

DOI: 10.19650/j.cnki.cjsi.J2108267

基于卷积神经网络与扩展卡尔曼滤波的 单目视觉惯性里程计*

林立雄¹, 郑佳春¹, 黄国辉², 蔡国玮³

(1. 集美大学海洋信息工程学院 厦门 361005; 2. 福州大学机械工程及自动化学院 福州 350000;
3. 福建船政交通职业学院 福州 350000)

摘要:针对单目相机采集室外图像易受环境光照影响、尺度存在不确定性的缺点,以及利用神经网络进行位姿估计不准确的问题,提出一种基于卷积神经网络(CNN)与扩展卡尔曼滤波(EKF)的单目视觉惯性里程计。采用神经网络取代传统里程计中基于几何约束的视觉前端,将单目相机输出的估计值作为测量更新,并通过神经网络优化EKF的误差协方差。利用EKF融合CNN输出的单目相机位姿和惯性测量单元(IMU)数据,优化CNN的位姿估计,补偿相机尺度信息与IMU累计误差,实现无人系统运动位姿的更新和估计。相比于使用单目图像的深度学习算法Depth-VO-Feat,所提算法融合单目图像和IMU数据进行位姿估计,KITTI数据集中09序列的平动、转动误差分别减少45.4%、47.8%,10序列的平动、转动误差分别减少68.1%、43.4%。实验结果表明所提算法能进行更准确的位姿估计,验证了算法的准确性和可行性。

关键词:视觉惯性里程计;神经网络;惯性测量单元;扩展卡尔曼滤波

中图分类号: TP242 TH74 文献标识码: A 国家标准学科分类代码: 510.80

Utilizing extended Kalman filter to improve convolutional neural networks based monocular visual-inertial odometry

Lin Lixiong¹, Zheng Jiachun¹, Huang Guohui², Cai Guowei³

(1. School of Information Engineering, Jimei University, Xiamen 361005, China; 2. School of Mechanical Engineering and Automation, Fuzhou University, Fuzhou 350000; 3. Fujian Chuanzheng Communications College, Fuzhou 350000, China)

Abstract: The outdoor images collected by monocular camera are easily affected by the light intensity. The scale of images is ambiguous. In addition, the pose estimation of convolutional neural networks (CNNs) is not accurate. To address these issues, a monocular vision-inertial odometry using CNN and the extended Kalman filter (EKF) is proposed. The CNN is used to replace the conventional odometry of the front-end vision based geometric constraints. The output of the monocular camera is used as the EKF measurement to correct the estimated pose of CNN. The error covariance of EKF is optimized by the CNN. The monocular camera pose data and the inertial measurement unit (IMU) data are fused in EKF to estimate the motion pose. The monocular scale informations and the cumulative errors of the IMU are compensated. Experimental results show that the proposed algorithm performs more precise pose estimation. The accuracy and feasibility of the algorithm are verified. Compared with the Depth-VO-Feat algorithm that relies on monocular images, the proposed algorithm combines monocular image and IMU data for pose estimation. The translation and rotation errors of the 09 sequence in KITTI dataset are reduced by 45.4% and 47.8%, respectively. The translation and rotation errors of 10 sequences are reduced by 68.1% and 43.4%, respectively.

Keywords: visual-inertial odometry; neural networks; inertial measurement unit; extended Kalman filter

收稿日期: 2021-07-19 Received Date: 2021-07-19

* 基金项目: 福建省自然科学基金(2019J05024)、国家自然科学基金(61803089)项目资助

0 引言

融合惯性测量单元 (inertial measurement unit, IMU) 的视觉惯性里程计 (visual-inertial odometry, VIO) 可以解决单目相机的不适定问题,随着基于微机电系统的高性能、低成本、轻量化的 IMU 大量投入市场,视觉结合 IMU 传感器的 VIO 导航方案逐渐成为同步定位与建图 (simultaneous localization and mapping, SLAM) 技术的发展趋势^[1]。VIO 算法是对给定测量值的最大似然估计,是一个状态优化问题^[2]。状态优化是 VIO 的核心内容,可分为两大类,分别是基于滤波^[3]的方法和基于优化^[4]的方法。滤波方法主要基于贝叶斯估计理论,包括卡尔曼滤波 (Kalman filter, KF)^[5]和粒子滤波^[2]两种方法,其中 KF 应用较为广泛。KF 是针对线性系统的无偏估计,而 SLAM 系统的观测方程和测量方程往往是非线性的,因此后续研究将 KF 扩展到非线性系统,其中忽略高阶项进行线性近似的方法称为扩展卡尔曼 (extended Kalman filter, EKF)^[6],以及使用无迹卡尔曼^[7]通过采样的方法近似非线性分布。Bloesch 等^[4]提出基于 EKF 的视觉惯性滤波框架 ROVIO,该框架采用像素强度误差,将多层图像块特征和 IMU 数据通过紧耦合方式结合到 EKF 滤波器。MSCKF^[8]使用多状态约束卡尔曼滤波器进行数据融合,与之相关的工作有 Kumar Robotics 实验室提出的基于 S-MSCKF^[9]的立体视觉惯性里程计。

单目 SLAM 系统主要采用的视觉方法是基于光学几何约束,如尺度不变特征变换^[10]、运动中恢复结构^[11],使用这些视觉方法估计图像深度时,相同的输入图像可以映射多个合理深度,这是一个不适定的问题。

基于几何理论的特征点法或直接法需要进行大量的优化运算,为避免复杂的计算过程,深度学习方法通过多层卷积神经网络 (convolution neural networks, CNN) 直接从图像中获取高级特征^[12-19]。Eigen 等^[20]提出将网络划分为全局和局部神经网络进行图像深度估计,首次使用 CNN 估计单目图像深度。Zhou 等^[21]从单目图像序列中学习深度和运动状态。与仅从图像中获取运动特征不同,ViNet^[22]将图像数据和 IMU 数据同时分别输入 CNN 和长短时记忆网络 (long short-term memory, LSTM)^[23],利用 LSTM 处理传感器频率不匹配问题,实现传感器数据融合。余洪山等^[24]利用 SuperPoint 网络在特征提取方面的强鲁棒性,使单目视觉惯性 SLAM 能够应用于更具挑战性的环境。现有大多 CNN 包含多层网络,模型参数大,进行单目图像位姿估计时会占用大量计算资源,使得深度学习方法较难应用到实际场景^[25]。

综上所述,通过对神经网络进行优化,结合深度学习

的单目视觉惯性里程计 (monocular VIO, Mono-VIO) 可以为无人系统室外自主定位问题提供有效的解决方案。单目相机成本低廉,存在尺度不确定性,而 IMU 在长时间使用中会产生漂移。Mono-VIO 算法的单目相机可以校正 IMU 的漂移误差,同时利用 IMU 数据弥补单目相机缺失的尺度信息,通过 EKF 滤波器融合单目视觉神经网络的输出信息和 IMU 数据,能提供更加精确的运动位姿估计。因此,本文提出融合 CNN 与 EKF 的 Mono-VIO 算法,以实现无人系统室外定位、导航等功能。本文的主要贡献如下:

1) 考虑到多目相机标定过程繁琐以及单目视觉的尺度不确定性,在视觉定位阶段提出使用端到端的轻量化单目视觉神经网络取代基于几何特征的 SLAM 前端。利用 CNN 算法直接对单目图像进行特征提取,快速获取单目相机的局部位姿。

2) 针对视觉图像容易产生运动模糊和受环境光照影响,以及 IMU 传感器存在累积误差问题,提出改进的 EKF 算法对单目相机的局部位姿进行校正,实现单目相机与 IMU 数据融合。利用 IMU 传感器数据补偿单目视觉神经网络的位姿估计误差,并建立异构传感器数据融合方程,输出优化后的精确全局位姿。

3) 在 KITTI 公开数据集^[26]中实现算法评估实验。实验从三维空间旋转角和 xyz 平面位移的估计精度两个方面,对所提出的结合 CNN 与 EKF 方法的异构传感器数据融合位姿估计方法进行对比验证。

