

基于深度学习位姿估计的多视结构光三维成像方法

崔海华^{1*}, 姜涛¹, 杜坤鹏², 国荣辉¹, 赵安安²¹南京航空航天大学机电学院, 江苏 南京 210016;²中航西安飞机工业集团股份有限公司, 陕西 西安 710089

摘要 多视角结构光测量是利用结构光测量系统从多个角度测量实现被测对象完整表达的过程, 所以多个视角下测量数据的拼接影响了被测对象的完整性。提出一种利用深度学习估计位姿并直接进行多视角数据拼接的方法。结构光测量模型为四步相移配合多频外差法, 实现单次高精度三维重建。使用只看一次(YOLO)网络识别被测对象的 3D 包围盒角点, 采用 n 点透视(PnP)算法进行目标位姿估计。由于测量系统和位姿估计的坐标系均统一到单目相机下, 多个视角的数据直接利用估计的位姿进行拼接。建立相邻点云的特征描述子, 利用迭代最近点(ICP)算法实现高精度拼接。结果表明, 提出的测量方法能够有效实现多视角结构光数据的拼接; 位姿估计的平移精度优于 3 mm, 旋转精度优于 1° , 拼接点云的平均偏差为 0.02 mm; 这与利用标志点拼接的结果具有可比较的精度水平。拼接方法适用于单次可完整估计位姿的多视角结构光测量, 提升了多视角拼接效率。

关键词 结构光测量; 多视角; 位姿估计; 深度学习; 数据拼接

中图分类号 TP751

文献标志码 A

doi: 10.3788/AOS202141.1712001

3D Imaging Method for Multi-View Structured Light Measurement Via Deep Learning Pose Estimation

Cui Haihua^{1*}, Jiang Tao¹, Du Kunpeng², Guo Ronghui¹, Zhao An'an²

¹ College of Mechanical and Electrical Engineering, Nanjing University of Aeronautics and Astronautics, Nanjing, Jiangsu 210016, China;

² AVIC Xi'an Aircraft Industry Group Co., Ltd., Xi'an, Shaanxi 710089, China

Abstract Multi-view structured light measurement is a process of utilizing structured light measurement system to achieve the complete expression of the measured object from multiple angles. Thus, the splicing of the measurement data from multiple perspectives affects the integrity of the measured object. In this paper, a new method is proposed to estimate the measurement pose using the deep learning and to directly align the multi-view data. The structure light measurement model is a four-step phase-shifting method combined with multi-frequency heterodyne method to realize the single high-precision three-dimensional reconstruction. In pose estimation, You only look once (YOLO) network is used to identify the 3D bounding box corner of the measured object, and perspective n point (PnP) algorithm is used to estimate the target pose. Since the coordinate systems of the measurement system and pose estimation are unified to the monocular camera, the data from multiple perspectives are directly spliced using the estimated pose. The feature descriptors of adjacent point clouds are established, and iterative closest point (ICP) algorithm is used to realize the high-precision stitching. The results show that the proposed measurement can effectively realize the multi-view structured light data splicing. The translation accuracy of pose estimation is better than 3 mm, the rotation accuracy is better than 1° , and the average deviation of stitching point cloud is 0.02 mm, which has a comparable accuracy level with that of using the method of the marker points. The proposed method is suitable for the multi-view structured light measurement with single pose estimation, which can improve the

收稿日期: 2020-12-29; 修回日期: 2021-02-14; 录用日期: 2021-03-23

基金项目: 国家重点研发计划(2019YFB1707501, 2019YFB2006100)、江苏省自然科学基金(BK20191280)

通信作者: *cuihh@nuaa.edu.cn

registration efficiency for multi-view measured data.

Key words structured light measurement; multiple view; pose estimation; deep learning; data splicing

OCIS codes 150.6910; 150.5758; 120.6650; 100.3008; 150.1488

1 引言

结构光测量技术是一种主动式光学测量方法,通过投射编码光栅获得被测物体的三维信息,在逆向工程、数字化建模、计算测量等领域研究广泛^[1]。随着计算机、计算光学等技术的发展,结构光在大范围、高动态、高反光等难测对象的测量方面均实现突破^[2-4]。结构光测量因其单次测量即可重建整个型面而被广泛应用。但是,对于复杂、大尺寸对象,由于光学成像受到测量视角的限制,单次测量无法获得物体的完整几何表达,多视角测量然后进行数据融合是主要的解决方案^[5]。

多视角结构光测量数据融合的核心问题是不同视角下测量视角的位姿估计。通常,数据拼接的过程包括粗拼接和全局优化。其中全局优化是以拼接后匹配点误差最小为目标优化变换矩阵,通常使用迭代最近点(ICP)算法。粗拼接方法较多,包括:基于精密运动控制^[6]或固定多测量系统^[7]的拼接、公共标志点转站拼接^[8-9]、不变特征转站拼接^[10]和跟踪测量拼接^[11-12]。基于精密运动控制或者固定多测量系统是利用硬件的精度进行数据拼接,多测量系统、运动机构需精确标定,大尺寸精密运动机构难以实现;公共标志点方法是在被测对象或固定位置粘贴标志点,结构光测量获得一个视角下被测对象型面数据的同时,检测当前视角的标志点。该方法依赖人工标志点识别和测量结果,公共区域内无标志点或少标志点将造成拼接精度下降;同样,基于不变特征的转站拼接方法依赖于被测对象自身特征,对于弱纹理的被测对象拼接效果差。三维数据直接拼接通常是基于点云的几何特征建立特征描述算子,然后进行特征匹配和数据拼接。采用深度学习建立三维点云识别算子,方法有效但相对基于图像的方法复杂程度高^[13];跟踪测量系统利用全局定位系统和终端扫描设备实现大部件局部扫描数据的整体拼接融合,该方法需两套设备,实施困难。另外,通过增加平面镜或透镜可提高单次测量的测量范围和对象的几何完整表达^[14-15],但是系统标定困难,操作不灵活,应用范围受限^[16-17]。

