

5-15-2020

## Non-contact Object Size Acquisition Based On Visual-Inertial Sensor Fusion

Kang Lai

*College of Systems Engineering, National University of Defense Technology, Changsha 410073, China;*

Yingmei Wei

*College of Systems Engineering, National University of Defense Technology, Changsha 410073, China;*

Jiang Jie

*College of Systems Engineering, National University of Defense Technology, Changsha 410073, China;*

Yuxiang Xie

*College of Systems Engineering, National University of Defense Technology, Changsha 410073, China;*

Follow this and additional works at: <https://dc-china-simulation.researchcommons.org/journal>



Part of the [Artificial Intelligence and Robotics Commons](#), [Computer Engineering Commons](#), [Numerical Analysis and Scientific Computing Commons](#), [Operations Research, Systems Engineering and Industrial Engineering Commons](#), and the [Systems Science Commons](#)

---

This Paper is brought to you for free and open access by Journal of System Simulation. It has been accepted for inclusion in Journal of System Simulation by an authorized editor of Journal of System Simulation.

---

## Non-contact Object Size Acquisition Based On Visual-Inertial Sensor Fusion

### Abstract

**Abstract:** In consideration of the visual data and the data from the inertial measurement unit (IMU) having the different characteristics, *the 6D pose of the visual sensor is recovered from the captured video by using the vision-based 3D reconstruction techniques, the obtained pose information and inertial measurements is fused via the spatial-temporal alignment, and based on the determined scale of the visual 3D reconstruction the non-contact measurement of a scene is performed.* The proposed approach achieves a relative measurement error of about 3% in the non-contact object size estimation experiment. As it does not require the additional calibration object or any special motion capture devices, the proposed method is flexible and is applicable in many potential areas.

### Keywords

visual-inertial sensing data, 3D reconstruction, spatial-temporal alignment, non-contact object size acquisition

### Recommended Citation

Kang Lai, Wei Yingmei, Jiang Jie, Xie Yuxiang. Non-contact Object Size Acquisition Based On Visual-Inertial Sensor Fusion[J]. Journal of System Simulation, 2020, 32(5): 892-900.

## 融合视惯传感数据的非接触式物体尺寸获取方法

康来, 魏迎梅, 蒋杰, 谢毓湘

(国防科技大学系统工程学院, 湖南 长沙 410073)

**摘要:** 从视觉数据和惯性测量单元(Inertial Measurement Unit, IMU)数据各自的特点出发, 利用计算机视觉三维重建技术从视觉数据中恢复视觉传感器的六维位姿信息, 分析计算惯性测量数据和视觉位姿数据的时空关系, 实现融合视惯传感数据的场景重建及尺度估计, 进而获取待测对象的绝对物理尺寸数据。上述方法在实验数据上取得了相对误差约为 3% 的非接触式物体尺寸估计结果, 由于该方法无需在待测场景中嵌入额外标定物或辅助定位装置, 具有较好的灵活性和适用范围。

**关键词:** 视惯传感数据; 三维重建; 时空配准; 非接触式物体尺寸获取

中图分类号: TP391

文献标识码: A

文章编号: 1004-731X (2020) 05-0892-09

DOI: 10.16182/j.issn1004731x.joss.19-0432

## Non-contact Object Size Acquisition Based On Visual-Inertial Sensor Fusion

Kang Lai, Wei Yingmei, Jiang Jie, Xie Yuxiang

(College of Systems Engineering, National University of Defense Technology, Changsha 410073, China)

**Abstract:** In consideration of the visual data and the data from the inertial measurement unit (IMU) having the different characteristics, the 6D pose of the visual sensor is recovered from the captured video by using the vision-based 3D reconstruction techniques, the obtained pose information and inertial measurements is fused via the spatial-temporal alignment, and based on the determined scale of the visual 3D reconstruction the non-contact measurement of a scene is performed. The proposed approach achieves a relative measurement error of about 3% in the non-contact object size estimation experiment. As it does not require the additional calibration object or any special motion capture devices, the proposed method is flexible and is applicable in many potential areas.

**Keywords:** visual-inertial sensing data; 3D reconstruction; spatial-temporal alignment; non-contact object size acquisition

## 引言

物体物理三维尺寸数据获取方法大致可分为接触式测量方法和非接触式测量方法 2 类<sup>[1]</sup>。一种典型的接触式测量工具是三坐标测量机<sup>[2]</sup>, 这种测量工具上安装有通过多自由度伺服装置控制的结

构测量探针, 通过伺服机械移动量获取探针基准点和测量点之间的距离。接触式测量具有较高的精度, 可以达到纳米级<sup>[3]</sup>, 但这种测量方式对待测物体的大小和形状有较大限制且测量效率较低, 此外接触式测量工具成本也较高。

非接触式测量在不接触待测对象的情况下获得对象的三维尺寸数据。非接触测量又分为主动式测量和被动式测量两类<sup>[1]</sup>。主动式测量向被测物体发射特定信号(如: 可见光、激光、电磁波、声波等), 通过对返回信号进行分析来计算待测物体表面的三



收稿日期: 2019-08-22 修回日期: 2020-02-17;  
基金项目: 国家自然科学基金面上项目(61873274);  
作者简介: 康来(1983-), 男, 四川安岳, 博士, 副教授, 研究方向为计算机视觉与模式识别、虚拟现实技术; 魏迎梅(1972-), 女, 陕西兰州, 博士, 教授, 研究方向为虚拟现实技术、数据可视化。