1 算法框架

1.1 总体方案

本文选择单目相机与 IMU 传感器融合方案,从输入的单目图像中获取无人系统局部位姿,借助传感器融合算法完成无人系统的位姿估计,算法框架如图 1 所示,整体算法使用 Python 编程语言。Mono-VIO 系统的核心是基于深度学习与 EKF 滤波器的异构多传感器融合算法,无人系统的局部位姿可以通过优化后的轻量化神经网络模型得到,通过 EKF 滤波器融合来自单目图像的位姿数据与 IMU 传感器数据,实现无人系统精确的位姿估计。

EKF 算法将 Mono-VIO 分为两个模块,分别是姿态预测模块和姿态更新模块。姿态预测模块基于 IMU 运动学进行状态传播,使用 IMU 数据对机器人自身位姿进行初始化,实现自身状态预测。姿态更新模块的核心为使用跨尺度连接、子像素卷积等技术的轻量化神经网络,该网络从单目图像中输出局部位姿并作为 EKF 的测量量,用于 EKF 状态更新,不断对无人系统的位姿状态进行更新,最终输出全局位姿,完成无人系统的运动状态估计。

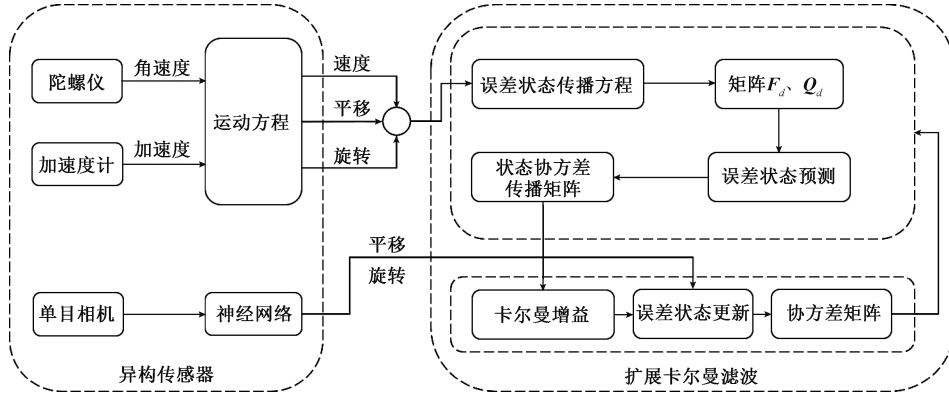


图 1 视觉里程计算法理论框架

Fig. 1 Block diagram of the proposed visual odometry algorithm

1.2 坐标定义

考虑一个无人系统平台,配备 IMU 传感器和单目相机,传感器相对于全局世界坐标系不断发生移动。Mono-VIO 的目标是使用单目图像和 IMU 数据估计无人系统平台的位置和方向, Mono-VIO 包括 3 个坐标系:1)世界坐标系 $\{W_1\}$,假设它是一个惯性系,作为单目相机和 IMU 的绝对位姿参考;2)IMU 坐标系 $\{I_2\}$,用于测量角速度和线性加速度,其原点位于 IMU 传感器表面;3)相机坐标系 $\{C_3\}$,其原点位于相机的光学中心。每组坐标系分别由单位正交基组成,坐标系的某一点可以由正交基乘以系数表示,两个坐标系变换关系表示成两组正交基的内积,即旋转矩阵 $SO(3)$:

$$SO(3) = [R | R \in \mathbb{R}^{3 \times 3}, R^T R = R R^T = I] \quad (1)$$

如图 2 所示, t_{12} 是坐标系 $\{I_2\}$ 相对于 $\{W_1\}$ 原点的平移向量, R_{12} 是将 $\{W_1\}$ 的向量分量转换到 $\{I_2\}$ 中分量的旋转矩阵,从坐标系 $\{W_1\}$ 到坐标系 $\{I_2\}$ 的位姿变换用 T_{12} 表示^[27]。 θ 是 z 轴的旋转角度,考虑从 $\{W_1\}$ 到 $\{I_2\}$ 再到 $\{C_3\}$ 的变换,包含两次旋转 R_{12} 、 R_{23} 和两次平移 t_{12} 、 t_{23} 。

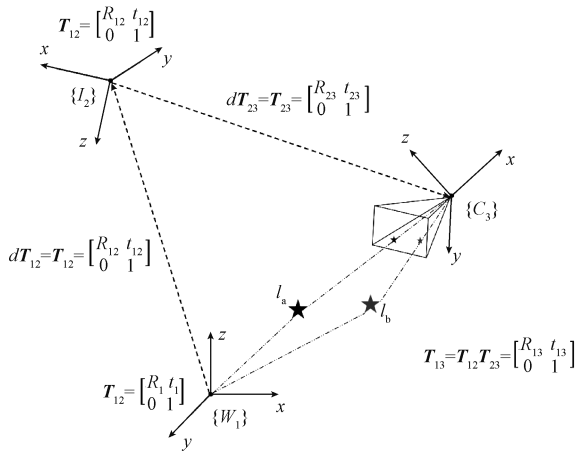


图 2 三维空间坐标系变换

Fig. 2 Three-dimensional coordinate transformation

1.3 状态参数化

KF 被设计用于线性状态空间系统,系统在离散时间 $k-1$ 到 k 的过程中,各个变量定义及维度大小如表 1 所示。所有噪声均服从零均值高斯分布, $w \sim N(0, Q)$, $\gamma \sim N(0, R)$ 。上标 $\hat{\cdot}$ 表示估计值,上标 $\sim$ 表示微分运算,上标 $\sim$ 表示向量误差,使用 α_x 表示任意变量 a 的斜对称矩阵。 ω_m 、 a_m 为 IMU 的角速度和加速度测量值, b_g 、 b_a 为 ω_m 、 a_m 的偏置, n_g 、 n_a 为陀螺仪和加速度计的高斯白噪声, n_{b_g} 、 n_{b_a} 为 b_g 、 b_a 的噪声。加速度 $g = [0, 0, 9.81]^T$ 。

表 1 离散时间系统变量

Table 1 Discrete-time system variable

变量	定义	维数
x	状态向量	$n_x \times 1$
z	输出向量	$n_y \times 1$
u	输入向量	$n_u \times 1$
w	过程噪声	$n_x \times 1$
γ	测量噪声	$n_y \times 1$
A	状态转移矩阵	$n_x \times n_x$
B	输入增益矩阵	$n_x \times n_u$
H	测量矩阵	$n_y \times n_x$

EKF 使用 IMU 传感器的陀螺仪和加速度计测量值作为初始值进行状态传播,总体状态变量由 IMU 姿态 X_i 和摄像头位置 P_i^c 组成, q_w^i 是四元数,表示 IMU 相对于世界坐标系的旋转, p_w^i 和 v_w^i 为 IMU 在世界坐标系下的平移和速度, λ 为常量, p_i^c 、 q_i^c 表示相机坐标系到 IMU 坐标系的平移和旋转,则移动机器人的状态向量 X 为:

$$X = [X_i \ P_i^c] = [\bar{q}_w^i \ p_w^i \ v_w^i \ b_g^T \ b_a^T \ \lambda \ p_i^c \ q_i^c] \quad (2)$$

