

DOI: 10.19650/j.cnki.cjsi.J2311336

高动态环境下的傅里叶梅林变换视觉 SLAM 算法*

栾添添, 吕奉坤, 班喜程, 孙明晓

(哈尔滨理工大学自动化学院 哈尔滨 150080)

摘要:针对视觉即时定位与地图构建(SLAM)中的静态假设限制其在动态现实环境中应用的问题,提出了一种在高动态环境下基于傅里叶梅林变换的视觉 SLAM 算法。首先,采用傅里叶梅林算法配准以补偿相机运动,应用帧间差分算法获取运动掩膜。同时利用短时密集连接网络进行语义分割确定潜在运动物体掩膜。结合运动掩膜与物体掩膜,获得最终的物体运动区域,剔除该区域的特征点。最后,依据稳定的静态特征点跟踪优化,提升位姿精度。实验结果表明,本算法的绝对轨迹误差与相对位姿误差相比于 ORB-SLAM2 减少了 95% 以上,相比于 DS-SLAM 减少了 30% 以上,验证其在复杂动态场景下具有良好的定位精度与鲁棒性,有效降低运动模糊与光照变化对运动检测的影响,同时,克服了传统动态 SLAM 难以检测非先验运动对象的弊端。

关键词: SLAM; 动态场景; 傅里叶梅林变换; 语义分割

中图分类号: TH85 TP242 **文献标识码:** A **国家标准学科分类代码:** 510.80

Fourier Merlin transform visual SLAM algorithm in highly dynamic environments

Luan Tiantian, Lyu Fengkun, Ban Xicheng, Sun Mingxiao

(School of Automation, Harbin University of Science and Technology, Harbin 150080, China)

Abstract: To address the limitations of the static assumption in visual SLAM for dynamic real-world applications, a visual SLAM algorithm is proposed, which is based on Fourier-Mellin transform for high dynamic environments. It involves Fourier-Mellin transform for motion compensation, employs frame differencing for motion mask generation, utilizes the short-term dense connection network for semantic segmentation to identify potential moving objects, combines motion and object masks to obtain the final object motion region, and eliminates the corresponding feature points in that region. Finally, the pose accuracy is optimized based on stable static feature points. Experimental results demonstrate a reduction of over 95% in absolute trajectory error and relative pose error compared to ORB-SLAM2, and over 30% compared to DS-SLAM. These evaluate its excellent localization accuracy and robustness in complex dynamic scenes. The impact of motion blur and lighting changes on motion detection is effectively mitigated, and the limitations of traditional dynamic SLAM in detecting non-prior motion objects are overcome.

Keywords: SLAM; dynamic scenes; Fourier-Merlin transform; semantic segmentation

0 引言

同步定位与地图构建技术(simultaneous location and mapping, SLAM)是指根据传感器数据实时构建周围环境地图,同时根据该地图推测自身的定位。基于视觉传感器的 SLAM 技术称为视觉同步定位与地图创建(visual

SLAM, VSLAM)技术^[1-2]。

在过去的 30 多年中,许多学者对 SLAM 进行了研究并取得了突出效果,如 ORB-SLAM2^[3]、RGBD-SLAM-V2^[4]等。但传统的 SLAM 大都基于静态环境假设,真实工作环境中存在动态对象,其不稳定特征点会干扰 SLAM 系统并导致性能下降^[5]。

收稿日期:2023-04-21 Received Date: 2023-04-21

* 基金项目:国家自然科学基金青年项目(62103120,51909049)、黑龙江省自然科学基金(LH2020E094)、黑龙江省博士后面上资助项目(LBH-Z22195, LBH-Z22197)、黑龙江省复合材料高效成型及智能装备技术创新中心开放课题面上项目(HPTIC202204)、黑龙江省普通本科高等学校青年创新人才培养计划项目(UNPYSCT-2020190)资助