<http://www.china-simulation.com>

维空间位置进而完成三维测量。被动式测量被动接收待测物体所反射的环境光, 通过分析待测物体的二维图像获得物体的三维尺寸数据。根据拍摄视角数量不同, 被动非接触式测量包括单目测量法、多视角测量法两类。其中, 单目测量法只需一台相机就可完成, 但通常需要借助先验知识才能实现三维重建, 而且精度不高<sup>[4]</sup>。多视角测量法通过在不同(至少两个)视角对待测物体进行拍摄, 然后利用多视图约束关系恢复场景的三维结构实现测量。近年来, 计算机视觉领域提出了许多基于图像的多视图三维重建方法, 比如: Bundler<sup>[5]</sup>、COLMAP<sup>[6]</sup>、PMVS<sup>[7]</sup>。被动非接触式测量方法对待测物体形状和尺寸限制较少, 且具有数据采集便捷、测量成本低、测量效率高等优势, 但如果仅仅依靠视觉数据, 被动非接触式测量方法仅能获取待测对象的相对尺寸。通常情况下, 被动非接触式测量方法主要采用在待测场景中嵌入已知尺寸标定物或通过辅助装置控制、记录相机位移两种手段来获得绝对物理尺寸信息。然而, 这两种方式均会降低非接触式测量的灵活性, 减小非接触式测量的适用范围。

本文提出一种融合视觉数据和惯性测量数据(简称为视惯传感数据)的非接触式物体尺寸获取方法。综合利用视觉数据和惯性测量单元(Inertial Measurement Unit, IMU)数据来恢复场景的物理尺度信息的思想来源于当前流行的即时定位和构图方法(Simultaneous Localization and Mapping, SLAM)<sup>[8-10]</sup>。此外, 本文描述的三维测量流程能处理更加一般化的场景, 且不需要在待测场景中嵌入额外标定物或添加额外辅助定位装置, 具有使用便捷灵活、适用范围广的特点。

## 1 基本流程

本文方法以视觉传感器数据和 IMU 数据为基础, 这两类数据采集传感器广泛存在于手机、平板等小型化智能设备。其中, 视觉数据为使用摄像头所采集的二维可见光视频数据, IMU 中的陀螺仪和加速度计可以测量 3 个轴的角速率和线性加速度。

视频数据通常具有较高的分辨率高但帧率相对较低, 且不包含物理尺寸信息; IMU 数据采集频率高且具有物理尺度信息, 但测量结果容易产生累积误差, 因此视觉数据和 IMU 数据各有优势和不足。本文从上述两类数据的特点出发, 通过视觉三维重建、数据时空配准等方法将视觉数据和 IMU 数据进行融合, 构建具有物理尺度信息的三维场景, 从而实现非接触式物体尺寸估计。本文考虑如下一般场景: 数据采集期间 IMU 与可见光摄像元件之间的相对位置固定, IMU 数据与视频数据无需时钟同步, 仅需确保 IMU 数据覆盖整个视频时间范围。

本文非接触式物体尺寸获取方法的基本流程如图 1 所示。

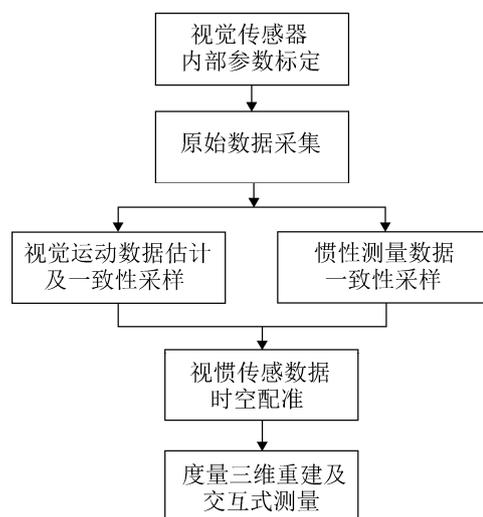


图 1 融合视惯传感数据的非接触式物体尺寸获取流程  
Fig. 1 Visual-inertial data fusion framework for non-contact measurement

各步骤的主要内容如下: 视觉传感器标定用于获取摄像装置的焦距、形变参数、主点位置等内部参数<sup>[11]</sup>。原始数据采集阶段获取视觉传感器在不同位置环绕被测物体所拍摄的影像数据, 同时采集拍摄过程中 IMU 的惯性测量数据。惯性测量数据包括三轴惯性测量角速率和三轴惯性测量线性加速度。重建影像数据的各视频帧相对于视觉传感器的位姿信息, 在此基础上计算视点移动的角速率和线性加速度, 本文称之为视觉角速率和视觉线性加速度。对惯性测量数据和计算出来的视觉位姿变化数据分

别进行插值和尺度变换, 获得一致性重采样数据。在此基础上, 将视觉运动数据和惯性测量运动数据进行时空配准, 并计算视点变化的物理尺度信息。根据上述信息可以实现场景的稠密三维点云重建, 并量算被测物体任意两点之间的物理距离。

