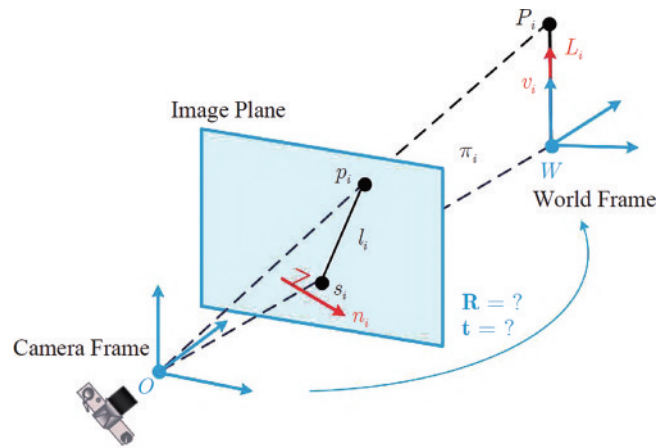


图 9 P n L 问题空间几何示意图^[34]

Fig. 9 Space geometry diagram of the P n L problem^[34]

根据直线对数量的不同,位姿求解问题可分为 P3L 问题和最小二乘 P n L 问题。Dhome 等^[35]最早采用多项式方法求解 P3L 问题的解析解,但此问题求解所获得的位姿不唯一。当直线对数量 $n \geq 4$ 时,可采用迭代优化方法、直接线性方法和多项式方法求解最小二乘 P n L 问题,P n L 方法的发展过程如图 10 所示。

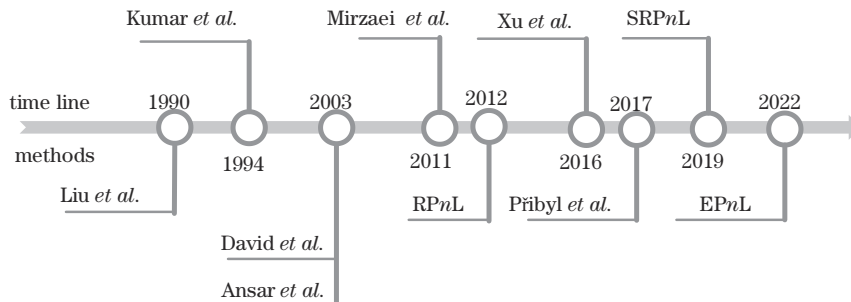


图 10 P n L 方法的发展过程

Fig. 10 Development of P n L methods

1) 迭代优化方法。Liu 等^[36]和 Kumar 等^[37]分别对位置和姿态进行分步和同步优化。而为提升复杂场景下线特征匹配和求解的鲁棒性, David 等^[38]基于 SoftPOSIT 算法^[8]所采用的假设-测试思想, 同时对直线特征 2D-3D 对应和位姿参数进行迭代优化。迭代优化方法受位姿初始影响较大, 不理想的位姿初值给迭代优化带来较大的时间成本, 且容易在优化过程中陷入局部最小。

2) 直接线性变换方法。Ansar 等^[39]通过直线匹配, 由直接线性变换(DLT)算法求解目标位姿。Xu 等^[40]将直线上 3D 点表示为重心坐标, 通过有效的零空间求解器获取摄像机坐标系下的控制点位置, 进而由坐标系对齐获取摄像机位姿。此后, Příbyl 等^[41]联合 DLT 算法和 Plücker 坐标表示, 获得较小的位置测量

误差和重投影误差。DLT 算法求解速率快, 但在直线对数量较少情况下求解不稳定。

3) 多项式方法。Mirzaei 等^[42]将位姿优化的最小二乘问题表示为多项式方程组求解, 通过乘法矩阵的特征分解有效获取方程组的解。为了进一步提升求解稳定性、准确性和实时性, 研究学者相继提出了 RPnL 算法^[43]、ASPnL 算法^[40]、SRPnL 算法^[44]和 EPnL 算法^[34]。多项式方法求解所需线对数量较少, 求解精度高, 但计算过程相对复杂。当前, 主流的 PnL 位姿求解方法以多项式方法为主, 未来该方法的发展将在保持精度和直线对数量不变的情况下, 进一步提升求解速率。主流 PnL 算法在 VGG 数据集上测试结果对比, 如表 2 所示。

表 2 PnL 方法在 VGG 数据集上的对比
Table 2 Comparison of PnL methods on VGG dataset

Method		Model-House	Corridor	Merton-College- I	Merton-College- II	Merton-College- III	University-Library	Wadham-College
Number of images		10	11	3	3	3	3	5
LPnL-Bar-LS ^[40]	$\Delta\theta / (^\circ)$	0.4135	0.1178	0.0241	0.0261	0.0652	0.3642	0.1526
	$\Delta t / m$	0.0403	0.0440	0.0099	0.0149	0.0233	0.1632	0.0909
AlgLS ^[42]	$\Delta\theta / (^\circ)$	0.4220	0.1983	3.6200	55.8037	3.7495	1.8838	60.0517
	$\Delta t / m$	0.0384	0.0888	1.1504	14.1879	1.3683	0.9519	9.8801
RPnL ^[43]	$\Delta\theta / (^\circ)$	0.5521	0.3652	1.0870	0.3249	1.7528	2.9731	0.4200
	$\Delta t / m$	0.0631	0.1150	0.3215	0.1660	0.9121	1.5613	0.1909
ASPnL ^[40]	$\Delta\theta / (^\circ)$	0.2265	0.0911	0.1141	0.1515	1.5584	3.6662	0.4227
	$\Delta t / m$	0.0162	0.0298	0.0314	0.0600	0.5571	1.6683	0.1955
SRPnL ^[44]	$\Delta\theta / (^\circ)$	0.2258	158.9520	0.4381	0.1151	36.4034	4.1848	0.0880
	$\Delta t / m$	0.0160	17.5570	0.1064	0.0495	3.9398	2.0632	0.0407
EPnL ^[34]	$\Delta\theta / (^\circ)$	0.2265	0.0969	0.0306	0.0170	0.0504	0.0871	0.0808
	$\Delta t / m$	0.0162	0.0252	0.0097	0.0123	0.0147	0.0343	0.0375

3.2.2 基于边缘/轮廓特征的位姿迭代方法

位姿迭代所用的线特征包括物体的直线边缘或曲线轮廓, 通过反复迭代将 3D 模型投影所得的边缘/轮廓与实时图像提取对应的边缘/轮廓进行对齐, 获取物

体的位姿参数。这里以基于 3D 模型轮廓的位姿迭代为例, 该方法中关键环节所对应的处理结果如图 11^[45]所示。基于边缘/轮廓特征的位姿迭代方法具体的发展过程如图 12 所示。

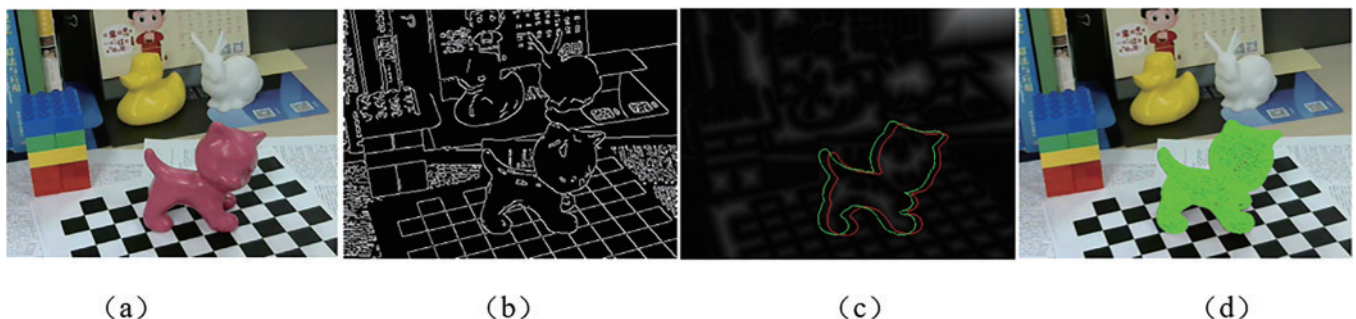


图 11 基于轮廓特征的位姿迭代过程^[45]。(a)RGB 图像;(b)轮廓提取结果;(c)初始位姿(红色)和优化位姿(绿色)对应的轮廓投影;(d)优化位姿对应的模型投影
Fig. 11 Process of pose iterative method based on contour features^[45]. (a) RGB image; (b) extracted contour features; (c) 2D contour projections corresponding to initial pose (red) and optimized pose (green); (d) model projection corresponding to optimized pose

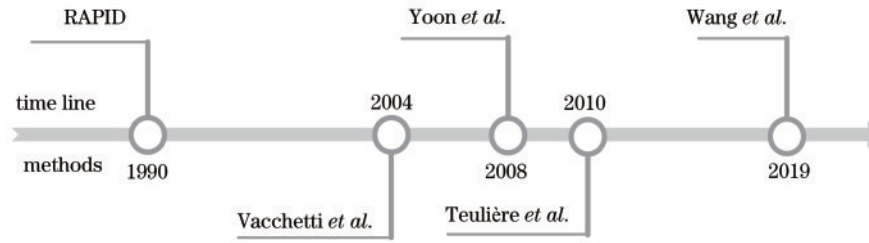


图 12 基于边缘/轮廓特征的位姿迭代方法的发展过程

Fig. 12 Development of pose iterative methods based on edge and contour feature

早在 1990 年, Harris 等^[46]提出了基于边缘的位姿迭代估计方法 RAPID, 此方法首先通过 3D 模型投影获取模型中所标注的投影边缘, 沿垂直于投影边缘方向寻找亮度梯度最大的图像边缘点。通过最小化投影边缘与图像边缘上对应点的距离之和, 进而求解物体位姿。实际迭代优化过程中, 尤其在相邻帧物体运动幅度较大时, 物体中所提取复杂线特征的外观相似性给匹配和位姿跟踪测量带来极大的挑战。为此, Vacchetti 等^[47]沿投影边缘法线方向寻找相似的图像边缘, 从特征层面建立多个匹配假设, 优化选取最小距离值所对应的匹配假设。Yoon 等^[48]采用卡尔曼滤波搜索物体位姿, 使投影边缘与实际边缘之间能够达最佳匹配。随着算力的不断提升, Teulière 等^[49]采用粒子滤波方法, 根据前一时刻的粒子状态和统计运动模型构建当前帧的多个位姿假设, 进而计算投影线特征与提取线特征的相似性对各粒子状态进行评价, 重复采样获取当前帧最优的粒子状态。除以上基于边缘/轮

廓对应点距离外, 为了避免因投影线特征与图像线特征匹配搜索的误差给位姿测量造成的影响, Wang 等^[45]根据图像像素相对于物体轮廓的距离值建立距离场, 根据不同距离场的差异性构建整体匹配的能量函数, 进而迭代求解物体位姿。

3.3 基于区域特征的位姿视觉测量方法

基于区域的位姿视觉测量方法相较于点、线等特征, 可以充分利用目标表面更多的特征, 对受遮挡或者具有对称性的目标位姿测量的鲁棒性更好。基于区域的位姿视觉测量方法可以分为两类: 1) 基于局部区域特征的经典位姿视觉测量方法, 这一类方法利用局部区域特征构建 2D-3D 对应匹配关系求解位姿; 2) 基于全局区域特征的位姿视觉迭代测量方法, 这一类方法通过目标函数最小化目标 3D 模型的投影区域和图像的重投影误差求得位姿最优解。如图 13 展示了基于区域特征的位姿视觉测量方法的发展过程。

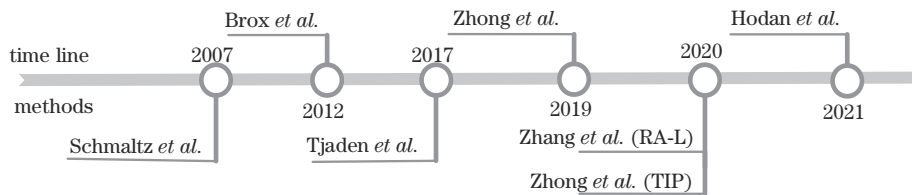


图 13 基于区域的位姿测量方法时间线

Fig. 13 Time line of region-based pose measurement method

3.3.1 基于局部区域特征的经典位姿视觉测量方法

对于基于局部区域特征 2D-3D 对应的方法, Hodaň 等^[50]提出了利用可控数量的局部表面区域表示一个目标, 如图 14 所示。该方法使用编码器-解码器卷积神经网络预测密集像素和局部表面之间的对应关系, 在每个像素上预测目标存在概率、给定目标的局部区域概率以及像素在局部区域内坐标系下的 3D 位置, 然后通过 PnP-RANSAC 算法的强大且高效的变体来估计可能存在的多个目标实例的位姿, 能够有效处理对称问题, 但局部表面数量如何确定有待探讨。在 Hodaň 等工作基础上, Wang 等^[27]提出几何引导直接回归网络, 通过区域对应和逐像素对应进行直接位姿回归, 区域对应作为辅助监督来解决对称目标的混淆问题。Pitteri 等^[51]提出将深度学习和 3D 几何结合在一

起, 通过基于局部嵌入将 CAD 模型与输入图像进行匹配。对于物体表面的点, 这种嵌入直接从 CAD 模型计算出来, 而对于图像位置通过神经网络预测。由于许多点可能具有相似的局部几何结构, 因此这些对应关系中有许多是不明确的。该方法通过 Mask-RCNN 以类不可知的方式检测新目标, 而无需对新目标进行局部表面嵌入训练, 因此该方法能够对未训练过的目标进行位姿测量, 泛化性能好。Shi 等^[52]提出从 RGB-D 图像中提取几何稳定块进行位姿估计。该方法提出首先通过检测和分割 RGB 图像, 然后裁剪相应位置的深度点云。根据对象标签, 根据对称性对目标进行分类。对于每个目标点云, 提取平面和圆柱面片, 并基于稳定性分析对一组几何稳定的面片组进行采样。基于目标几何稳定的面片组, 将每个面片的几何特征和面片间

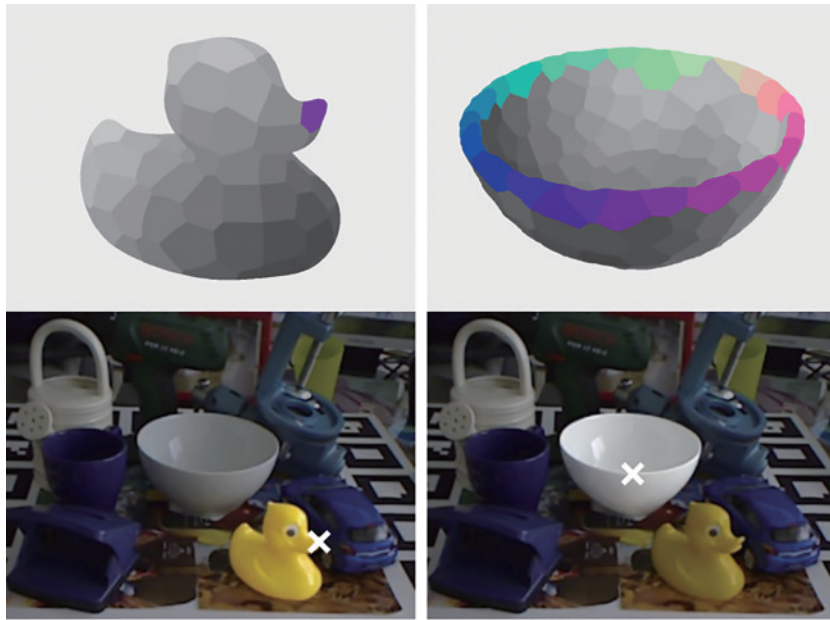


图 14 目标表面局部区域对应^[50]

Fig. 14 Local area correspondence of target surface^[50]

的上下文特征串联输出多层感知器中获得位姿。该方法能够解决因对称引起的位姿模糊问题,并能够泛化到未见目标。

3.3.2 基于全局区域特征的位姿视觉迭代测量方法

除了前述利用局部区域特征构建 2D-3D 对应的位姿测量方法,基于全局区域的特征在位姿迭代优化方法中也获得广泛使用。在深度学习未大规模发展前,该方法主要采用基于水平集的目标分割方法获得目标区域,构建目标函数优化位姿。Schmaltz 等^[53]提出将三维形状先验知识集成到基于水平集的图像分割和基于轮廓的三维姿态跟踪的变分模型中,通过分割方法提取的物体轮廓来估计物体的三维姿态参数,能够在一定程度上处理部分遮挡和表面反光。在此基础上,Brox 等^[54]提出直接定义在位姿参数上的能量函数优化位姿而不需要现实地提取目标轮廓,通过使用当前位姿投影出的区域更新目标和背景区域的概率分布模型来优化位姿,进一步提高了算法