其中, Ω 表示四元数矩阵乘法运算,估计值的运动学

方程为:

$$\dot{\hat{\mathbf{q}}}_w^i(t) = \frac{1}{2}\Omega(\boldsymbol{\omega}_m - \hat{\mathbf{b}}_g)\hat{\mathbf{q}}_w^i \quad (3)$$

$$\dot{\hat{\mathbf{p}}}_w^i(t) = \hat{\mathbf{v}}_w^i(t) \quad (4)$$

$$\dot{\hat{\mathbf{v}}}_w^i(t) = (\hat{\mathbf{R}}_w^i)^T(\mathbf{a}_m - \hat{\mathbf{b}}_g) - \mathbf{g} = \mathbf{C}_{\hat{\mathbf{q}}_w^i}^T(\mathbf{a}_m - \hat{\mathbf{b}}_g) - \mathbf{g} \quad (5)$$

$$\dot{\hat{\mathbf{b}}}_g(t) = \mathbf{0}_{3 \times 1} \quad \dot{\hat{\mathbf{b}}}_a(t) = \mathbf{0}_{3 \times 1} \quad (6)$$

$$\lambda = 0 \quad \dot{\hat{\mathbf{p}}}_i^c(t) = \mathbf{0}_{3 \times 1} \quad \dot{\hat{\mathbf{q}}}_i^c(t) = \mathbf{0}_{7 \times 1} \quad (7)$$

其中, $\mathbf{C}_{(q)}$ 表示四元数 $\mathbf{q}$ 对应的旋转矩阵, 四元数 $\hat{\mathbf{q}}$

与估计值 $\hat{\mathbf{q}}$ 的误差 $\Delta\boldsymbol{\theta}_w^T$ 满足以下关系:

$$\mathbf{C}_{\hat{\mathbf{q}}_w^i}^T \approx (\mathbf{I}_3 - \Delta\boldsymbol{\theta}_w^T) \mathbf{C}_{\hat{\mathbf{q}}_w^i}^T \quad (8)$$

相机姿态为四元数, 因此采用四元数的乘法误差模型, 符号 Δ 表示误差, 状态误差 $\tilde{\mathbf{x}}(t)$ 为:

$$\tilde{\mathbf{x}}(t) = [\Delta\boldsymbol{\theta}_w^T(t) \quad \Delta\mathbf{p}_w^T(t) \quad \Delta\mathbf{v}_w^T(t) \quad \Delta\mathbf{b}_g^T(t) \quad \Delta\mathbf{b}_a^T(t) \quad \Delta\lambda \quad \Delta\mathbf{p}_i^T(t) \quad \Delta\boldsymbol{\theta}_i^T(t)] \quad (9)$$

使用 $\delta\hat{\mathbf{q}}_w^i$ 表示实际姿态与估计姿态间的旋转变换:

$$\delta\hat{\mathbf{q}}_w^i = \hat{\mathbf{q}}_w^i \otimes \hat{\mathbf{q}}_w^i \approx \begin{bmatrix} \frac{1}{2}\Delta\boldsymbol{\theta}_w^T \\ 1 \end{bmatrix} \quad (10)$$

$$\delta\hat{\mathbf{q}}_i^c = \hat{\mathbf{q}}_i^c \otimes \hat{\mathbf{q}}_i^c \approx \begin{bmatrix} \frac{1}{2}\delta\boldsymbol{\theta}_i^T \\ 1 \end{bmatrix}$$

进一步拓展式(3)~(7), $[\hat{\mathbf{a}}_x]_\times, [\hat{\boldsymbol{\omega}}_x]_\times$ 为估计值 $\hat{\mathbf{a}}, \hat{\boldsymbol{\omega}}$ 的斜对称矩阵, 则连续时间误差状态的运动学方程为:

$$\Delta\dot{\mathbf{p}}_w^i(t) = \Delta\mathbf{v}_w^i \quad (11)$$

$$\Delta\dot{\mathbf{v}}_w^i(t) = -\mathbf{C}_{\hat{\mathbf{q}}_w^i}^T[\hat{\mathbf{a}}_x]_\times\delta\boldsymbol{\theta} - \mathbf{C}_{\hat{\mathbf{q}}_w^i}^T\Delta\mathbf{b}_a - \mathbf{C}_{\hat{\mathbf{q}}_w^i}^T\mathbf{n}_a \quad (12)$$

$$\Delta\dot{\boldsymbol{\theta}}_w^i(t) = -[\hat{\boldsymbol{\omega}}_x]_\times\delta\boldsymbol{\theta} - \Delta\mathbf{b}_g - \mathbf{n}_g \quad (13)$$

$$\Delta\dot{\mathbf{b}}_g(t) = \mathbf{n}_{wg} \quad \Delta\dot{\mathbf{b}}_a(t) = \mathbf{n}_{wa} \quad (14)$$

$$\Delta\lambda = 0 \quad \Delta\dot{\mathbf{p}}_i^c(t) = 0 \quad \Delta\dot{\boldsymbol{\theta}}_i^c(t) = 0 \quad (15)$$

2 视觉-惯性数据融合

本文提出将经典的 EKF 状态估计器与 CNN 结合, 融合来自单目图像输出的位姿和 IMU 传感器数据, 实现无监督学习的 Mono-VIO。

2.1 基于 CNN 的单目视觉测量

如图 3 所示, 将一组连续的单目图像输入 CNN, 利用 CNN 从训练集的源视图 I_s 重建出目标视图 I_t ^[19], $\mathbf{K}$ 为相机内参矩阵, I_t 的像素点 p_t 到 I_s 平面的像素点 p_s 投影变换可以通过预测的图像深度 $\hat{D}_t$ 和运动位姿 $\hat{T}_{t \rightarrow s}$ 求得, 即:

$$\mathbf{p}_s = \mathbf{K}\hat{T}_{t \rightarrow s}\hat{D}_t\mathbf{K}^{-1} \quad (16)$$

本文对 Hu 等^[28]提出的神经网络加以改进, 并参考 Zhan 等^[17]的研究, 进一步提高神经网络位姿估计精度。

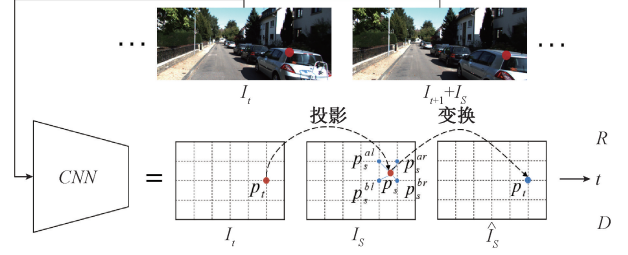


图 3 神经网络估计原理

Fig. 3 Neural network estimation principle

神经网络采用 Encoder-Decoder 结构, 使用 Resnet-50 作为编码层, CNN 整体框架如图 4 所示, 分为编码层 (encoder)、ASPP (atrous spatial pyramid pooling)、门控自适应解码层 (gated adaptive decoder, GAD)、多尺度特征融合 (multi-scale feature fusion, MFF) 以及重建模块 (reconstruction, R)。其中 Encoder 用于提取低分辨率单目图像的基本特征信息, 对图像进行下采样, 图像大小压缩至原图像的 1/32; ASPP 处理压缩后的图像特征, 从不同尺度放大图像, 并融合不同尺度特征; GAD 完成压缩图像的放大操作, 在子像素卷积的上采样过程中加入门控模块自适应筛选来自上一层解码器层和相应 Encoder 的特征信息, 弥补图像压缩中丢失的细节信息, 输出高精度的特征数据; MFF 作为 R 的输入单元, 将下采样过程中生成的不同分辨率特征图像进行子像素卷积上采样; R 接收来自 GAD 和 MFF 的高级特征, 进一步细化特征信息, 输出深度图像和相对位姿, 完成整个估计过程, 具体可参考文献[29]。

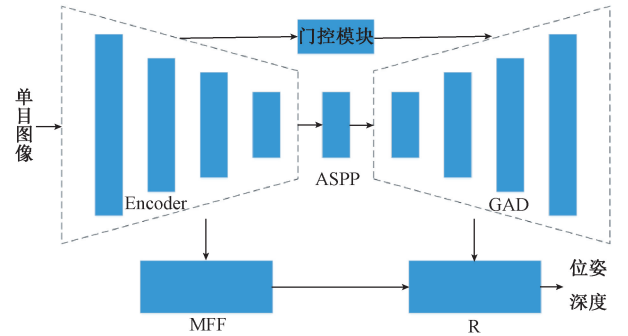


图 4 神经网络结构

Fig. 4 Neural network structure

2.2 误差状态传播

连续时间误差状态传播方程为:

$$\dot{\tilde{\mathbf{x}}}(t) = \mathbf{F}_c(t)\tilde{\mathbf{x}}(t) + \mathbf{G}_c(t)\mathbf{n}(t) \quad (17)$$