随着深度学习技术的快速发展,深度学习同样被用于结构光测量优化方面,比如鲁棒相位求解,快速投影和传感器融合^[1]。利用深度学习可直接进行目

标识别、位姿估计、深度估计等任务,尤其对于测量依赖稳定环境的情况,深度学习不依赖环境和光照,提高了测量的稳定性和可靠性。利用单幅图像估计被测对象位姿成为研究热点,传统的单目位姿估计依赖被测对象的几何约束,采用 n 点透视(PnP)算法进行求解,在没有几何约束时位姿估计失效^[18]。利用卷积神经网络(CNN)进行位姿估计如6自由度物体位姿检测器(DPOD)、只看一次(YOLO)6自由度位姿检测(YOLO6D)可实现被测对象位姿直接估计,对于多目标位姿估计的准确率在80%以上^[19-20]。在多视角测量的数据粗拼接阶段,相比较利用公共标志点或公共特征进行数据融合方法,直接估计位姿的方式省略了特征点匹配、拼接位姿解算等过程,且位姿估计过程不依赖任何特征标志,易于实现自动化,提升了整体的测量效率^[21]。由于深度学习位姿估计精度受限于训练样本和网络模型,深度学习的方法适用于粗拼接阶段,结合ICP算法可实现多视角数据的快速高精度融合。

综上所述,本文提出基于测量视角位姿估计的多视角结构光测量数据拼接方法。利用深度学习方法进行目标特征的位姿估计,建立被测对象和测量系统的相对变换。由于测量数据同样位于测量坐标系,故可实现多视角数据的直接拼接。最后利用ICP算法进行全局测量数据优化调整。该方法提升了多视角结构光测量的智能性和测量效率,并可推广至多视角摄影测量、激光扫描等系统。

2 多视角结构光数据融合策略

图1为所提的多视角结构光测量数据融合策略。如图1所示,测量系统主要包含三方面内容:第一是结构光测量平台,本测量系统由单目相机配合投影模块构成,投影模块投射事先加载的正弦条纹,相机采集后进行相位求解,计算被测型面的三维坐标。第二是基于深度学习的被测目标估计。利用YOLO深度学习网络,估计被测对象的3D包围盒,然后利用PnP算法求解相机的位姿。第三是多视角测量数据拼接与融合。单视角结构光测量数据及对象的目标位姿同时获取,多个视角的测量数据利用每个视角估计的位姿进行坐标变换,实现粗拼接,然后基于ICP算法进行多视角数据的高精度拼接。

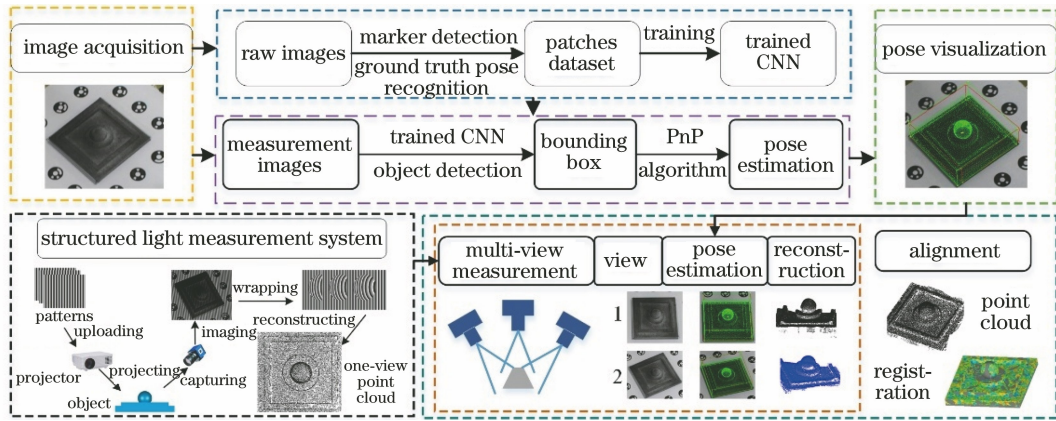


图 1 多视角结构光测量数据拼接策略

Fig. 1 Proposed data alignment strategy for multi-view structured light measurement

2.1 结构光三维重建方法

采用单目结构光测量模型,利用四步相移算法和多频外差原理求解包裹相位和绝对相位^[22],通过标定单目相机和投影模块坐标系相对关系,实现相位与被测型面的三维坐标对应求解。采用标准的四步相移算法,投射获得的图像灰度值可表示为

$$\begin{cases} I_0(x,y) = A(x,y) + B(x,y)\cos[\varphi(x,y)] \\ I_1(x,y) = A(x,y) + B(x,y)\cos[\varphi(x,y) + \pi/2] \\ I_2(x,y) = A(x,y) + B(x,y)\cos[\varphi(x,y) + \pi] \\ I_3(x,y) = A(x,y) + B(x,y)\cos[\varphi(x,y) + 3\pi/2] \end{cases}, \quad (1)$$

式中: $I_i(x,y)$ 表示第*i*($i=0,1,2,3$)幅条纹图像在图像坐标(x,y)的像素灰度值; $A(x,y)$ 表示图像的背景光强; $B(x,y)$ 表示条纹干涉获得的图像在(x,y)处光强变化的幅值, $\varphi(x,y)$ 表示所求的初始包裹相位。四步相移求解的包裹相位公式为

$$\varphi(x,y) = \arctan \frac{I_1(x,y) - I_3(x,y)}{I_0(x,y) - I_2(x,y)}. \quad (2)$$

获得包裹相位图后,利用多频外差原理计算绝对相位。对于双频外差,求解的绝对相位 ϕ_1, ϕ_2 可表示为

$$\begin{cases} \phi_1 = \varphi_1 + 2\pi \times \text{floor} \left\{ \frac{p_2}{p_2 - p_1} \left[N_{12} + \alpha + \frac{\varphi_1 - \varphi_2}{2\pi} \right] \right\} \\ \phi_2 = \varphi_2 + 2\pi \times \text{floor} \left\{ \frac{p_1}{p_2 - p_1} \left[N_{12} + \alpha + \frac{\varphi_1 - \varphi_2}{2\pi} \right] \right\} \end{cases}, \quad (3)$$

式中: $\text{floor}(\cdot)$ 表示向下取整; φ_1, φ_2 表示两个频率相位的包裹相位; α 为取 0 或 1 的参数, $\varphi_1 < \varphi_2$ 时, $\alpha=1, \varphi_1 \geq \varphi_2$ 时, $\alpha=0$ 。 p_1, p_2 表示两个频率的正弦光栅的周期(节距)且有 $p_1 < p_2$;两条正弦条纹的级数分别为 $n_g = N_g + \varphi_g / 2\pi, N_g \in \mathbb{Z}, g=1,2$,其