本文后续内容重点介绍视觉运动信息估计、运动数据一致性采样、视惯传感数据的时空配准方法以及度量重建和距离测量方法。

## 2 主要方法

惯性测量数据可以通过 IMU 直接获取, 包括角速率和线性加速度数据, 而视觉传感器仅能捕获场景的视觉信息。本文从视频帧序列中恢复相机的位姿信息, 然后将其转换为角速率和线性加速度, 并通过与 IMU 采集的数据进行融合获取场景尺度信息。为了便于叙述, 本文将涉及到的 4 类运动数据分别称为惯性测量角速率、惯性测量线性加速度、视觉角速率、视觉线性加速度。记  $M$  为惯性测量原始数据的数据采样点数量,  $N$  为视频帧数量, IMU 3 个轴分别为  $X_{imu}$ ,  $Y_{imu}$ ,  $Z_{imu}$ , 视觉传感器的 3 个坐标轴  $X_{cam}$ ,  $Y_{cam}$ ,  $Z_{cam}$ 。关于  $X_{imu}$ ,  $Y_{imu}$ ,  $Z_{imu}$  的惯性测量角速率分别为  $\alpha^{imu}$ ,  $\beta^{imu}$ ,  $\gamma^{imu}$ , 惯性测量线性加速度分别为  $a^{imu}$ ,  $b^{imu}$ ,  $c^{imu}$ , 关于  $X_{cam}$ ,  $Y_{cam}$ ,  $Z_{cam}$  的视觉角速率分别为  $\alpha^{cam}$ ,  $\beta^{cam}$ ,  $\gamma^{cam}$ , 视觉线性加速度分别为  $a^{cam}$ ,  $b^{cam}$ ,  $c^{cam}$ 。

### 2.1 视觉角速率和线性加速度估计

本文采用运动推断结构(Structure from Motion, SfM)<sup>[5][6]</sup>技术计算各视频帧所对应相机视点的六维位姿(即位置和姿态)信息, 其中姿态指的是摄像头坐标系下各视点的三维旋转参数。本文采用  $\mathbf{x} = (x, y, z)^T$  表示视点在三维空间中的位置, 采用单位四元数  $\mathbf{q} = (q_0, q_1, q_2, q_3)^T$  表示某帧图像对应的旋转变量<sup>[11]</sup>。角速率描述旋转参数在各坐标轴上的变化情况。为了获取角速率, 首先将四元数转换为关于坐标轴  $X_{cam}$ ,  $Y_{cam}$ ,  $Z_{cam}$  的离散角度数据  $\phi^{cam}$ 、 $\theta^{cam}$ 、 $\varphi^{cam}$ , 计算公式如式(1):

$$\begin{cases} \phi^{cam} = \arctan\left(\frac{2(q_0q_1 + q_2q_3)}{q_0^2 - q_1^2 - q_2^2 + q_3^2}\right) \\ \theta^{cam} = \arcsin(2(q_0q_2 - q_1q_3)) \\ \varphi^{cam} = \arctan\left(\frac{2(q_0q_3 + q_1q_2)}{q_0^2 + q_1^2 - q_2^2 - q_3^2}\right) \end{cases} \quad (1)$$

通过对各视频帧对应的角度数据序列  $\{\phi_j^{cam}\}$  ( $j=1, 2, \dots, N$ ) 求导可以计算关于坐标轴  $X_{cam}$  的视觉角速率  $\{\alpha_j^{cam}\}$ 。采用类似的方法可以计算关于坐标轴  $Y_{cam}$ ,  $Z_{cam}$  的角速率  $\{\beta_j^{cam}\}$  和  $\{\gamma_j^{cam}\}$ 。通过对各视频帧对应的视点位置数据分量求导计算关于各坐标轴的视觉线性速度, 对该视觉线性速度序列求导即可获得相应的线性加速度。比如: 对关于坐标轴  $X_{cam}$  的位置分量数据  $\{x_j^{cam}\}$  求导可以计算视觉线性速度序列  $\{v_j^{cam}\}$ , 然后通过对序列  $\{v_j^{cam}\}$  求导计算坐标轴  $X_{cam}$  的视觉线性加速度序列  $\{a_j^{cam}\}$ 。采用类似的方法计算关于坐标轴  $Y_{cam}$ ,  $Z_{cam}$  的视觉线性加速度序列  $\{b_j^{cam}\}$  和  $\{c_j^{cam}\}$ 。

### 2.2 运动数据一致性重采样

由于视觉传感器和 IMU 数据采集频率不一致, 且视觉重建位置数据没有尺度信息(因此视觉线性加速度只有相对大小, 但角速率有绝对大小), 不利于对二者进行时空融合。因此需要对惯性测量数据和视觉运动数据进行插值和拉伸, 实现运动数据的一致性采样。重采样后的视觉运动数据和惯性测量数据的采样密度在时间上应保持一致。本文将视觉运动数据和惯性测量数据重采样点数量分别为  $rN/r_{cam}$  和  $rM/r_{imu}$ 。其中,  $r_{cam}$  为视频数据帧率,  $r_{imu}$  为 IMU 数据帧率,  $r$  为  $r_{cam}$  和  $r_{imu}$  的最小公倍数。具体而言, 重采样视觉角速率  $\alpha^{camr}$ ,  $\beta^{camr}$  和  $\gamma^{camr}$  与原始视觉角速率  $\alpha^{cam}$ ,  $\beta^{cam}$ ,  $\gamma^{cam}$  的关系为:

$$\begin{cases} \alpha_k^{camr} = r_{cam} \left( \alpha_{w_1(k)}^{cam} + w_3(k) \left( \alpha_{w_2(k)}^{cam} - \alpha_{w_1(k)}^{cam} \right) \right) \\ \beta_k^{camr} = r_{cam} \left( \beta_{w_1(k)}^{cam} + w_3(k) \left( \beta_{w_2(k)}^{cam} - \beta_{w_1(k)}^{cam} \right) \right) \\ \gamma_k^{camr} = r_{cam} \left( \gamma_{w_1(k)}^{cam} + w_3(k) \left( \gamma_{w_2(k)}^{cam} - \gamma_{w_1(k)}^{cam} \right) \right) \end{cases} \quad (2)$$