的鲁棒性。Bibby 等^[55]提出了一种基于逐像素后验概率的二维视觉跟踪算法,该方法忽略图像中目标区域和背景区域各自内部像素的相关性,建立了图像的生成式成像模型,提升了位姿跟踪的实时性。Prisacariu 等^[56]基于上述的工作提出同时进行 2D 分割和 2D 至 3D 位姿跟踪的概率框架,通过像素级后验隶属概率(与似然相反)定义该区域及其周围背景的能量,根据 3D 模型的位姿参数推导出投影区域和图像区域的能量的差异,进而通过最优化方法搜索正确的位姿,该方法简化了计算并利用 GPU 加速进一步提高了算法的实时性能。Tjaden 等^[57]和 Liu 等^[58]都是一种局部分割模型,如图 15(a)所示,通过附着在目标表面的多个圆形区域内标记前景和背景概率,并推导出高斯-牛顿优化方案,该方案可大大加快收敛速度,并具有高度精确且鲁棒的跟踪性能。之后,进一步发展了基于深度学习的方法。Zhong 等^[59]结合了基于深度学习的语义分割框架和基于优化的位姿迭代模块,分割网络采

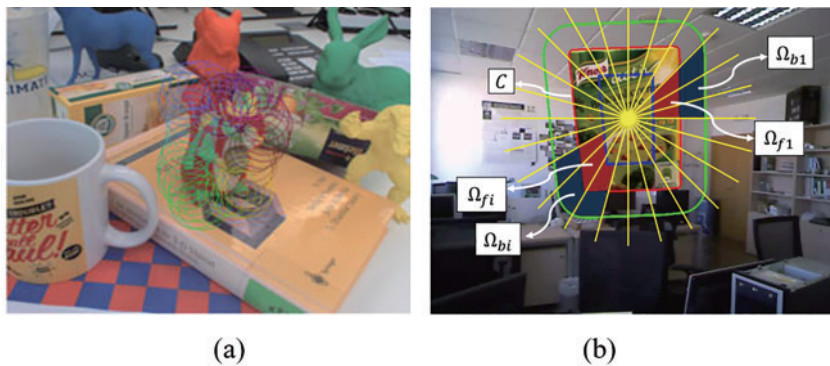


图 15 基于不同形状的局部区域估计前景和背景概率。(a)圆形区域^[57]; (b)带状区域^[61]

Fig. 15 Estimation of foreground and background probability based on different shape of areas. (a) Local circular area^[57]; (b) banded area^[61]

用 DeepLabV3+^[60] 预测物体包括被遮挡部分的准确完整掩膜,在预测的全掩膜的指导下执行非线性的 6 自由度的位姿优化方法。该方法在处理重度遮挡方面优于 Tjaden 等^[57] 提出的方法,并且可以处理以前方法失败的极端情况。在 Tjaden 等^[57] 工作基础上,Zhong 等^[61] 在算法中将依附在模型表面的圆形区域改为依附在轮廓两侧的带状区域,同时融合目标表面的“描述子场”,如图 15(b)所示,进一步提升了算法的稳定性。在此基础上,Zhong 等^[62] 提出在目标被部分遮挡的条件下,结合目标表面的颜色和图像中的边缘分布,通过给区域损失附加权重的方式,降低目标遮挡对优化的影响,提高了遮挡情况下的算法鲁棒性。

基于区域的方法相比基于点、线等特征的位姿测量方法,能够利用更多目标表面特征,具有更强的鲁棒性和更高的精度,但是区域特征的表达形式仍不能够充分发挥这类方法的优势,表 3 对比了近些年基于区域的位姿测量方法的精度、速度、适用性。如图 13 所示,早期的基于区域的位姿测量方法主要采用概率统计方法构建局部区域的描述,近些年来在基于区域的位姿测量方法中逐步引入深度学习提取和表示特征,进一步提高了基于区域特征的位姿测量方法的性能。由于基于区域的方法需要对所有目标表面像素进行处理,这降低了这类方法的处理速度。如何提高基于区域的方法的处理速度和鲁棒的区域特征表示方法仍有待进一步研究。

表 3 基于区域特征的位姿视觉测量方法对比

Table 3 Comparison of pose measurement methods based on region-based feature

Method	Year	Accuracy	Speed	Feasibility
Schmaltz <i>et al.</i> ^[53]	2007	Common	Common	Good application effect in weak texture, just for asymmetric targets
Brox <i>et al.</i> ^[54]	2012	Better	Common	Better application effect for occluded or symmetric targets, better robustness
Tjaden <i>et al.</i> ^[57]	2017	Better	Good	Better application effect for occluded or symmetric targets, good robustness and better real-time performance
Zhong <i>et al.</i> ^[59]	2020	Better	Better	Better robustness and more suitable for symmetric targets
Hodañ <i>et al.</i> ^[50]	2021	Good	Common	Better suitable for occluded or symmetric targets, better robustness

3.4 基于多特征融合的位姿视觉测量方法

近年来,视觉位姿测量方法不断地更新优化,均利用了物体的几何信息作为计算特征,如点、线、边缘、区域、轮廓等。但从分析来看,目前大多数方法往往仅用上述单一特征实现位姿测量,这使得在遮挡、光照变化、物体高度对称等情况下位姿测量效果不佳,容易失效。实际上,在不同场景下,不同特征对场景变化的敏

感度不同,因而结合物体的多种特征能够在不同场景下更充分更完整地表达物体的位姿信息,且可更有效地抵抗遮挡、光照变化、旋转尺度变化等挑战性因素对位姿测量鲁棒性和精度的影响,有效提升位姿测量的整体效果。图 16 给出了基于多特征融合的位姿视觉测量方法的发展过程。

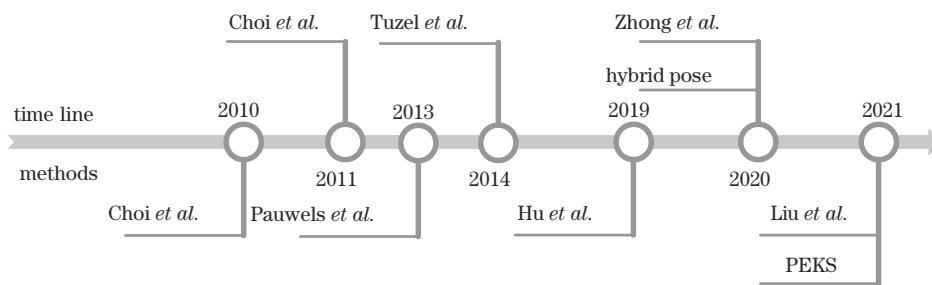


图 16 基于多特征融合的位姿视觉测量方法发展过程

Fig. 16 Development of object pose measurement methods based on multiple features

3.4.1 基于多特征融合的传统位姿视觉测量方法

Choi 等^[63-64] 提出了一种融合边缘和点两种特征的位姿迭代估计方法。该方法使用特征点在全局位姿估计过程中估计初始位姿,然后在局部位姿估计过程中使用边缘特征对位姿进行迭代优化。Pauwels 等^[65] 利用双目视觉方法,采用对被测目标的表面点特征、颜色特征和光流信息等多特征融合的方法对位姿进行迭代估计,并且在弱纹理的目标上取得了

很好的成绩,如图 17 所示。Tuzel 等^[66] 对目标提取不同的点对特征,包括表面到表面的点对特征、表面到边界的点对特征以及边界到边界的点对特征,通过对模型进行训练来获取任务中不同特征的重要程度,然后对这些特征采取分级策略,实现了在提高准确度的同时降低了运算成本。Liu 等^[58] 针对飞机飞行过程中的位姿测量问题,提出了一种融合区域和轮廓特征的位姿估计方法,设计了基于距离变换、赫

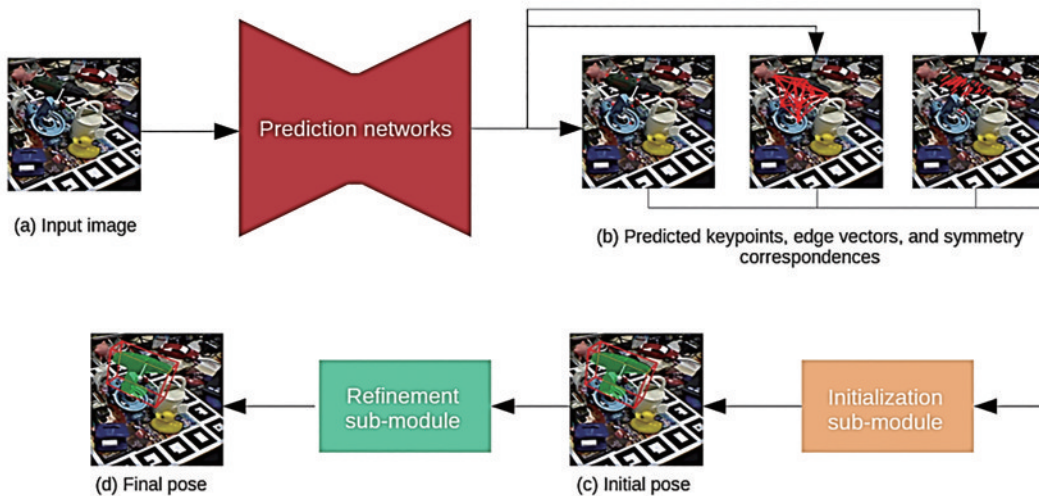
图 17 强遮挡情况下的物体位姿测量结果^[65]Fig. 17 Object pose measurement results under strong occlusion^[65]

维赛德变换的投影目标区域归属概率计算方法和基于相对熵差异度量的位姿优化求解算法,并结合全局稠密关键点特征进行位姿联合优化解算,能够有效避免对称目标的二义、多义性,在精度和鲁棒性方面都表现优异。

3.4.2 基于多特征融合的深度学习位姿视觉测量方法

Hu 等^[67]运用深度学习的方法进行物体位姿测量,设计了一种双流网络结构,网络中的分割流和回归流分别针对于图像的区域特征和点特征,结合两种特征输出更加可靠的位姿信息,能够有效地处理多个相互遮挡的纹理较差的对象。Zhong 等^[62]提出了一种基于极坐标的本地区域分割方法,在检测遮挡部分的像素时使用了区域距离特征和颜色特征融合的方式,对部分遮挡情况的目标进行位姿测量时有较好的鲁棒性。

Song 等^[68]提出一种基于混合特征表示的位姿估计方法,如图 18 所示,首先对三种物体特征进行提取和融合,然后利用改进的 EPnP 算法进行位姿求解,有效地提升了位姿测量的精度和稳定性。该方法中定义了关键点、边缘和对称信息三种物体特征,用于表征物体局部几何信息、物体不同部分之间的关系和物体内部的结构关系,但由于边缘特征设定为两个关键点之间的连线,边缘特征的提取依赖于关键点特征的提取,仍有进一步提升的空间。Fan 等^[21]针对飞机目标,基于飞机形状结构特点,选取了两种具有针对性的特征:关键点和骨架,其中,关键点表征飞机局部的集合信息,骨架代表全局的拓扑结构,并利用堆叠沙漏网络对特征进行预测。该方法较现有的基于关键点的方法显著提升了精度,并且对室外远距离场景复杂天气下的位姿测量有很好的鲁棒性,如图 19 所示。

图 18 混合表示下的位姿测量方法^[68]Fig. 18 Pose measurement method under hybrid representations^[68]

基于多特征融合的位姿测量方法,由于不同特征对不同场景变化的敏感度不同,相对于单一特征类的位姿测量方法有更好的鲁棒性。在抵抗遮挡、光照变化、旋转尺度变化等场景有效地提升了位姿测量的整

体效果。此类方法的效果对比如表 4 所示。在不同场景下,如何更具针对性地利用不同特征,实现不同特征的有效结合,是未来研究的重点。

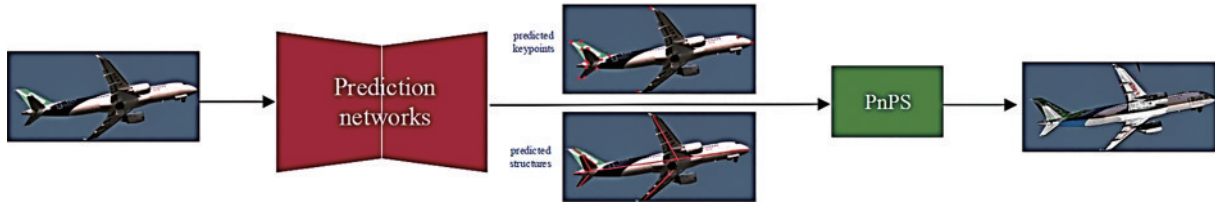


图 19 基于关键点和骨架的飞机姿态位姿测量方法^[21]

Fig. 19 Aircraft pose estimation approach with keypoints and structures^[21]

表 4 基于多特征的物体姿态测量方法比较

Table 4 Comparison on object pose measurement methods based on multiple feature

Method	Year	Accuracy	Speed	Feasibility
Choi <i>et al.</i> ^[63]	2010	Common	Good	Need to know the 3D coordinates of each keypoint
Choi <i>et al.</i> ^[64]	2011	Good	Common	Need to know the 3D coordinates of each keypoint
Pauwels <i>et al.</i> ^[65]	2013	Common	Good	Difficult to extend to the articulated scenario
Tuzel <i>et al.</i> ^[66]	2014	Good	Common	Identifies and ranks important features on 3D object
Liu <i>et al.</i> ^[58]	2021	Good	Better	Suitable for application
Fan <i>et al.</i> ^[21]	2021	Good	Common	Need the 3D aircraft model
Hu <i>et al.</i> ^[67]	2019	Common	Good	Suitable for severe occlusion
Zhong <i>et al.</i> ^[62]	2020	Good	Common	Still fail for heavy occlusions
Song <i>et al.</i> ^[68]	2020	Good	Better	No efficient 3D object detection module for pose initialization Limited by image resolution

3.5 基于三维重构的位姿视觉测量方法

基于三维重构的位姿视觉测量是指利用前述双目(也可以多目)视觉测量系统实现目标特征三维重构,进而求解目标位姿参数。基于双目三维重构能够有效解决单目视觉位姿测量方法由于深度信息缺失导致的

三维空间感知二义性问题,提供更加精准的位置和姿态测量结果。按照三维特征重构的稀疏程度可以将这类方法分为基于关键特征三维重构的方法和基于稠密三维点云重构的方法。图 20 给出了基于三维重构的位姿视觉测量方法的发展过程。

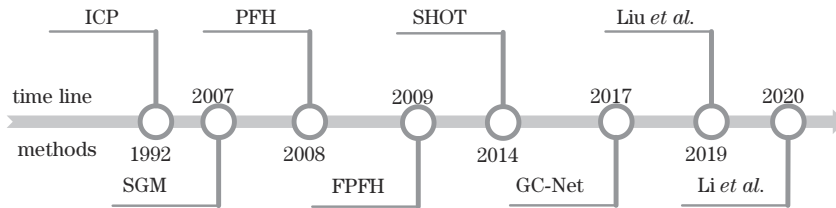


图 20 基于三维重构的位姿视觉测量方法发展过程

Fig. 20 Development of object pose measurement methods based on 3D reconstruction

3.5.1 基于关键特征三维重构的视觉位姿测量方法

关键特征是指事先定义的能够在三维空间中完整表示目标三维位姿的稀疏特征。基于关键特征的方法首先需要从双目(或多目)视觉系统得到的两幅或多幅图像中提取稀疏关键特征,并根据三维重构模型对其进行三维重构,进而利用这些特征求解目标位姿参数。经典方法如前述点、线特征等提取算法在特征提取基础上,利用特征描述相似度^[7]和对极几何约束^[69]等实现立体匹配。随后利用所得立体匹配关系,在世界坐标系下重构这些特征的三维信息。由于事先已知这些特征在目标三维 CAD 模型坐标系下的三维信息,因此可以直接利用奇异值分解(SVD)^[70]求得二者之间的刚体变换即为目标位姿参数。然而,关键特征提取和

立体匹配易受图像质量影响,且求解难度高、运算量较大,因此这类经典方法的鲁棒性通常较差。当前,深度学习技术快速发展并展现了其强大的鲁棒性。如图 21 所示,一些学者进一步研究利用神经网络实现关键特征的提取和匹配^[71-72]。由于特征提取和匹配不可避免存在误差,一些方法进一步利用 RANSAC 算法^[17]发现并消除粗大误差的影响。另一种方案则是建立关键特征三维投影误差与二维图像重投影误差目标优化函数,运用 Bundle Adjustment^[73]方法实现平差,获得位姿参数的最优测量结果。