为了解决动态对象带来的 SLAM 性能的下降的问题, Kim 等^[6]提出通过深度场景计算背景模型,通过背景减除来区分场景中的动静态对象。Sun 等^[7]由随机抽样一致性(random sample consensus, RANSAC)^[8]方法计算相邻图片之间的变换并通过背景减除获得初始运动物体区域,通过深度图进行分割以区分动态与静态对象。然而,这些传统方法易受光照变化和大面积运动对象的影响而失效。利用深度学习的方法对传统运动检测算法进行约束从而提高运动检测算法的鲁棒性变成了可能^[9]。Yu 等^[10]提出了 DS-SLAM 方案, DS-SLAM 先利用 RANSAC 算法计算离群点从而得到动态特征点,再利用 SegNet^[11]获得语义信息和运动特征点过滤动态部分,从而提高姿态估计的准确性。Bescos 等^[12]提出 DynaSLAM 利用 Mask-RCNN^[13]分割出所有可能的先验动态对象,然后根据深度差与角度差信息剔除所有的非先验运动对象。Zhong 等^[14]的 Detect-SLAM,首次将视觉 SLAM 与多目标检测算法集成,实现动静态特征点的区分。深度学习虽然改善了先验运动物体在运动场景中的漏检误检问题,但是当场景中潜在运动对象占据较大比例时,直接移除所有潜在运动对象会导致特征点过少,从而导致轨迹丢失^[15]。同时,深度学习需要大量的计算资源,在实时性要求较高的 SLAM 问题中,使用基于深度学习的运动检测方法并不是一个理想的选择。

考虑到 SLAM 场景的运动检测与传统的运动检测不同,其相机本身也在运动,因此需要对相机输入的前后两

帧进行配准,以消除相机自运动带来的影响^[16]。相机配准主要分为基于特征与频域两种^[17]。基于特征的图像配准算法通过提取图像中的点、线、边缘等特征信息,并计算特征描述子来实现配准^[18]。比较经典的有 Yi 等^[19]的 SR-SIFT,其提出了一个尺度变换标准,以提高 SIFT 特征点^[20]在图像配准时精度。但是基于特征点的图像配准在环境出现光照变化、遮挡与运动模糊时缺乏鲁棒性。基于频域的图像配准方法的主要特征为将图像先通过傅里叶变换转换为频域,再通过相位相关等操作实现配准。比较经典的有 Reddy 等^[21]提出的通过傅里叶梅林变换(Fourier-Merlin transform, FMT)来计算图像间的平移、旋转与缩放。在环境光照变化、遮挡导致的特征缺失的环境中,基于频域的图像配准方法被证明比基于特征的图像配准方法更准确和更快^[22]。

本文利用 FMT 在图像配准方面的独特优势结合帧间差分(temporal difference, TD)算法,实现了高鲁棒的运动检测,结合 ORB-SLAM2 和实时语义分割(short-term dense concatenate, STDC)网络,提出了一种在动态场景下基于 FMT 的视觉 SLAM 算法。

1 基于 FMT 的视觉 SLAM

1.1 算法框架

选择 ORB-SLAM2 作为主体框架,如图 1 所示。

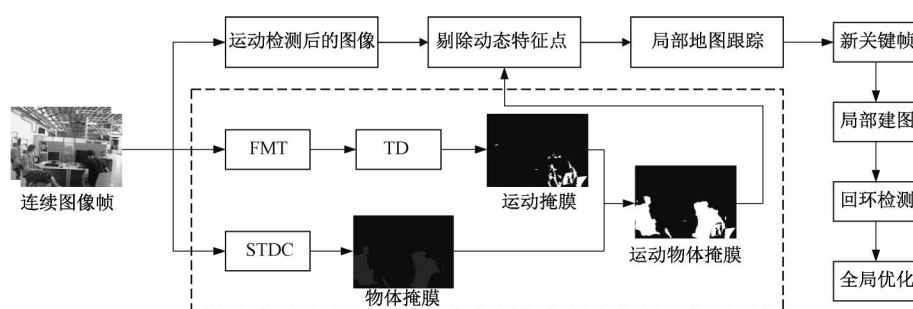


图 1 SLAM 系统的总体流程图

Fig. 1 General flow diagram of the SLAM system

1) 系统将深度相机获得的彩色图同时输入 ORB-SLAM2 的跟踪线程与运动检测线程中。对输入图像提取 ORB 特征点,并等待运动检测线程获取运动物体区域。使用傅里叶-梅林方法对当前帧与上一帧配准,根据配准结果对当前帧进行变换,实现运动补偿。