式中:  $k \in \{1, 2, \dots, rN/r_{\text{cam}}\}$ , 且

$$\begin{cases} w_1(k) = \left[ \frac{(kN - k - N)r_{\text{cam}} + rN}{rN - r_{\text{cam}}} \right] \\ w_2(k) = \left[ \frac{(kN - k - N)r_{\text{cam}} + rN}{rN - r_{\text{cam}}} \right] \\ w_3(k) = \frac{(kN - k - N)r_{\text{cam}} + rN}{rN - r_{\text{cam}}} - w_1(k) \end{cases} \quad (3)$$

注意, 公式(2)中将视频数据帧率  $r_{\text{cam}}$  作为拉伸系数, 其原因是原始视觉角速率通过求导计算, 因此反应的是两帧之间的差异, 而不是时间上的角速率, 故需转换为以弧度/s 为统一单位。类似的, 重采样视觉线性加速度  $a^{\text{camr}}$ ,  $b^{\text{camr}}$  和  $c^{\text{camr}}$  与原始视觉角速率  $a^{\text{cam}}$ ,  $b^{\text{cam}}$ ,  $c^{\text{cam}}$  的关系为:

$$\begin{cases} a_k^{\text{camr}} = r_{\text{cam}}^2 \left( a_{w_1(k)}^{\text{cam}} + w_3(k) \left( a_{w_2(k)}^{\text{cam}} - a_{w_1(k)}^{\text{cam}} \right) \right) \\ b_k^{\text{camr}} = r_{\text{cam}}^2 \left( b_{w_1(k)}^{\text{cam}} + w_3(k) \left( b_{w_2(k)}^{\text{cam}} - b_{w_1(k)}^{\text{cam}} \right) \right) \\ c_k^{\text{camr}} = r_{\text{cam}}^2 \left( c_{w_1(k)}^{\text{cam}} + w_3(k) \left( c_{w_2(k)}^{\text{cam}} - c_{w_1(k)}^{\text{cam}} \right) \right) \end{cases} \quad (4)$$

式中:  $r_{\text{cam}}^2$  作为拉伸系数, 因为加速度是通过 2 次求导计算而来。

对于惯性测量数据而言, 由于 IMU 输出的是单位统一的绝对数值, 因此无需施加拉伸系数。具体而言, 重采样惯性测量角速率  $\alpha^{\text{imur}}$ ,  $\beta^{\text{imur}}$  和  $\gamma^{\text{imur}}$  计算如式(5)所示:

$$\begin{cases} \alpha_l^{\text{imur}} = \alpha_{w_4(l)}^{\text{imu}} + w_6(l) \left( \alpha_{w_5(l)}^{\text{imu}} - \alpha_{w_4(l)}^{\text{imu}} \right) \\ \beta_l^{\text{imur}} = \beta_{w_4(l)}^{\text{imu}} + w_6(l) \left( \beta_{w_5(l)}^{\text{imu}} - \beta_{w_4(l)}^{\text{imu}} \right) \\ \gamma_l^{\text{imur}} = \gamma_{w_4(l)}^{\text{imu}} + w_6(l) \left( \gamma_{w_5(l)}^{\text{imu}} - \gamma_{w_4(l)}^{\text{imu}} \right) \end{cases} \quad (5)$$

重采样惯性测量线性加速度  $a^{\text{imur}}$ ,  $b^{\text{imur}}$  和  $c^{\text{imur}}$  计算如式(6)所示:

$$\begin{cases} a_l^{\text{imur}} = \left( a_{w_4(l)}^{\text{imu}} + w_6(l) \left( a_{w_5(l)}^{\text{imu}} - a_{w_4(l)}^{\text{imu}} \right) \right) \\ b_l^{\text{imur}} = \left( b_{w_4(l)}^{\text{imu}} + w_6(l) \left( b_{w_5(l)}^{\text{imu}} - b_{w_4(l)}^{\text{imu}} \right) \right) \\ c_l^{\text{imur}} = \left( c_{w_4(l)}^{\text{imu}} + w_6(l) \left( c_{w_5(l)}^{\text{imu}} - c_{w_4(l)}^{\text{imu}} \right) \right) \end{cases} \quad (6)$$

式中:  $l \in \{1, 2, \dots, rM/r_{\text{imu}}\}$ , 且

$$\begin{cases} w_4(l) = \left[ \frac{(lM - l - M)r_{\text{imu}} + lM}{rM - r_{\text{imu}}} \right] \\ w_5(l) = \left[ \frac{(lM - l - M)r_{\text{imu}} + lM}{rM - r_{\text{imu}}} \right] \\ w_6(l) = \frac{(lM - l - M)r_{\text{imu}} + lM}{rM - r_{\text{imu}}} - w_4(l) \end{cases} \quad (7)$$

### 2.3 视惯运动数据时空配准优化

经过一致性重采样处理之后的运动数据具有相同的采样频率, 但由于数据记录延迟, 视觉数据和 IMU 数据在时间轴上并未精确配准。另外, 根据视觉数据恢复的位置没有绝对尺度信息, 因此视觉线性加速度描述了传感器的相对运动关系, 视觉线性加速度和惯性测量线性加速度之间理论上存在一个常量关联系数。获取该关联系数就能恢复移动的绝对距离, 进而恢复三维场景的绝对尺度。为此, 本质上需要解决视觉运动数据和惯性测量数据在时间上和空间上的配准问题。

