

文章编号 1004-924X(2021)07-1720-11

基于编码标记点的高精度运动估计

迟书凯, 叶 旋, 高 翔*, 解则晓, 陶冬冬
(中国海洋大学 工程学院, 山东 青岛 266100)

摘要:为实现前景运动的精确估计,本文提出了一种基于编码标记点的高精度运动估计方法。该方法在测量环境内的前背景中粘贴环状编码标记点,并在每次前景运动后采集若干图像,通过同时估计所有相机位姿以及背景固定标记点、前景运动标记点的空间坐标,获取前景运动六自由度参数(旋转、平移)。为实现上述目标,本文提出了包括基于随机抽样一致性的编码标记点识别与提取,图聚类的编码标记点与相机分组,增量式从运动恢复结构的背景点与相机位姿初始化,图优化的前景点与前景运动初始化以及广义光束平差的全局优化等在内的一套高精度运动估计流程。实验结果表明,本文提出的前景运动估计方法精度在 0.3 mm 左右,可以满足高精度前景运动估计的需求。

关键词:高精度运动估计;编码标记点;基于图聚类的分组;基于广义光束平差的全局优化

中图分类号:TP391.4 **文献标识码:**A **doi:**10.37188/OPE.20212907.1720

Coded marker-based high-accuracy motion estimation

CHI Shu-kai, YE Xuan, GAO Xiang*, XIE Ze-xiao, TAO Dong-dong

(College of Engineering, Ocean University of China, Qingdao 266100, China)

* Corresponding author, E-mail: xgao@ouc.edu.cn

Abstract: In order to achieve accurate foreground motion estimation, a coded marker-based high-accuracy motion estimation method is proposed in this paper. First, several circular coded markers are pasted on the foreground and background of the measurement environment. Then, several images are captured after each foreground motion. Finally, based on the captured images, all the camera poses and the spatial coordinates of the fixed markers on the background and the moving markers on the foreground are estimated simultaneously to obtain the 6 Degrees of Freedom (DoF) parameters (rotation and translation) of foreground motion. To this end, a high-accuracy motion estimation pipeline is proposed in this paper, which includes RANDOM SAMPLE CONSENSUS (RANSAC)-based coded marker detection and recognition, graph clustering-based coded marker and camera partitioning, incremental Structure from Motion (SfM)-based background marker and camera initialization, graph optimization-based foreground marker and foreground motion initialization, and generalized Bundle Adjustment (BA)-based global optimization. The experimental results show that the accuracy of the proposed foreground motion estimation method is approximately 0.3 mm, which is satisfactory for high-accuracy foreground motion estimation.

Key words: High-accuracy motion estimation; coded marker; graph clustering-based partitioning; generalized BA-based global optimization

收稿日期:2020-12-25;修订日期:2021-02-25.

基金项目:国家自然科学基金(No. 62003319, No. 42076192, No. 62076026);山东省自然科学基金资助项目(No. ZR2020QF075);中国博士后科学基金资助项目(No. 2020M682239);中央高校基本科研业务费专项(No. 201964013)

1 引言

为实现前景运动的精确估计,本文提出了一种基于编码标记点的高精度运动估计方法。前景运动估计指的是估计动态场景(前景)相对于静态场景(背景)的运动参数,包括三自由度的旋转参数以及三自由度的平移参数。其在智能监控、目标跟踪、姿态估计等领域有着广泛的应用。现有的前景运动估计方法主要包括两类:基于RGB图像的前景运动估计以及基于RGB-D图像的前景运动估计。

基于RGB图像的前景运动估计方法主要有光流法^[1-2]、目标跟踪法^[3-4]、直接估计法^[5-6]等。光流法通过估计帧间光流的方式估计前景运动,当前景运动幅度较大时,会因不满足光流假设导致光流估计的失败;目标跟踪法采用基于视觉的目标跟踪方法对前景目标进行跟踪以估计前景运动,这类方法通常只是给出前景目标的包围盒且估计的是目标在图像上的二维运动;直接估计法基于深度神经网络,通过在二维图像上直接估计前景目标的六自由度位姿参数实现前景运动估计,尽管这类方法可以估计六自由度的前景运动,但是这类基于图像的姿态估计方法通常精度相对较低,通常难以满足前景运动估计的精度要求。

另外,对于基于RGB-D图像的前景运动估计,Li等人^[7]提出了一种同时重建静态背景与非刚体前景的方法,该方法采用实时实例分割算法“你只看系数”(You Only Look At CoefficientTs, YOLACT)^[8]实现前背景分割,然后采用刚体与非刚体迭代最近点(Iterative Closest Point, ICP)算法,实现背景与前景的跟踪。上述方法虽然能够实现前景运动估计,但是由于RGB-D图像中深度图噪声较大且有效范围较小,难以精确捕捉较大范围的前景三维运动,在一些对精度要求较高的场景(如机器人末端运动精确估计)中应用受限。

由上述分析可知,现有前景运动估计方法存在维度缺失、范围受限、精度不足等问题,在对精度要求较高的大范围六自由度运动(如机械臂拾放、机器人搬运等)进行运动估计时,现有方法难

以胜任。对此,本文提出了一种面向上述特定工业应用场景运动估计方法,该方法借助编码标记点实现前景运动的高精度估计。相对于场景本身的纹理特征,编码标记点这类人工标志可以提高测量系统的鲁棒性,在场景结构复杂、纹理缺乏等外界条件下仍可提供较高精度的观测量(图像特征点位置),且几乎不会存在误匹配现象。该方法在测量环境内的前背景中均粘贴编码标记点,所采用的标记点为环状编码标记点,不同的编码标记点均可通过环状编码带确定唯一的身份信息^[9]。在此基础上,本文提出了一套基于环状编码标记点的高精度运动估计流程,包括编码标记点的提取与识别、与相机的分组、背景点与相机位姿初始化、前景点与前景运动初始化、全局优化在内的五个步骤,可以同时估计相机位姿以及(前背景)标记点三维坐标,实现了前景运动的全自动、高精度估计。本文将图聚类方法引入编码标记点与相机的分组过程,实现了前景运动的全局最优划分;将广义光束平差方法引入编码标记点坐标与相机位姿的优化过程,实现了前景运动的全局最优估计。