中 N_g 表示条纹级数的整数部分;定义 $N_{1,2}$ 表示两条纹合成条纹的级数 $n_{1,2}$ 中的整数部分, $p_{1,2}$ 表示合成条纹周期(节距),以此类推。如果 p_1, p_2 选择恰当,恰好使得叠加的条纹覆盖整个视场,则 $N_{1,2}=0$,但此时三维重建的细节特征显得模糊,采用三频外差进行求解,即通过增加一条固定周期的正弦条纹求解 $N_{1,2}$ 的值。若第三条条纹的节距为 p_3 ,且 $p_{1,2} < p_3$,重叠条纹节距为 $p_{1,2,3}$,对应的级数整数和小数部分分别为 $N_{1,2,3}$ 和 $\Delta n_{1,2,3}$ 。由此可求解出双频条纹重叠光栅的级数

$$n_{1,2} = \frac{p_3}{p_3 - p_{1,2}} [N_{1,2,3} + \alpha + \Delta n_{1,2} - \Delta n_3]. \quad (4)$$

则有

$$\begin{cases} N_{1,2} = \text{floor}(n_{1,2}) \\ \Delta n_{1,2} = \begin{cases} \Delta n_1 - \Delta n_2, \Delta n_1 - \Delta n_2 \leq 0 \\ \Delta n_1 - \Delta n_2 + 1, \Delta n_1 - \Delta n_2 > 0 \end{cases} \end{cases}, \quad (5)$$

式中, $\Delta n_1, \Delta n_2, \Delta n_3$ 分别表示三个条纹级数的小数部分, $\Delta n_{1,2}$ 表示 1,2 两个条纹的合成条纹级数的小数部分。当 $\Delta n_3 > \Delta n_{1,2}$ 时, $\alpha=1$;反之, $\alpha=0$ 。使用三频外差使重叠光栅的周期能够覆盖整个视场,则 $N_{1,2,3}=0$,根据(5)式求解 $N_{1,2}$,根据(3)式求解绝对相位。

在单目结构光系统中,投影模块的投射过程可看作相机成像的逆过程。即空间一个点在相机坐标系下的坐标可通过小孔成像确定,同样,空间点在投影仪“成像”坐标下的坐标可通过绝对相位表示,即

$$u_p = \frac{\phi(x,y) \times W}{2\pi N_v}, \quad (6)$$

式中: u_p 表示空间点在投影平面的水平方向坐标; W 为投影仪图像宽度; N_v 表示条纹周期数; $\phi(x,y)$ 为绝对相位。而对于空间点(x,y,z)在相机成

像过程将产生两个方程,即

$$s \begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \begin{bmatrix} f_u & 0 & u_0 \\ 0 & f_v & v_0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} R & t \\ 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \\ 1 \end{bmatrix}, \quad (7)$$

式中: f_u, f_v, u_0, v_0 分别表示相机的主距和主点,通过相机标定获得, (u, v) 为图像平面坐标; R, t 为相机外参,包括旋转矩阵和平移向量,通过标定板建立世界坐标系进行求解。通过检测图像点,并计算

其对应的绝对相位值,利用(6)式和(7)式可构建三个方程,即可反求出空间点的三维坐标。

2.2 基于深度学习的位姿估计

利用深度学习进行位姿估计的基本思路是:利用 YOLO 网络识别被测对象的三维包围盒在图像上的成像点,然后利用 PnP 算法求解对象的位姿。该方法可以实现单幅图像位姿的快速求解,且求解精度较高。图 2 为位姿估计的总体框图。

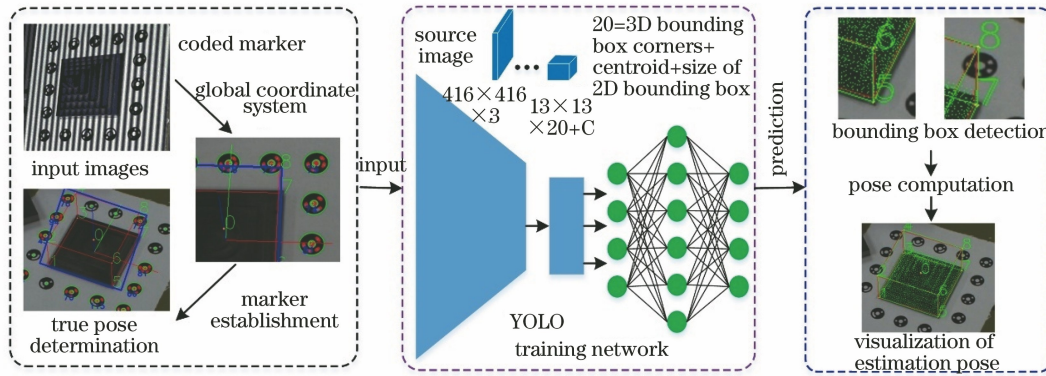


图 2 利用 YOLO 网络进行位姿估计的过程

Fig. 2 Pose estimation based on YOLO network

网络直接在图像中识别包围盒角点的像素坐标,损失函数 L 定义为

$$L = \lambda_p \sum_{h=1}^9 \| \mathbf{p}_h - \tilde{\mathbf{p}}_h \|^2 + \lambda_c L_c, \quad (8)$$

式中: $\mathbf{p}, \tilde{\mathbf{p}}$ 分别表示真实角点和预测的角点; $\| \cdot \|^2$ 表示 2-范数运算符,物理意义为计算两个向量之间的欧氏距离; λ_p 表示对应的权重,设置值为 1; λ_c 为可信度的权重; L_c 表示可信度的损失函数。对于目标物体的质心在图像上的投影与预测的投影坐标之间存在偏差,同样利用欧氏距离求解偏差值。若偏差距离为 D ,定义距离阈值为 d ,则可信度函数定义为

$$L_c = \begin{cases} \exp\left(1 - \frac{D}{d}\right), & D < d \\ 0, & \text{otherwise} \end{cases}, \quad (9)$$

结合(8)式,根据网络,训练时对图像进行划分元胞,如果质心投影点在元胞内,则 $\lambda_c = 5$,否则 $\lambda_c = 0.1$ 。

网络的输入为原始图像,输出为目标位姿,同时可输出目标对象包围盒的角点图像坐标、2D 包围盒的长宽、以及目标类别。文中被测对象只有 1 个,类别为 1。如图 2 所示,网络的真值输入与输出数据准备过程如下:

采集图像,手动分割出图像中的目标物体作为掩模二值图;利用圆形编码点构建全局坐标系,被测

目标摆放在编码坐标范围内,根据实际位置约束计算全局坐标与被测对象坐标的相对关系,求解位姿真值。根据损失函数,将位姿的真值转化为图像包围盒角点坐标和质心坐标。由于不能直接从图像中获得被测对象的 3D 包围盒,所以利用(7)式重投影的方式获得包围盒角点的图像坐标。首先识别图像中的编码点,根据实际设定的码点的位置约束计算相机位置。若编码点的图像坐标为 $\mathbf{m}_l(u_l, v_l)$,设定对应的空间坐标 $\mathbf{M}_l(X_l, Y_l, Z_l)$,根据本文 3D 包围盒角点数量,设定此处 $l = 0, 1, 2, \dots, 8$,根据(7)式相机成像模型

$$s \begin{bmatrix} u_l \\ v_l \\ 1 \end{bmatrix} = \begin{bmatrix} r_{11} & r_{12} & r_{13} & t_1 \\ r_{21} & r_{22} & r_{23} & t_2 \\ r_{31} & r_{32} & r_{33} & t_3 \end{bmatrix} \begin{bmatrix} X_l \\ Y_l \\ Z_l \\ 1 \end{bmatrix}, \quad (10)$$

其中, $\mathbf{R} = \begin{bmatrix} r_{11} & r_{12} & r_{13} \\ r_{21} & r_{22} & r_{23} \\ r_{31} & r_{32} & r_{33} \end{bmatrix}$, $\mathbf{t} = \begin{bmatrix} t_1 \\ t_2 \\ t_3 \end{bmatrix}$ 。将(10)式展开

并消除 s , 得 $u_l = \frac{r_{11}X_l + r_{12}Y_l + r_{13}Z_l + t_1}{r_{31}X_l + r_{32}Y_l + r_{33}Z_l + t_3}$, $v_l = \frac{r_{21}X_l + r_{22}Y_l + r_{23}Z_l + t_2}{r_{31}X_l + r_{32}Y_l + r_{33}Z_l + t_3}$ 。若位姿矩阵 $\mathbf{T}(\mathbf{R}, \mathbf{t})_{3 \times 4} = [\mathbf{r}_1, \mathbf{r}_2, \mathbf{r}_3]^T$, \mathbf{r}_j 为 \mathbf{T} 的第 j ($j = 1, 2,$

3) 行的转置向量。则(8)式可表示为

$$\begin{cases} \mathbf{r}_1^T \mathbf{M}_l - \mathbf{r}_3^T \mathbf{M}_l u_l = 0 \\ \mathbf{r}_2^T \mathbf{M}_l - \mathbf{r}_3^T \mathbf{M}_l v_l = 0 \end{cases} \quad (11)$$

利用编码点识别的图像坐标及其对应的空间三维坐标,建立线性方程组

$$\begin{bmatrix} \mathbf{M}_0^T & 0 & -u_0 \mathbf{M}_0^T \\ 0 & \mathbf{M}_0^T & -v_0 \mathbf{M}_0^T \\ \vdots & \vdots & \vdots \\ \mathbf{M}_8^T & 0 & -u_8 \mathbf{M}_8^T \\ 0 & \mathbf{M}_8^T & -v_8 \mathbf{M}_8^T \end{bmatrix} \begin{bmatrix} \mathbf{r}_1 \\ \mathbf{r}_2 \\ \mathbf{r}_3 \end{bmatrix} = \mathbf{0}. \quad (12)$$

通过奇异值分解法对系数矩阵进行分解,获得 \mathbf{T} 的最小二乘解。编码的相对位置为事先设定,与相机标定中的标定板相似,编码点的圆心几何关系已知,三维坐标点可人为设定。接着,按照相机投影模型便可以将被测对象包围盒角点及形心投影到图像上,得到角点的真值。根据网络模型,图像 2D 包围盒利用 8 个角点拟合一个最小包围矩形,矩形的长宽也作为网络的训练输出值,故共计需要 21 个参数。综上所述,训练数据集包括:原始图像、二值掩模图像、三维模型(.ply 格式)、相机标定参数以及上述 21 个数值。测试网络的输入为原始图像,其通过在随机生成的视点处拍摄得到,输出为上述的 21 个参数。网络每一次迭代均生成 8 个包围盒角点坐标,利用(10)~(12)式进行位姿估计,然后将理论的空间包围盒角点重投影到图像上,计算损失函数值并调整网络权重。位姿的估计精度用平移精度 E_T 和旋转精度 E_A 定义为

$$\begin{aligned} E_T &= \frac{1}{N} \sum_{q=1}^N \|\mathbf{T}_{q_e} - \mathbf{T}_{q_0}\|^2 \\ E_A &= \frac{1}{N} \sum_{q=1}^N \|\mathbf{A}_{q_e} - \mathbf{A}_{q_0}\|^2 \end{aligned} \quad (13)$$

式中: $\mathbf{T}_{q_e}, \mathbf{T}_{q_0}$ 分别表示在第 q 次迭代后估计的平移向量和对应的真实平移向量; $\mathbf{A}_{q_e}, \mathbf{A}_{q_0}$ 分别表示第 q 次迭代后估计的欧拉角和对应的真实的欧拉角; N 为测试集个数。

2.3 多视角数据融合与优化

利用 2.1 节结构光测量方法可获得当前测量视角下被测对象的三维型面坐标。若在第 a, b 个测量视角下获得的点云数据分别记为 \mathbf{M}_a 和 \mathbf{M}_b 。同时,在两个视角下,利用 2.2 节深度学习方法直接估计被测对象的拍摄姿态分别记为 \mathbf{T}_a 和 \mathbf{T}_b 。测量的数据可以利用估计姿态进行转换,即

$$\mathbf{M}'_b = \mathbf{T}_a^{-1} \mathbf{T}_b \mathbf{M}_b. \quad (14)$$

由于测量坐标系和对象位姿坐标系均在相机坐标系下,则变换后的测量数据可直接进行拼接融合,即 $\mathbf{M} = \{\mathbf{M}_a, \mathbf{M}'_b\}$ 。以此类推,全局测量数据的初始拼接结果可表示为 $\mathbf{M} = \{\mathbf{M}_0, \mathbf{T}_0^{-1} \mathbf{T}_1 \mathbf{M}_1, \dots, \mathbf{T}_0^{-1} \mathbf{T}_n \mathbf{M}_n\}$,其中 n 为全部测量视角个数。