3.5.2 基于稠密三维重构的视觉位姿测量方法

稠密三维重构是指对每个参考图像像素的位置均重构其三维坐标信息。重构所得的稠密三维信息可以

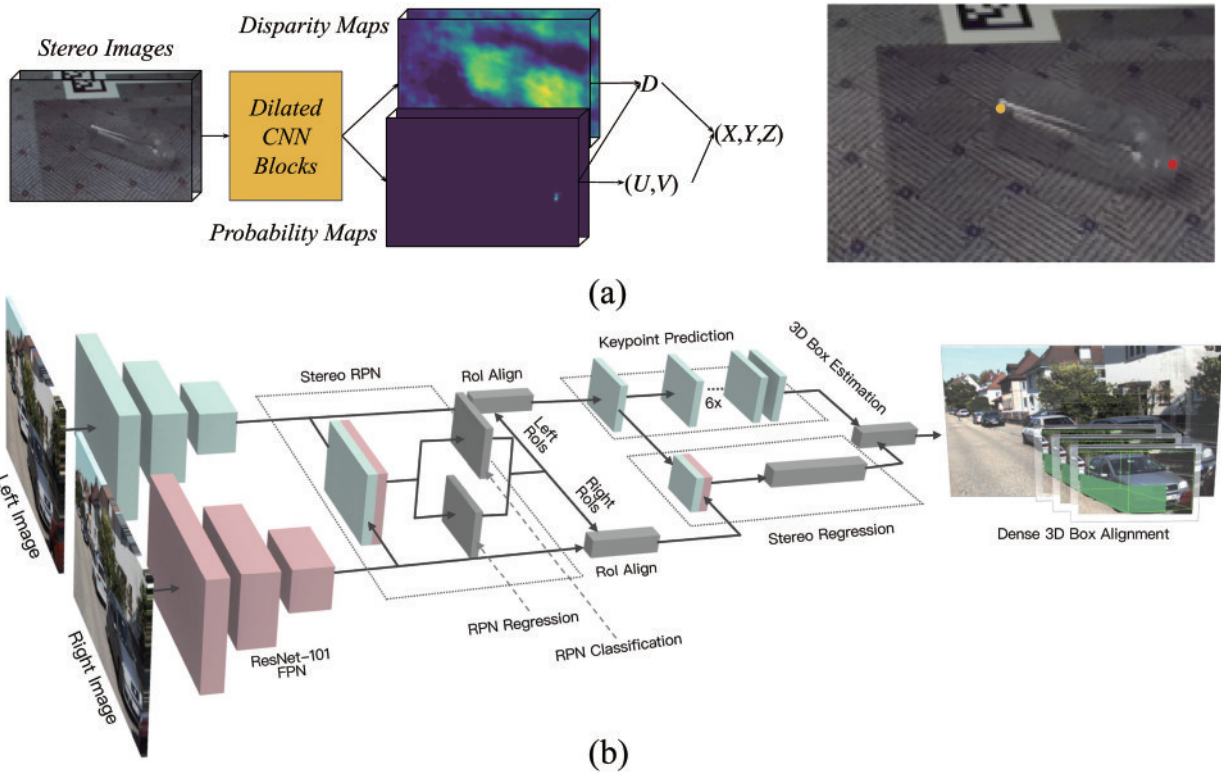


图 21 双目立体视觉重构与姿态测量方法。(a)基于概率图表示关键点的双目立体视觉重构与姿态测量^[71];(b)基于 3D 检测框角点的双目立体视觉重构与姿态测量方法^[72]

Fig. 21 Multi-view pose measurement methods. (a) Multi-view 3D keypoints estimation^[71]; (b) multi-view 3D bounding box estimation^[72]

表示为三维点云或进一步为其添加面元信息形成三维网格。由于所重构的稠密三维信息来自于图像像素,而非与目标三维 CAD 模型一一对应,因此需要将所得三维信息与目标三维 CAD 模型进行配准,在此过程中可同时求解目标位姿参数。稠密重构需要建立图像间

稠密立体匹配关系即视差图,经典算法常采用局部、全局、半全局等匹配策略得到视差^[74]。图 22(a)为一种经典的半全局方法,将二维全局匹配最优搜索简化至一维,以此达到测量精度和速度的平衡。然而,经典方法往往难以应对弱纹理、透明、反光物体等因素干

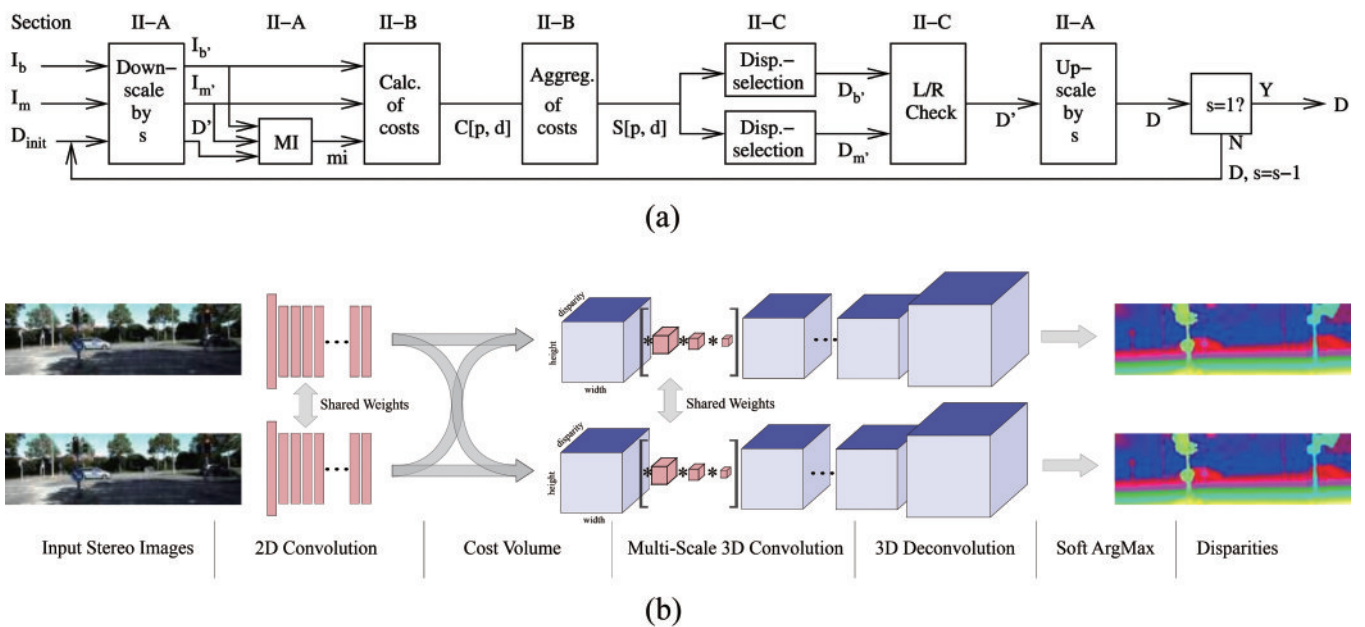


图 22 视差估计方法。(a)SGM 方法^[74];(b)GC-Net 深度视差回归^[75]

Fig. 22 Disparities estimation method. (a) Semiglobal matching (SGM) stereo method^[74]; (b) GC-Net: geometry and context for deep stereo regression^[75]

扰,因此后续的相关研究大多采用了深度学习技术。如图 22(b)所示, Kendall 等^[75]提出 GC-Net 使用三维卷积预测稠密视差。利用视差可以重构得到稠密三维点云,进而与目标三维 CAD 模型进行配准,得到目标三维位姿参数。配准过程通常分为全局配准(粗配准)和局部配准(细配准)两个步骤。全局配准是指待检测点云与目标点云相差较大时,将待检测点云进行粗略变换使其和目标点云大致地匹配上,在全局配准后再使用局部配准对待检测点云进行调整优化。通常使用特征匹配来完成全局配准,其中常用于提取点云特征的特征算子有 PFH^[76]、基于其改进的 FPFH^[77],以及 SHOT^[78]等。在待检测点云和目标点云上提取显著全局点云特征,寻找待检测点云与目标点云之间的对应关系进行全局配准。全局配准后的待检测点云与目标点云之间的位姿误差缩小,还需再通过局部配准经典的迭代最近点(ICP)^[79]等方法,对

全局配准后的待检测点云进行迭代优化,以获得更优位姿参数。

表 5 比对了几种不同的基于点特征的位姿视觉测量方法的精度、速度、适用性。基于三维重构的位姿视觉测量方法能够通过整合与提取图像中信息,恢复当前场景的深度信息,进行三维重构,进而求解目标的位姿参数。经典的三维重构方法受环境因素影响较大,无论是关键特征的提取还是图像间视差的计算,都难以应对弱纹理、光线暗弱等因素导致的图像质量较差的问题。因此,当前的研究重点在于利用深度学习的方法提高三维重构过程中的匹配精度,从而获得更加准确的位姿信息。而在引入深度学习之后,确实有效提高了三维重构的位姿测量精度,但是该方法的计算量显著增大,并且需要大量的三维模型来进行训练,并且由于三维模型的维度较高,训练难度显著提升,这也是该方法当前的研究重点。

表 5 基于三维重构的位姿视觉测量方法比对

Table 5 Comparison on object pose measurement methods based on 3D reconstruction

Method	Year	Accuracy	Speed	Feasibility
SGM <i>et al.</i> ^[74]	2007	Common	Good	Poor application effect in weak texture, transparent, reflective objects scenes
FPFH <i>et al.</i> ^[77]	2009	Better	Better	The reduction of complexity compared to PFH ensures the real-time performance
SHOT <i>et al.</i> ^[78]	2014	Good	Good	Better robustness, efficiency and description for point clouds with noise, clutter and uneven density
GC-Net <i>et al.</i> ^[75]	2017	Better	Better	More suitable for multi-target tracking in occluded scenes

4 多源数据融合的位姿视觉测量方法

从信息获取的传感器层面而言,实现位姿测量的五大典型技术途径包括惯导、卫星、雷达、星敏和视觉。从获取的图像维度层面而言,主要来源包括 RGB 相机和 RGB-D 相机,其中 RGB-D 相机同时可提供深度信息。不同的技术途径、不同的相机具有不同的优势,对不同的信息敏感度不同,但同样也各自存在原理和使

用条件上的一些不足。多源数据融合是通过不同传感器获取的信息(如惯导+视觉),结合不同维度的特征(如二维图像信息+深度信息),通过特定的技术手段实现信息的融合,从而实现不同来源信息之间的优势互补,能够有效地提升位姿测量的精度和鲁棒性,使得位姿测量系统更加地稳定。多源数据融合位姿测量方法分类如表 6 所示。

表 6 多源数据融合位姿测量方法分类

Table 6 Multi-source data fusion pose measurement method

Classification	Combination	Characteristic	Application scenario
	Visual system + inertial navigation system	Inertial navigation system has low cost and good stability, but there are accumulated errors. Visual measurement has good autonomy and high precision in close range	Flight process of aircraft, automatic pilot and so on
Multi-sensor data fusion	Visual system + LiDAR	LiDAR is used to acquire depth information, and the visual system is used to acquire other information in the scene	SLAM
	Visual system + other sensors	GPS is used for positioning, and UWB is used to obtain position information, etc., which can assist the visual system to obtain more accurate position and pose	Flight process of aircraft, human-computer interaction and so on
Multi-dimensional feature fusion	2D information + deep information	Common features (points, lines, regions, etc.) can be acquired in two-dimensional images, and depth information can make up for the deficiency of two-dimensional information and obtain more abundant object features	Posture measurement of static objects

4.1 “惯导+视觉”的多源数据融合位姿测量方法

在多种传感器输入层面的多源融合位姿视觉测量中,“惯导+视觉”的组合方式使用频率最高,在同步定位与建图(SLAM)、飞行器位姿测量、人体位姿测量等领域,众多学者和机构采用多传感器输入的多源数据融合方式来提升位姿测量的精度。惯导成本低、稳定性好,但存在累积误差。视觉测量自主性好,无累积误差,近场精度高,采用“惯导+视觉”的方式能够实现两种传感器的优势互补,有效提升测量精度。

耿明志^[80]基于视觉/惯导组合,利用视觉系统对图像特征进行提取,并借助惯导系统提供的姿态初值,解算出无人机(UAV)相对于跑道的姿态。张洲宇^[81]基于惯导/视觉融合完成了舰机相对位姿的估计,利用惯导系统得到姿态的旋转矩阵初值,然后通过视觉测量中的正交迭代算法,对姿态参数进行修正和优化,最终得到舰机的相对位姿,通过仿真结果验证融合算法能够有效地降低单一算法的误差。Mourikis等^[82]提出了实现精确航天器着陆的视觉辅助

惯导算法。利用视觉系统获取下降轨迹中和地面之间的关系,并利用扩展卡尔曼滤波来将视觉系统中获取的特征和惯导系统数据结合,结合资源自适应和实时能力的方式,准确地计算出航天器着陆过程中相对于地面的位姿和速度。李建军等^[83]采用视觉/惯导的方式,在火星着陆段的下降段利用视觉系统对着陆区域进行成像和特征提取,然后与轨道段获得的图像进行匹配,得到着陆器相对于着陆点的位姿信息,然后基于误差卡尔曼模型,将视觉系统的数据和惯导数据结合,最终得到更加精准的位姿信息,如图 23 所示。余新中^[84]通过扩展卡尔曼滤波算法将视觉信息和惯性测量单元信息相融合用于实现人体头部姿态估计,并应用于特种车辆辅助驾驶随动系统中。王妮^[85]基于“视觉+惯导+雷达高度计”来实现全球卫星导航系统(GNSS)拒止条件下的无人机姿态估计。利用雷达高度计提供高度信息,视觉系统测得的位姿信息来修正惯导输出的位姿信息,利用卡尔曼滤波的方式实现数据的融合。

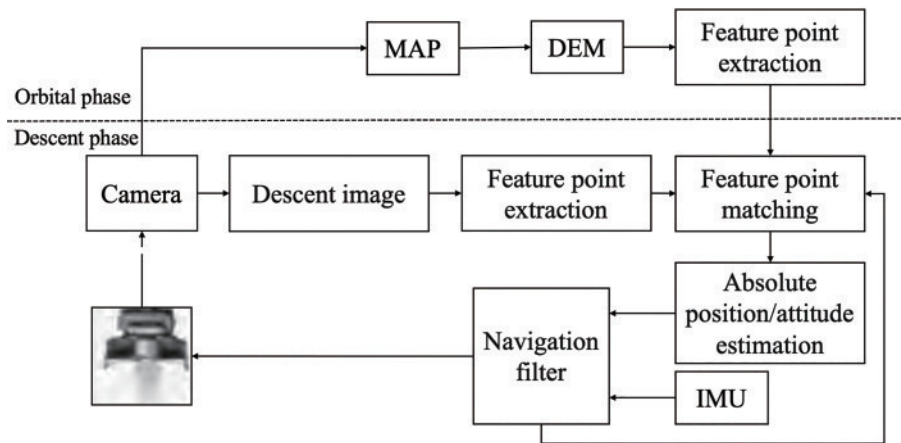


图 23 基于视觉/惯导的多源融合位姿测量算法^[83]

Fig. 23 Multi-source fusion pose measurement algorithm based on vision/IMU^[83]

4.2 “其他测量方式+视觉”的位姿测量方法

惯导+视觉的方法虽然较为常用,但是仅将这两类数据进行多源数据融合的测量方式,仍然无法有效适用大范围、远距离、多变天气等的复杂场景高精度位姿视觉测量的需求。因此,雷达和卫星等也常被用于和惯导、视觉进行结合,从而实现更多传感源数据的融合以有效增强目标位姿参数测量的精度和鲁棒性。

在广泛研究的 SLAM 领域中,通常采用激光雷达来获取场景实时深度信息并与视觉姿态测量相融合,构建的“视觉+雷达”系统能够输出更为精确的场景即时位姿信息。Zuo 等^[86]采用激光雷达/惯导/视觉相结合的方式来提升视觉里程计输出的位姿信息的精度,采用滑动窗平面特征跟踪的方式来处理 3D 激光雷达信息,然后将激光雷达信息、惯导信息和视觉系统获得的稀疏视觉特征信息采用紧耦合的操作,最终输出精度高、鲁棒性强的姿态信息。

如图 24 所示,辛琪等^[87]利用加速度计/磁强计/视觉/GPS 多路传感信息,通过低维混合扩展卡尔曼滤波对多路传感数据建立起状态估计,然后利用马尔可夫估计对多路信息进行结合,最终实现了比单路测量更为精确的无人机飞行姿态估计。

张胜凯^[88]将视觉、超宽带(UWB)信号和惯导信号融合在一个系统中,利用 UWB 的相对测距和角度来获取更加良好的视觉信息,同时还增加惯导来获取一些丢失的视觉里程,最终实现了弱纹理场景下微型无人机姿态的估计。Cheng 等^[89]提出了一种基于惯导/GPS/视觉/超宽带的多传感器融合的无人机着陆位姿估计、定位算法,算法框架如图 25 所示。超宽带信息用于弥补 GPS 和视觉传感器在提供位置信息方面的不足。在位姿测量前建立起舰机相对的坐标系关系,然后收集每个传感器得到的数据,通过坐标传递转换到同一坐标系,利用基于马氏距离的卡方检验来验证