系统噪声 $\mathbf{n} = [\mathbf{n}_g^T, \mathbf{n}_{bg}^T, \mathbf{n}_a^T, \mathbf{n}_{ba}^T]^T$, $\mathbf{F}_c$ 为连续时间误差状态转移矩阵, 而 $\mathbf{G}_c$ 是输入噪声矩阵, 由下式给出^[30]:

$$F_c = \begin{bmatrix} -[\hat{\omega}_x] & -I_3 & 0_3 & 0_3 & 0_3 & 0_3 \\ 0_3 & 0_3 & 0_3 & 0_3 & 0_3 & 0_3 \\ -C_{\hat{q}_w}^T [\hat{a}_x] & 0_3 & 0_3 & -C_{\hat{q}_w}^T & 0_3 & 0_3 \\ 0_3 & 0_3 & 0_3 & 0_3 & 0_3 & 0_3 \\ 0_3 & 0_3 & I_3 & 0_3 & 0_3 & 0_3 \\ 0_3 & 0_3 & 0_3 & 0_3 & 0_3 & 0_3 \end{bmatrix} \quad (18)$$

$$G_c = \begin{bmatrix} -I_3 & 0_3 & 0_3 & 0_3 \\ -0_3 & I_3 & 0_3 & 0_3 \\ -I_3 & 0_3 & -C_{\hat{q}_w}^T & 0_3 \\ -I_3 & 0_3 & 0_3 & I_3 \\ -I_3 & 0_3 & 0_3 & 0_3 \\ 0_3 & 0_3 & 0_3 & 0_3 \end{bmatrix} \quad (19)$$

系统噪声为零均值高斯白噪声。式(17)为 IMU 测量的连续时间传播模型,但在 EKF 的实际应用中,需要计算离散时间状态转换矩阵 F_d [31]:

$$F_d = \exp(F_c \Delta t) = I_d + F_c \Delta t + \frac{1}{2!} F_c^2 \Delta t^2 + H. O. T \quad (20)$$

状态转移矩阵 F_d 可以进一步表示为:

$$F_d = \begin{bmatrix} \Theta & \Psi \\ 0_3 & I_{d3} \end{bmatrix} \quad (21)$$

对于 $|\omega| \rightarrow 0$, 利用斜对称矩阵 $[\hat{\omega}_x]$ 的性质,根据洛必达定则,矩阵 Θ 可以写成:

$$\lim_{|\omega| \rightarrow 0} \Theta = I_d - \Delta t [\hat{\omega}_x] + \frac{\Delta t^2}{2} [\hat{\omega}_x]^2 \quad (22)$$

同理,矩阵 Ψ 可以写成:

$$\lim_{|\omega| \rightarrow 0} \Theta = I_d \Delta t + \frac{\Delta t^2}{2} [\hat{\omega}_x] - \frac{\Delta t^3}{6} [\hat{\omega}_x]^2 \quad (23)$$

最终 F_d 为:

$$F_d = \begin{bmatrix} I_{d3} & \Delta t & A & B & -C_{\hat{q}_w}^T \frac{\Delta t^2}{2} & 0_{3 \times 7} \\ 0_3 & I_{d3} & C & D & -C_{\hat{q}_w}^T \Delta t & 0_{3 \times 7} \\ 0_3 & 0_3 & E & F & 0_3 & 0_{3 \times 7} \\ 0_3 & 0_3 & 0_3 & I_{d3} & 0_3 & 0_{3 \times 7} \\ 0_3 & 0_3 & 0_3 & 0_3 & I_{d3} & 0_{3 \times 7} \\ 0_{7 \times 3} & 0_{7 \times 3} & 0_{7 \times 3} & 0_{7 \times 3} & 0_{7 \times 3} & I_{d7} \end{bmatrix}$$

$$A = -C_{\hat{q}_w}^T [\hat{a}_x] \left(\frac{\Delta t^2}{2} - \frac{\Delta t^3}{3!} [\omega_x] + \frac{\Delta t^4}{4!} [\omega_x]^2 \right)$$

$$B = -C_{\hat{q}_w}^T [\hat{a}_x] \left(-\frac{\Delta t^3}{3!} + \frac{\Delta t^4}{4!} [\omega_x] - \frac{\Delta t^5}{5!} [\omega_x]^2 \right)$$

$$C = -C_{\hat{q}_w}^T [\hat{a}_x] \left(\Delta t - \frac{\Delta t^2}{2!} [\omega_x] + \frac{\Delta t^3}{3!} [\omega_x]^2 \right) \quad (24)$$

$$D = -A$$

$$E = I_d - \Delta t [\omega_x] + \frac{\Delta t^2}{2!} [\omega_x]^2$$

$$F = -\Delta t + \frac{\Delta t^2}{2!} [\omega_x] - \frac{\Delta t^3}{3!} [\omega_x]^2$$

假设离散时间系统的噪声为高斯白噪声,且相互独立,则噪声协方差 Q_d 计算如下:

$$Q_d = \int_{\Delta t} F_d(\tau) G_c Q_c G_c^T F_d^T(\tau) d\tau = \int_{\Delta t} F_d(\tau) \begin{bmatrix} \sigma_{n_a}^2 \cdot I_{3 \times 3} & 0_{3 \times 3} & 0_{3 \times 3} & 0_{3 \times 3} \\ 0_{3 \times 3} & \sigma_{n_{wa}}^2 \cdot I_{3 \times 3} & 0_{3 \times 3} & 0_{3 \times 3} \\ 0_{3 \times 3} & 0_{3 \times 3} & \sigma_{n_g}^2 \cdot I_{3 \times 3} & 0_{3 \times 3} \\ 0_{3 \times 3} & 0_{3 \times 3} & 0_{3 \times 3} & \sigma_{n_{wg}}^2 \cdot I_{3 \times 3} \end{bmatrix} d\tau \quad (25)$$

最终得到 EKF 的状态协方差矩阵:

$$P_{k+1|k} = F_d P_{klk} F_d^T + Q_d \quad (26)$$

2.3 测量更新

在 EKF 更新阶段,使用单目相机测量姿态和位置,并进行 EKF 状态更新,将外部测量结果与滤波器估计值融合。神经网络利用式(16)中的投影变换从单目图像中得到位姿 $\tilde{T}_{t \rightarrow s}$, 定义 q_w^c, p_w^c 为 $\{C_3\}$ 坐标系到 $\{W_1\}$ 坐标系的旋转和平移。 Δt 时刻内, CNN 输出的位姿 $\tilde{T}_{\Delta t} = [\tilde{q}_w^c, \tilde{p}_w^c]$, 不确定噪声的协方差为 R_{k+1} , $\tilde{\varphi}_w^{c,k}$ 为旋转矩阵所对应的向量, e^r 为平移误差, e^θ 为四元数旋转误差,则单目相机的测量误差 e_z 为:

$$e_z = \begin{bmatrix} e^\theta \\ e^r \end{bmatrix} = \begin{bmatrix} \ln(\exp(\tilde{\varphi}_w^{c,k}) C_{p_w^c}^T)^V \\ \tilde{p}_w^c - p_w^c \end{bmatrix} \quad (27)$$

其中, $\tilde{q}_w^c$ 为四元数估计值, $\tilde{\varphi}_w^c, \tilde{p}_w^c$ 为旋转和平移测量值, 利用 BCH (Baker-Campbell-Hausdorff) 公式将式(27)转换为 [32]:

$$e^\theta = \ln(\exp(\tilde{\varphi}_w^{c,k}) C_{p_w^c}^T)^V \approx \tilde{\varphi}_w^c - \varphi_w^c \quad (28)$$