本文采用基于角速率的时间配准思路。具体而言, 视觉角速率和惯性测量角速率之间的差异包括时间轴上的偏移量  $N_t$ 、三维旋转偏移量  $\mathbf{R}$  以及角速率固有内部偏差  $\delta_1$ , 时间配准问题可以转化如下优化问题:

$$\{N_t^*, \mathbf{R}^*, \delta_1^*\} = \arg \min_{\{N_t, \mathbf{R}, \delta_1\}} \sum_{k=1}^{k=rN/r_{\text{cam}}} \left\| \mathbf{R} \omega_k^{\text{cam}} - (\omega_{k+N_t}^{\text{imu}} + \delta_1) \right\|_2^2 \quad (8)$$

式中:  $\omega_k^{\text{cam}} = (\alpha_k^{\text{camr}}, \beta_k^{\text{camr}}, \gamma_k^{\text{camr}})^T$  为视觉角速率数值向量序列,  $\omega_k^{\text{imu}} = (\alpha_k^{\text{imur}}, \beta_k^{\text{imur}}, \gamma_k^{\text{imur}})^T$  为惯性测量角速率数值向量序列,  $N_t$  的搜索范围限定为  $[0, rM/r_{\text{imu}} - rN/r_{\text{cam}}]$  (不失一般性, 假设 IMU 数据点多于视觉运动数据点)。本文利用黄金分割搜索方法(任意固定  $N_t$  对应于一个最小二乘优化问题)求解上述问题获取最佳估计  $N_t^*$ ,  $\mathbf{R}^*$ ,  $\delta_1^*$ , 从而完成时间配准。

空间配准即估计视觉线性加速度和惯性测量线性加速度之间的绝对尺度  $\tau$ , 此外本文还考虑了线性加速度固有内部偏差  $\delta_2$ , 将空间配准转化为如下线性最小二乘优化问题:

$$\{\tau^*, \delta_2^*\} = \arg \min_{\{\tau, \delta_2\}} \sum_{k=1}^{k=N/r_{\text{cam}}} \left\| \tau \mathbf{R}^* \mathbf{v}_k^{\text{camr}} - (\mathbf{v}_{k+N_r}^{\text{imu}} + \delta_2) \right\|_2^2 \quad (9)$$

式中:  $\mathbf{v}_k^{\text{camr}} = (a_k^{\text{camr}}, b_k^{\text{camr}}, c_k^{\text{camr}})^T$  为视觉线性加速度数值向量序列,  $\mathbf{v}_k^{\text{imur}} = (a_k^{\text{imur}}, b_k^{\text{imur}}, c_k^{\text{imur}})^T$  为惯性测量线性加速度数值向量序列。本文利用奇异值分解(Singular Value Decomposition, SVD)<sup>[12]</sup>方法求解上述优化问题获取  $\tau^*$ ,  $\delta_2^*$  的最优估计, 从而完成空间配准。

## 2.4 稠密度量三维重建和非接触式物体尺寸获取

完成空间配准之后, 进行度量三维重建。首先, 在获得视觉位姿数据的基础上, 利用基于图像的稠密三维点云重建方法构建待测量物体的稠密三维点云模型。对于视频帧中被测物体上的任意对象点  $A$ 、 $B$ , 在三维稠密点云模型中对应点为  $A'$  和  $B'$ , 根据三维点的坐标可以计算得出  $A'$  和  $B'$  之间的距离  $d^{\text{cam}} = |A'B'|$ , 两个测量点之间的实际物理距离测量结果为:

$$\hat{d} = \tau^* d^{\text{cam}} \quad (10)$$

若真实长度为  $d$ , 则百分比相对误差为  $\frac{100|\tau^* d^{\text{cam}} - d|}{d}$ 。通过在重建物体上进行量算, 可以验证度量重建的精度。

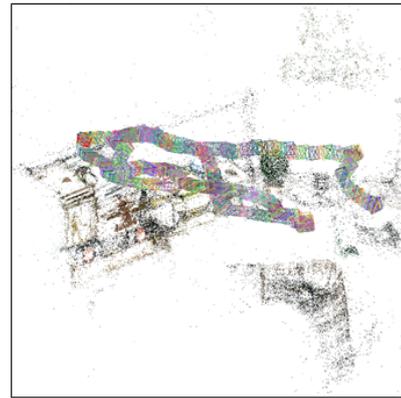
## 3 实验结果及分析

为了验证本文方法的有效性, 我们分别在 Android 操作系统和 Ubuntu 操作系统中实现了原始数据采集程序和数据处理算法。图 2 展示了本实验采集的部分视觉数据和视觉三维重建结果。其中, 图 2(a)为一帧视频数据样例; 图 2(b)~(c)为运动推断结构方法重建的视觉位姿数据以及三维稀疏点云两个视角的可视化结果。本文采用的运动推断结构方法基于 COLMAP<sup>[6]</sup>实现。原始视觉数据和惯性测量数据通过 Xiaomi 5 智能手机中集成的摄像头、加速度计和陀螺仪采集, 智能手机的操作系统为 Android 6.0, 数据采集程序在 Android

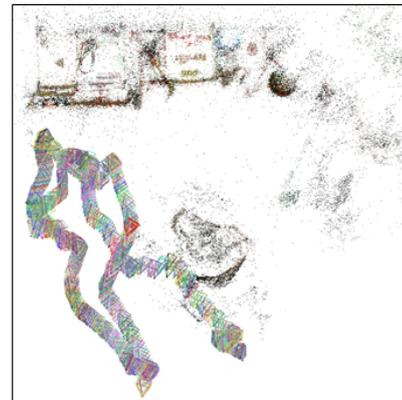
Studio 开发环境中实现, 程序设计语言为 Java。本实验采集的测试数据中影像视频数据的分辨率为 1 280×720 像素, 视频帧率为 30 Hz, 惯性测量数据采样频率为 50 Hz, 数据采集的持续时间约为 30 s。其它关键算法在 Ubuntu 16.0 操作系统中实现, 程序开发语言为 Python。测试用计算配置如下: Intel(R) Core(TM) i7-7700HQ CPU, NVIDIA GeForce GTX 1060 GPU, 16 GB 物理内存。