2 方法

2.1 方法概述

本文提出的基于标记点的高精度运动估计方法的输入为自由拍摄的图像,输出为前景运动(旋转、平移)的估计结果。为区分运动前景与静止背景,在测量场景中粘贴编码标记点,并将位于前景运动目标上与背景静止环境下的编码标记点分别称为前景与背景标记点。该方法主要包括五个部分:1)进行编码标记点的识别与提取,获取图像中各点的中心位置以及码值信息;2)采用基于图聚类的方法,实现编码标记点与相机的分组,即自动区分前景与背景标记点以及不同次前景运动分别对应的相机集合;3)采用增量式从运动恢复结构的方式,对背景点三维坐标以及相机位姿进行初始化;4)根据估计的相机位姿,采用三角测量的方式初始化前景点的三维坐标,并通过图优化的方式初始化前景运动;5)对已估计的前背景点三维坐标、相机位姿、前景运动进行全局优化,以获取前景运动的最终估计结

果。该方法的流程图如图 1 所示,下文将对上述五个部分进行具体介绍。

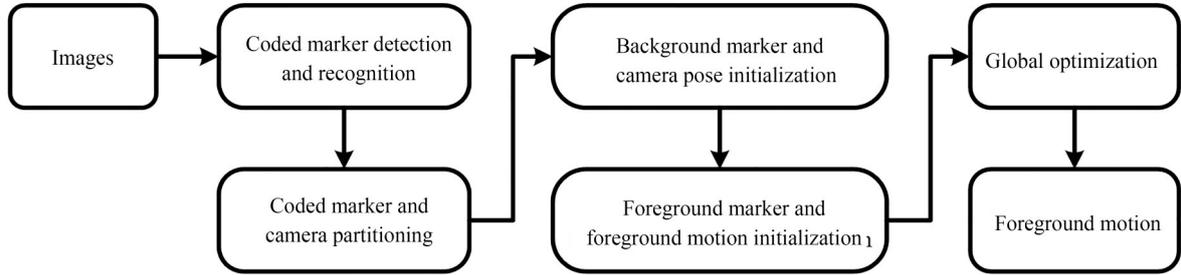


图 1 基于编码标记点的前景运动估计方法的流程图

Fig. 1 Flow chart of the proposed coded marker-based foreground motion estimation method

2.2 编码标记点的提取与识别

在进行编码点的提取时,采用 Xie 等人^[10]的方法,获取标记点的近似成像中心以及像素级边缘点,并利用 Zernike 正交矩^[11]实现边缘点的亚像素化。在此基础上,采用随机抽样一致性(Random Sample Consensus, RANSAC)算法^[12]以及非线性最小二乘优化,对标记点成像椭圆参数进行初始化与优化。上述优化的目标函数定义如下:

$$C^* = \arg \min \frac{1}{2} \sum_{i=1}^N \delta_i d^2(\mathbf{p}_i, C), \quad (1)$$

其中: $C = \begin{pmatrix} a & d & e \\ d & b & f \\ e & f & c \end{pmatrix}$ 为椭圆系数矩阵^[13]; $\{\mathbf{p}_i =$

$(u_i, v_i)^T \mid i = 1, 2, \dots, N\}$ 为用于椭圆拟合的亚像素级边缘点集合; δ_i 为边缘点内外点标签, $\delta_i = 1$ 表示 \mathbf{p}_i 为基于 RANSAC 的椭圆拟合对应的内点, 否则 $\delta_i = 0$; $d(\mathbf{p}_i, C)$ 表示点 m_i 到椭圆 C 的距离, 在此采用 Polar-N-Direction 距离^[14]。此处及下文的非线性最小二乘优化问题均采用 Ceres Solver 进行求解。

在获取各编码标记点的成像椭圆系数矩阵 C 后, 椭圆方程可写作:

$$au^2 + buv + cv^2 + du + ev + f = 0. \quad (2)$$

进而可通过下式计算对应的椭圆中心 $\mathbf{p}_c = (u_c, v_c)^T$:

$$\begin{cases} u_c = \frac{be - df}{d^2 - ab} \\ v_c = \frac{af - de}{d^2 - ab} \end{cases} \quad (3)$$

基于编码标记点的提取结果, 采用 Xie 等人提出的等椭圆内切圆的方法^[9], 获取标记点的码值信息, 实现标记点的识别。

2.3 编码标记点与相机的分组

为实现相机位姿以及前景运动的估计, 需要先进行编码标记点与相机的分组, 即自动区分前景与背景标记点以及不同次前景运动分别对应的相机集合。基于背景标记点数量远小于位于前景标记点的假设, 当对一对分属不同次前景运动的图像利用匹配(同一码值)的标记点进行基于 RANSAC 的基本矩阵估计^[15]时, 对应的标记点匹配对内点为背景点, 外点为前景点。基于上述观察, 可以采用如下的朴素算法进行编码标记点与相机的分组: 对输入图像集合中的任意一对图像进行基于 RANSAC 的基本矩阵估计并记录其中各匹配对的内点情况; 将各标记点在所有图像对中属于内点的比例作为标记点分组依据, 背景点的该比例应高于前景点; 将各图像对中匹配标记点的内点比例作为相机分组依据, 属于同一次前景运动的图像对该比例应高于分属不同次前景运动的图像对。

然而, 上述朴素分组算法存在许多不足: (1) 由于共外极线的情况, 镜头畸变的存在以及距离阈值的不当, 会导致部分内外点的错误归类; (2) 对于不同的实验配置, 上述算法中内点比例的取值分布会存在区别, 因此标记点与相机分组的阈值难以确定; (3) 上述算法仅利用了各标记点及