2) 将运动补偿后的图像与上一帧图像进行帧间差分与阈值化处理,得到最终的运动掩膜。在语义分割线程中,将当前帧输入 STDC 模型,得到物体掩膜。将运动掩膜与物体掩膜相结合,得到运动物体区域,该区域输入 ORB-SLAM2 的跟踪线程中。

3) 跟踪线程得到了运动物体区域后,对当前帧中落在运动物体区域的特征点进行剔除,利用静态点估计位姿,将位姿信息与关键帧信息传送至局部建图线程以及回环检测线程,利用光束平差法(bundle adjustment, BA)对地图点和位姿信息进行局部与全局优化。

1.2 运动检测线程

1) 图像配准

图像配准是变换图像以使其与参考图像对准的过程, FMT 最早是由 Reddy 等^[21]用于图像配准,其为相位相关法的一种扩展,可以同时恢复图像间的旋转、尺度与

平移变换。该方法的流程如图 2 所示。其中 DFT 为离散傅里叶变换, HPF 为高通滤波, LPT 为对数极坐标变

换, CPS 为相位相关操作, IDFT 为逆傅里叶变换, argmax 为返回最大值索引。

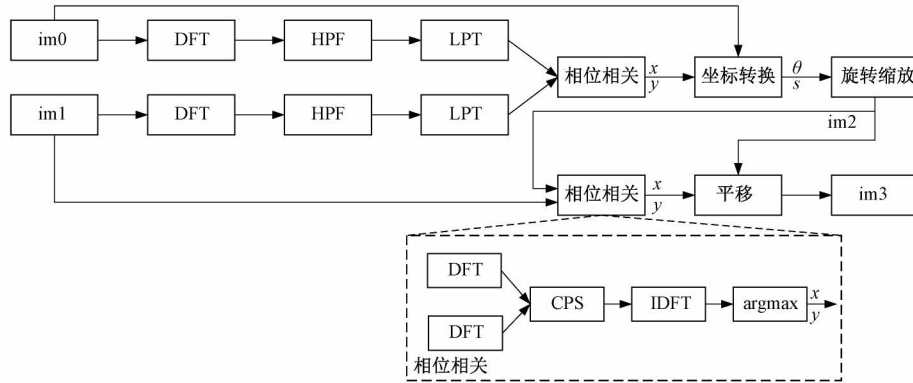


图 2 FMT 图像配准流程

Fig. 2 Flow of FMT image alignment

使用 FMT 进行图像配准包含两个步骤:估计旋转和缩放与估计平移。在第 1 步中,相位相关操作的输入是要配准的灰度图像 $im0$ 和 $im1$ 的离散傅里叶变换后又对数极坐标变换后的图像。对数极坐标图像 (x, y) 的估计位移被转换为旋转 θ 与尺度 s , 并用于变换 $im0$ 。对 $im0$ 变换后得到 $im2$, $im2$ 与 $im1$ 是第 2 个相位相关操作的输入。来自第 2 个相位相关操作的估计位移被用于再次变换 $im2$ 得到 $im3$ 。 $im3$ 即经过旋转、缩放、平移后与 $im1$ 对齐的 $im0$ 图像。

假设参考图像为 $f(x, y)$ 与待配准图像 $g(x, y)$, 它们之间存在相对平移 $(\Delta x, \Delta y)$ 、相对旋转 $\Delta\theta$ 、相对缩放 s 则其关系如式(1)所示。

$$g(x, y) = f(s(x \cos \Delta\theta + y \sin \Delta\theta) - \Delta x, s(-x \sin \Delta\theta + y \cos \Delta\theta) - \Delta y) \quad (1)$$

其经过傅里叶变换在频域中的表示为:

$$G(u, v) = e^{-2\pi i(u\Delta x + v\Delta y)} \cdot s^{-2} F[s^{-1}(u \cos \Delta\theta + v \sin \Delta\theta), s^{-1}(-u \sin \Delta\theta + v \cos \Delta\theta)] \quad (2)$$

其中, $G(u, v)$ 与 $F(u, v)$ 分别为 $g(x, y)$ 与 $f(x, y)$ 的傅里叶变换。为了减少低频噪声对配准的影响, 对 $G(u, v)$ 与 $F(u, v)$ 分别执行高通滤波, 设滤波函数为 $H(u, v)$, 截止频率 $D_0 = 30$, 有:

$$H(u, v) = \begin{cases} 0, & \sqrt{u^2 + v^2} < D_0 \\ 1, & \sqrt{u^2 + v^2} > D_0 \end{cases} \quad (3)$$

$$\begin{aligned} G(u, v) &= H(u, v) \cdot G(u, v) \\ F(u, v) &= H(u, v) \cdot F(u, v) \end{aligned} \quad (4)$$