当获得初始拼接结果后,为消除位姿估计的误差,使得多视角测量数据融合更加准确,进一步进行数据拼接优化。利用具有距离约束的快速点特征直方图描述算法对点云的特征点进行描述。首先建立局部坐标系,记第 a 个视角下的点云第 b 个被描述的特征点为 \mathbf{P}_{ab} ,其 k 个邻域点集记作 \mathbf{P}_{ab}^k 。以 \mathbf{P}_{ab} 为中心, \mathbf{P}_{ab}^k 内任意两个点可以建立局部坐标系 $\{\mathbf{P}_{ab}\}$;同时,取 \mathbf{P}_{ab} 中另一 m 个邻域点 \mathbf{P}_{ab}^m 为中心,同样建立局部坐标系 $\{\mathbf{P}_{ab}^m\}$ 。若点 \mathbf{P}_{ab} 和 \mathbf{P}_{ab}^m 对应的法向矢量分别为 \mathbf{n}_0 和 \mathbf{n}_m ,则可以构建特征描述算子 $\mathbf{D}_{ab} = \{d_{ab}, \alpha_{ab}, \varphi_{ab}, \theta_{ab}\}$,具体为

$$\begin{cases} d_{ab} = \|\mathbf{P}_{ab}^k - \mathbf{P}_{ab}\|, k = 1, 2, \dots \\ \alpha_{ab}^k = \mathbf{r}_y \mathbf{n}_m \\ \varphi_{ab} = \frac{\mathbf{r}_x (\mathbf{P}_{ab}^k - \mathbf{P}_{ab})}{\|\mathbf{P}_{ab}^k - \mathbf{P}_{ab}\|} \\ \theta_{ab} = \arctan(\mathbf{r}_z \mathbf{n}_m, \mathbf{r}_x \mathbf{n}_m) \end{cases}, \quad (15)$$

式中: $(\mathbf{r}_x, \mathbf{r}_y, \mathbf{r}_z)$ 表示坐标系 $\{\mathbf{P}_{ab}\}$ 三个方向的单位向量; d_{ab} 表示中心点到其余点的欧氏距离; $\alpha_{ab}^k, \varphi_{ab}$ 表示所求两个向量的投影; θ_{ab} 表示两向量的夹角。对每个视角下的点云构建特征描述算子,然后利用特征描述向量的欧氏距离作为判断准则进行特征点的相似度判断依据。接下来即是调整拼接矩阵使得每个视角下拼接后的点云对应点之间的偏差最小。建立目标函数如下

$$\begin{aligned} &(\mathbf{T}_0, \mathbf{T}_a, \mathbf{T}_s) = \\ &\sum_a \sum_b \sum_s \sum_t x_{abst} \|\mathbf{T}_0^{-1} \mathbf{T}_a \mathbf{M}_{ab} - \mathbf{T}_0^{-1} \mathbf{T}_s \mathbf{M}_{st}\|, \end{aligned} \quad (16)$$

式中: x_{abst} 为 0 或 1 的整数,表示第 a 个视角下的点云中第 b 个点是否与第 s 个视角下点云的第 t 个点为相似点,其通过特征描述子(15)式来判断; \mathbf{T}_0 为初始视角相机位姿, $\mathbf{T}_a, \mathbf{T}_s$ 为第 a 和第 s 视角的相机位姿; \mathbf{M}_{ab} 为第 a 视角第 b 个点坐标; \mathbf{M}_{st} 为 s 视角 t 个点坐标。目标函数是保证全局范围内,所有的相似点之间的距离达到全局最小,利用 ICP 算法进行求解。

3 实验结果与分析

3.1 位姿估计网络数据获取与结果分析

为验证融合位姿估计的多视角结构光测量的有

效性,搭建实验平台。硬件包括:商用投影仪(NEC NP43+)和工业相机(DMK 31BU03)。相机的分辨率为 1024 pixel×768 pixel。图 3 为系统的硬件及网络训练结果图。深度学习网络在 Windows 10 平台运行,配置 i7-9750 CPU 和 NVIDIA GeForce GTX 1660 Ti 显卡。其中结构光测量界面基于 Qt 进行编写,为显示与可视化使用 C++ 语言,主要采用 Boost

(C++ 扩展库)、Eigen(矩阵运算库)、OpenCV(计算机视觉库)等第三方开源库实现。在实验中,随机采集 1600 幅图像,通过检测有效的圆形编码点对应的图像筛选出 1443 个有效图像。随机选择 1155 张图像作为训练数据集,剩下 288 张图像作为测试集。准备训练所需的真值数据,包括 3D 包围盒角点及型心的归一化图像坐标、2D 包围盒长宽、类别。

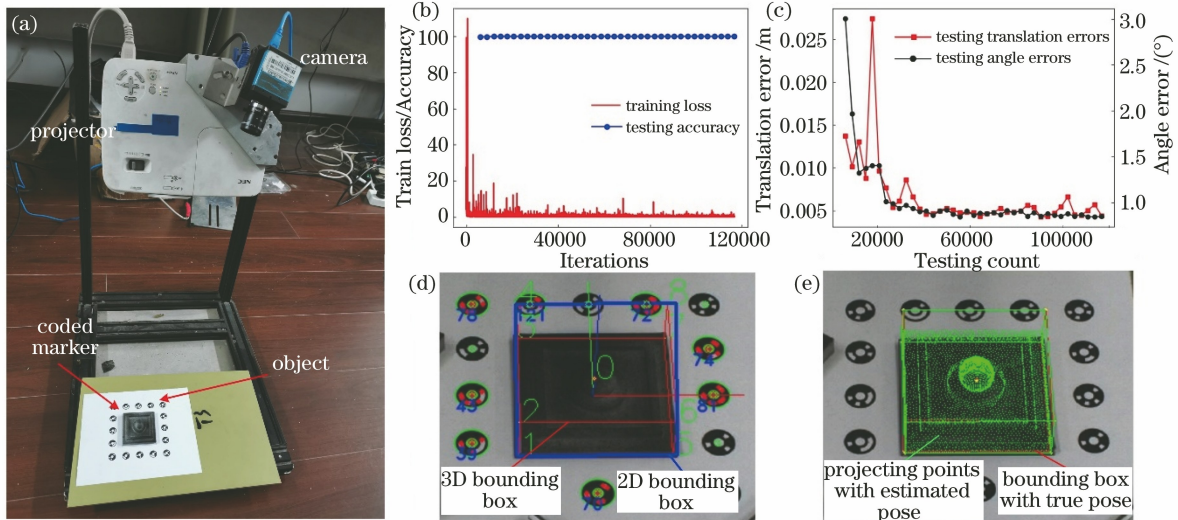


图 3 实验设备及结果。(a)实验设备;(b)训练损失及测试精度;(c)平移和旋转误差;
(d)真值位姿确定可视化结果;(e)预测位姿的可视化结果

Fig. 3 Experimental results. (a) Setup; (b) training loss and testing accuracy; (c) translation and rotation error; (d) true pose determination; (e) pose estimation visualization