各图像对的内点比例这类局部信息,而未利用各标记点之间及各相机之间的全局信息,使得分组算法鲁棒性较差。

针对上述问题,本文提出了一种基于图聚类(Graph Clustering)的自动编码标记点与相机的分组算法,从图论的角度对该分组问题进行建模,将每幅图像及每个标记点分别作为顶点,建立相机图结构及标记点图结构,再通过Dhillon等人的图聚类算法^[16],实现编码标记点与相机的自动分组,该方法可以有效利用全局信息,有着更强的鲁棒性。

具体来说,对于相机分组,首先建立相机图结构(Graph Structure),记为 $G^C = \{V^C, E^C\}$,其中 $v_i^C \in V^C$ 与 $e_{i,j}^C \in E^C$ 分别表示该相机图的各项点与边,边 $e_{i,j}^C$ 上的权重 $w_{i,j}^C$ 为图像对 $v_i^C \leftrightarrow v_j^C$ 之间的标记点匹配内点对数。从统计意义上讲,属于同一次前景运动的图像对的匹配内点对数高于分属不同次前景运动的图像对,因此给定前景运动次数(即相机组数),对上述加权相机图进行图聚类,可以实现相机的自动分组。对于标记点分组,同样先建立标记点图结构,记为 $G^M = \{V^M, E^M\}$,其中 $v_i^M \in V^M$ 与 $e_{i,j}^M \in E^M$ 分别表示该相机图的各项点与边。与相机图不同,由于标记点图中边 $e_{i,j}^M$ 上的权重 $w_{i,j}^M$ 用于衡量标记点对 $v_i^M \leftrightarrow v_j^M$ 是否同组,需要给出 $w_{i,j}^M$ 的合理计算方法。在此,首先对于每个标记点 $v_i^M \in V^M$ 通过公式(4)计算其特征矩阵,可知,若 $v_i^M \leftrightarrow v_j^M$ 属于同组时,对应的特征矩阵 $F_i^M \leftrightarrow F_j^M$ 也会较为接近。因此,为使标记点图上边的权重在两点同组时更大且取值均为正,本文采用 $w_{i,j}^M = 2|V^M|^2 - 2|V^M| - \sum_{m=1}^{|V^M|} \sum_{n=1}^{|V^M|} |f_i^M(m,n) - f_j^M(m,n)|$ 方式对其进行计算。

$$F_i^M = \begin{pmatrix} f_i^M(1,1) & \cdots & f_i^M(1,|V^M|) \\ \vdots & \ddots & \vdots \\ f_i^M(|V^M|,1) & \cdots & f_i^M(|V^M|,|V^M|) \end{pmatrix}.$$

$$\text{where } f_i^M(m,n) = \begin{cases} 1 & v_i^M \text{ is inlier for } v_m^C \leftrightarrow v_n^C \\ 0 & v_i^M \text{ is invisible for } v_m^C \text{ or } v_n^C \\ -1 & v_i^M \text{ is outlier for } v_m^C \leftrightarrow v_n^C \end{cases} \quad (4)$$

在此基础上,对上述加权标记点图进行图聚

类,即可实现编码标记点的自动(前背景)分组。

2.4 背景点与相机位姿初始化

经过上一节的编码标记点分组后,可以利用背景点在各幅图像上的投影椭圆中心测量值,借助从运动恢复结构技术^[17-18],估计各背景点的三维坐标以及各相机的位姿(旋转、平移)。由于本文涉及的相机与标记点数量相对较少且标记点的匹配关系精确已知,因此采用一种简化的增量式从运动恢复结构的方法进行背景点与相机位姿的初始化,具体过程简述如下:

Step 1:找到公共可见的编码标记点匹配对数最多的一对图像并根据其匹配标记点成像椭圆中心估计图像对之间的本质矩阵并对其分解获取相对旋转、平移^[19];

Step 2:通过三角测量算法^[20],估计匹配标记点的三维坐标,实现初始图像对重建;

Step 3:通过迭代的形式添加新的图像并进行场景扩展,直至完成所有相机的位姿估计。

在迭代添加相机时,每次选取与当前已获取三维坐标的标记点公共可见对数最多的相机,给定的二维-三维点对应关系,采用基于RANSAC的透视三点(Perspective-3-Point, P3P)算法^[21]估计新添加相机的位姿,并采用三角测量算法估计之前未获取的公共可见标记点的三维坐标,实现场景扩展。另外,在初始图像对重建及每次添加图像与扩展场景后,均采用光束平差(Bundle Adjustment, BA)技术^[22]对标记点三维坐标,相机内参数及其位姿进行优化。

由于本文中各图像均由同一设备采集,在背景点与相机位姿初始化的过程中各幅图像之间共享内参数。本文采用的相机模型为含有径向及切向畸变的相机模型,相机内参数包括焦距 $\{f_x, f_y\}$,主点 $p_0 = (u_0, v_0)$,径向畸变系数 $\{k_1, k_2, k_3\}$,切向畸变系数 $\{p_1, p_2\}$ 。其中,焦距初始值从相机的可交换图像文件格式(Exchangeable Image file Format, EXIF)读取,主点与畸变系数的初始值分别置为图像中心与0,在对相机内参进行优化时固定主点,只优化焦距与畸变参数。根据采用的上述相机模型,若某椭圆实际成像中心为 $p_c = (u_c, v_c)^T$,则其理想成像中心 $\bar{p}_c = (\bar{u}_c, \bar{v}_c)^T$ 可通过下式进行计算:

$$\begin{cases} \bar{u}_c = f_x \bar{x} + u_0 \\ \bar{v}_c = f_y \bar{y} + v_0 \end{cases} \quad (5)$$