对执行完高通滤波的 $G(u, v)$ 与 $F(u, v)$ 取模, 以分离相对平移分量避免其对旋转与缩放估计的影响:

$$m^{(G)}(u, v) = s^{-2} m^{(F)}[s^{-1}(u \cos \Delta\theta + v \sin \Delta\theta), s^{-1}(-u \sin \Delta\theta + v \cos \Delta\theta)] \quad (5)$$

其中, $m^{(G)}$ 与 $m^{(F)}$ 分别为 $G(u, v)$ 与 $F(u, v)$ 的幅值

谱。将其转换为对数极坐标下的表示:

$$m_{LP}^{(G)}(\lambda, \theta) = s^{-2} m_{LP}^{(F)}(\lambda - \Delta\lambda, \theta - \Delta\theta) \quad (6)$$

其中, λ 为极坐标中半径的对数, $\Delta\lambda = \log s$, 这样, 频域中由旋转缩放引入的乘积项被转换成了加减项。因此, 可以利用相位相关操作来对旋转与缩放进行估计。即:

$$Q_{LP}(\alpha, \beta) = \frac{\overline{M_{LP}^{(F)}(\alpha, \beta)} \cdot M_{LP}^{(G)}(\alpha, \beta)}{|M_{LP}^{(F)}(\alpha, \beta) \cdot M_{LP}^{(G)}(\alpha, \beta)|} e^{-2\pi i(\alpha\Delta\lambda + \beta\Delta\theta)} \quad (7)$$

其中, Q_{LP} 称为 FMT 的互功率谱, $M_{LP}^{(G)}$ 和 $M_{LP}^{(F)}$ 分别为 $m_{LP}^{(G)}$ 与 $m_{LP}^{(F)}$ 的 FMT, 也称为 $g(x, y)$ 与 $f(x, y)$ 的傅里叶梅林变换, 上划线表示共轭。对该互功率谱进行傅里叶逆变换, 得到互相关函数 C_{LP} :

$$C_{LP}(\lambda, \theta) = \mathcal{F}^{-1}[Q_{LP}(\alpha, \beta)] \approx \delta(\lambda - \Delta\lambda, \theta - \Delta\theta) \quad (8)$$

其中, 互相关函数 C_{LP} 理论上是个脉冲函数, 其在脉冲点 (λ, θ) 处表示 $f(x, y)$ 相对 $g(x, y)$ 的相对缩放因子的对数与相对旋转, 其相对缩放因子为 $s = e^\lambda$ 。

在消除旋转与缩放后, 对于图像中的相对平移 $(\Delta x, \Delta y)$, 其在频域中的表示为:

$$G(u, v) = e^{-2\pi i(u\Delta x + v\Delta y)} \cdot F(u, v) \quad (9)$$

其中, i 为虚数单位, 对其进行相位相关操作, 就得到了相对平移的估计值。图像配准算法伪代码如算法 1 所示。

对图像进行配准的示例如图 3 所示, 其中, 图 3(a) 与 (b) 为待配准图像 $im0$ 与 $im1$, 图 3(c) 为配准融合图。为了能够更明显的展示配准的效果, 选择了 TUM 数据集 freiburg3_walking_rpy 序列的第 n 帧, 以及第 $n+30$ 帧作为配准的输入图像, 并将完成变换的配准图 $im3$ 与 $im1$

算法1 傅里叶梅林变换算法

```

1: function FOURIER MELLIN TRANSFORM
   ALGORITHHE(im0, im1)
2: F_im0 = DFT(im0)
3: F_im1 = DFT(im1)
4: filtered_F_im0 = HPF(F_im0)
5: filtered_F_im1 = HPF(F_im1)
6: LP_im0 = LPT(filtered_F_im0)
7: LP_im1 = LPT(filtered_F_im1)
8: (x, y) = 相位相关(LP_im1, LP_im0)
9: rotation_angle, scaling_factor = 坐标转换(x, y)
10: im2 = 旋转缩放(im0, rotation_angle, scaling_factor)
11: (x, y) = 相位相关(im1, im2)
12: im3 = 平移(im2, x, y)
13: return im3
14: end function
15: function 相位相关(im1, im0)
16: F_im0 = DFT(im0)
17: F_im1 = DFT(im1)
18: Q_LP = CPS(F_im0, F_im1)
19: C_LP = IDFT(Q_LP)
20: x, y = argmax(C_LP)
21: return x, y
22: end function