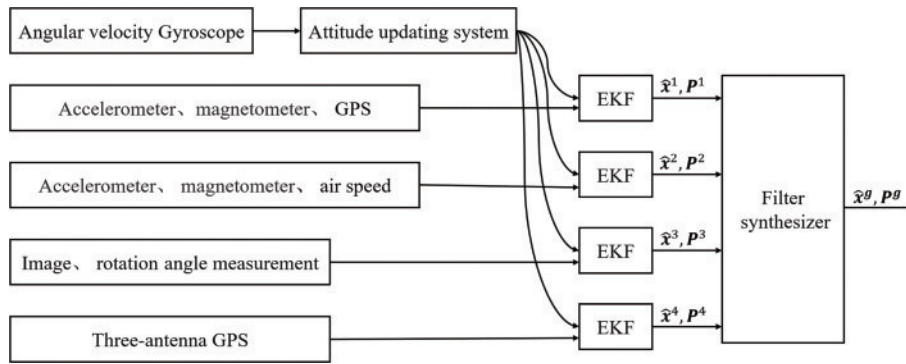


图 24 基于 GPS/视觉的多源融合位姿测量算法^[87]

Fig. 24 Multi-source fusion pose measurement algorithm based on GPS/vision^[87]

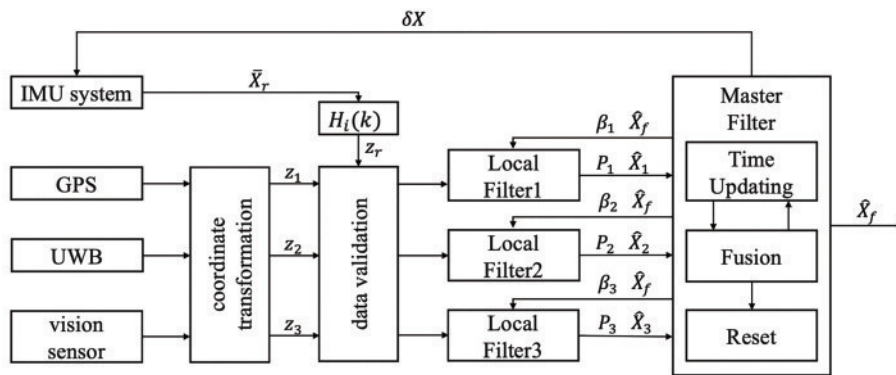


图 25 基于惯导/GPS/视觉/超宽带的多源融合位姿测量算法^[89]

Fig. 25 Multi-source fusion pose measurement algorithm based on IMU/GPS/vision/UWB^[89]

数据的有效性,然后基于联合卡尔曼滤波进行数据融合,最终输出舰机相对位姿等信息。

4.3 基于多维度特征的多源数据融合的位姿视觉测量方法

在图像的特征维度层面,目前常用于位姿测量的图像包括 RGB 图像和 RGB-D 图像。较 RGB 图像而言,RGB-D 图像可以输出深度信息。深度摄像机虽然比一般的摄像机更为昂贵,但在某些场景下,深度信息的加入能够为位姿测量精度贡献很大的力量。He 等^[90]将 RGB 图像中的外观信息和深度图像的几何信息作为互补数据,通过构建双向融合模块,提取

得到融合点的特征,用于后续的位姿估计求解,如图 26 所示。He 等^[91]基于 Hough 投票的思想,结合关键点的二维信息和深度信息,利用投票的方式选出三维关键点,然后利用最小二乘拟合得到位姿信息。Wang 等^[92]构造了一个端到端的模型,对 RGB-D 图像中的图像信息和深度信息进行融合。将图像信息通过 CNN 提取颜色嵌入,将深度信息通过 PointNet 获得几何嵌入,将两者融合获得稠密特征再进行位姿估计。

基于多源数据的位姿视觉测量方法近年来取得了较大的发展,但由于需对来源不同的特征进行提取、匹

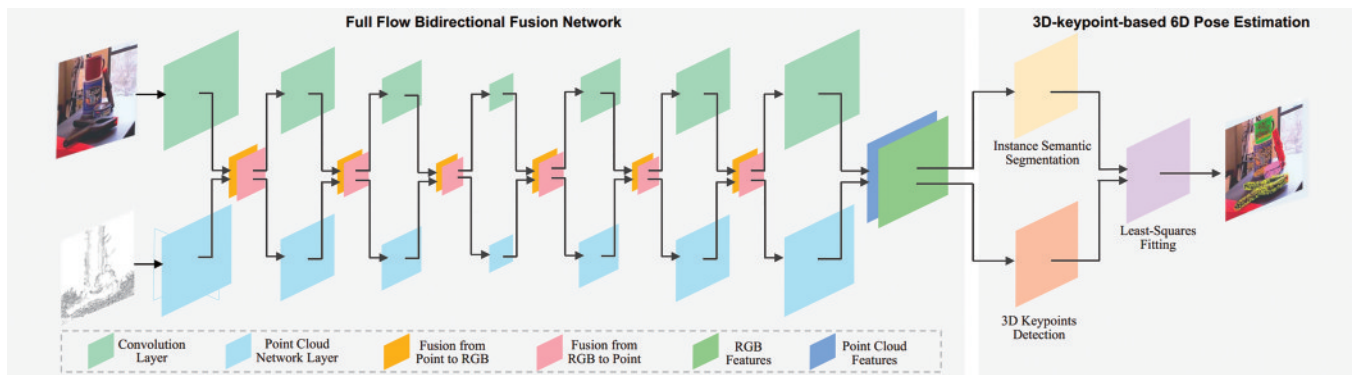


图 26 采用 RGB-D 图像作为输入的 FFB6D 算法框架^[90]

Fig. 26 FFB6D algorithm framework using RGB-D image as input^[90]

配和融合,耗时较长,且精度与融合方法密切相关。因此,如何实现多源数据时间同步、如何有效地实现数据选取和融合、如何提升模型的处理效率都值得进一步研究。

5 位姿视觉测量典型应用领域

5.1 飞行器飞行位姿视觉测量应用

位姿信息是反映飞行器当前状态的基础信息,对飞行器在飞行过程中进行准确的位姿测量在飞行器测试过程中具有重要的意义。随着视觉传感器和计算机视觉技术的迅速发展,基于视觉的位姿估计方法在飞行器飞行位姿测试中得到了较好的应用。

针对飞机着陆过程中的位姿测量研究,早在 1998 年 Chatterji 等^[93]就提出了机载位姿测量的方案,在飞机着陆过程中使用了视觉位姿测量的方法。该方法融合机场灯光模型的信息、机载摄像机获取的视频图像信息以及滚转传感器获得的侧倾姿态角信息,用于在夜间着陆期间确定飞机相对于跑道的位姿、偏航和俯仰方向,并通过仿真结果表明了所提方法的可行性。Oszust 等^[94]采用类似的方法进行模拟实验,从机载图像采集装置所获得的机场图像中提取 PAPI 系统和 ALSF-2 系统的指示灯,并构建如图 27 所示的坐标系确定飞机着陆过程中的位姿状态。

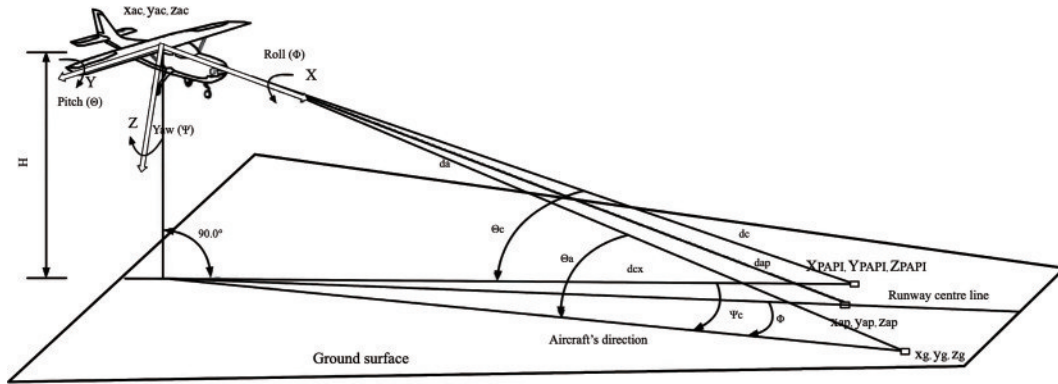


图 27 模拟实验中坐标系设定^[94]

Fig. 27 Coordinate system setting in simulation experiment^[94]

Gibert 等^[95-96]针对民用飞机着陆过程提出了相对姿态估计方法,并将所提出的方法利用空客模拟器进行了实验验证。所提出的方法主要从视觉传感器所获取的图像中提取的特征来简化视觉模型,并且假设跑道知识未知,用飞机的惯导信息作为辅助实现几何重构,最终解决 3D 位姿估计问题,该方法设定的飞机着陆过程模型如图 28 所示。Meng 等^[97]设计了一套视觉、雷达、惯导相结合的飞机着陆引导系统,并通过仿真实验的方式进行验证,实验场景如图 29(a)所示,现场实验布置如图 29(b)所示。该方法通过视觉系统解

算船上的地标坐标,通过管道系统获得飞机的运动信息,通过雷达系统获取飞机与船只的相对位置。通过所获得的飞机和船只的相对运动信息获得来引导降落指令。空客公司在其自主滑行及起降(ATTOL)项目中,利用机载图像识别技术,通过基于视觉的全自动飞行测试实现了民用飞机的自主滑行、起飞和降落^[98]。

在飞行器飞行位姿测量应用中,非机载的位姿视觉测量系统由于不受机载功耗、体积、重量、空间等限制,具有显著优势。此类系统一般固定布置在地面机场、舰船甲板等参考平台上,允许部署更充足的算力与更复杂的方法。一般情况下非机载位姿视觉测量系统依靠所获取的飞机图像特征与飞机三维 CAD 模型特征之间的 2D-3D 匹配关系测量飞机位姿参数。马德里理工大学使用带合作标记的三目视觉系统实现无人直升机位姿参数测量^[99]。如图 30 所示,通过在无人机底座安装四个颜色不同的球形标记辅助相机识别显著特征,进而实现位姿测量。实验结果表明,在 3 m 范围内,定位精度优于 10 cm,未定量说明姿态精度。国防科技大学 Ma 等^[100]在无人机自主着陆的测试中,设计了一种仿人视觉系统的地面立体视觉制导系统。如图 31 所示,两个安装在云台上的相机组成双目视觉系统,跟踪拍摄飞机的降落。飞机被视为单个点,通过三角测量来获得其位置。但在简化过程中丢失了姿态信息。

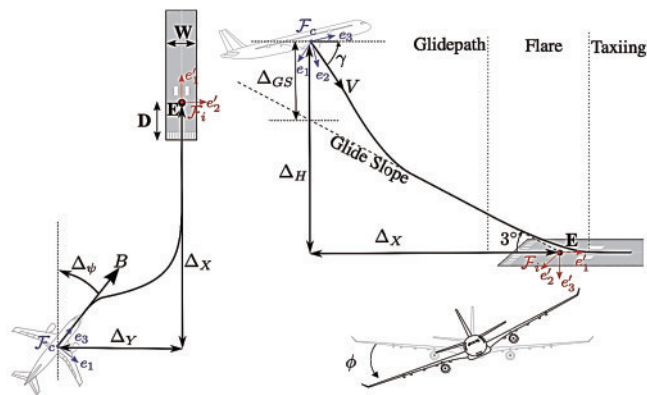


图 28 飞机着陆过程模型^[95]

Fig. 28 Model of aircraft landing process^[95]

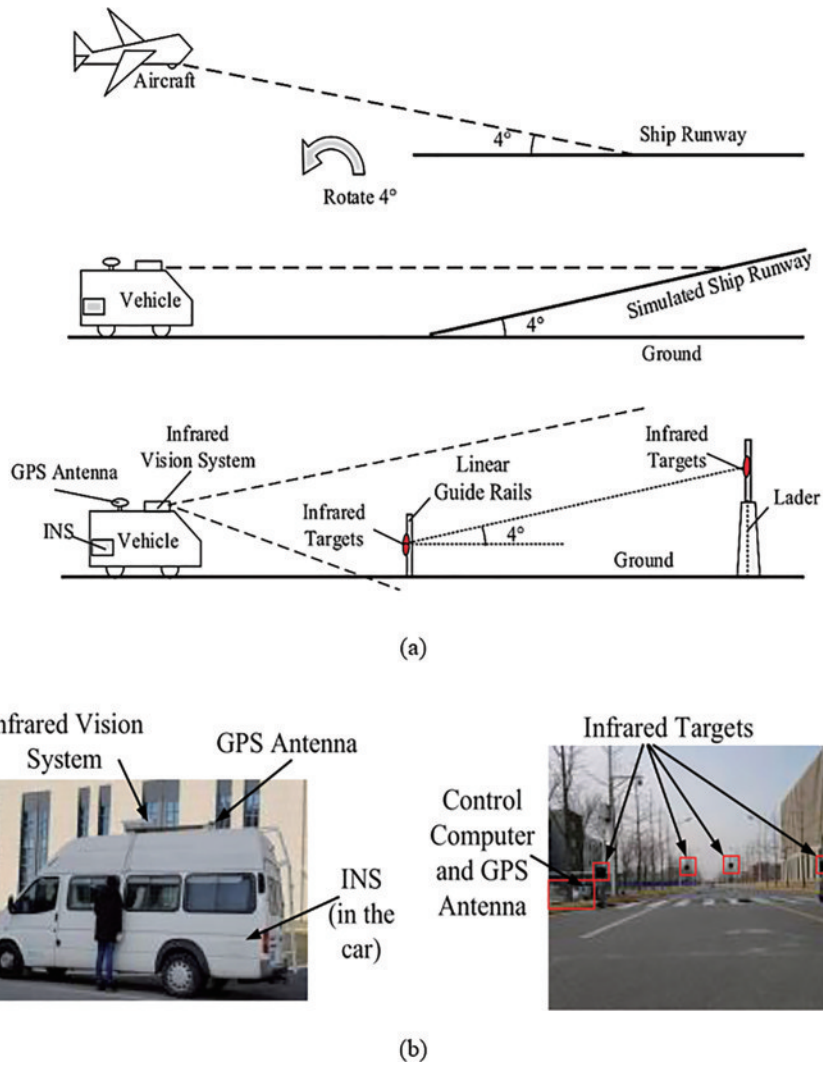


图 29 飞机着陆引导系统实验^[97] (a) 实验场景示意图; (b) 现场实验布置

Fig. 29 Experiment of aircraft landing guidance system^[97]. (a) Schematic diagram of the experiment; (b) experimental setup

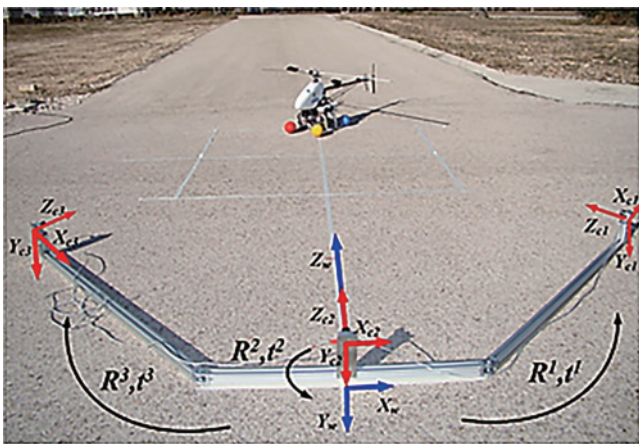


图 30 马德里理工大学三目视觉位姿测量系统^[99]

Fig. 30 Trinocular visual system of Madrid University of Technology^[99]

目前已有实际应用于舰船上的飞机飞行位姿视觉测量的非机载测量系统,这些系统多集中于对飞机位置参数的测量。法国 DNCS 公司开发的一种舰载无

人机自动着舰系统 D2AD,于 2008 年启动,致力于降低自动起降系统风险,且能够保证垂直起降无人机在舰上的安全使用,不依赖任何卫星定位系统就能够实现无人机的自动着舰引导。葡萄牙海军研究中心研究组利用机载和舰载结合的手段实现固定翼飞机在快速巡逻舰上的网式回收^[101],如图 32 所示,其中舰载设备包含一台安装在舰船尾部的相机以及计算机处理系统,实现了固定翼飞机的跟踪和姿态测量。Santos 等^[102]在无人机自主着舰测试中设计了一套舰载视觉系统。该系统使用 RGB 相机获取无人机图像,利用部署在工作站中的位姿估计算法求取无人机位姿,控制无人机降落在规定区域。该系统使用基于三维 CAD 模型的视觉方法估计无人机在着舰过程中的位姿,基于目标轮廓和内外颜色分布之间的色差,通过最大化似然函数来估计飞机姿态。在我国重点飞机试飞测试任务中,北京航空航天大学的张广军院士团队研制了用于固定翼飞机相对位姿测量的单目非机载视觉系统并应用于实际试飞任务中^[58],具体测量系统模型如