对式(27)进行展开:

$$e^\theta \approx \tilde{\varphi}_w^c - \ln(C_{p_w^c}^T) \approx \tilde{\varphi}_w^c - \varphi_w^c - J_r(\varphi_w^c)^{-1} \delta \varphi_w^c \quad (29)$$

$$e^r \approx \tilde{p}_w^c - p_w^c - \delta p_w^c \quad (30)$$

$$e_z^\theta \approx \tilde{\varphi}_w^c - \hat{\varphi}_w^c \quad (31)$$

$$e_z^r \approx \tilde{p}_w^c - \hat{p}_w^c \quad (32)$$

用 $\tilde{z} = H\tilde{x} + \gamma$ 表示 EKF 的观测方程, 推导测量矩阵 H :

$$H = \frac{\partial e_z}{\partial x} = \begin{bmatrix} 0 & -J_r(\hat{\varphi}_w^c) & 0 & 0 \\ 0 & 0 & -I & 0 \end{bmatrix} \quad (33)$$

测量矩阵 H 对 EKF 进行周期性的状态更新, 得到状态估计值, 步骤如下:

1) 计算测量误差 $\tilde{e}_z$:

$$\tilde{e}_z = \begin{bmatrix} \tilde{e}_z^\theta \\ \tilde{e}_z^r \end{bmatrix} \quad (34)$$

2) 计算误差的协方差 S :

$$S = HP_{k+1|k}H^T + R_{k+1} \quad (35)$$

3) 计算卡尔曼增益 K_{k+1} :

$$K_{k+1} = P_{k+1|k}H^TS^{-1} \quad (36)$$

4) 更新估计值 $\hat{x}$:

$$\hat{x} = K_{k+1}\tilde{e}_z \quad (37)$$

5) 更新误差状态协方差矩阵 $P_{k+1|k+1}$:

$$P_{k+1|k+1} = (I_d - K_{k+1}H)P_{k+1|k} + K_{k+1}R_{k+1}K_{k+1}^T \quad (38)$$

3 实验结果与分析

3.1 数据处理

CNN 训练过程中将单目图像分辨率调整为 128×416 , 相机频率为 10 Hz。而 IMU 数据通过计算可得位移和欧拉角, 由于 KITTI Odometry 数据集不能直接得到 IMU 数据, 因此使用 pykitti 工具包读取 Raw Data 原始数据集中 OXTS 文件提供的 GPS/IMU 数据集合, 由此获得速度与加速度, 并对速度和加速度添加随机噪声模拟 IMU 的真实输出。添加噪声后的数据变换更为剧烈, 更加考验算法的可靠性。

CNN 算法使用 Python 编程语言编写, 在 Ubuntu 16.04 系统中搭建 Pytorch 深度学习框架, 模型训练中使用的显卡型号为 NVIDIA GeForce RTX 1660Ti, 显存为 6 G。CNN 算法在 KITTI 数据集集中的单目深度估计图如图 5 所示, 图 5(a)、(b)、(c)、(d)、(e) 分别表示 RGB 图像、真实值、Zhou 等^[21]方法、Yin 等^[33]方法、本文方法得到的单目深度图像。从图 5 中可以看到, 图 5(c)、(d) 列的深度图无法正常显示物体轮廓, 重建的深度图像细节

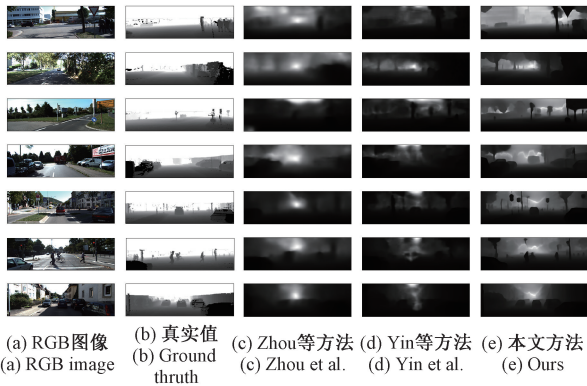


图 5 不同算法在 KITTI 数据集集中的单目深度估计

Fig. 5 Monocular depth estimation of different algorithms in KITTI dataset

不如本文方法。如图 6、7 所示, 展示了 IMU 数据的噪声误差, 实线为未经处理的数据, 圆点 Noised 表示添加噪声后的数据, 其中图 6(a) 为速度曲线, 图 6(b) 为加速度曲线。

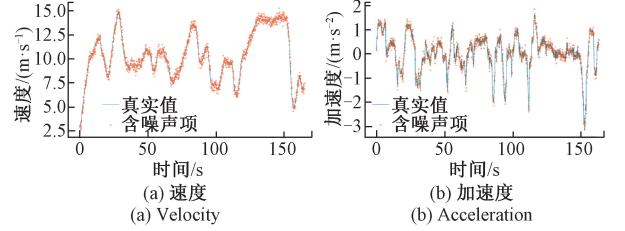


图 6 添加噪声后的 IMU 速度、加速度 (09 序列)

Fig. 6 IMU velocity and acceleration with noise (sequence 09)

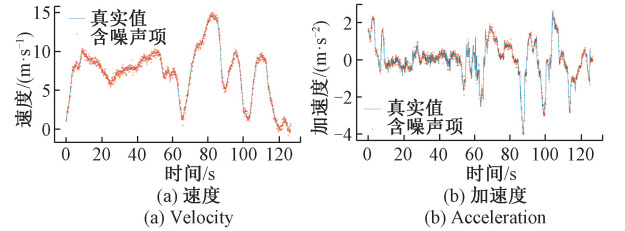


图 7 添加噪声后的 IMU 速度、加速度 (10 序列)

Fig. 7 IMU velocity and acceleration with noise (sequence 10)

为更直观对比添加噪声前后的数据变换, 如图 8、9 所示, 图 9(a)、(b)、(c) 分别为未添加噪声的俯仰角速度、速度、加速度真实值, 图 9(d)、(e)、(f) 分别为添加噪声后的俯仰角速度、速度、加速度曲线。

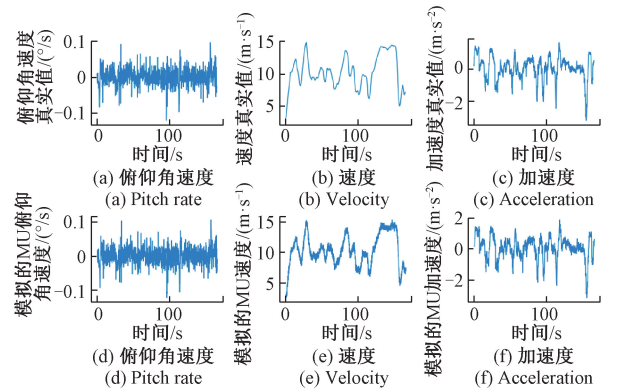


图 8 KITTI 数据集真实值及噪声项 (09 序列)

Fig. 8 Groundtruth and noise items of KITTI dataset (sequence 09)

3.2 旋转误差

EKF 融合算法使用包含噪声的 IMU 传感器数据对 CNN 的估计位姿进行修正, 如图 10、11 所示, EKF 算法使用含噪声项作为输入数据, 可分别得到翻滚角 (图 10

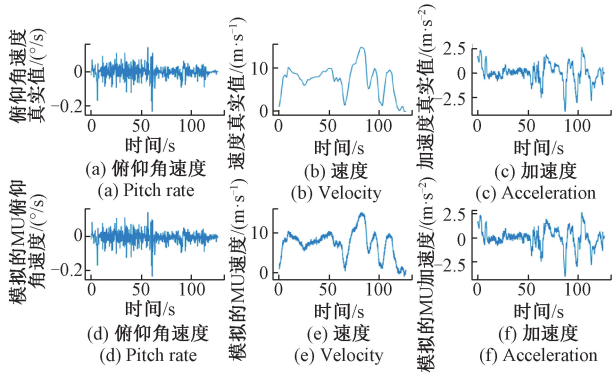


图9 KITTI数据集真实值及噪声项(10序列)