```



(a) 待配准图 *im0* (b) 待配准图 *im1* (c) 配准图
(a) Reference image *im0* (b) Reference image *im1* (c) Registered image

图3 图像配准示例图

Fig. 3 Example diagram of image alignment

进行融合。对于大部分静态物体,如显示器、桌子与货架等,改进的FMT能够准确的进行配准,对于动态对象,该变换也能够较好的对其动态对象的静态部分。这表明改进的FMT能够在环境中存在动态对象时较好的完成了配准任务。

2) 运动检测

使用帧间差分算法进行运动检测,并结合阈值化、边缘检测和轮廓聚类等操作最终输出一张运动掩膜。

输入帧间差分的图像包括 *im1* 和已经完成配准的 *im3* 图像,通过计算 *im3* 与 *im1* 之间的像素强度的绝对差来执行帧间差分,可以得到差值图像。帧间差分的公式

如式(10)所示,其中, $D_i(x, y)$ 为第 i 帧差分图像, $f_i(x, y)$ 为第 i 帧图像,即 $im1$, $f_{i+1}(x, y)$ 为第 $i+1$ 帧图像,即 $im3$ 。

$$D_i(x, y) = |f_i(x, y) - f_{i+1}(x, y)| \quad (10)$$

对差值图像进行二值化,在阈值以上的像素值设置为 255(白色),以下设置为 0(黑色)。即:

$$R_i(x, y) = \begin{cases} 255, & D_i(x, y) > T \\ 0, & D_i(x, y) < T \end{cases} \quad (11)$$

其中, $R_i(x, y)$ 为第 i 帧的阈值图, T 为二值化阈值。经多轮实验,在阈值设置为 40 时,本算法效果最佳。

为了消除配准误差导致的阈值图像中存在部分静止物体的残像,对阈值图进行边缘检测以构造边缘掩膜,并依据该掩膜剔除静止物体残像。首先对阈值图应用 Canny 边缘检测构造边缘掩膜,将定向矩形框拟合到轮廓,若矩形框的纵横比非常低,则将其分类为边缘,并对其进行剔除。剔除残像后的阈值图即为最终的运动掩膜。运动检测算法伪代码如算法 2 所示。

算法2 运动检测算法

```

Input: Current frame, im1; Alignment frame, im3
Output: Motion masks, motion Mask
1 begin
2   Difference image,  $D = \text{abs}(im1 - im3)$ 
3   for each pixel position (i, j) in D do
4     if  $D(i, j) \geq T$  then
5        $D(i, j) = 255$ 
6     else
7        $D(i, j) = 0$ 
8     end if
9   end for
10  CannyOutput = Canny(D)
11  ContoursList = findContours(CannyOutput)
12  filteredContours = []
13  for each contour in ContoursList do
14    rect = minAreaRect(contour)
15    aspectRatio = CalculateAspectRatio(rect)
16    if aspectRatio > threshold then
17      filteredContours.push_back(contour)
18    end if
19  end for
20  motionMask = createNewMask(im1.size)
21  for each contour in filteredContours do
22    motionMask = AreaSetOne(contour, motionMask)
23  end for
24 end

```

对图像进行帧间差分的示例图如图4所示,其中,图4(a)~(c)为输入的RGB图像,图4(d)~(f)为经过配准与前帧进行帧间差分后得到的运动掩膜图。可以看到,尽管有部分噪音,但仍然较好的完成了运动检测的要求。

1.3 语义分割线程

使用傅里叶梅林算法与帧间差分进行运动检测可能会存在运动目标边界不清晰、对运动目标检测不完整的问题,因此还需要语义分割线程来对其进行完善。

在语义分割线程中本文使用轻量级语义分割网络STDC网络,STDC网络由文献[23]提出,其在保证分割精度的情况下,仍能够满足实时性的要求。与主流的双边分割网络^[24]不同的是,STDC网络设计了一种新的短时密集连接模块来提取具有可扩展感受野和多尺度信息的深度特征,从而实现了目前阶段在精度与时间上的最佳平衡。