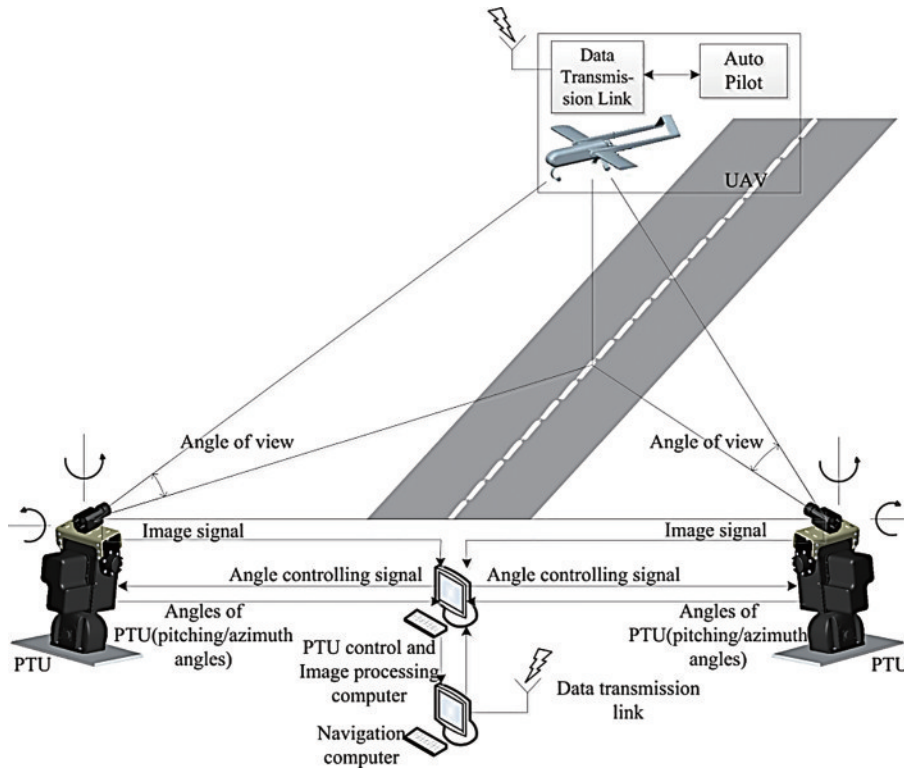


图 31 无人机自主着陆的测试中立体视觉制导系统^[100]

Fig. 31 Stereo vision guidance system in the test of UAV autonomous landing^[100]



图 32 葡萄牙海军舰基无人机视觉测量回收系统^[101]

Fig. 32 Visual measurement and recovery system of Portuguese Naval ship-based UAV^[101]

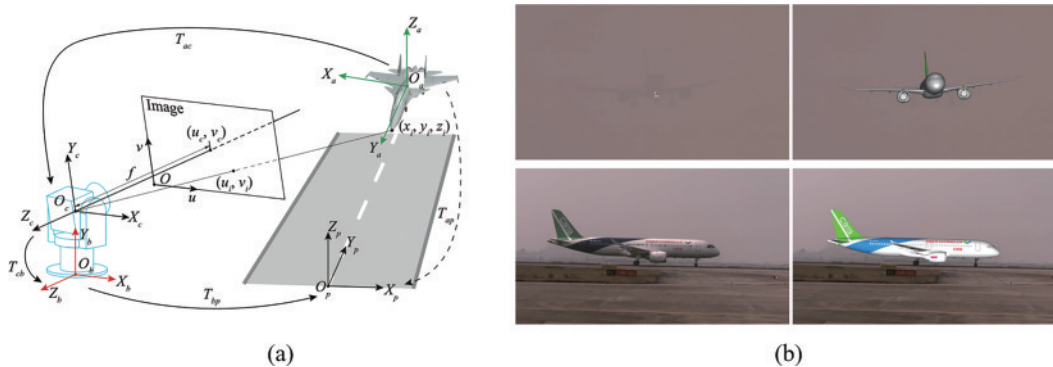


图 33 飞机试飞测试任务中的单目位姿视觉测量系统^[68]。(a)单目位姿测量系统模型；(b)实时位姿测量结果

Fig. 33 Monocular pose measurement system applied to aircraft test task^[68]. (a) Monocular pose measurement system model;

(b) real-time pose results

图 33(a)所示,采用可见光长变焦成像系统,远距离大范围内对飞机进行实时跟踪成像,并采用基于模型的位姿视觉测量方法,以 30 Hz 数据率实时测量飞机相

对于参考平台的相对位姿参数,如图 33(b)所示。

位姿视觉测量技术及其系统不仅在航空飞行器领域有着广泛的应用,在航天飞行器领域同样发挥了很

重要的作用,在各国的航天任务中受到青睐。欧洲多国联合开发的 SMART-OLEV 系统^[103]旨在帮助航天器恢复功能,延长地球静止卫星的寿命。SMART-OLEV 航天器上搭载有立体视觉系统,用于捕获用户航天器的图像,然后在地面工作站中进行处理、计算,能够得到 SMART-OLEV 航天器与用户航天器的相对位姿,以便后续使用,如图 34 所示。

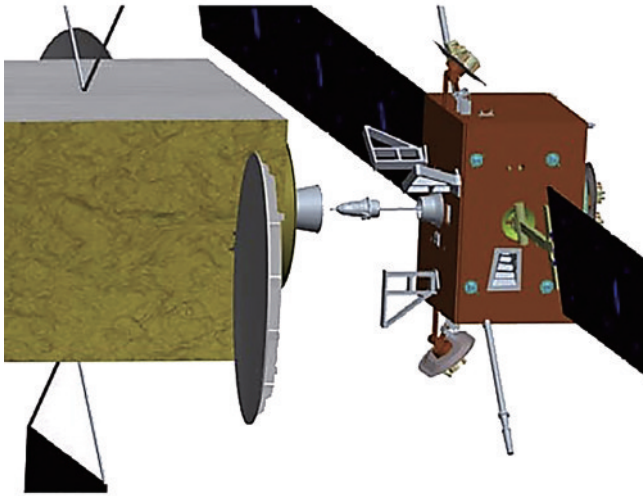


图 34 正在和客户卫星对接的 SMART-OLEV^[103]

Fig. 34 SMART-OLEV docking with the client satellite^[103]

5.2 武器靶试与投放位姿视觉测量应用

在试验场武器靶试测量应用中,对火炮发射、机载武器投放、弹头侵彻靶标过程等进行记录,对待测试样的各阶段位姿进行判读测量,能够获取武器装备发射过程中的直观数据,对武器的设计、试验和改良提供重要数据支撑。

在火炮类武器研制过程中,火炮发射时炮口振动的测试是提升大口径火炮密集度的重要参考。然而,由于发射时炮口处于极端高温、气流等严酷环境,因此对炮口振动进行准确测量十分困难。针对这一难题,李世立等^[104]利用视觉传感器非接触测量的特性,有效克服了测量环境的限制。如图 35 所示,通过高速摄像机拍摄弹丸出炮口时刻炮口合作标记点的图像,进而提取标记点亚像素高精度图像特征,对炮口振动位移、速度、加速度等参数进行准确测量。该系统能够对 0.2 mm 位移量级的炮口振动进行测量。

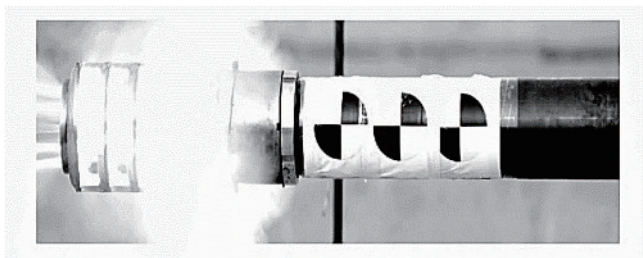


图 35 炮口振动视觉测量^[104]

Fig. 35 Visual measurement of muzzle vibration^[104]

机载导弹、炸弹等武器是新一代战斗机应对超视距作战需求的必然选择,而机载武器投放是决定其能否发挥作用的关键过程之一。在武器投放系统设计和验证中,机弹分离相容性研究需要对投放后武器位姿进行测量记录。宋威等^[105]搭建了风洞投放试验,如图 36 所示,利用高速摄像机拍摄不同气流速度、载机攻角等因素下弹机分离过程序列图像,通过视觉算法测量弹体垂直方向运动轨迹和俯仰姿态角,为投放系统设计提供重要参考。

各类武器撞击目标时着靶姿态和侵彻过程是指导武器设计,分析其破坏力等重要参考。由于武器撞击过程中的产生高温高压、强烈气流冲击等干扰因素,导致弹载传感器难以应对测量需求。相比之下,视觉测量方法的非接触测量特征使其很好的胜任这一任务。吴榕榕等^[106]利用可见光高速摄像机记录破冰弹侵彻冰靶体过程中的飞行位姿,如图 37(a)所示。同时利用红外传感器记录弹体在冰靶体内的运动位姿和温度变化,如图 37(b)所示。

除了可以研究武器的威力外,撞击试验还能为安全防护提供重要参考。刘晶波等^[107]利用火箭撬试验了大型客机撞击核设施的过程,如图 38 所示。在试验过程中,高速摄像机拍摄下飞机模型与混凝土安全壳模型撞击的过程,通过视觉算法分析出模型撞击过程中位姿参数,为试验提供了重要数据参考。

5.3 自动化装配对接位姿视觉测量应用

现代工业生产过程中传统的人工装配对接技术在装配过程中需要不断试错调整,导致加工生产效率相对较低,难以保证装配的精度和产品性能的一致性。自动化装配对接已成为提高现代工业效率和产品质量不可或缺的关键环节。单纯由机器人按照规划路径进行的自动化装配,往往缺乏主动的认知感应能力,在外部因素的干扰下,难以保证装配正常进行,智能化程度较低。随着位姿视觉测量技术的发展,为自动化装配对接提供了新思路和新手段,促进了工业生产制造高质量发展。

在航空航天领域,位姿视觉测量技术已初步满足了重要部件的自动化装配需求,尤其在航天器对接方面应用较为成熟。图 39 为空客总装线中的飞机装配现场, Marguet 等^[108]融合激光传感器和视觉传感器搭建了辅助装配系统,为空客总装线中飞机机身、机翼等大型部件间的装配进行引导。图 40 为安装在“发现号”航天飞机上的 TriDAR (Triangulation LIDAR) 系统^[109]。加拿大 NEPTEC 公司所研发的 LCS (Laser Camera System) 系统^[110]和 TriDAR 系统融合激光测距和立体视觉测量获取目标的三维点云数据,并通过点云配准方法解算相对位姿信息,目前已多次完成“发现号”航天飞机与国际空间站的交会对接任务。为实现航天器舱段自动化对接,陈冠宇等^[111]研制了基于激光轮廓传感器和 CCD 图像传感器的舱段自动对接装置,

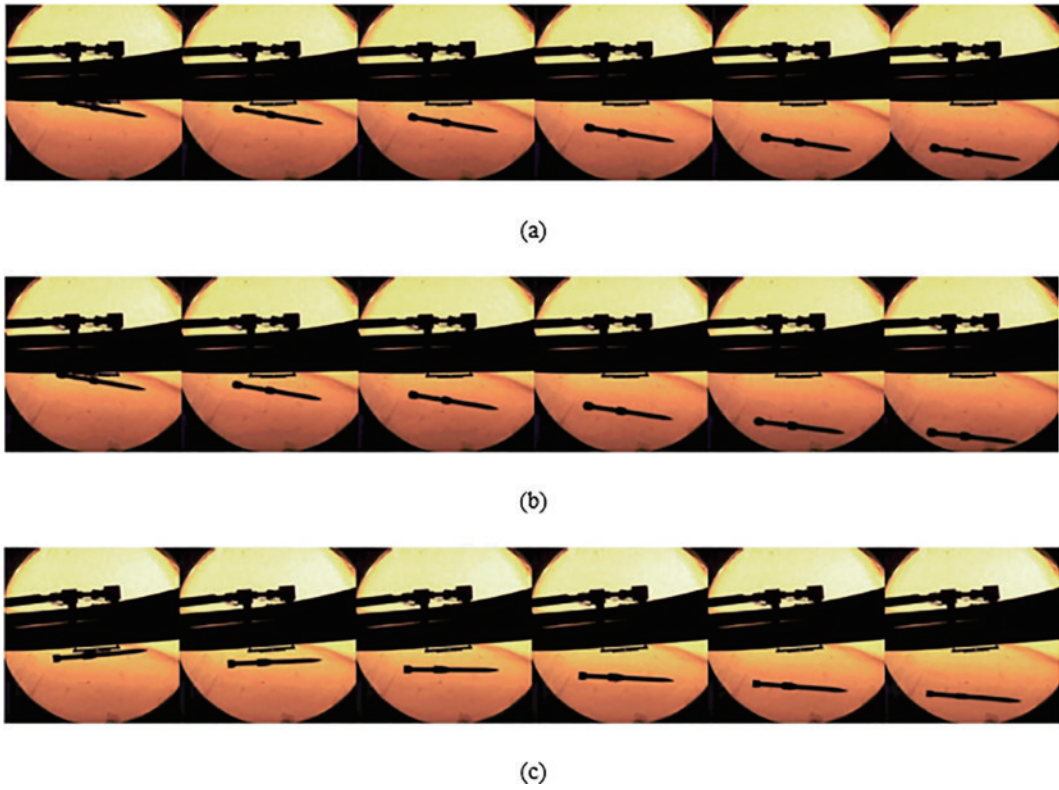


图 36 不同载机攻角下内埋导弹分离图像序列^[105]。(a)攻角为 0°; (b)攻角为 2°; (c)攻角为 3°
 Fig. 36 Image sequence of embedded missile separation under different attack angles of carrier aircraft^[105]. (a) Attack angles is 0°; (b) attack angle is 2°; (c) attack angle is 3°

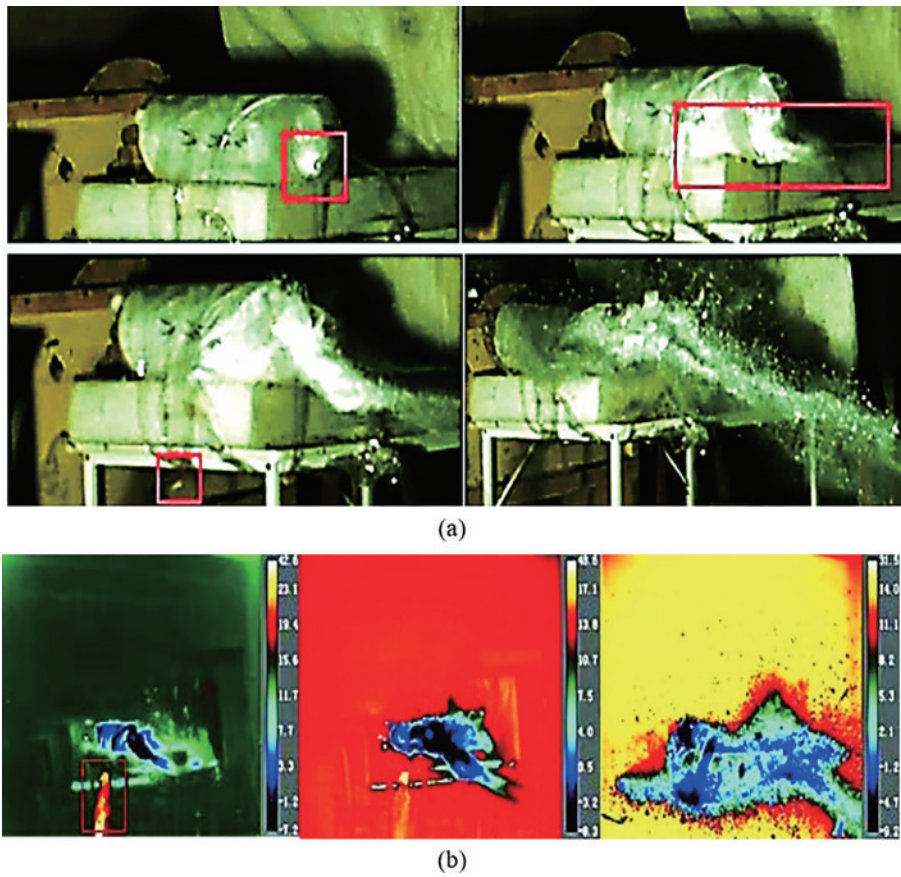


图 37 弹体侵彻冰靶标试验^[106]。(a)可见光高速摄像机拍摄图像; (b)红外高速摄像机拍摄图像
 Fig. 37 Test of projectile penetrating ice target^[106]. (a) Imaging with normal high speed camera; (b) imaging with high speed infrared camera