如图 3(a)所示,结构光测量系统由一个摄像机和光学投影仪组成。使用 3D 打印件作为实验对象,模型来自诺丁汉大学制造测量团队^[23]。图 3(b)显示出了训练损失和测试精度,其中训练损失定义为真实和估计图像点之间的距离偏移,测试精度设定为偏移值小于 5 pixel 的点的数量占估计点总数的百分比。随着迭代次数的增加,训练损失迅速下降,测试精度保持 100%。从图 3(b)曲线变化趋势可以看出,该网络对于目标姿态估计稳定且精度高。由图 3(c)可知,三个方向的总平移误差小于 3 mm,旋转总误差小于 1°,满足初始数据拼接的精度需求。图 3(d)展示了一幅图像用于真值姿态计算的

编码点检测、三维和二维边界框确定的过程。图 3(e)显示了投影模型、估计的 3D 包围盒和被测对象的型心投影点。从图 3(e)可以看出,考虑到角点偏移和质心均小于 5 pixel,目标物体投影点与被测对象吻合得很好。为判断位姿识别网络的鲁棒性和适用性,进一步对更多的 3D 打印件进行网络训练和位姿估计实验,并利用估计的位姿将目标对象的理论数模重投影到图像中,可视化的结果如图 4 所示。

图 4 中每一个对象进行位姿估计均需要进行一次网络训练,故相对于利用标志点的拼接方法,位姿估计的方法在前期训练方面耗时较大。但是由于其不需要粘贴标志点,且训练完成后对被测对象摆放

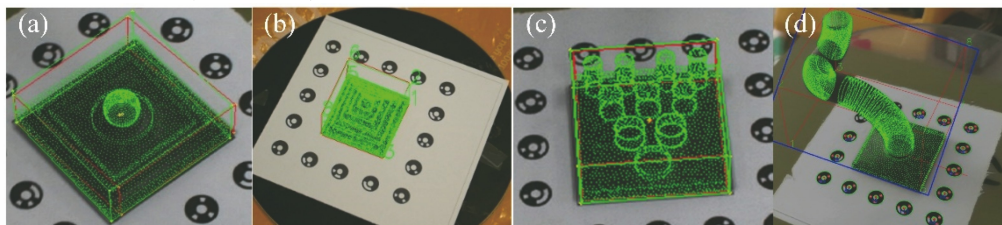


图 4 利用更多模型进行位姿估计验证。(a)球体;(b)金字塔;(c)圆柱;(d)弯管

Fig. 4 More cases presentation of object pose estimation. (a) Sphere; (b) pyramid; (c) pillars; (d) elbow

位姿没有任何要求,使得位姿估计的方法更加灵活、便捷。对应的估计位姿估计误差如表 1 所示,误差计算方法为(13)式。

表 1 位姿估计误差计算

Object	Sphere	Pyramid	Pillars	Elbow
Mean re-projecting error /pixel	2.526	0.846	1.753	3.216
Translation error /mm	3.541	1.897	3.019	5.187
Angle error /($^{\circ}$)	0.85	0.42	0.55	1.05

图 4 表明,从不同的角度来看,估计的姿势与真实姿势保持一致。对球体、金字塔、圆柱和弯管等不同形状的位姿估计的分析,证明了该网络在多形状姿态估计中的鲁棒性和有效性。从表 1 可以看出,根据估计的 8 个角点和真实角点之间的间隙确定的像素误差在 0.846~3.216 pixel 之间变化,证明了网络在 3D 目标检测中的高精度。此外,通过估计

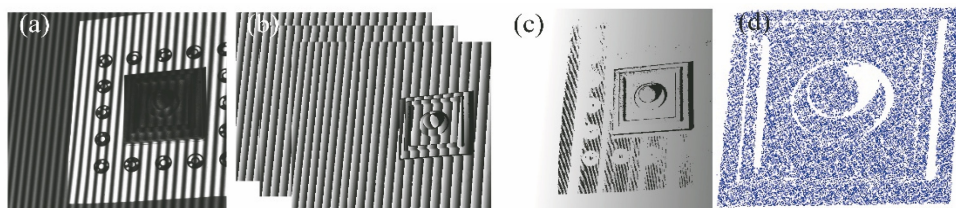


图 5 利用所搭建的系统进行单视角结构光重建。(a)投影图像;(b)包裹相位;(c)绝对相位;(d)三维点云
Fig. 5 Single-view structured light reconstruction based on the proposed system. (a) Projection image;
(b) wrapping phase; (c) absolute phase; (d) 3D point cloud

图 5 展示了投影图像,三种频率的包裹相位,绝对相位以及重建的三维点云。可以看出,单视角无法获得完整的球体表面形貌,只能获得少部分三维数

据。进一步,在另一个视角下进行三维测量,从采集的图像中估计当前视角下目标物体的位姿,然后利用估计的位姿进行拼接和融合。拼接的结果如图 6 所示。

3.2 结构光测量与数据融合结果分析

采用四步相移和三频外差法,三种相移的周期分别为 $T_1=12$ pixel、 $T_2=13$ pixel、 $T_3=14$ pixel,则有 $T_{12}=156$ pixel、 $T_{23}=182$ pixel,以及 T_{12} 、 T_{23} 叠加光栅相位差的周期为 $T_{123}=1092$ pixel,即系统可以实现像素宽度为 1092 pixel 的测量范围。投影仪共投射 13 张条纹,并触发相机采集图像计算绝对相位,其中 12 张用于三维重建,第 13 张不投条纹图,用于位姿估计。可实现固定视角下的三维重建。图 5 展示了利用所搭建的结构光测量系统进行单视角重建的过程,此处仍然使用球体进行验证。