其中, $(\bar{x}, \bar{y})^T$ 为椭圆理想成像中心归一化坐标, 计算如下:

$$\begin{cases} \bar{x} = x(1 + k_1 r^2 + k_2 r^4 + k_3 r^6) + 2p_1 xy + p_2(r^2 + 2x^2) \\ \bar{y} = y(1 + k_1 r^2 + k_2 r^4 + k_3 r^6) + p_1(r^2 + 2y^2) + 2p_2 xy \end{cases} \quad (6)$$

其中, $r^2 = x^2 + y^2$, $(x, y)^T$ 为椭圆实际成像中心归一化坐标, 计算如下:

$$\begin{cases} x = (u_c - u_0)/f_x \\ y = (v_c - v_0)/f_y \end{cases} \quad (7)$$

经过上述流程, 可以估计各背景点的三维坐标以及各相机的旋转、平移, 实现背景点与相机位姿初始化。

2.5 前景点与前景运动初始化

在估计各相机位姿以后, 可以通过三角测量算法计算各前景点在不同次前景运动中分别的三维坐标, 进而通过各前景点在不同次前景运动中的坐标变换关系表征各次前景运动。需要注意, 在对前景点进行三角测量时需要按照相机分组结果, 对每次前景运动分别进行。在此基础上, 可以通过 Arun 等人^[23]提出的方法估计各次前景运动之间的相对旋转、平移, 并将其记为 $\{R_{i,j}, t_{i,j}\}$, 其中 $R_{i,j}$ 与 $t_{i,j}$ 分别表示第 i 次与第 j 次前景运动之间的相对旋转与相对平移。然后, 根据获取的前景运动的相对旋转与平移构建位姿图, 图中顶点表示各次前景运动的绝对位姿, 边表示两次前景运动之间的相对位姿。给定位姿图中的相对位姿, 采用位姿图优化^[24]的方式, 估计各次前景运动的绝对位姿初始值, 实现前景运动的初始化。上述位姿图优化的目标函数定义如下:

$$\begin{aligned} & \{R_i^*, t_i^*\} = \\ & \arg \min \frac{1}{2} \sum_{\substack{p_{i,j} \in E^p \\ p_{i,j} \in E^p}} \left(d_{R^3}(t_{i,j}, t_j - R_j R_i^T t_i)^2 + \right. \\ & \left. d_{\text{SO}(3)}(R_{i,j}, R_j R_i^T)^2 \right) \end{aligned} \quad (8)$$

其中, $d_{R^3}(t_i, t_j)$ 与 $d_{\text{SO}(3)}(R_i, R_j)$ 分别表示在实数空间与特殊正交群上的距离度量。具体如下式

$$\begin{aligned} & \{R_{i,j}^{C*}, t_{i,j}^{C*}\}, \{X_{i,k}^{F*}\}, \{X_l^{B*}\}, \{R_i^{F*}, t_i^{F*}\} = \arg \min E(R_{i,j}^C, t_{i,j}^C, X_{i,k}^F, X_l^B, R_i^F, t_i^F) = \\ & \arg \min (E_{2D}^B(R_{i,j}^C, t_{i,j}^C, X_l^B) + E_{2D}^F(R_{i,j}^C, t_{i,j}^C, X_{i,k}^F) + E_{3D}^F(X_{i,k}^F, R_i^F, t_i^F) + E_{3D}^S(X_l^B)) \end{aligned} \quad (9)$$

上述优化问题中含有四类误差项, 包括背景点与前景点的反投影误差项 $E_{2D}^B(R_{i,j}^C, t_{i,j}^C, X_l^B)$ 与

所示:

$$\begin{aligned} d_{R^3}(t_i, t_j) &= \|t_i - t_j\|_{L_2}, \\ d_{\text{SO}(3)}(R_i, R_j) &= \arccos \frac{\text{tr}(R_j R_i^T) - 1}{2}. \end{aligned}$$

2.6 全局优化

经上述步骤, 本文对相机位姿、(前背景)标记点三维坐标以及前景运动进行了初始化, 此小节需要利用上述初值, 通过非线性最小二乘优化的方式, 对上述变量进行全局优化, 以实现前景运动的高精度估计。

本文提出的前景运动估计方法的前景运动估计方法的系统整体示意图以及相关符号标记如图 2 所示。其中 $\{R_{i,j}^C, t_{i,j}^C\}$ 表示相机位姿, $R_{i,j}^C$ 与 $t_{i,j}^C$ 分别为第 i 组第 j 个相机的绝对旋转与平移; $\{X_{i,k}^F\}$ 表示前景标记点三维坐标, $X_{i,k}^F$ 为第 k 个前景点在第 i 次前景运动时的三维坐标; $\{X_l^B\}$ 表示背景标记点三维坐标, X_l^B 为第 l 个背景点的三维坐标; $\{R_i^F, t_i^F\}$ 表示前景运动, R_i^F 与 t_i^F 分别为第 i 次前景运动的绝对旋转与平移; 同一颜色的相机及前景标记点属于同一次前景运动。

另外, 由于从运动恢复结构技术在进行相机标定与场景重建时存在尺度不确定性, 需要在进行全局优化之前借助标尺将上述待优化变量恢复至真实尺度。此处的标尺指的是若干对已知真实距离的背景编码标记点对, 将尺度比 r^S 记为上述标记点对真实距离与测量距离比值的均值, 则通过 $\{R_{i,j}^C, t_{i,j}^C\} \leftarrow \{R_{i,j}^C, r^S t_{i,j}^C\}$, $\{X_{i,k}^F\} \leftarrow \{r^S X_{i,k}^F\}$, $\{X_l^B\} \leftarrow \{r^S X_l^B\}$, $\{R_i^F, t_i^F\} \leftarrow \{R_i^F, r^S t_i^F\}$ 进行真实尺度恢复。

在此基础上, 即可通过求解下式中的最优化问题实现上述参数的全局优化:

$E_{2D}^F(R_{i,j}^C, t_{i,j}^C, X_{i,k}^F)$ 以及前景运动与标尺长度的空间误差项 $E_{3D}^F(X_{i,k}^F, R_i^F, t_i^F)$ 与 $E_{3D}^S(X_l^B)$, 下面将分

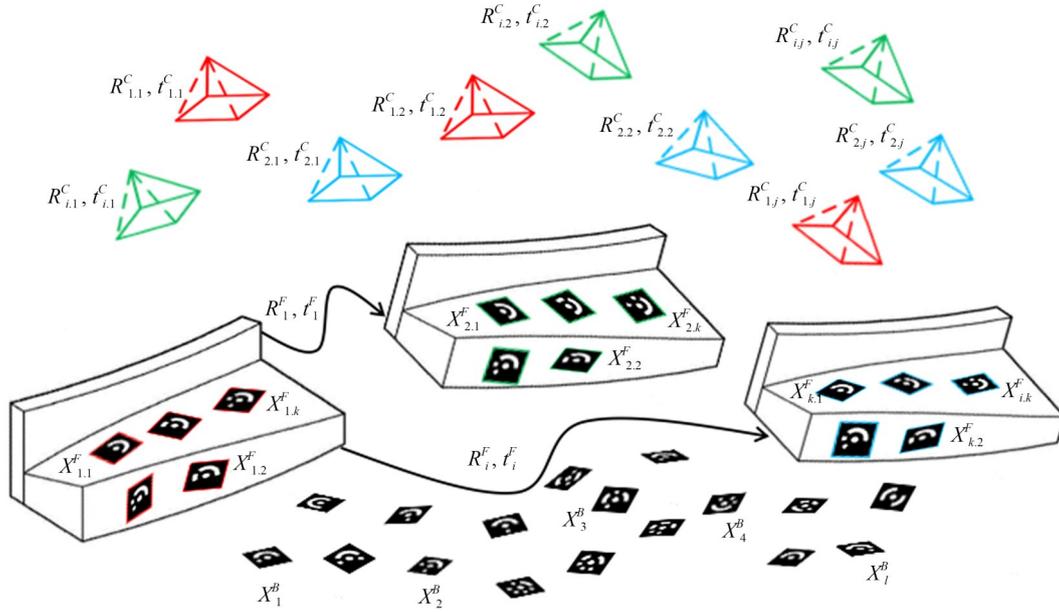


图 2 本文前景运动估计系统及符号标记示意图

Fig. 2 Schematic diagram of our proposed foreground motion estimation system and symbolic representation

别介绍上述四类误差项的定义:

$$E_{2D}^B(R_{i,j}^C, t_{i,j}^C, X_l^B) = \sum_i \sum_j \sum_l \delta_{i,j,l}^B \left\| \frac{p_{i,j,l}^B - \gamma(K, R_{i,j}^C, t_{i,j}^C, X_l^B)}{\gamma(K, R_{i,j}^C, t_{i,j}^C, X_l^B)} \right\|_{L_2}^2, \quad (10)$$

其中: $\delta_{i,j,l}^B$ 为可见标记, $\delta_{i,j,l}^B = 1$ 表示背景标记点 X_l^B 在第 i 组第 j 个相机可见, 否则 $\delta_{i,j,l}^B = 0$; $p_{i,j,l}^B$ 表示标记点 X_l^B 在第 i 组第 j 幅图像中的成像椭圆中心坐标, γ 表示投影方程, K 表示相机内参矩阵;

$$E_{2D}^F(R_{i,j}^C, t_{i,j}^C, X_{i,k}^F) = \sum_i \sum_j \sum_k \delta_{i,j,k}^{F-2D} \left\| \frac{p_{i,j,k}^F - \gamma(K, R_{i,j}^C, t_{i,j}^C, X_{i,k}^F)}{\gamma(K, R_{i,j}^C, t_{i,j}^C, X_{i,k}^F)} \right\|_{L_2}^2, \quad (11)$$

上式中符号与式(10)较为类似, 在此不再赘述, 仅说明一点, 前景标记点在进行投影时仅投影至同组相机, 即 $R_{i,j}^C, t_{i,j}^C, X_{i,k}^F$ 三个变量下标的第一个元素(相机组别 i)相同;

$$E_{3D}^F(X_{i,k}^F, R_i^F, t_i^F) = \sum_i \sum_j \sum_k \delta_{i,j,k}^{F-3D} \left\| \begin{pmatrix} (R_i^F)^T (X_{i,k}^F - t_i^F) \\ (R_j^F)^T (X_{j,k}^F - t_j^F) \end{pmatrix} \right\|_{L_2}^2, \quad (12)$$

其中, $\delta_{i,j,k}^{F-3D}$ 为重建标记, $\delta_{i,j,k}^{F-3D} = 1$ 表示第 k 个前

景标记点在第 i 次与第 j 次前景运动中均通过三角测量获得了三维坐标, 否则 $\delta_{i,j,k}^{F-3D} = 0$;

$$E_{3D}^S(X_l^B) = \sum_l \sum_m \delta_{l,m}^S \left(d_{l,m} - \|X_l^B - X_m^B\|_{L_2} \right)^2, \quad (13)$$

其中, $\delta_{l,m}^S$ 为标尺标记, $\delta_{l,m}^S = 1$ 表示第 l 个前景标记点与第 m 个前景标记点之间存在标尺, 否则 $\delta_{l,m}^S = 0$, $d_{l,m}$ 表示标尺的实际长度。

基于上述介绍, 采用 Gao 等人提出的广义光束平差方法^[25], 对式(5)进行优化求解, 同时最小化其中的反投影误差以及空间误差, 以获取前景运动的最终估计结果 $\{R_i^{F*}, t_i^{F*}\}$ 。