Fig. 9 Groundtruth and noise items of KITTI dataset (sequence 10)

(d)、俯仰角(图10(e))、偏航角(图10(f)),而图10(a)、(b)、(c)分别为所对应的欧拉角真实值。

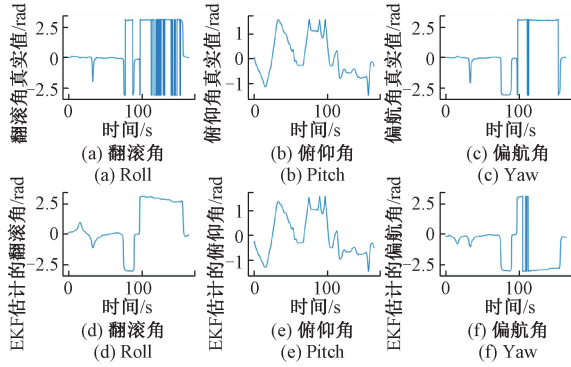


图10 欧拉角真实值及估计值(09序列)

Fig. 10 Groundtruth and estimation of Euler angle (sequence 09)

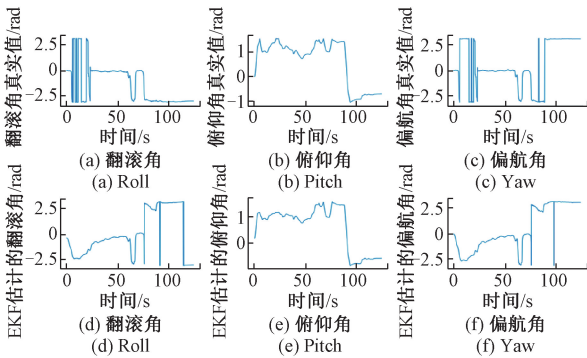


图11 欧拉角真实值及估计值(10序列)

Fig. 11 Groundtruth and estimation of Euler angle (sequence 10)

欧拉角估计值中,以俯仰角为例。如图8(d)、9(d)所示,EKF融合算法输入的俯仰角角速度包含随机噪声。

相比于图8(a)、9(a)中未添加噪声的角速度,添加噪声后的数据变换更为剧烈,数据融合时需对噪声进行有效的处理。

EKF融合算法估计的俯仰角及误差如图12、13所示,其中图12(a)为俯仰角估计值,图12(b)为估计值的误差,点线为俯仰角真实值,虚线为EKF融合算法估计得到的俯仰角。由图12、13可知,即使在俯仰角变换剧烈的情况下,俯仰角估计值均能较好地跟踪真实值的变换。从右边的误差图可以看到,俯仰角在起始和终点处的估计误差较大,中间时刻的估计误差较小,整体角度偏差小,平均误差保持在较小的数值范围内,能保证相对可靠的估计,可以满足实际应用需求。

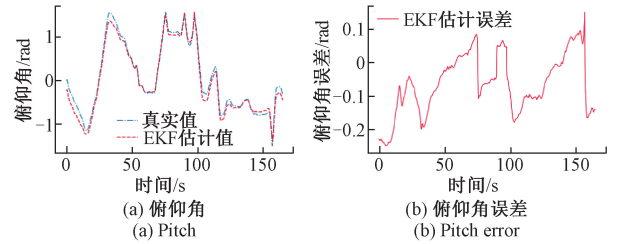


图12 俯仰角估计值及误差(09序列)

Fig. 12 Pitch angle estimation and error (sequence 09)

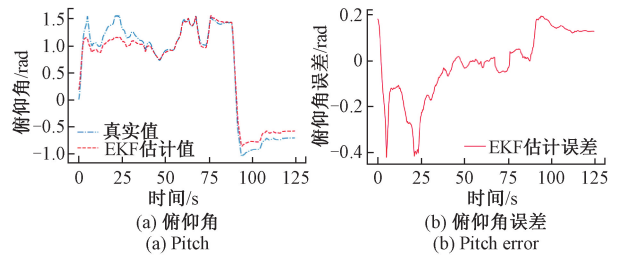


图13 俯仰角估计值及误差(10序列)

Fig. 13 Pitch angle estimation and error (sequence 10)

3.3 平移误差

KITTI数据集的09、10序列中,EKF融合算法输出的平面轨迹如图14、15所示。从图中可以看到经本文改进后的CNN算法比Depth-VO-Feat更贴近真实轨迹,表2中的位姿估计精度也可以看出改进的CNN算法的优越性。同时,CNN算法仅使用单目相机进行位姿估计,相比于单个单目相机的估计值,融合单目相机与IMU的EKF算法整体误差更小,在拐角处也能进行良好的位姿估计。

从粗虚线圆圈标记处可看到,EKF算法能对CNN算法的估计值进行有效的偏移校正。表2给出本文算法与Depth-VO-Feat算法的位姿估计精度对比, t_{err} 表示100~800 m长度上的平均平移均方根误差百分比, r_{err} 表示100~800 m长度上每100 m的平均转动均方根误差。

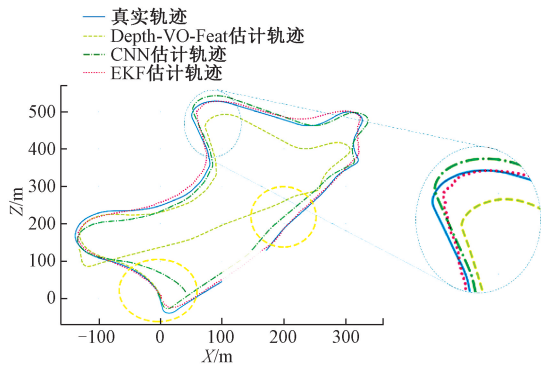


图 14 多传感器融合算法轨迹图(09 序列)

Fig. 14 Trajectory of multi-sensor fusion algorithm (sequence 09)

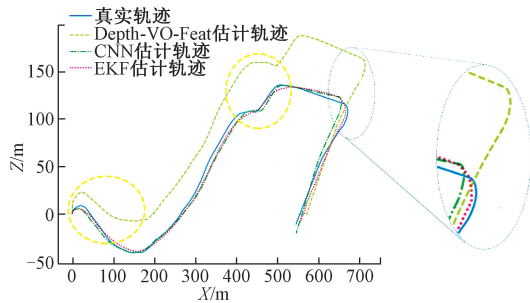


图 15 多传感器融合算法轨迹图(10 序列)

Fig. 15 Trajectory of multi-sensor fusion algorithm (sequence 10)

表 2 KITTI 数据集的位姿估计精度对比

Table 2 Comparison of pose estimation accuracy in the KITTI odometry dataset

方法	Seq. 09		Seq. 10	
	t_{err} %	$r_{err}/$ (°/100 m)	t_{err} %	$r_{err}/$ (°/100 m)
Depth-VO-Feat ^[17]	11.93	3.91	12.45	3.46
本文算法 (Monocular)	8.28	3.50	8.93	2.67
本文算法 (Monocular +IMU)	6.51	2.04	3.97	1.96

如图 16、17,展示了在 X 轴和 Z 轴方向本文 EKF 融合算法与真实位移的对比,图 16(a)表示 X 轴,图 16(b)表示 Z 轴,点线为真实值,虚线为传感器融合得到的估计值。图中可以看到, X 轴和 Z 轴的估计值与真实值的误差较小,位姿估计的全过程均能有效贴近真实值。

如图 18、19 所示,将本文所提出的方法与经典的 ORB-SLAM 算法^[34]、SfMLearner^[21]和 Depth-VO-Feat 等深度学习方法进行对比,相较于 SfMLearner、Depth-VO-Feat 这两种基于深度学习的方法,所提出的 EKF 融合方法能更好地进行位姿估计,保证位姿精度。