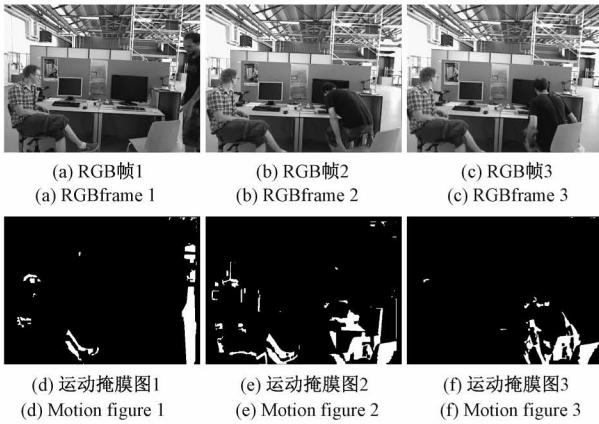


图4 帧间差分示例图

Fig. 4 Example diagram of inter-frame difference

根据STDC网络获得的物体掩膜与运动检测线程中的运动掩膜对物体进行运动判定,假设物体掩膜中第*i*个对象的总像素个数为 M_i 个,运动掩膜对应区域的总像素个数为 F_i 个,可以得到该物体的运动概率 ρ_i :

$$\rho_i = \frac{F_i}{M_i} \quad (12)$$

设置一个阈值 ε ,若运动概率 ρ_i 大于阈值 ε 则将该物体视为运动物体,并将其传入跟踪线程对落在其范围内的特征点进行剔除,否则,视为静态物体。经过多轮实验,阈值 ε 设置为0.1,本算法效果最好。

未在数据集中出现的非先验物体,仍要考虑其对于SLAM的影响,因此,对于落在运动掩膜中的特征点,仍执行剔除操作。即对于动态特征点的判定,只需要其落在运动掩膜中或落在运动物体掩膜中即可。

图5为对图像进行语义分割后与运动掩膜融合所得的运动物体掩膜示例图。其中,图5(a)为运动掩膜,图5(b)为进行语义分割后得到的物体掩膜,图5(c)为二者融合后得到的运动物体掩膜,而图5(d)为基于运动物体掩膜剔除特征点后得到的特征点提取图。

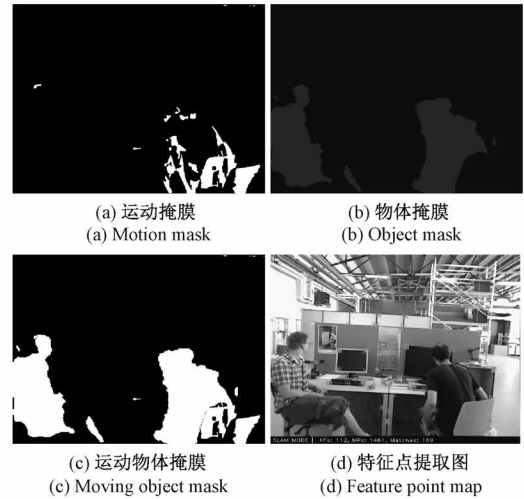


图5 运动物体掩膜示例图

Fig. 5 Example diagram of moving object mask

由于使用仅包含人类的自建数据集对STDC网络进行训练,因此语义分割算法仅能对人类进行分割。但因为运动掩膜的存在,本算法也能对非先验的运动物体,如椅子,这有效的保证本算法的鲁棒性。

2 实验验证

2.1 实验环境与实验数据集

算法测试环境为Ubuntu20.04,CPU为Intel Core i7-12700H,显卡为Nvidia GeForce RTX3060,内存为16 GB,本文基于TUM数据集对算法进行测试。

本文选取TUM数据集中5个经典的动态场景进行测试,分别为freiburg3_sitting_static、freiburg3_walking_static、freiburg3_walking_halfsphere、freiburg3_walking_rpy、freiburg3_walking_xyz,其中,freiburg3为传感器名称,sitting与walking为场景中人的动作,static、halfsphere、rpy、xyz为传感器采集数据时的运动状态。对于场景中对象运动程度较大的序列,称为高动态序列,反之,称为低动态序列。对于上述5个序列,freiburg3_sitting_static为低动态序列,其余4个序列为高动态序列。为方便起见,本文用fr3表示传感器名称,s与w表示场景中的人的动作.static、half、rpy、xyz为传感器采集数据时的运动状态。