3 实验结果

为验证本文提出的基于编码标记点的前景运动估计方法的有效性, 现对其中的一个关键步骤编码标记点与相机分组, 以及前景运动估计的最终结果进行实验验证。本文采用的图像采集设备为 Apple iPhone 7, 镜头焦距为 4 mm, 采集的图像分辨率为 4032×3024 。

3.1 编码标记点与相机分组实验

为验证本文提出的编码标记点与相机分组方法的有效性, 将本文方法与 2.3 节介绍的朴素

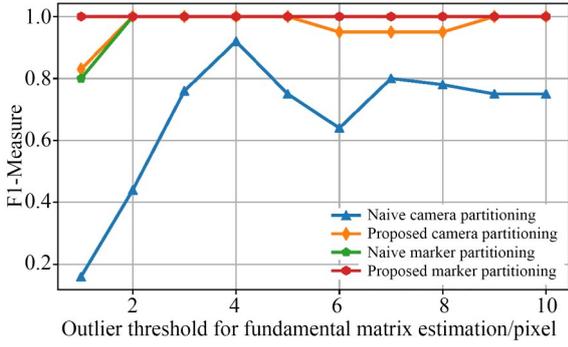


图3 编码标记点与相机分组对比实验结果

Fig. 3 Comparative results on coded marker and camera partitioning

分组方法进行对比。在进行对比时,为评价分组效果,需要先确定评价指标。在此,本文采用 F1-Measure 进行分组效果评价。具体来说,假设将人工标记的分组结果(真值)与某待评测方法的分组结果分别记为 $\{P_j | j=1, 2, \dots, N\}$ 与 $\{C_i | i=1, 2, \dots, M\}$, 其中, P_j 与 C_i 分别表示人工标记的分组结果的第 j 组与待评测方法的分组结果的第 i 组元素集合。首先,先根据 $\{P_j\}$ 与 $\{C_i\}$ 计算 F1-Measure 矩阵:

$$F_1 = \begin{pmatrix} F_1(P_1, C_1) & \dots & F_1(P_1, C_M) \\ \vdots & F_1(P_j, C_i) & \vdots \\ F_1(P_N, C_1) & \dots & F_1(P_N, C_M) \end{pmatrix}, \quad (14)$$

F1-Measure 矩阵的第 j 行,第 i 列的元素计算方式为:

$$F_1(P_j, C_i) = \frac{2P(P_j, C_i)R(P_j, C_i)}{P(P_j, C_i) + R(P_j, C_i)}, P(P_j, C_i) = \frac{|P_j \cap C_i|}{|C_i|}, R(P_j, C_i) = \frac{|P_j \cap C_i|}{|P_j|}.$$

在此基础上,即可获取待

评测方法 F1-Measure 的值: $F_1 = \sum_{j=1}^N \omega_j F_1(j)$, 其

$$\text{中, } \omega_j = \frac{|P_j|}{\sum_{k=1}^N |P_k|}, F_1(j) = \max_{1 \leq i \leq M} F_1(P_j, C_i).$$

对于本文分组方法与朴素分组方法来说,给定不同的基本矩阵估计外点阈值,可获得不同的分组结果。根据上述分组评价指标,即可对两种方法进行分组效果对比,结果如图 3 所示。

由图可知,对于标记点分组,两种方法效果接近其都较为准确,而对于相机分组而言,本文提出的基于图聚类的分组方法效果明显优于朴素分组方法。另外,随着基本矩阵估计的外点阈值的变化,本文提出的相机分组方法效果变化不明显而朴素分组方法效果变化剧烈,说明本文方法对外点阈值变化更加不敏感,鲁棒性更强。在实际测量过程中,用于基于 RANSAC 的基本矩阵估计的外点阈值取为 4 像素。

3.2 前景运动估计精度对比实验

为对本文提出的前景运动估计方法的精度进行评价,首先需要确定评价指标。在此本文借助坐标测量机 (Coordinate Measuring Machine, CMM) 提供位置真值进行精度评价。具体来说,在坐标测量机静止基座(背景)与移动末端(前景)分别布置若干标记点,操作末端移动,拍摄若干幅图像如图 4 所示,并记录其各次移动后的空间坐标作为前景位置真值 $\{c_i^{GT}\}$ 。另外,通过本文方法估计的前景位置为 $\{c_i^F | c_i^F = -(R_i^F)^T t_i^F\}$ 。采用 Arun 等人^[22]提出的方法估计前景位置真值与估计值之间的相对旋转、平移,并记为 $\{R, t\}$, 则各次前景运动的估计误差计算如下: $e_i = \|c_i^{GT} - R c_i^F - t\|_{L_2}$ 。在实际实验验证时,本文采



图4 前景运动估计精度评价实验示例图像

Fig. 4 Example images of foreground motion estimation accuracy evaluation experiment

用的实验配置为背景标记点共 28 个,前景标记点共 8 个。另外,为兼顾实验效果与测试效率,本文在进行实验时共操作 CMM 进行了 4 次前景运动,并通过图像采集设备共采集了 40 幅图像,各次前景运动采集的图像数分别为 9, 10, 10, 11。

基于上述评价指标与实验数据,我们首先验证了本文提出的基于 RANSAC 与 Polar-N-Direction 距离的标记点成像椭圆拟合方法的有效

性。具体来说,我们将采用此方法拟合椭圆与基于最小二乘拟合椭圆^[9-10]分别得到的椭圆成像中心坐标用于进行前景运动估计,通过比较最终前景运动估计的精度来衡量椭圆拟合的效果,结果如表 1 所示。由表 1 可知,采用本文方法进行椭圆拟合获取的椭圆中心精度高于基于最小二乘的椭圆拟合方式,能够实现更为精确的前景运动估计。

表 1 标记点提取精度对比实验结果

Tab. 1 Comparative experiment results on coded maker extraction accuracy

(mm)