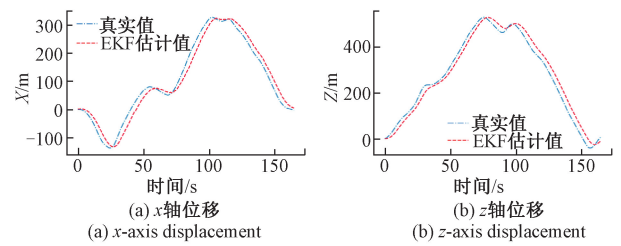
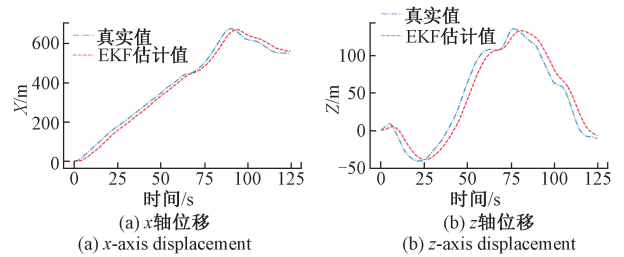
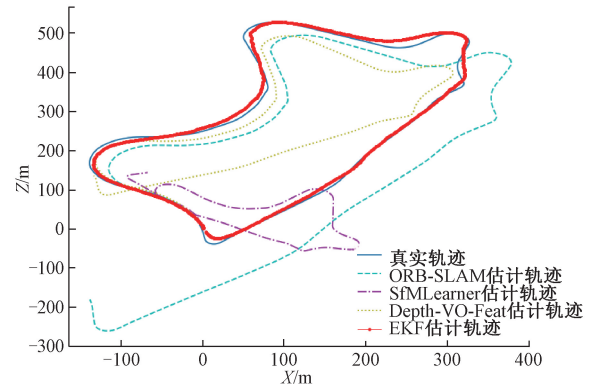
图 16 X 轴和 Z 轴方向估计值与真实位移对比图(09 序列)Fig. 16 Comparison of the proposed approach versus groundtruth in X and Z -axis directions (sequence 09)图 17 X 轴和 Z 轴方向估计值与真实位移对比图(10 序列)Fig. 17 Comparison of the proposed approach versus groundtruth in X and Z -axis directions (sequence 10)

图 18 不同算法的轨迹对比图(09 序列)

Fig. 18 Trajectory comparison of different algorithms (sequence 09)

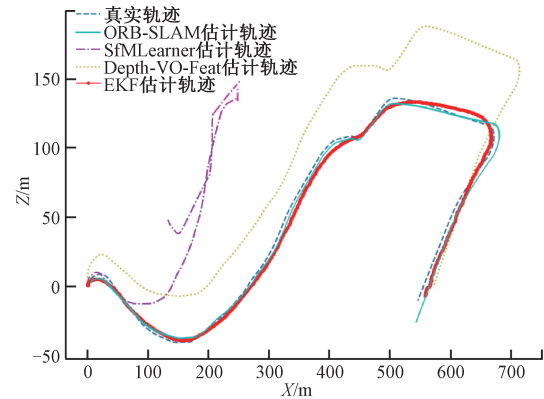


图 19 不同算法的轨迹对比图(10 序列)

Fig. 19 Trajectory comparison of different algorithms (sequence 10)

本文使用异构多传感器融合的 EKF 位姿估计方法,对 CNN 估计的单目相机位姿进行校正。图 18、19 中的本文提出的方法,相较于其它方法,在 09、10 序列中本文方法与真实轨迹的差距最小,表现远好于 SfMLearner 和 Depth-VO-Feat 方法,表明本文方法具有更高精度和实用性。

表 3 给出不同算法在 KITTI 数据集中的位姿估计精度,从表 3 可知,相比于 Depth-VO-Feat 算法,本文方法在 09 序列中的平动误差下降 45.4%,转动误差下降 47.8%;在 10 序列中的平动误差下降 68.1%,转动误差下降 43.4%。

表 3 KITTI 数据集的位姿估计精度对比

Table 3 Comparison of pose estimation accuracy in the KITTI dataset

方法	Seq. 09		Seq. 10	
	t_{err}	$r_{err}/$	t_{err}	$r_{err}/$
	%	($^{\circ}/100\text{ m}$)	%	($^{\circ}/100\text{ m}$)
ORB-SLAM ^[34]	15.30	0.26	3.68	0.48
SfMLearner ^[21]	17.84	6.78	37.91	17.78
Depth-VO-Feat ^[17]	11.93	3.91	12.45	3.46
本文算法	6.51	2.04	3.97	1.96

此外,由于实时性是衡量算法性能的一项重要指标,因此将所提出算法的处理效率与现有算法进行比较。位姿估计过程中仅使用 CPU 运行,未使用 GPU 加速,CPU 型号为 Intel Core i5-9300H。如表 4 所示,表中计算了不同算法在位姿估计过程中处理每张图像所需的运行时间,与 ORB-SLAM、SfMLearner、Geonet^[33] 算法不同的是,本文算法还需要额外处理 IMU 数据。为更加全面评估算法效率,分别在 KITTI 数据集的 09 和 10 序列中进行算法测试。从表中可以看出,尽管本文算法需要同时处理图像和 IMU 数据,但本文算法的整体运行时间仍优于其它同类算法。

表 4 KITTI 数据集的位姿估计运行时间对比

Table 4 Comparison of pose estimation runtime in the KITTI dataset

方法	Seq. 09	Seq. 10
ORB-SLAM ^[34]	0.114	0.126
SfMLearner ^[21]	0.130	0.127
Geonet ^[33]	0.097	0.098
本文算法	0.078	0.081

4 结 论

本文提出使用 EKF 算法改进基于 CNN 的位姿估计方法,能替代仅依靠视觉的定位方案,小型、低功耗、低成本的 IMU 可以完成自身位姿预测,姿态更新是从一个或多个测量图像中获取。单目相机和 IMU 的互补模态和误差特性使两者可以相互结合,以高保真度估算自身运动。其中单目图像输入 CNN 获得位姿测量量,IMU 数据作为 EKF 预测阶段的输入,通过状态更新完成数据融合,完成室外条件下移动机器人更加准确的自主定位。

参考文献

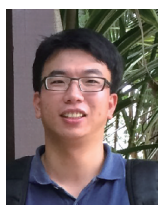
- [1] CHENG C, ZHU H, LI M, et al. A review of visual-inertial simultaneous localization and mapping from filtering-based and optimization-based perspectives[J]. Robotics, 2018, 7(3): 45.
- [2] MUR-ARTAL R, TARDOS J D. Visual-inertial monocular SLAM with map reuse[J]. IEEE Robotics and Automation Letters, 2017, 2(2): 796-803.
- [3] GAO X, ZAHNG T, LIU Y, et al. 14 lectures on visual slam[M]. Beijing: Electronic Industry Press, 2017: 141-161.
- [4] BLOESCH M, OMARI S, HUTTER M, et al. Robust visual inertial odometry using a direct EKF-based approach[C]. IEEE/RSJ International Conference on Intelligent Robots and Systems. Piscataway, USA: IEEE, 2015: 298-304.
- [5] LEUTENEGGER S, LYNEN S, BOSSE M, et al. Keyframe-based visual-inertial odometry using nonlinear optimization[J]. The International Journal of Robotics Research, 2015, 34(3): 314-334.
- [6] KALMAN R E. A new approach to linear filtering and prediction problems[J]. Transactions of the ASME-Journal of Basic Engineering, 1960, 82(1): 35-45.
- [7] MCELHOE B A. An assessment of the navigation and course corrections for a manned flyby of mars or venus[J]. IEEE Transactions on Aerospace and Electronic Systems, 1966, 2(4): 613-623.
- [8] JULIER S J, UHLMANN J K. Unscented filtering and nonlinear estimation[J]. Proceedings of the IEEE, 2004, 92(3): 401-422.
- [9] LI M, MOURIKIS A I. High-precision, consistent EKF-based visual-inertial odometry[J]. The International Journal of Robotics Research, 2013, 32(6): 690-711.
- [10] SUN K, MOHTA K, PFROMMER B, et al. Robust stereo visual inertial odometry for fast autonomous