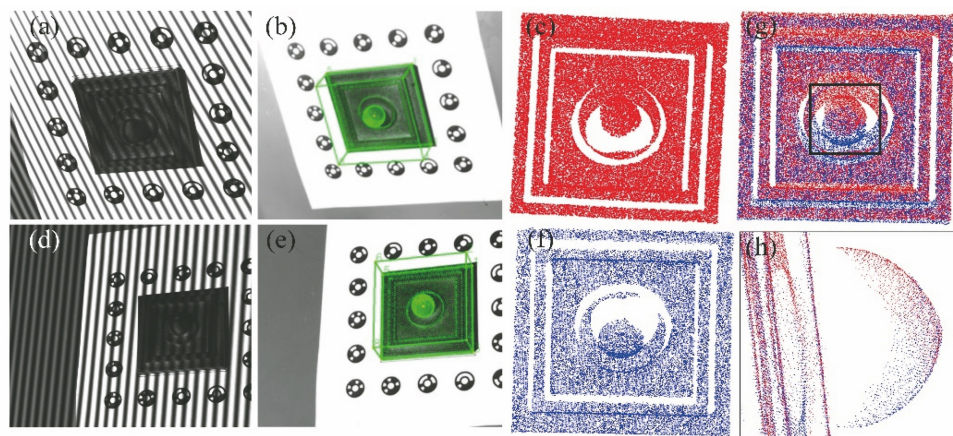


图 6 利用估计位姿进行点云拼接。(a)(d)两个视角下的投影图像;(b)(e)两个视角位姿估计结果;
(c)(f)两个视角重建的点云;(g)点云拼接结果;(h)图(g)中黑色框放大图

Fig. 6 Point cloud splicing using estimated pose. (a)(d) Projection images in two views; (b)(e) pose estimation results;
(c)(f) point clouds in two views; (g) data splicing result; (h) zoom-in view of box in Fig. (g)

图 6 展示了利用多视角重建策略进行两视图测量的过程,包括位姿估计和三维重建。图 6(g)为两个视角的点云拼接的结果,图 6(h)表明拼接算法是有效的。为进一步验证拼接结果的精度,利用更多的算例,将拼接后的点云与理论的数模相对比,比较点的偏差。具体如图 7 所示。

图 7 显示了不同视角下的测量数据与位姿估计结果。图 7(a)是两个不同视图的三维支柱的拼接结果,其中变换矩阵是通过深度学习估计的。图 7(b)是使用 ICP 算法对图 7(a)进行的改进。为了评估对准精度,图 7(b)中的点云与理论 CAD 模型进

行配准。图 7(c)表示两视图对准点云的误差分布。此外,如图 7(c)所示,使用估计的姿势,配准误差的最小值和最大值分别为 -0.176 mm 和 0.2 mm。平均误差为 0.103 mm,标准偏差为 0.1 mm。图 7(d)是使用估计的位姿进行凹台模型的粗略对齐。图 7(a)和图 7(d)均是采用估计的位姿直接进行数据拼接的结果,其不依赖特征点真实值,由于多个角度下估计的位姿结果存在偏差,两片点云数据偏差是单视角位姿估计结果的综合偏差。精细配准和误差分布分别如图 7(e)和图 7(f)所示,表明了使用深度学习的多视图数据配准的有效性。

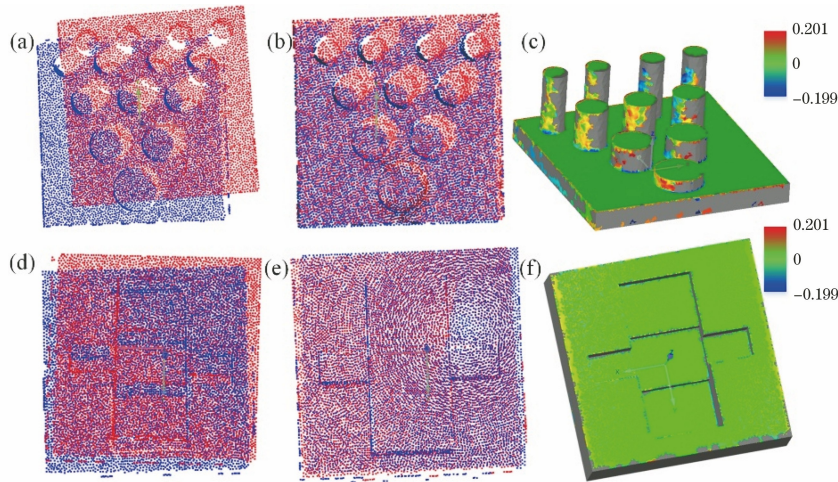


图 7 用估计姿态进行数据配准。(a)(d)采用基于深度学习的姿态估计方法,对立柱和凹台进行两视图配准;(b)(e)基于初始刚性变换的 ICP 算法的全局求精;(c)(f)融合数据在图(b)、(e)和 CAD 模型中最终配准的误差分布,其中误差是利用点到模型的距离计算的

Fig. 7 Data registration with estimated pose. (a)(d)Two-view registration of pillars and recess, with deep learning-based pose estimation; (b)(e) global refinement using ICP algorithm based on rough rigid transformation; (c)(f) error distributions of the final registration of fused data in Figs. (b), (e) and CAD model, where the error is determined by the point-to-model distance

利用控制标记点对两个视图的点云进行融合,平均误差结果如表 2 所示。

表 2 利用标志点和深度学习方法数据融合误差对比
Table 2 Error comparison of data fusion using markers and deep learning

Object	Sphere	Pyramid	Pillars
Pose estimation /mm	0.065	0.058	0.067
Coded marker /mm	0.044	0.038	0.046
Error /mm	0.021	0.020	0.023

表 2 对比了利用提出的方法和利用标志点方法进行两视角数据拼接的误差。可以看出平均误差的偏差在 0.02 mm 左右,利用估计变换得到的误差参数也与基于标记的拼接结果接近,验证了基于深度学习的多视点结构光测量与现有基于特征的配准结果的可比性。现有配置下,单幅图像

位姿估计平均时间为 0.05 s,两视角标志点检测与匹配优化时间为 0.08 s,15 个视角下的特征点光束平差优化时间为 1.6 s,平均单幅图像减小时间 0.056 s,拍摄视角越多,直接位姿估计的方法在测量过程中无需粘贴标志点,整体测量效率更高,且深度学习保证了被测对象任意摆放的情况下均能稳定估计位姿,促进了测量智能化水平提升。结合图 7,所提的多视角测量数据融合的策略最终的精度依赖 ICP 算法,但提出的方法的优势在于提升粗配准的效率和灵活性,减少标志点检测和匹配的误差,从而避免了多视角特征的误检测与误匹配,以及光束平差等复杂计算问题,且可实现实时位姿估计。