Foreground motion estimation error	1st	2nd	3rd	4th	average	standard deviation
Least squares-based	0.570	0.648	0.511	0.414	0.535	0.097
RANSAC-based (ours)	0.287	0.286	0.335	0.231	0.285	0.043

另外,为验证本文采用的相机模型(见 2.4 节)的有效性,基于上述评价指标与实验数据,进行了优化不同内参数时前景运动估计精度对比实验,结果如表 2 所示。由表 2 可知,本文采用的相机模型(固定主点,优化焦距、畸变)精度最高,对切向

畸变系数以及主点坐标进行优化可以分别提升及降低前景运动估计精度。因此,对于本文选用的图像采集设备及其采集的具体测试数据,我们最终选择优化的相机内参为包括焦距、径向、切向畸变在内的七参数模型: $\{f_x, f_y, k_1, k_2, k_3, p_1, p_2\}$ 。

表 2 内参优化对比实验结果

Tab. 2 Comparative experiment results on optimized intrinsic parameters

(mm)

Foreground motion estimation error	1st	2nd	3rd	4th	average	standard deviation
f_x, f_y, k_1, k_2, k_3	0.356	0.330	0.461	0.304	0.363	0.069
$f_x, f_y, k_1, k_2, k_3, u_0, v_0$	0.477	0.501	0.450	0.381	0.452	0.052
$f_x, f_y, k_1, k_2, k_3, p_1, p_2, u_0, v_0$	0.349	0.336	0.427	0.295	0.352	0.055
$f_x, f_y, k_1, k_2, k_3, p_1, p_2$ (ours)	0.287	0.286	0.335	0.231	0.285	0.043

为验证本文提出的前景运动估计方法的有效性,基于上述评价指标与实验数据,在此进行了前景运动估计精度对比实现,将本文方法(proposed method)与(1)无全局优化的前景运动估计方法(W/O Global Optimization), (2)朴素的前景运动估计方法(Naive Method), (3)以及基于场景本身纹理特征的前景运动估计方法(Texture-Based Method)进行对比。对于无全局优化方法,直接将 2.5 节通过图优化获取的前景运动初值作为前景运动的最终估计结果。对于该朴素方法,在采用本文 2.3 节提出的方法对标记点与相机进行自动分组之后,对于同属一组的相机而

言,所有标记点均可看作背景点,因此可以采用本文 2.4 节的方法计算各组内部的相机位

姿与标记点坐标,然后根据各组计算得到的实际背景点的坐标将各组测量结果统一至同一基准之下,在此基础上实现前景运动估计。对于基于纹理的方法,我们将采集的图像导入至当前最具代表性的基于图像纹理特征的三维建模系统 COLMAP 中,获取测试场景的稠密点云,结果如图 5 所示。由图 5 可知,背景固定场景以及 4 次前景运动场景均可成功重建。在此基础上先根据标尺将进行场景尺度恢复,然后人为划分出不同次前景运动对应的三维点云并采用带颜色的

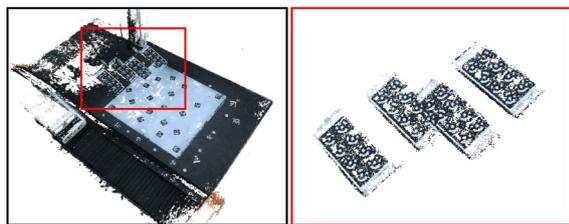


图 5 测试场景 COLMAP 稠密重建结果

Fig. 5 Dense reconstruction result on the test scene by COLMAP

ICP算法^[26]估计各次前景运动。

本文方法与两种对比方法的对比实验结果如表 3 所示。由表 3 可知,相对于基于图优化的前景运动初始化结果,基于广义光束平差的全局优化能够获得更为精确的前景运动估计,且本文提出的前景运动估计方法在精度与稳定性方面均高于朴素方法与基于纹理的方法,估计误差的均值小于 0.3 mm,可以满足大多数场景的测量精度要求。

表 3 前景运动估计精度对比实验结果

Tab. 3 Comparative experiment results on foreground motion estimation accuracy (mm)

Foreground motion estimation error	1st	2nd	3rd	4th	average	standard deviation
w/o global optimization	0.696	0.702	0.892	0.728	0.755	0.093
Naive method	0.344	0.419	0.336	0.290	0.347	0.053
Texture-based method	0.844	0.850	1.103	0.875	0.918	0.124
Proposed method	0.287	0.286	0.335	0.231	0.285	0.043

4 结 论

本文提出了一种基于编码标记点的高精度运动估计方法。该方法通过在测量环境中前背景上粘贴环状编码标记点,并在每次前景运动后采集若干图像,之后同时估计所有相机位姿以及背景固定标记点、前景运动标记点空间坐标,实现前景运动的高精度估计。针对上述目标,提出

了一套高精度运动估计流程,包括编码标记点的提取与识别、编码标记点与相机的分组、背景点与相机位姿初始化、前景点与前景运动初始化、全局优化五个步骤。实验结果表明,本文提出的运动估计方法精度在 0.3 mm 左右,可以满足高精度前景运动估计的需求。另外,本文通过实验验证了提出的编码标记点与相机的分组方法以及运动估计整体流程的有效性。

参考文献:

- [1] SULTANAL M, MAHMOOD A, JAVED S, *et al.* Unsupervised deep context prediction for background estimation and foreground segmentation[J]. *Machine Vision Applications*, 2019, 30: 375 - 395.
- [2] 王向军, 张继龙, 阴雷. 光流法运动估计在 FPGA 上的实现与性能分析[J]. *光学精密工程*, 2019, 27(1): 211-220.
WANG X J, ZHANG J L, YIN L. Implementation and performance analysis of optical flow motion estimation on FPGA[J]. *Opt. Precision Eng.*, 2019, 27(1): 211-220. (in Chinese)
- [3] YU Y C, XIONG Y L, HUANG W L, *et al.* Deformable Siamese attention networks for visual object tracking[C]. *In IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2020: 6728-6737.
- [4] DANELLJAN M, GOOL L V, TIMOFTE R. Probabilistic regression for visual tracking[C]. *In IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2020: 7183-7192.
- [5] SONG C, SONG J R, HUANG Q X. Hybrid-Pose: 6D object pose estimation under hybrid representations[C]. *In IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2020: 431-440.
- [6] CHEN W, JIA X, CHANG H J, *et al.* G2L-Net: Global to local network for real-time 6D pose estimation