- flight[J]. IEEE Robotics and Automation Letters, 2018, 3(2): 965-972.
- [11] LOWE D G. Distinctive image features from scale-invariant keypoints[J]. International Journal of Computer Vision, 2004, 60(2): 91-110.
- [12] SCHÖNBERGER J L, RADENOVIC F, CHUM O, et al. From single image query to detailed 3D reconstruction[C]. IEEE Conference on Computer Vision and Pattern Recognition, Piscataway, USA: IEEE, 2015: 5126-5134.
- [13] MA F, KARAMAN S. Sparse-to-dense: Depth prediction from sparse depth samples and a single image[C]. IEEE International Conference on Robotics and Automation. Piscataway, USA: IEEE, 2018: 4796-4803.
- [14] DOSOVITSKIY A, FISCHER P, ILG E, et al. FlowNet: Learning optical flow with convolutional networks[C]. IEEE International Conference on Computer Vision. Piscataway, USA: IEEE, 2015: 2758-2766.
- [15] KENDALL A, GRIMES M, CIPOLLA R. PoseNet: A convolutional network for real-time 6-dof camera relocalization[C]. IEEE International Conference on Computer Vision. Piscataway, USA: IEEE, 2015: 2938-2946.
- [16] MAYER N, ILG E, HÄUSSER P, et al. A large dataset to train convolutional networks for disparity, optical flow, and scene flow estimation[C]. IEEE Conference on Computer Vision and Pattern Recognition, Piscataway, 2016: 4040-4048.
- [17] ZHAN H, GARG R, WEERASEKERA C S, et al. Unsupervised learning of monocular depth estimation and visual odometry with deep feature reconstruction[C]. IEEE Conference on Computer Vision and Pattern Recognition. Piscataway, USA: IEEE, 2018: 340-349.
- [18] KENDALL A, CIPOLLA R. Geometric loss functions for camera pose regression with deep learning[C]. IEEE Conference on Computer Vision and Pattern Recognition, Piscataway, 2017: 6555-6564.
- [19] 潘林豪,田福庆,应文健,等. 单目相机-IMU 外参自动标定与在线估计的视觉-惯导 SLAM[J]. 仪器仪表学报, 2019, 40(6): 56-67.
- PAN L H, TIAN F Q, YING W J, et al. VI-SLAM algorithm with camera-IMU extrinsic automatic calibration and online estimation[J]. Chinese Journal of Scientific Instrument, 2019, 40(6): 56-67.
- [20] EIGEN D, PUHRSCH C, FERGUS R. Depth map prediction from a single image using a multi-scale deep network[C]. International Conference on Neural Information Processing Systems, La Jolla, USA: Neural Information Processing Systems Foundation, 2014: 2366-2374.
- [21] ZHOU T, BROWN M, SNAVELY N, et al. Unsupervised learning of depth and ego-motion from video[C]. IEEE Conference on Computer Vision and Pattern Recognition. Piscataway, 2017: 6612-6619.
- [22] RONALD C, SEN W, WEN H, et al. VINet: Visual-inertial odometry as a sequence-to-sequence learning problem[C]. Thirty-First AAAI Conference on Artificial Intelligence. Palo Alto, USA: AAAI, 2017: 3995-4001.
- [23] SHI X, CHEN Z, WANG H, et al. Convolutional LSTM network: A machine learning approach for precipitation nowcasting[M]. International Conference on Neural Information Processing Systems, Montreal, Canada: MIT Press, 2015: 802-810.
- [24] 余洪山,郭丰,郭林峰,等. 融合改进 SuperPoint 网络的鲁棒单目视觉惯性 SLAM[J]. 仪器仪表学报, 2021, 42(1): 116-126.
- YU H SH, GUO F, GUO L F, et al. Robust monocular visual-inertial SLAM based on the improved SuperPoint network[J]. Chinese Journal of Scientific Instrument, 2021, 42(1): 116-126.
- [25] CADENA C, CARLONE L, CARRILLO H, et al. Past, present, and future of simultaneous localization and mapping: Toward the robust-perception age[J]. IEEE Transactions on Robotics, 2016, 32(6): 1309-1332.
- [26] GEIGER A, LENZ P, URTASUN R. Are we ready for autonomous driving? The kitti vision benchmark suite[C]. IEEE Conference on Computer Vision and Pattern Recognition. Piscataway, USA: IEEE, 2012: 3354-3361.
- [27] FORSTER C, ZHANG Z, GASSNER M, et al. SVO: Semidirect visual odometry for monocular and multicamera systems[J]. IEEE Transactions on Robotics, 2017, 33(2): 249-265.
- [28] HU J, OZAY M, ZHANG Y, et al. Revisiting single image depth estimation: Toward higher resolution maps with accurate object boundaries[C]. IEEE Winter Conference on Applications of Computer Vision. Piscataway, USA: IEEE, 2019: 1043-1051.
- [29] LIN L, HUANG G, CHEN Y, et al. Efficient and high-quality monocular depth estimation via gated multi-scale network[J]. IEEE Access, 2020, 8: 7709-7718.
- [30] TRAWNY N, ROUMELIOTIS S. Indirect Kalman filter for 3D attitude estimation[R]. USA: University of Minnesota, Department of Computer Science and Engineering, 2005.
- [31] LI C, WASLANDER S L. Towards end-to-end learning of

visual inertial odometry with an EKF[C]. Conference on Computer and Robot Vision. Cham, Switzerland: Springer, 2020: 190-197.

- [32] WEISS S, SIEGWART R. Real-time metric state estimation for modular vision-inertial systems[C]. IEEE International Conference on Robotics and Automation, Piscataway, 2011: 4531-4537.
- [33] YIN Z, SHI J. GeoNet: Unsupervised learning of dense depth, optical flow and camera pose [C]. IEEE Conference on Computer Vision and Pattern Recognition, 2018: 1983-1992.
- [34] MUR-ARTAL-R, MONTIEL J M M, TARDOS J D. ORB-SLAM: A versatile and accurate monocular SLAM system [J]. IEEE Transactions on Robotics, 2015, 31(5): 1147-1163.

作者简介



林立雄, 2008 年于华侨大学获得学士学位, 2011 年于厦门大学获得硕士学位, 2016 年于厦门大学获得博士学位, 现为集美大学副教授, 主要研究方向为多源组合导航和无人机控制系统。

E-mail: elelinlixiong@139.com

Lin Lixiong received his B. Sc. degree from Huaqiao University in 2008, and received his M. Sc. degree and Ph. D. degree both from Xiamen University in 2011 and 2016, respectively. He is currently an associate professor at Jimei University. His main research interests include integrated navigation and UAV control.



郑佳春 (通信作者), 1986 年于大连海事大学获得学士学位, 现为集美大学教授, 主要研究方向为卫星通信导航和动态定位与跟踪。

E-mail: jchzheng@jmu.edu.cn

Zheng Jiachun (Corresponding author) received his B. Sc. degree from Dalian Maritime University in 1986. He is currently a professor at Jimei University. His main research interests include satellite communication navigation and dynamic positioning and tracking.



黄国辉, 2018 年于福建农林大学获得学士学位, 现为福州大学研究生, 主要研究方向为多源组合导航。

E-mail: guohuihang@126.com

Huang Guohui received his B. Sc. degree from Fujian Agriculture and Forestry University in 2018. He is currently a graduate student at Fuzhou University. His main research interest is integrated navigation.



蔡国玮, 2002 年于天津大学获得学士学位, 2009 年于新加坡国立大学获得博士学位, 现为福建船政交通职业学院教授, 主要研究方向为多源组合导航和无人机控制系统。

E-mail: feihu.cgw@foxmail.com

Cai Guowei received his B. Sc. degree from Tianjin University in 2002, and received his Ph. D. degree in 2009 from National University of Singapore. He is currently a professor at Fujian Chuazheng Communications College. His main research interests include integrated navigation and UAV control.