4 结 论

提出了融合深度学习位姿估计的多视角结构光测量方法,利用位姿估计网络直接估计当前测量视角下被测对象的6D姿态,可用于多视角三维测量的初始拼接,相比于传统基于辅助转台、贴点及跟踪测量等方法具有便捷性和灵活性,与基于特征识别的拼接方法相比,提出的方法的适用性更广。采用提出的策略进行多视角测量数据拼接的精度与理论数模相比平均误差为0.02 mm,与基于编码点的多视角拼接具有相近的拼接效果。该方法解决了传统多视测量方法对辅助装置和特征信息的依赖性问题,更重要的是该方法可适用于一些特殊的应用场景比如不允许粘贴多余物的零部件原位测量等。但是,提出的方法相对标志点拼接方法对前期训练信息具有依赖性,另外,提出的方法对于单视角不能包含全部对象的目标重建,需进行位姿估计网络的升级,进一步研究位姿估计模型,实现大部件无特征点的测量数据拼接和整体三维重建。

参 考 文 献

- [1] Marrugo A G, Gao F, Zhang S. State-of-the-art active optical techniques for three-dimensional surface metrology: a review[J]. *Journal of the Optical Society of America A*, 2020, 37(9): B60-B77.
- [2] Zhang S. High-speed 3D shape measurement with structured light methods: a review[J]. *Optics and Lasers in Engineering*, 2018, 106: 119-131.
- [3] Cui H H, Jiang T, Cheng X S, et al. A general gamma nonlinearity compensation method for structured light measurement with off-the-shelf projector based on unique multi-step phase-shift technology[J]. *Journal of Modern Optics*, 2019, 66(15): 1579-1589.
- [4] Li Z J, Cui H H, Liu C Y, et al. A shape measurement method based on automatic multiple exposure surface structured light[J]. *Acta Optica Sinica*, 2018, 38(11): 1112004.
李兆杰, 崔海华, 刘长毅, 等. 一种基于自动多次曝光面结构光的形貌测量方法[J]. *光学学报*, 2018, 38(11): 1112004.
- [5] Garrido-Jurado S, Muñoz-Salinas R, Madrid-Cuevas F J, et al. Simultaneous reconstruction and calibration for multi-view structured light scanning[J]. *Journal of Visual Communication and Image Representation*, 2016, 39: 120-131.
- [6] Lang W, Xue J P, Li C H, et al. Splicing of multi-view point clouds based on calibrated parameters of turntable[J]. *Chinese Journal of Lasers*, 2019, 46(11): 1104003.
郎威, 薛俊鹏, 李承杭, 等. 基于旋转台参数标定实现多视角点云拼接[J]. *中国激光*, 2019, 46(11): 1104003.
- [7] Jin M. Study on structured light multi-view measurement[D]. Qingdao: Ocean University of China, 2008.
金明. 结构光多视角测量技术的研究[D]. 青岛: 中国海洋大学, 2008.
- [8] Barone S, Paoli A, Razonale A V. Three-dimensional point cloud alignment detecting fiducial markers by structured light stereo imaging[J]. *Machine Vision and Applications*, 2012, 23(2): 217-229.
- [9] Liu W, Lan Z G, Zhang Y, et al. Global data registration technology based on dynamic coded points[J]. *IEEE Transactions on Instrumentation and Measurement*, 2018, 67(2): 394-405.
- [10] Tang M H, Cheng X S, Cui H H, et al. A high stability mapping stitching method for structured light three-dimensional measurement[J]. *Acta Optica Sinica*, 2014, 34(11): 1112006.
汤明辉, 程筱胜, 崔海华, 等. 面向结构光三维测量的高稳定映射拼接方法[J]. *光学学报*, 2014, 34(11): 1112006.
- [11] Jiang T, Cheng X, Cui H, et al. Combined shape measurement based on locating and tracking of an optical scanner[J]. *Journal of Instrumentation*, 2019, 14(1): P01006.
- [12] Jiang T, Cui H H, Cheng X S. Accurate calibration for large-scale tracking-based visual measurement system[J]. *IEEE Transactions on Instrumentation and Measurement*, 2021, 70: 1-11.
- [13] Chang W C, Pham V T. 3-D point cloud registration using convolutional neural networks[J]. *Applied Sciences*, 2019, 9(16): 3273.
- [14] Psota P, Tang H M, Pooladvand K, et al. Multiple angle digital holography for the shape measurement of the unpainted tympanic membrane[J]. *Optics Express*, 2020, 28(17): 24614-24628.
- [15] Zeng Z. The research of 3D information acquisition technology of object overall shape[D]. Shenzhen: Shenzhen University, 2017.
曾增. 物体全面形三维信息获取技术研究[D]. 深圳: 深圳大学, 2017.
- [16] Yu L P, Pan B. Single-camera stereo-digital image correlation with a four-mirror adapter: optimized design and validation[J]. *Optics and Lasers in Engineering*, 2016, 87: 120-128.
- [17] Xu J, Wang P, Yao Y Y, et al. 3D multi-directional

- sensor with pyramid mirror and structured light[J]. *Optics and Lasers in Engineering*, 2017, 93: 156-163.
- [18] Zhang X F, Liu H B, Shang Y. Robust orthogonal iteration algorithm for single camera pose estimation[J]. *Acta Optica Sinica*, 2019, 39(9): 0915004.
张雄锋, 刘海波, 尚洋. 单目相机位姿估计的稳健正交迭代方法[J]. *光学学报*, 2019, 39(9): 0915004.
- [19] 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.
- [20] Jiang T, Cui H H, Cheng X S, et al. Viewpoint planning for robot photogrammetry based on initial pose estimation via deep learning[J]. *Acta Automatica Sinica*, 2020, 46(x): 1-12.
姜涛, 崔海华, 程筱胜, 等. 基于深度学习初始位姿估计的机器人摄影测量视点规划[J]. *自动化学报*, 2020, 46(x): 1-12.
- [21] 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, Korea (South). New York: IEEE Press, 2019: 1941-1950.
- [22] Chen S L, Zhao J B, Xia R B. Improvement of the phase unwrapping method based on multi-frequency heterodyne principle[J]. *Acta Optica Sinica*, 2016, 36(4): 0412004.
陈松林, 赵吉宾, 夏仁波. 多频外差原理相位解包裹方法的改进[J]. *光学学报*, 2016, 36(4): 0412004.
- [23] Eastwood J, Sims-Waterhouse D, Piano S, et al. Autonomous close-range photogrammetry using machine learning [C/OL]. Proceedings of the 14th International Symposium of Measurement Technology and Intelligent Instruments. September 1-4, 2019, Niigata, Japan. [2020-12-20]. https://www.researchgate.net/publication/335082725_Autonomous_close-range_photogrammetry_using_machine_learning.