(a) 视频帧样例



(b) 根据视觉数据恢复的三维重建结果(视图 1)



(c) 根据视觉数据恢复的三维重建结果(视图 2)

图 2 测试视觉数据和视觉三维重建结果  
Fig. 2 Experimental visual data and vision-based 3D reconstruction results

图 3 为重采样视觉角速率和重采样惯性测量角速率的时空配准结果。这里仅展示关于 Z 轴的角速率, 视觉角速率和惯性测量角速率分别采用虚线和实线表示。其中, 图 3(a)为配准前的角速率对比。从图中可以看出, 在数据配准之前, 难以发现视觉角速率和惯性测量角速率之间关联关系, 这是由于视觉传感器和 IMU 具有不同的坐标系, 因此关于各自坐标轴的角速率在空间上不具备统一的比较基准。此外, 由于数据采样记录时间存在延迟, 采

样数据也不具备统一的时间轴。图 3(b)为空间配准、时间不配准的角速率对比。从图中容易看出, 视觉传感角速率和惯性测量角速率经过空间配准之后, 数据被转换到一致的空间里, 数据差异仅存在时间轴上的一个整体偏移。图 3(c)为空间配准和时间配准后的角速率对比, 视觉角速率和惯性测量角速率被转换到统一的时空坐标系下, 数据状态的变化呈现出较好的一致性。

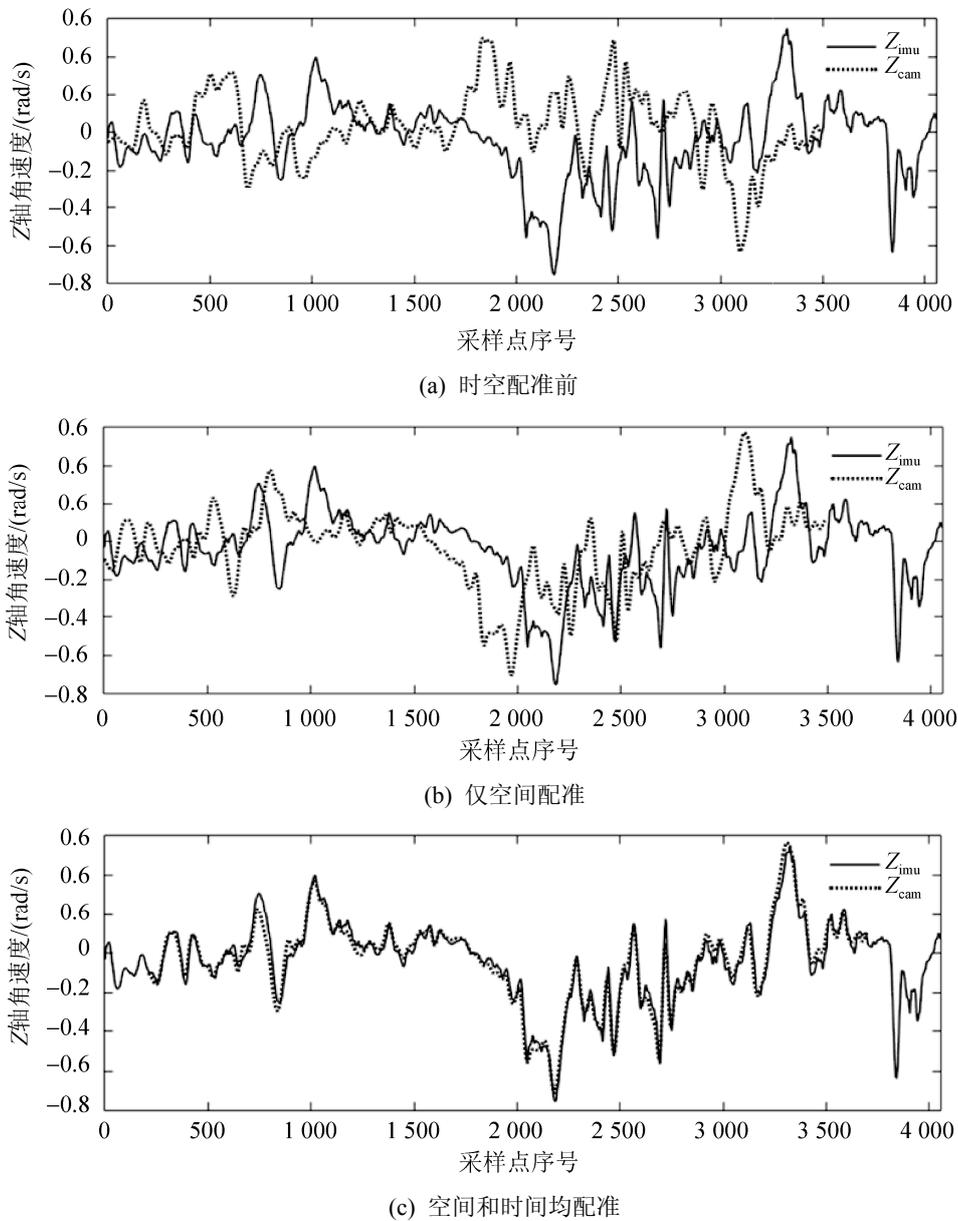


图 3 视觉角速率(虚线)和惯性测量角速率(实线)时空配准。

Fig. 3 Spatial-temporal alignment between visual angular velocity (dashed line) and inertial angular velocity (solid line)