- tion with embedding vector features [C]. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2020: 4233-4242.
- [7] LI Y, ZHANG T W, NAKAMURA Y, *et al.* SplitFusion: Simultaneous tracking and mapping for non-rigid scenes [C]. In *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2020.
- [8] BOLYA D, ZHOU C, XIAO F Y. YOLACT: Real-time instance segmentation [C]. In *International Conference on Computer Vision (ICCV)*, 2019: 9157-9166.
- [9] 解则晓, 高翔, 朱瑞新. 环状编码标记点的高效提取与鲁棒识别算法 [J]. *光电子·激光*, 2015, 26(3): 559-566.
- XIE Z X, GAO X, ZHU R X. Efficient extraction and robust recognition algorithm of circular coded target [J]. *Journal of Optoelectronics·Laser*, 2015, 26(3): 559-566. (in Chinese)
- [10] 解则晓, 高翔, 崔健. 移动式三维测量用圆形标记点提取算法 [J]. *中国激光*, 2013, 40(12): 1208002.
- XIE Z X, GAO X, CUI J. Extraction algorithm of circular targets used for mobile three-dimensional measurement [J]. *Chinese Journal of Lasers*, 2013, 40(12): 1208002. (in Chinese)
- [11] GHOSAL S, MEHROTRA R. Orthogonal moment operators for subpixel edge detection [J]. *Pattern Recognition*, 1993, 26(2): 295-306.
- [12] FISCHLER M A, BOLLES R C. Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography [J]. *Communications of the ACM*, 1981, 24(6): 381-395.
- [13] 解则晓, 王晓敏. 平面标定靶标标记点的圆心提取 [J]. *光学精密工程*, 2019, 27(2): 440-449.
- XIE Z X, WANG X M. Extraction of the center of the mark point of the plane calibration target [J]. *Opt. Precision Eng.*, 2019, 27(2): 440-449. (in Chinese)
- [14] WU Y H, WANG H R, TANG F L, *et al.* Efficient conic fitting with an analytical Polar-N-Direction geometric distance [J]. *Pattern Recognition*, 2019, 90: 415-423.
- [15] HARTLEY, R I. In defense of the eight-point algorithm [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1997, 19(6): 580-593.
- [16] DHILLON I S, GUAN Y Q, KULIS B. Weighted graph cuts without eigenvectors: A multilevel approach [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2007, 29(11): 1944-1957.
- [17] CUI Z P, TAN P. Global structure-from-motion by similarity averaging [C]. In *International Conference on Computer Vision (ICCV)*, 2015: 864-872.
- [18] SCHÖNBERGER J L, FRAHM J M. Structure-from-motion revisited [C]. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016: 4104-4113.
- [19] Nistér D. An efficient solution to the five-point relative pose problem [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2004, 26(6): 756-770.
- [20] HARTLEY R I, ZISSERMAN A. *Multiple View Geometry in Computer Vision* [M]. Cambridge: Cambridge University Press, 2004.
- [21] GAO X S, HOU X R, TANG J L, *et al.* Complete solution classification for the perspective-three-point problem [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2003, 25(8): 930-943.
- [22] AGARWAL S, SNAVELY N, SEITZ S M, *et al.* Bundle adjustment in the large [C]. In *European Conference on Computer Vision (ECCV)*, 2010: 29-42.
- [23] ARUN K S, HUANG T S, BLOSTEIN S D. Least-squares fitting of two 3-D point sets [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1987, 9(5): 698-700.
- [24] CARLONE L, TRON R, DANILIDIS K, *et al.* Initialization techniques for 3D SLAM: A survey on rotation estimation and its use in pose graph optimization [C]. In *IEEE International Conference on Robotics and Automation (ICRA)*, 2015: 4597-4604.
- [25] GAO X, SHEN S H, ZHU L J, *et al.* Complete

scene reconstruction by merging images and laser scans[J]. *IEEE Transactions on Circuits and Systems for Video Technology*, 2020, 30(10): 3688-3701.

[26] PARK J, ZHOU Q, KOLTUN V. Colored point cloud registration revisited[C]. *In IEEE International Conference on Computer Vision (ICCV)*, 2017: 143-152.

作者简介:



迟书凯(1980—),男,山东烟台人,博士,高级工程师,2002年于青岛海洋大学获得学士学位,2007年于中国海洋大学获得硕士学位,2017年于中国海洋大学获得博士学位,现为中国海洋大学自动化教学实验中心主任,主要从事水下机器视觉、智能仪器仪表、水下无人系统方面的研究。E-mail:chishukai@ouc.edu.cn



解则晓(1968-),男,山东临沂人,博士,教授,博士生导师,1991年于青岛大学获得学士学位,1997年于大连理工大学获得硕士学位,2000年于天津大学获得博士学位,2000年至2002年从事博士后科研工作,现为中国海洋大学自动化及测控系教授,主要从事机器视觉与视觉测量、逆向工程、水下三维探测技术方面的研究。E-mail:xiezexiao@ouc.edu.cn

通讯作者:



高翔(1989—),男,山东临沂人,博士,讲师,2012年于中国海洋大学获得学士学位,2015年于中国海洋大学获得硕士学位,2019年于中国科学院大学获得博士学位,现为中国海洋大学工程学院讲师,主要从事三维计算机视觉、基于图像的大规模场景三维重建等方面的研究。E-mail:xgao@ouc.edu.cn