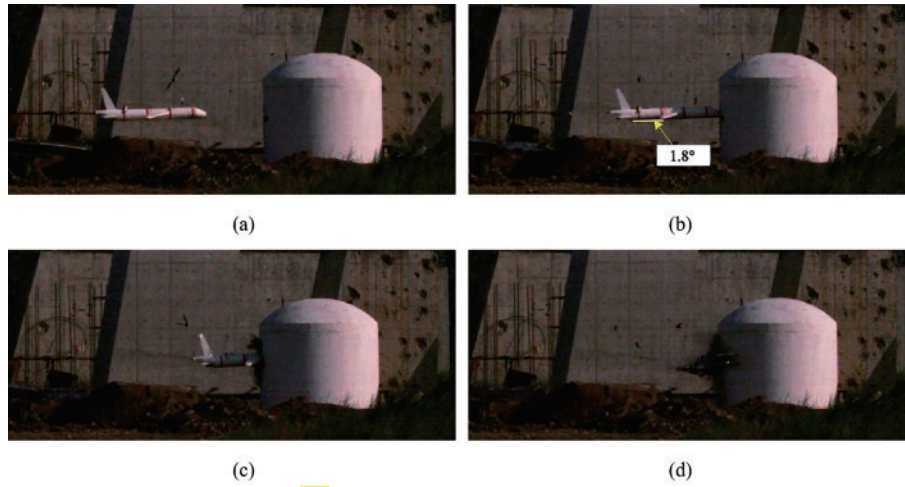


图 38 着靶高速摄像与位姿判读^[107]。(a)着靶前;(b)着靶后 0 s;(c)着靶后 0.015 s;(d)着靶后 0.0875 s
Fig. 38 High speed camera and position and attitude interpretation^[107]. (a) Before hitting the target; (b) the hitting moment; (c) 0.015 s after; (d) 0.0875 s after



图 39 空客总装线中的飞机舱段装配^[108]
Fig. 39 Aircraft cabin assembly in airbus assembly line^[108]

如图 41 所示。该装置采用单目位姿视觉测量技术获取舱段对接孔位置和角度偏差,位姿调整精度和效率可满足航天器舱体的自动化对接需求。针对狭小空间内轴孔装配位姿难以测量的难题,王稼祥等^[112]结合双目位姿测量技术,研制出高精度航空发动机低压涡轮轴自动化装配系统,如图 42 所示。该系统实现了涡轮轴装配的轴孔自动调姿对接、安装止口自动加热测温、螺母自动拧紧等自动化装配。

装配微型元件的精准装配同样可采用视觉位姿测量方法进行辅助,Liu 等^[113]研制高精度微观视觉自动装配系统,如图 43 所示。系统通过正交布置的显微成像装置对所需对接的毫米级尺寸组件进行成像,并根据组件的关键点和边缘特征测量两组件之间的相对位置和姿态,进而引导装配机构完成自动装配。Tamadazte 等^[114]分别使用 CCD 摄像机、光学显微镜和扫描电子显微镜对不同尺度微型组件进行成像,在微型元件 CAD 模型的辅助下使用基于边缘/轮廓特征的位姿迭代算法测量元件位姿,实现自动化精准装配。

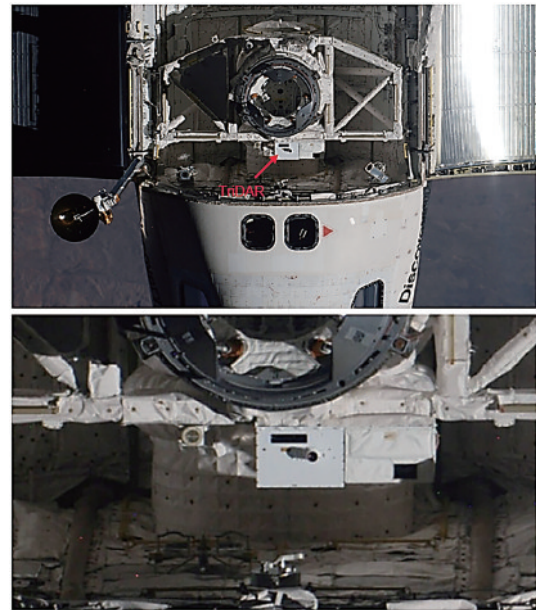


图 40 “发现号”航天飞机上的 TriDAR 系统^[109]
Fig. 40 TriDAR on board Space Shuttle Discovery^[109]

杜甫^[115]针对加速度计关键零部件的装配问题,采用显微视觉测量系统对微型零部件边缘和轮廓特征进行提取,实现其位姿测量。

5.4 人体运动姿态视觉估计应用

随着社会需求的变化和技术的发展,人体姿态估计作为计算机视觉研究领域的基础任务之一,受到了大量学者的关注,并在业界得到了广泛的应用。人体姿态估计中将人体的头、手肘、脚踝等关键部位定义为人体的关节点,从 2D 或 3D 图像中定位到人体的关节点,并预测得到关节点的二维空间位置坐标信息或三维空间位置坐标和角度等信息,从而计算出人体当前的姿态信息,最终应用于动作识别、人体跟踪、人机交互和增强/虚拟现实等领域。



图 41 舱段自动化对接平台^[111]

Fig. 41 Cabin automatic docking platform^[111]

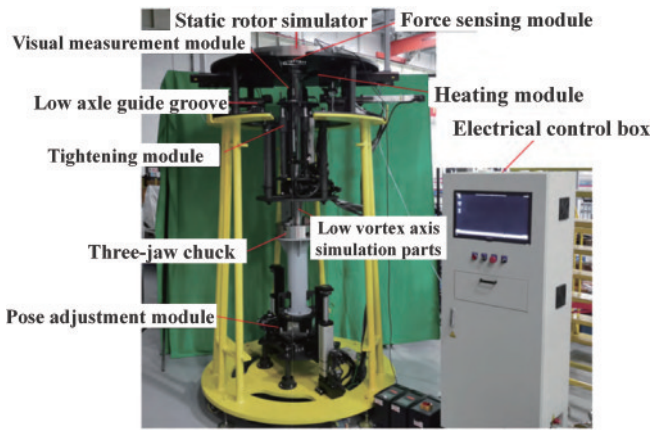


图 42 航空发动机低压涡轮轴自动化装配设备^[112]

Fig. 42 Automatic assembly equipment for low pressure turbine shaft of aero-engine^[112]

人体姿态估计最初常用图模型^[116]和全局特征^[117]的方式来解算人体姿态,DeepPose^[118]首批将深度学习引入人体姿态估计后,出现了许多优秀的基于深度学习的人体姿态估计算法,如 CPM^[119]、CPN^[120]、HigherHRNet^[121]、MSPN^[122]等,为人体姿态估计的应

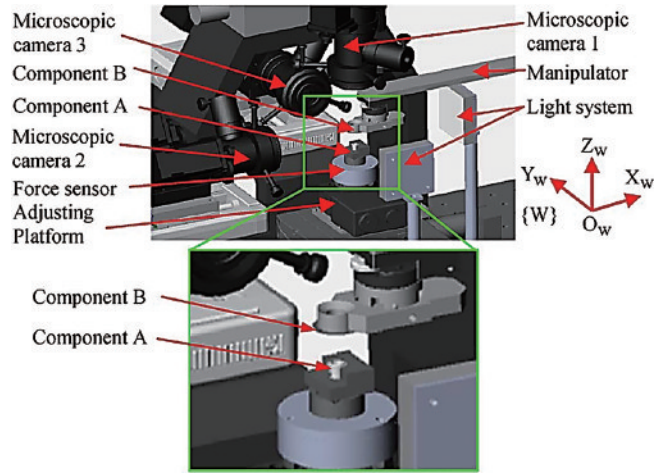


图 43 高精度微观视觉自动装配系统^[113]

Fig. 43 High precision micro-vision automatic assembly system^[113]

用提供了扎实的理论基础。

根据人体姿态估计应用整体任务的目的,将人体姿态估计的应用分为两类:将姿态估计作为整体任务中的一环,如人体动作识别、人体行为分析、人体跟踪等,将人体姿态估计作为关键环节,并与其他技术相结合实现任务;直接将人体姿态估计应用于各领域,如各种运动场景、交通领域、安防领域、医疗领域等。

在人体动作识别中,通过了解前后多幅图像的人体姿态,来识别人体的动作,是人体姿态估计的常用应用之一。Liu 等^[123]首先对图像中的人体姿态进行估计,然后将姿态的 heatmap 用于增强基于骨架的动作识别,如图 44 所示。Luvizon 等^[124]采用多任务框架从静止图像中实现 2D 和 3D 的姿态估计,然后从视频序列中完成动作识别。Yan 等^[125]引入时间姿态卷积,有效地编码多个姿态的模态,用于动作识别。

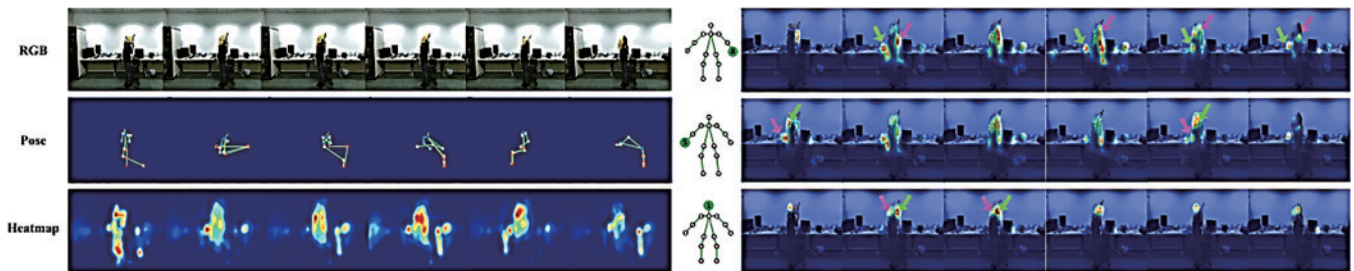


图 44 人体动作识别^[123]

Fig. 44 Human action recognition^[123]

在人体跟踪领域,同样也是以人体姿态估计为基础。Wang 等^[126]提出了一种改进 HRNet 的人体姿态估计和跟踪方法,如图 45 所示,采用剪辑跟踪网络,通过利用视频前后帧的关系来提升估计精度和抗遮挡能力。VoxelTrack^[127]基于多视图对人体进行姿态估计和跟踪,将多视图的信息合成三维体素表征,能够有效

地抗遮挡,提升人体姿态估计和跟踪的精确度。

在人体行为分析中,首先对人体的姿态进行估计,然后对姿态进行分类,实现人体行为的分析。唐永正^[128]提出的基于骨架的 FP-Net 实现了抗遮挡人体正面姿态估计,并将其与人体跟踪算法结合,应用于教室里学生的行为分析。



图 45 人体姿态跟踪^[126]

Fig. 45 Human pose tracking^[126]

在运动分析领域,Leeds Sports Pose (LSP) 数据集^[129]专门针对运动中的人体姿态,包含田径、羽毛球、棒球、体操等项目。在 2022 北京冬奥会中,英特尔 3DAT 运动员追踪系统能够对速度滑冰运动员的关键

部位进行姿态识别^[130],如图 46(a)所示,用于运动员训练的评估和指导调整;我国自研的花样滑冰 AI 辅助评分系统 1.0^[131],如图 46(b)所示,能够对花滑运动员进行姿态估计,用于辅助裁判评分。

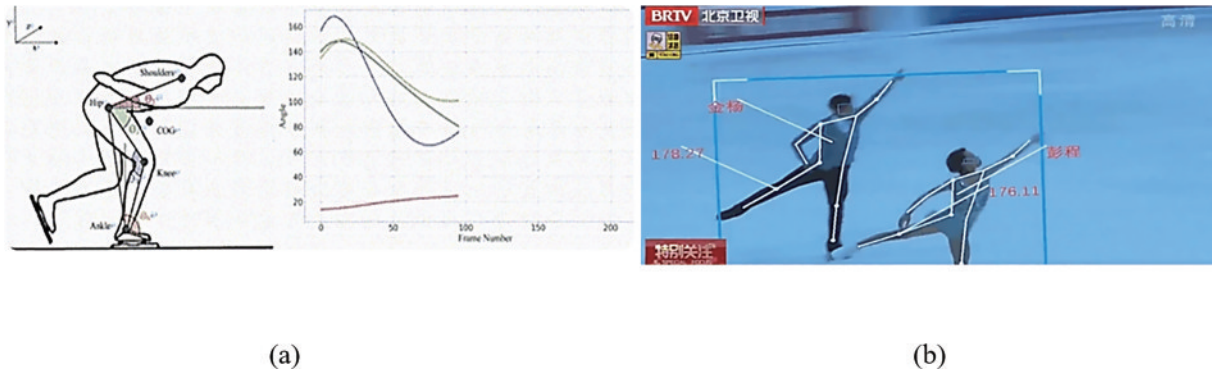


图 46 2022 北京冬奥会中的人体姿态估计。(a)3DAT 运动员追踪系统^[130];(b)运动员打分体系^[131]

Fig. 46 Human pose estimation in the 2022 Beijing Winter Olympics. (a) 3DAT athlete tracking system^[130]; (b) athlete scoring system^[131]

在交通领域,人体姿态估计对象包括驾驶员和行人。自动驾驶的安全性与环境、行人的姿态密切相关,夏盼^[132]针对驾驶场景,基于 DS-CNN 实现了人体头部和身体姿态的联合估计用于对行人姿态的估计,为保证自动驾驶的安全提供基础。Zhao 等^[133]基于人体姿态回归实现了精确的行人检测,采用关节索引特征和对称索引特征来提升检测器的可分辨性,并使姿态估计和行人检测有效协作,提升了行人检测的精确性。刘楠^[134]对 OpenPose 进行改进,降低模型计算

量,实现了快速的飞行器驾驶舱内驾驶员姿态估计,对驾驶员状态进行实时监测,如图 47 所示。

在医疗领域,Wu 等^[135]提出了一个新的 3D 人体姿态估计方法用于平躺姿势的医疗诊断和修复,首先利用当前姿态估计方法实现 2D 全身关键点的识别,然后结合深度信息和坐标变换得到 3D 姿态信息。Cao 等^[136]将多模态条件变分自编码器和 HRNet 相结合用于卧床人体姿态估计。Rohan 等^[137]利用人体姿态估计进行实时步态分析,以帮助医生更好地理解步态异常。

图 47 驾驶员姿态估计^[134]Fig. 47 Pilot pose estimation^[134]

6 结束语

视觉位姿测量技术具有全自主、可视化、抗电磁干扰、无累积误差、近场精度高等突出特点,不仅可以实现运动目标位置与姿态参数的同步动态实时精确测量,还可以结合测量场景视觉信息进行场景态势的有效感知,比如场景的突发异常和危险物介入等,实现具有智能感知能力的视觉位姿测量,这是惯导、卫星、雷达、星敏感器等位姿测量技术所不具备的。同时,在国防、军事等特殊需求领域,当考虑到战时拒止环境、无线电静默限制等极端条件适应性时,比如在航母远洋作战时的舰载机着舰位姿测量,卫星的使用必然受到干扰,而雷达也常常不能正常开机,而视觉位姿测量可进行全自主测量的特点,使得其与惯导的有机融合既可以发挥惯导测量的连续性优势,又可以利用近距离视觉测量精度高的优势,往往成为最佳位姿测量途径。未来,视觉位姿测量技术与惯导、卫星、雷达、星敏感器等位姿测量技术的全方位融合,将是满足飞机试飞、导弹靶试、卫星对地观测等不同领域多情形位姿测量需求的研究重点。