通过上述角速率时空联合配准方法,能够获得视觉传感数据和惯性测量传感数据在时间轴上的最佳偏移估计。基于该时间偏移,可以将视觉线性加速度和惯性测量线性加速度进行时间配准。图 4(a)为线性加速度时间配准结果,可以看出视觉线性加速度和惯性测量线性加速度的变化具有明显

的正相关特性,峰值和低谷均较好吻合,验证了前述角速率时间配准的正确性。图 4(b)为空间配准后的角速率对比,经过一个全局尺度调整,视觉线性加速度和惯性测量线性加速度之间的差异得以最小化,该全局尺度即作为度量三维重建和非接触式物体尺寸估计的基准。

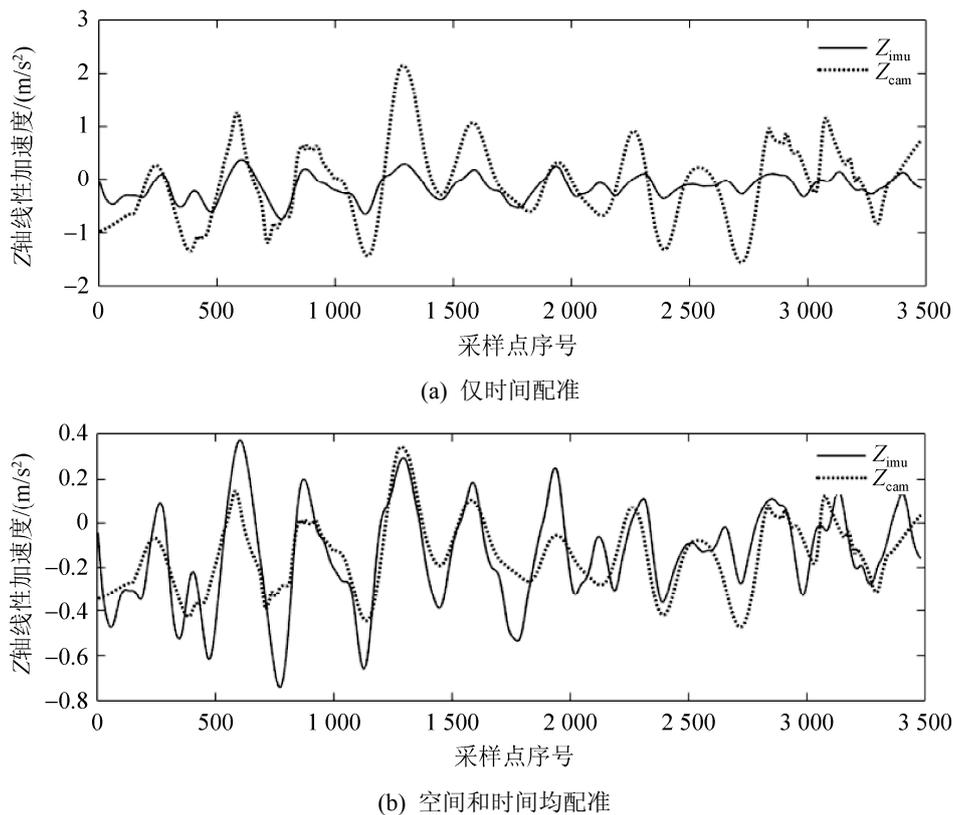


图 4 视觉线性加速度(虚线)和惯性测量线性加速度(实线)时空配准。

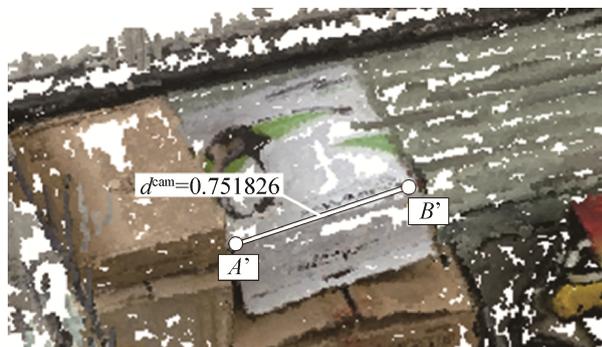
Fig. 4 Spatial-temporal alignment between visual linear acceleration (dashed line) and inertial linear acceleration (solid line)

基于上述结果,继续开展度量三维重建和非接触式物体尺寸获取实验。图 5 展示了一个稠密三维重建和物体尺寸量测示例,该实验采用 PMVS 方法<sup>[7]</sup>实现稠密点云生成。其中,图 5(a)为基于视觉的稠密三维重建结果,度量信息可通过线性加速度时空对齐结果获取。图 5(b)为非接触式物体尺寸获取结果,图像空离量算转化为度量三维重建模型的量算。例如,在图 5(b)中选取对象点  $A$ ,  $B$ , 分别对应于图 5(b)三维稠密点云模型中的  $A'$  和  $B'$ 。本实例中,时空配准阶段获得的视觉线性加速度和惯性测量线性加速度之间的绝对尺度  $\tau^*=0.219$ 。不含度量信息

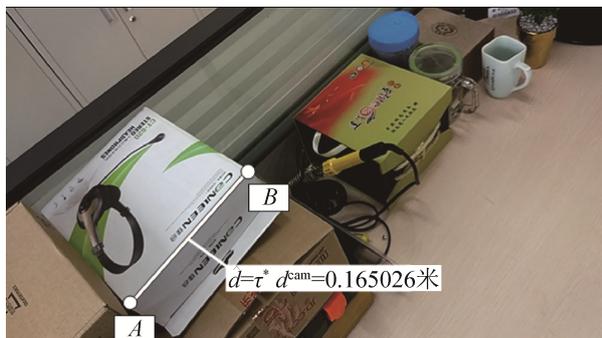
的三维模型中对应的距离(即  $|AB|$  的长度)  $d^{\text{cam}} \approx 0.752$ , 因此测量距离为  $\hat{d} = \tau^* d^{\text{cam}} \approx 0.165$  m。实际长度为 0.17 m, 相对测量误差为 2.93%。

本文提出的数据一致性采样方法和时空配准方法(基于 SVD 求解)均具有较低的计算复杂度,对于上述 30 s 测试数据计算耗时小于 5 s。本文非接触式物体尺寸获取流程中视觉数据稀疏三维重建和稠密三维重建采用开源软件(COLMAP 和 PMVS)实现,对于上述 30 s 测试数据计算耗时约为 30 min。实际应用中,数据采集完成后离线执行场景重建和空间配准,在此基础上可以测量场景中

任意两点之间的距离, 而且可以多次测量。本文非接触式物体尺寸获取方法的基本出发点是利用 IMU 数据来恢复场景的物理尺度信息, 这种基本思想源于即时定位和构图方法<sup>[8-10]</sup>。与上述已有方法相比, 本文方法主要有两方面的优势: 一是上述已有方法均要求 IMU 数据和视觉数据基本对齐, 而本文方法仅需确保 IMU 数据覆盖整个视频时间范围, 无需 IMU 和视觉传感器时钟同步; 二是上述方法均为稀疏三维重建, 而本文采用了 PMVS 三维重建方法, 因此能够获得更加稠密的三维重建结果。上述两方面的优势能够有效提高非接触式物体尺寸获取方法的适用范围。



(a) 基于视觉的稠密三维重建结果



(b) 非接触式物体尺寸获取结果

图 5 稠密三维重建和非接触式物体尺寸获取示例  
Fig. 5 An example of dense 3D reconstruction and non-contact measurement application

## 4 结论

本文提出一种融合视觉数据和惯性测量数据的非接触式物体尺寸获取方法。针对视觉数据分辨率高但采样率相对较低, IMU 采样率高但累积数

值误差较大的特点, 利用计算机视觉三维重建技术恢复视觉传感器的位姿信息, 然后通过分析计算惯性测量数据和视觉位姿数据的时空关系, 获取待测对象的绝对物理尺寸数据。本文方法不需要在待测场景中嵌入额外标定物, 也不需要添加额外辅助定位装置, 使用便捷灵活、适用范围广。非接触式物体尺寸获取在虚拟现实模型构建、虚拟仿真等领域具有较好的应用前景, 下一步工作可从方法在大规模场景中的可扩展性和鲁棒性两方面开展更加深入的研究。

## 参考文献:

- [1] 丁少闻, 张小虎, 于起峰, 等. 非接触式三维重建测量方法综述[J]. 激光与光电子学进展, 2017, 54(7): 27-41. Ding Shaowen, Zhang Xiaohu, Yu Qifeng, et al. Overview of Non-Contact 3D Reconstruction Measurement Methods[J]. Laser & Optoelectronics Progress, 2017, 54(7): 27-41.
- [2] Krajewski G, Woźniak A. Simple master artefact for CMM dynamic error identification[J]. Precision Engineering (S0141-6359), 2014, 38(1): 64-70.
- [3] Huang Q, Yu F, Gong E, et al. Nano-CMM stage with zero Abbe error and its error analysis[J]. Optics and Precision Engineering (S1004-924X), 2013, 21(3): 664-671.
- [4] Oswald M R, Töppe E, Cremers D. Fast and Globally Optimal Single View Reconstruction of Curved Objects [C]// Proceedings of Conference on Computer Vision and Pattern Recognition (CVPR) 2012. Providence, RI, USA: IEEE, 2012: 534-541.
- [5] Snavely N, Seitz S M, Szeliski R. Modeling the World from Internet Photo Collections[J]. International Journal of Computer Vision (S0920-5691), 2008, 80(2): 189-210.
- [6] Schönberger J L, Frahm J M. Structure-from-Motion Revisited[C]// Proceedings of Conference on Computer Vision and Pattern Recognition (CVPR) 2016. Las Vegas, NV, USA: IEEE, 2016: 4104-4113.
- [7] Furukawa Y, Ponce J. Accurate, Dense, and Robust Multi-View Stereopsis[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence (S0162-8828), 2010, 32(8): 1362-1376.
- [8] Mur-Artal R, Tardós J D. Visual-Inertial Monocular SLAM with Map Reuse[J]. IEEE Robotics and Automation Letters (S23773766), 2017, 2(2): 796-803.

- [9] Zhou D, Dai Y, Li H. Reliable scale estimation and correction for monocular Visual Odometry [C]// 2016 IEEE Intelligent Vehicles Symposium (IV). Gothenburg, Sweden: IEEE, 2016: 490-495.
- [10] Li P, Qing T, Hu B, et al. Monocular Visual-Inertial State Estimation for Mobile Augmented Reality[C]// 2017 IEEE International Symposium on Mixed and Augmented Reality (ISMAR). Nantes, France: IEEE, 2017: 11-21.
- [11] Hartley R, Zisserman A. Multiple View Geometry in Computer Vision[M]. New York: Cambridge University Press, 2003.
- [12] Kalman D. A Singularly Valuable Decomposition: The SVD of a Matrix[J]. The College Mathematics Journal (S0746-8342), 1996, 27(1): 2-23.