参 考 文 献

- [1] 毕道明, 黄辉, 范静, 等. 非线性移动路径跟踪及着舰控制应用[J]. 北京航空航天大学学报, 2021, 47(1): 45-55.
Bi D M, Huang H, Fan J, et al. Nonlinear moving path following control and its application to carrier landing[J]. Journal of Beijing University of Aeronautics and Astronautics, 2021, 47(1): 45-55.
- [2] 魏振忠. 舰载机着舰位姿视觉测量技术概述[J]. 测控技术, 2020, 39(8): 2-6.
Wei Z Z. Overview of visual measurement technology for landing position and attitude of carrier-based aircraft[J]. Measurement & Control Technology, 2020, 39(8): 2-6.
- [3] Li D F, Zhang Y C, Liu C, et al. Review of photoacoustic imaging for microrobots tracking *in vivo*[J]. Chinese Optics Letters, 2021, 19(11): 111701.
- [4] Zhou L Y, Huang X W, Fu Q, et al. Fine edge detection in single-pixel imaging[J]. Chinese Optics Letters, 2021, 19(12): 121101.
- [5] Gao Z R, Su Y, Zhang Q C. Single-event-camera-based 3D trajectory measurement method for high-speed moving targets[J]. Chinese Optics Letters, 2022, 20(6): 061101.
- [6] Harris C G, Stephens M. A J combined corner and edge detector[C]//Proceedings of the Alvey Vision Conference 1988, August 31-September 2, 1988, Manchester, UK. Sheffield: Alvey Vision Club, 1988: 147-151.
- [7] Lowe D G. Distinctive image features from scale-invariant keypoints[J]. International Journal of Computer Vision, 2004, 60(2): 91-110.
- [8] David P, DeMenthon D, Duraiswami R, et al. SoftPOSIT: simultaneous pose and correspondence determination[J]. International Journal of Computer Vision, 2004, 59(3): 259-284.
- [9] Ravela S, Draper B, Lim J, et al. Adaptive tracking and model registration across distinct aspects[C]//Proceedings 1995 IEEE/RSJ International Conference on Intelligent Robots and Systems, August 5-9, 1995, Pittsburgh, PA, USA. New York: IEEE Press, 2002: 174-180.
- [10] Bay H, Ess A, Tuytelaars T, et al. Speeded-up robust features (SURF) [J]. Computer Vision and Image Understanding, 2008, 110(3): 346-359.
- [11] Rublee E, Rabaud V, Konolige K, et al. ORB: an efficient alternative to SIFT or SURF[C]//2011 International Conference on Computer Vision, November 6-13, 2011, Barcelona, Spain. New York: IEEE Press, 2012: 2564-2571.
- [12] Haralick R M, Lee D, Ottenburg K, et al. Analysis and solutions of the three point perspective pose estimation problem[C]//Proceedings of 1991 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, June 3-6, 1991, Maui, HI, USA. New York: IEEE Press, 2002: 592-598.
- [13] Horaud R, Conio B, Le Boulleux O, et al. An analytic solution for the perspective 4-point problem[J]. Computer Vision, Graphics, and Image Processing, 1989, 47(1): 33-44.
- [14] Lepetit V, Moreno-Noguer F, Fua P. EPnP: an accurate $O(n)$ solution to the PnP problem[J]. International Journal

- of Computer Vision, 2009, 81(2): 155-166.
- [15] Li S Q, Xu C, Xie M. A robust $O(n)$ solution to the perspective-n-point problem[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2012, 34(7): 1444-1450.
- [16] Zheng Y Q, Kuang Y B, Sugimoto S, et al. Revisiting the PnP problem: a fast, general and optimal solution [C]//2013 IEEE International Conference on Computer Vision, December 1-8, 2013, Sydney, NSW, Australia. New York: IEEE Press, 2014: 2344-2351.
- [17] Fischler M A, Bolles R C. Random sample consensus[J]. Communications of the ACM, 1981, 24(6): 381-395.
- [18] Rad M, Lepetit V. BB8: a scalable, accurate, robust to partial occlusion method for predicting the 3D poses of challenging objects without using depth[C]//2017 IEEE International Conference on Computer Vision (ICCV), October 22-29, 2017, Venice, Italy. New York: IEEE Press, 2017: 3848-3856.
- [19] Pavlakos G, Zhou X W, Chan A, et al. 6-DoF object pose from semantic keypoints[C]//2017 IEEE International Conference on Robotics and Automation (ICRA), May 29-June 3, 2017, Singapore. New York: IEEE Press, 2017: 2011-2018.
- [20] Tekin B, Sinha S N, Fua P. Real-time seamless single shot 6D object pose prediction[C]//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, June 18-23, 2018, Salt Lake City, UT, USA. New York: IEEE Press, 2018: 292-301.
- [21] Fan R Z, Xu T B, Wei Z Z. Estimating 6D aircraft pose from keypoints and structures[J]. Remote Sensing, 2021, 13(4): 663.
- [22] Peng S D, Liu Y, Huang Q X, et al. PVNet: pixel-wise voting network for 6DoF pose estimation[C]//2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), June 15-20, 2019, Long Beach, CA, USA. New York: IEEE Press, 2020: 4556-4565.
- [23] Li Z G, Wang G, Ji X Y. CDPN: coordinates-based disentangled pose network for real-time RGB-based 6-DoF object pose estimation[C]//2019 IEEE/CVF International Conference on Computer Vision (ICCV), October 27-November 2, 2019, Seoul, Republic of Korea. New York: IEEE Press, 2020: 7677-7686.
- [24] Zakharov S, Shugurov I, Ilic S. DPOD: 6D pose object detector and refiner[C]//2019 IEEE/CVF International Conference on Computer Vision (ICCV), October 27-November 2, 2019, Seoul, Republic of Korea. New York: IEEE Press, 2020: 1941-1950.
- [25] Hu Y L, Fua P, Wang W, et al. Single-stage 6D object pose estimation[C]//2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), June 13-19, 2020, Seattle, WA, USA. New York: IEEE Press, 2020: 2927-2936.
- [26] Chen H S, Wang P C, Wang F, et al. EPro-PnP: generalized end-to-end probabilistic perspective-n-points for monocular object pose estimation[C]//2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), June 18-24, 2022, New Orleans, LA, USA. New York: IEEE Press, 2022: 2771-2780.
- [27] Wang G, Manhardt F, Tombari F, et al. GDR-net: geometry-guided direct regression network for monocular 6D object pose estimation[C]//2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), June 20-25, 2021, Nashville, TN, USA. New York: IEEE Press, 2021: 16606-16616.
- [28] Duda R O, Hart P E. Use of the Hough transformation to detect lines and curves in pictures[J]. Communications of the ACM, 1972, 15(1): 11-15.
- [29] von Gioi R G, Jakubowicz J, Morel J M, et al. LSD: a fast line segment detector with a false detection control [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2010, 32(4): 722-732.
- [30] Akinlar C, Topal C. EDLines: a real-time line segment detector with a false detection control[J]. Pattern Recognition Letters, 2011, 32(13): 1633-1642.
- [31] Xue N, Bai S, Wang F D, et al. Learning attraction field representation for robust line segment detection[C]//2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), June 15-20, 2019, Long Beach, CA, USA. New York: IEEE Press, 2020: 1595-1603.
- [32] Wang Z H, Wu F C, Hu Z Y. MSLD: a robust descriptor for line matching[J]. Pattern Recognition, 2009, 42(5): 941-953.
- [33] Zhang L L, Koch R. Line matching using appearance similarities and geometric constraints[M]//Pinz A, Pock T, Bischof H, et al. Joint DAGM (German Association for Pattern Recognition) and OAGM symposium. Lecture notes in computer science. Heidelberg: Springer, 2012, 7476: 236-245.
- [34] 王平, 何卫隆, 张爱华, 等. EPnL: 一种高效且精确的 PnL 问题求解算法[J]. 自动化学报, 2022, 48(10): 2600-2610.
- Wang P, He W L, Zhang A H, et al. EPnL: an efficient and accurate algorithm to the PnL problem[J]. Acta Automatica Sinica, 2022, 48(10): 2600-2610.
- [35] Dhome M, Richetin M, Lapreste J T, et al. Determination of the attitude of 3D objects from a single perspective view[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 1989, 11(12): 1265-1278.
- [36] Liu Y, Huang T S, Faugeras O D. Determination of camera location from 2-D to 3-D line and point correspondences[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 1990, 12(1): 28-37.
- [37] Kumar R, Hanson A R. Robust methods for estimating pose and a sensitivity analysis[J]. CVGIP: Image Understanding, 1994, 60(3): 313-342.
- [38] David P, DeMenthon D, Duraiswami R, et al. Simultaneous pose and correspondence determination using line features[C]//2003 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, June 18-20, 2003, Madison, WI, USA. New York: IEEE Press, 2003: 424-431.
- [39] Ansar A, Daniilidis K. Linear pose estimation from points or lines[J]. IEEE Transactions on Pattern Analysis

- and Machine Intelligence, 2003, 25(5): 578-589.
- [40] Xu C, Zhang L L, Cheng L, et al. Pose estimation from line correspondences: a complete analysis and a series of solutions[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2017, 39(6): 1209-1222.
- [41] Příbyl B, Zemčík P, Čadík M. Absolute pose estimation from line correspondences using direct linear transformation[J]. *Computer Vision and Image Understanding*, 2017, 161: 130-144.
- [42] Mirzaei F M, Roumeliotis S I. Globally optimal pose estimation from line correspondences[C]//2011 IEEE International Conference on Robotics and Automation, May 9-13, 2011, Shanghai, China. New York: IEEE Press, 2011: 5581-5588.
- [43] Zhang L L, Xu C, Lee K M, et al. Robust and efficient pose estimation from line correspondences[M]//Lee K M, Matsushita Y, Rehg J M, et al. *Computer vision-ACCV 2012. Lecture notes in computer science*. Heidelberg: Springer, 2013, 7726: 217-230.
- [44] Wang P, Xu G L, Cheng Y H, et al. Camera pose estimation from lines: a fast, robust and general method [J]. *Machine Vision and Applications*, 2019, 30(4): 603-614.
- [45] Wang B, Zhong F, Qin X Y. Robust edge-based 3D object tracking with direction-based pose validation[J]. *Multimedia Tools and Applications*, 2019, 78(9): 12307-12331.
- [46] Harris C, Stennett C. RAPID-a video rate object tracker [C]//*Proceedings of the British Machine Vision Conference*, September, 1990, Oxford. London: British Machine Vision Association, 1990: 1-6.
- [47] Vacchetti L, Lepetit V, Fua P. Combining edge and texture information for real-time accurate 3D camera tracking[C]//*Third IEEE and ACM International Symposium on Mixed and Augmented Reality*, November 5, 2004, Arlington, VA, USA. New York: IEEE Press, 2005: 48-56.
- [48] Yoon Y, Kosaka A, Kak A C. A new Kalman-filter-based framework for fast and accurate visual tracking of rigid objects[J]. *IEEE Transactions on Robotics*, 2008, 24(5): 1238-1251.
- [49] Teulière C, Marchand E, Eck L. Using multiple hypothesis in model-based tracking[C]//2010 IEEE International Conference on Robotics and Automation, May 3-7, 2010, Anchorage, AK, USA. New York: IEEE Press, 2010: 4559-4565.
- [50] Hodaň T, Baráth D, Matas J. EPOS: estimating 6D pose of objects with symmetries[C]//2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), June 13-19, 2020, Seattle, WA, USA. New York: IEEE Press, 2020: 11700-11709.
- [51] Pitteri G, Bugeau A, Ilic S, et al. 3D Object detection and pose estimation of unseen objects in color images with local surface embeddings[M]//Ishikawa H, Liu C L, Pajdla T, et al. *Computer Vision-ACCV 2020. Lecture notes in computer science*. Cham: Springer, 2021, 12622: 38-54.
- [52] Shi Y F, Huang J W, Xu X, et al. StablePose: learning 6D object poses from geometrically stable patches[C]//2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), June 20-25, 2021, Nashville, TN, USA. New York: IEEE Press, 2021: 15217-15226.
- [53] Schmalz C, Rosenhahn B, Brox T, et al. Region-based pose tracking[M]//Martí J, Benedi J M, Mendonça A M, et al. *Iberian conference on pattern recognition and image analysis. Lecture notes in computer science*. Heidelberg: Springer, 2007, 4478: 56-63.
- [54] Brox T, Rosenhahn B, Gall J, et al. Combined region and motion-based 3D tracking of rigid and articulated objects[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2010, 32(3): 402-415.
- [55] Bibby C, Reid I. Robust real-time visual tracking using pixel-wise posteriors[M]//Forsyth D, Torr P, Zisserman Z. *Computer vision-ECCV 2008. Heidelberg: Springer*, 2008, 5303: 831-844.
- [56] Prisacariu V A, Reid I D. PWP3D: real-time segmentation and tracking of 3D objects[J]. *International Journal of Computer Vision*, 2012, 98(3): 335-354.
- [57] Tjaden H, Schwanecke U, Schömer E. Real-time monocular pose estimation of 3D objects using temporally consistent local color histograms[C]//2017 IEEE International Conference on Computer Vision (ICCV), October 22-29, 2017, Venice, Italy. New York: IEEE Press, 2017: 124-132.
- [58] Liu F L, Wei Z Z, Zhang G J. An off-board vision system for relative attitude measurement of aircraft[J]. *IEEE Transactions on Industrial Electronics*, 2022, 69(4): 4225-4233.
- [59] Zhong L S, Zhang Y, Zhao H, et al. Seeing through the occluders: robust monocular 6-DOF object pose tracking via model-guided video object segmentation[J]. *IEEE Robotics and Automation Letters*, 2020, 5(4): 5159-5166.
- [60] Chen L C, Zhu Y K, Papandreou G, et al. Encoder-decoder with atrous separable convolution for semantic image segmentation[M]//Ferrari V, Hebert M, Sminchisescu C, et al. *Computer vision-ECCV 2018. Lecture notes in computer science*. Cham: Springer, 2018, 11211: 833-851.
- [61] Zhong L S, Zhang L. A robust monocular 3D object tracking method combining statistical and photometric constraints[J]. *International Journal of Computer Vision*, 2019, 127(8): 973-992.
- [62] Zhong L S, Zhao X L, Zhang Y, et al. Occlusion-aware region-based 3D pose tracking of objects with temporally consistent polar-based local partitioning[J]. *IEEE Transactions on Image Processing*, 2020, 29: 5065-5078.
- [63] Choi C, Christensen H I. Real-time 3D model-based tracking using edge and keypoint features for robotic manipulation[C]//2010 IEEE International Conference on Robotics and Automation, May 3-7, 2010, Anchorage, AK, USA. New York: IEEE Press, 2010: 4048-4055.

- [64] Choi C, Christensen H I. Robust 3D visual tracking using particle filtering on the SE(3) group[C]//2011 IEEE International Conference on Robotics and Automation, May 9-13, 2011, Shanghai, China. New York: IEEE Press, 2011: 4384-4390.
- [65] Pauwels K, Rubio L, Díaz J, et al. Real-time model-based rigid object pose estimation and tracking combining dense and sparse visual cues[C]//2013 IEEE Conference on Computer Vision and Pattern Recognition, June 23-28, 2013, Portland, OR, USA. New York: IEEE Press, 2013: 2347-2354.
- [66] Tuzel O, Liu M Y, Taguchi Y, et al. Learning to rank 3D features[M]//Fleet D, Pajdla T, Schiele B, et al. Computer vision-ECCV 2014. Lecture notes in computer science. Cham: Springer, 2014, 8689: 520-535.
- [67] Hu Y L, Hugonot J, Fua P, et al. Segmentation-driven 6D object pose estimation[C]//2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), June 15-20, 2019, Long Beach, CA, USA. New York: IEEE Press, 2020: 3380-3389.
- [68] Song C, Song J R, Huang Q X. HybridPose: 6D object pose estimation under hybrid representations[C]//2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), June 13-19, 2020, Seattle, WA, USA. New York: IEEE Press, 2020: 428-437.
- [69] Hartley R, Zisserman A. Multiple view geometry in computer vision[M]. Cambridge: Cambridge University Press, 2000.
- [70] Wahba G. A least squares estimate of satellite attitude[J]. SIAM Review, 1965, 7(3): 409.
- [71] Liu X Y, Jonschkowski R, Angelova A, et al. KeyPose: multi-view 3D labeling and keypoint estimation for transparent objects[C]//2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), June 13-19, 2020, Seattle, WA, USA. New York: IEEE Press, 2020: 11599-11607.
- [72] Li P L, Chen X Z, Shen S J. Stereo R-CNN based 3D object detection for autonomous driving[C]//2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), June 15-20, 2019, Long Beach, CA, USA. New York: IEEE Press, 2020: 7636-7644.
- [73] Triggs B, McLauchlan P F, Hartley R I, et al. Bundle adjustment: a modern synthesis[M]//Triggs B, Zisserman A, Szeliski R. Vision algorithms: theory and practice. Lecture notes in computer science. Heidelberg: Springer, 2000, 1883: 298-372.
- [74] Hirschmüller H. Stereo processing by semiglobal matching and mutual information[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2008, 30(2): 328-341.
- [75] Kendall A, Martirosyan H, Dasgupta S, et al. End-to-end learning of geometry and context for deep stereo regression[C]//2017 IEEE International Conference on Computer Vision (ICCV), October 22-29, 2017, Venice, Italy. New York: IEEE Press, 2017: 66-75.
- [76] Rusu R B, Marton Z C, Blodow N, et al. Persistent point feature histograms for 3D point clouds[C]//10th International Conference on Intel Autonomous System (IAS-10), July 24, 2008, Baden-Baden, Germany. Amsterdam: IOS PRESS, 2008: 119-128.
- [77] Rusu R B, Blodow N, Beetz M. Fast point feature histograms (FPFH) for 3D registration[C]//2009 IEEE International Conference on Robotics and Automation, May 12-17, 2009, Kobe, Japan. New York: IEEE Press, 2009: 3212-3217.
- [78] Salti S, Tombari F, Di Stefano L. SHOT: Unique signatures of histograms for surface and texture description [J]. Computer Vision and Image Understanding, 2014, 125: 251-264.
- [79] Besl P J, McKay N D. Method for registration of 3-D shapes[J]. Proceedings of SPIE, 1992, 1611: 586-606.
- [80] 耿明志. 无人机自动着陆过程中的视觉导航技术研究[D]. 南京: 南京航空航天大学, 2007: 74-80.
- [81] 张洲宇. 基于多源信息融合的无人机自主着舰引导技术研究[D]. 南京: 南京航空航天大学, 2017: 51-67.
- [82] 耿明志. The application of visual navigation in auto-landing of unmanned aerial vehicle[D]. Nanjing: Nanjing University of Aeronautics and Astronautics, 2007: 74-80.
- [83] Zhang Z Y. Research on multi-source information based guidance technology for unmanned aerial vehicle autonomous carrier-landing[D]. Nanjing: Nanjing University of Aeronautics and Astronautics, 2017: 51-67.
- [84] Mourikis A I, Trawny N, Roumeliotis S I, et al. Vision-aided inertial navigation for spacecraft entry, descent, and landing[J]. IEEE Transactions on Robotics, 2009, 25(2): 264-280.
- [85] 李建军, 王大轶. 一种图像辅助火星着陆段自主导航方法[J]. 宇航学报, 2016, 37(6): 687-694.
- [86] Li J J, Wang D Y. An image-based autonomous navigation method for precise landing on Mars[J]. Journal of Astronautics, 2016, 37(6): 687-694.
- [87] 余新中. 基于多传感器的头部姿态估计与随动控制研究[D]. 成都: 电子科技大学, 2022: 34-47.
- [88] Yu X Z. Research on head pose estimation and follow-up control based on multi-sensor[D]. Chengdu: University of Electronic Science and Technology of China, 2022: 34-47.
- [89] 王妮. 基于多源信息融合的无人机视觉导航技术[D]. 西安: 西安电子科技大学, 2019: 55-62.
- [90] Wang N. Vision navigation technology based on multi-source information fusion for UAV[D]. Xi'an: Xidian University, 2019: 55-62.
- [91] Zuo X X, Yang Y L, Geneva P, et al. LIC-fusion 2.0: LiDAR-inertial-camera odometry with sliding-window plane-feature tracking[C]//2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), October 24-January 24, 2021, Las Vegas, NV, USA. New York: IEEE Press, 2021: 5112-5119.
- [92] 辛琪, 史忠科. 基于多源信息的飞行姿态估计方法[J]. 飞行力学, 2012, 30(6): 527-531.
- [93] Xin Q, Shi Z K. Flight attitude determination base on multiple measurements[J]. Flight Dynamics, 2012, 30(6): 527-531.
- [94] 张胜凯. 无线信号感知与惯性导航融合的微型无人机状

- 态估计研究[D]. 武汉: 华中科技大学, 2021: 68-92.
- Zhang S K. State estimation for micro aerial vehicles via wireless-inertial fusion[D]. Wuhan: Huazhong University of Science and Technology, 2021: 68-92.
- [89] Cheng J Y, Yi J J, He L, et al. Improvement of shipboard landing performance of shipborne UAV using multi-sensor fusion[C]//Proceedings of 2019 International Conference on Computer Science, Communications and Big Data, March 24-25, 2019, Beijing, China. [S.l.: s.n.], 2019: 304-309.
- [90] He Y S, Huang H B, Fan H Q, et al. FFB6D: a full flow bidirectional fusion network for 6D pose estimation [C]//2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), June 20-25, 2021, Nashville, TN, USA. New York: IEEE Press, 2021: 3002-3012.
- [91] He Y S, Sun W, Huang H B, et al. PVN3D: a deep point-wise 3D keypoints voting network for 6DoF pose estimation[C]//2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), June 13-19, 2020, Seattle, WA, USA. New York: IEEE Press, 2020: 11629-11638.
- [92] Wang C, Xu D F, Zhu Y K, et al. DenseFusion: 6D object pose estimation by iterative dense fusion[C]//2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), June 15-20, 2019, Long Beach, CA, USA. New York: IEEE Press, 2020: 3338-3347.
- [93] Chatterji G B, Menon P K, Sridhar B. Vision-based position and attitude determination for aircraft night landing[J]. *Journal of Guidance, Control, and Dynamics*, 1998, 21(1): 84-92.
- [94] Oszust M, Kapuscinski T, Warchol D, et al. A vision-based method for supporting autonomous aircraft landing [J]. *Aircraft Engineering and Aerospace Technology*, 2018, 90(6): 973-982.
- [95] Gibert V, Burlion L, Chriette A, et al. New pose estimation scheme in perspective vision system during civil aircraft landing[J]. *IFAC-PapersOnLine*, 2015, 48 (19): 238-243.
- [96] Gibert V, Plestan F, Burlion L, et al. Visual estimation of deviations for the civil aircraft landing[J]. *Control Engineering Practice*, 2018, 75: 17-25.
- [97] Meng Y, Wang W, Han H, et al. A vision/radar/INS integrated guidance method for shipboard landing[J]. *IEEE Transactions on Industrial Electronics*, 2019, 66 (11): 8803-8810.
- [98] 空客完成全自主滑行起降飞行测试[J]. *航空维修与工程*, 2020(7): 110.
- Airbus completed the flight test of fully autonomous taxiing and landing[J]. *Aviation Maintenance & Engineering*, 2020(7): 110.
- [99] Martínez C, Campoy P, Mondragón I, et al. Trinocular ground system to control UAVs[C]//2009 IEEE/RSJ International Conference on Intelligent Robots and Systems, October 10-15, 2009, St. Louis, MO, USA. New York: IEEE Press, 2009: 3361-3367.
- [100] Ma Z W, Hu T J, Shen L C. Stereo vision guiding for the autonomous landing of fixed-wing UAVs: a saliency-inspired approach[J]. *International Journal of Advanced Robotic Systems*, 2016, 13(2): 43.
- [101] Santos N P, Lobo V, Bernardino A. Autoland project: fixed-wing UAV landing on a fast patrol boat using computer vision[C]//OCEANS 2015-Seattle, October 27-31, 2019, Seattle, WA, USA. New York: IEEE, 2019: 1-5.
- [102] Santos N P, Lobo V, Bernardino A. Unmanned aerial vehicle tracking using a particle filter based approach[C]//2019 IEEE Underwater Technology (UT), April 16-19, 2019, Kaohsiung, Taiwan, China. New York: IEEE Press, 2019.
- [103] Kaiser C, Sjöberg F, Delcura J M, et al. SMART-OLEV: an orbital life extension vehicle for servicing commercial spacecrafts in GEO[J]. *Acta Astronautica*, 2008, 63(1/2/3/4): 400-410.
- [104] 李世立, 郭旻, 杜文斌, 等. 基于高亚像素精度的图像处理测量坦克炮口振动位移[J]. *测试技术学报*, 2017, 31(2): 131-136.
- Li S L, Guo M, Du W B, et al. Measurement of tank gun muzzle vibration based on image processing with high sub-pixel precision[J]. *Journal of Test and Measurement Technology*, 2017, 31(2): 131-136.
- [105] 宋威, 艾邦成, 蒋增辉, 等. 内埋武器投放分离相容性的风洞投放试验预测与评估[J]. *航空学报*, 2020, 41(6): 523415.
- Song W, Ai B C, Jiang Z H, et al. Prediction and assessment of drop separation compatibility of internal weapons by wind tunnel drop-test[J]. *Acta Aeronautica et Astronautica Sinica*, 2020, 41(6): 523415.
- [106] 吴榕榕, 王健, 王英霖. 弹体高速侵彻冰体研究[J]. *弹箭与制导学报*, 2022, 42(4): 74-80.
- Wu R R, Wang J, Wang Y L. Study on high speed penetration of projectile into ice body[J]. *Journal of Projectiles, Rockets, Missiles and Guidance*, 2022, 42 (4): 74-80.
- [107] 刘晶波, 王菲, 孙运轮. 大飞机撞击钢筋混凝土核安全壳模型试验研究[J]. *建筑结构学报*, 2022, 43(8): 185-195.
- Liu J B, Wang F, Sun Y L. Experimental study on a large aircraft impacting reinforced concrete nuclear containment structure model[J]. *Journal of Building Structures*, 2022, 43(8): 185-195.
- [108] Marguet B, Ribere B. Measurement-assisted assembly applications on airbus final assembly lines[J]. *SAE Transactions*, 2003, 112: 372-375.
- [109] Ruel S, Luu T. STS-128 on-orbit demonstration of the TriDAR targetless rendezvous and docking sensor[C]//2010 IEEE Aerospace Conference, March 6-13, 2010, Big Sky, MT, USA. New York: IEEE Press, 2010.
- [110] Samson C, English C, Deslauriers A, et al. The neptec three-dimensional laser camera system: from space mission STS-105 to terrestrial applications[J]. *Canadian Aeronautics and Space Journal*, 2004, 50(2): 115-123.
- [111] 陈冠宇, 成群林, 张解语, 等. 基于多传感器测量的航天器舱段自动对接位姿调整方法[J]. *北京航空航天大学*

- 学报, 2019, 45(6): 1232-1239.
- Chen G Y, Cheng Q L, Zhang J Y, et al. Multi-sensor measurement based position and pose adjustment method for automatic docking of spacecraft cabins[J]. Journal of Beijing University of Aeronautics and Astronautics, 2019, 45(6): 1232-1239.
- [112] 王稼祥, 陈坤勇, 江录春, 等. 基于视觉引导的航空发动机低压涡轮轴自动化装配系统研究[J]. 机械设计与研究, 2021, 37(3): 94-100, 104.
- Wang J X, Chen K Y, Jiang L C, et al. Research on automatic assembly system of aero engine low pressure turbine shaft based on visual guidance[J]. Machine Design & Research, 2021, 37(3): 94-100, 104.
- [113] Liu S, Xu D, Zhang D P, et al. High precision automatic assembly based on microscopic vision and force information[J]. IEEE Transactions on Automation Science and Engineering, 2016, 13(1): 382-393.
- [114] Tamadazte B, Marchand E, Dembélé S, et al. CAD model-based tracking and 3D visual-based control for MEMS microassembly[J]. International Journal of Robotics Research, 2010, 29(11): 1416-1434.
- [115] 杜甫. 挠性组件自动组装夹具与装配作业控制[D]. 大连: 大连理工大学, 2021: 34-44.
- Du F. Automatic assembly fixture and assembly operation control for flexible components[D]. Dalian: Dalian University of Technology, 2021: 34-44.
- [116] Rogez G, Rihan J, Ramalingam S, et al. Randomized trees for human pose detection[C]//2008 IEEE Conference on Computer Vision and Pattern Recognition, June 23-28, 2008, Anchorage, AK, USA. New York: IEEE Press, 2008.
- [117] 代钦, 石祥滨, 乔建忠, 等. 结合遮挡级别的人体姿态估计方法[J]. 计算机辅助设计与图形学学报, 2017, 29(2): 279-289.
- Dai Q, Shi X B, Qiao J Z, et al. Articulated human pose estimation with occlusion level[J]. Journal of Computer-Aided Design & Computer Graphics, 2017, 29(2): 279-289.
- [118] Toshev A, Szegedy C. DeepPose: human pose estimation via deep neural networks[C]//2014 IEEE Conference on Computer Vision and Pattern Recognition, June 23-28, 2014, Columbus, OH, USA. New York: IEEE Press, 2014: 1653-1660.
- [119] Wei S H, Ramakrishna V, Kanade T, et al. Convolutional pose machines[C]//2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), June 27-30, 2016, Las Vegas, NV, USA. New York: IEEE Press, 2016: 4724-4732.
- [120] Chen Y L, Wang Z C, Peng Y X, et al. Cascaded pyramid network for multi-person pose estimation[C]//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, June 18-23, 2018, Salt Lake City, UT, USA. New York: IEEE Press, 2018: 7103-7112.
- [121] Cheng B W, Xiao B, Wang J D, et al. HigherHRNet: scale-aware representation learning for bottom-up human pose estimation[C]//2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), June 13-19, 2020, Seattle, WA, USA. New York: IEEE Press, 2020: 5385-5394.
- [122] Pavlakos G, Zhou X W, Derpanis K G, et al. Coarse-to-fine volumetric prediction for single-image 3D human pose[C]//2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), July 21-26, 2017, Honolulu, HI, USA. New York: IEEE Press, 2017: 1263-1272.
- [123] Liu M Y, Yuan J S. Recognizing human actions as the evolution of pose estimation maps[C]//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, June 18-23, 2018, Salt Lake City, UT, USA. New York: IEEE Press, 2018: 1159-1168.
- [124] Luvizon D C, Picard D, Tabia H. 2D/3D pose estimation and action recognition using multitask deep learning[C]//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, June 18-23, 2018, Salt Lake City, UT, USA. New York: IEEE Press, 2018: 5137-5146.
- [125] Yan A, Wang Y L, Li Z F, et al. PA3D: pose-action 3D machine for video recognition[C]//2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), June 15-20, 2019, Long Beach, CA, USA. New York: IEEE Press, 2020: 7914-7923.
- [126] Wang M C, Tighe J, Modolo D. Combining detection and tracking for human pose estimation in videos[C]//2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), June 13-19, 2020, Seattle, WA, USA. New York: IEEE Press, 2020: 11085-11093.
- [127] Zhang Y F, Wang C Y, Wang X G, et al. VoxelTrack: multi-person 3D human pose estimation and tracking in the wild[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2023, 45(2): 2613-2626.
- [128] 唐永正. 复杂场景下的人体姿态估计与动作识别研究[D]. 北京: 北京工业大学, 2021: 21-44.
- Tang Y Z. Research on human pose estimation and action recognition in complex scenes[D]. Beijing: Beijing University of Technology, 2021: 21-44.
- [129] Johnson S, Everingham M. Clustered pose and nonlinear appearance models for human pose estimation[C]//Proceedings of the British Machine Vision Conference 2010, August 31-September 3, Aberystwyth. London: British Machine Vision Association, 2010: 1-11.
- [130] 机器之心 Pro. 3DAT 助力冬奥开幕式及赛前训练: 这究竟是什么「神秘黑科技」?[EB/OL]. (2022-03-05)[2022-12-27]. <https://baijiahao.baidu.com/s?id=1728261204867089230&wfr=spider&for=pc>.
- Machine Heart Pro. 3DAT helps the opening ceremony and pre-competition training of the Winter Olympics: what is this 'mysterious black technology'? [EB/OL]. (2022-03-05)[2022-12-27]. <https://baijiahao.baidu.com/s?id=1728261204867089230&wfr=spider&for=pc>.
- [131] 工信微报.【冬奥进行时】惊艳! 这 23 项信息技术开创了冬奥历史[EB/OL]. (2022-02-14)[2022-12-27]. https://view.inews.qq.com/k/20220214A03KXR00?web_channel=wap&openApp=false.

- Industry and Information Micro News. 'When the winter olympics are in progress' amazing! These 23 information technologies made the history of the Winter Olympics [EB/OL]. (2022-02-14) [2022-12-27]. https://view.inews.qq.com/k/20220214A03KXR00?web_channel=wap&openApp=false.
- [132] 夏盼. 面向智能车主动安全的行人姿态估计[D]. 成都: 电子科技大学, 2019: 40-72.
- Xia P. Pedestrian posture estimation for smart car active safety[D]. Chengdu: University of Electronic Science and Technology of China, 2019: 40-72.
- [133] Zhao Y, Yuan Z J, Chen B D. Accurate pedestrian detection by human pose regression[J]. IEEE Transactions on Image Processing, 2019, 29: 1591-1605.
- [134] 刘楠. 人体姿态估计与驾驶员姿态追踪[D]. 上海: 上海交通大学, 2020: 27-49.
- Liu N. Human pose recognition and pilot posture tracking [D]. Shanghai: Shanghai Jiao Tong University, 2020: 27-49.
- [135] Wu Q Q, Xu G H, Zhang S C, et al. Human 3D pose estimation in a lying position by RGB-D images for medical diagnosis and rehabilitation[C]//2020 42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society, July 20-24, 2020, Montreal, Canada. New York: IEEE Press, 2020: 5802-5805.
- [136] Cao T, Armin M A, Denman S, et al. In-bed human pose estimation from unseen and privacy-preserving image domains[C]//2022 IEEE 19th International Symposium on Biomedical Imaging (ISBI), March 28-31, 2022, Kolkata, India. New York: IEEE Press, 2022.
- [137] Rohan A, Rabah M, Hosny T, et al. Human pose estimation-based real-time gait analysis using convolutional neural network[J]. IEEE Access, 2020, 8: 191542-191550.