2.2 算法评价指标

本算法使用绝对轨迹误差 (absolute trajectory error, ATE) 和相对位姿误差 (relative pose error, RPE) 来评估性能,采用均方根误差 (root mean square error, RMSE) 和标准差 (standard deviation, SD) 作为评价指标。绝对轨迹误差指估计位姿与真实位姿之间的直接差值,反映算法的精度和轨迹的全局一致性。相对位姿误差则主要描述固定时间间隔内两帧之间位姿误差的精度,反映算法在平移和旋转方面的漂移情况。

2.3 定量分析

本算法基于 ORB-SLAM2 开发,与其进行比较可以很好的验证本算法性能。作为动态 SLAM 领域最为经典的工作之一,DS-SLAM 是本算法的另一个比较对象。本算法与 ORB-SLAM2、DS-SLAM 算法绝对轨迹误差比较结果如表 1 和 2 所示。值得注意的是,表 2 中的改进结果是由表中对比较算法对应的指标值减去本算法的指标值,再除以被比较算法的指标值得来的。

表 1 绝对轨迹误差对比结果
Table 1 Comparison results for absolute trajectory error

序列	ORB-SLAM2		DS-SLAM		本算法	
	RMSE	SD	RMSE	SD	RMSE	SD
fr3/s/static	0.007 941	0.003 906	0.006 051	0.003 162	0.006 523	0.003 288
fr3/w/static	0.337 815	0.124 654	0.009 313	0.006 131	0.007 780	0.004 305
fr3/w/half	0.896 690	0.336 069	0.034 131	0.020 673	0.020 509	0.009 639
fr3/w/rpy	0.756 718	0.390 958	0.042 984	0.032 459	0.026 988	0.014 872
fr3/w/xyz	0.825 030	0.328 485	0.020 571	0.012 721	0.013 303	0.007 258

表 2 绝对轨迹误差的改进结果
Table 2 Improved results of ATE %

序列	ORB-SLAM2		DS-SLAM	
	RMSE	SD	RMSE	SD
fr3/s/static	17.86	15.82	-7.80	-3.98
fr3/w/static	97.70	96.55	16.46	29.78
fr3/w/half	97.71	97.13	39.91	53.37
fr3/w/rpy	96.43	96.20	37.21	54.18
fr3/w/xyz	98.39	97.79	35.33	42.94

从表 1 与 2 中可以看出,本算法的绝对轨迹误差的均方根误差与标准差都明显优于 ORB-SLAM2 与 DS-SLAM,本算法在高动态序列中具有更高的精度与更紧凑的误差分布。在 fr3/w/xyz 序列中,本算法的均方根误差与标准差相对于 ORB-SLAM2 降低了约 98.39% 和 97.79%,相对于 DS-SLAM 降低了约 35.33% 和 42.94%。在 fr3/w/half 序列中,本算法的均方根误差于标准差相对于 ORB-SLAM2 降低了约 97.71% 和 97.13%,相对于 DS-SLAM 降低了约 39.91% 和 53.37%。在动态场景下,本算法具有更好的鲁棒性。

但在低动态序列中,本算法的绝对轨迹误差的均方根误差与标准差相对于 ORB-SLAM2 降低仅为 17.86% 和 15.82%,相对于 DS-SLAM 提高了 7.80% 与 3.98%,这是因为在低动态序列中,运动对象并不是时刻运动,导致

运动对象静止时其特征点被用于定位,在运动时被剔除,影响了系统的精度。

表 3~5 所示为与 ORB-SLAM2 与 DS-SLAM 算法的相对位姿误差的对比结果。在高动态序列中,本算法平移部分相对位姿误差的均方根误差较 ORB-SLAM2 均降低 98% 以上,较 DS-SLAM 均降低 60% 以上,标准差较 ORB-SLAM2 均降低 97% 以上,较 DS-SLAM 均降低 60% 以上。在低动态序列中,本算法平移部分相对位姿误差的均方根误差较 ORB-SLAM2 均降低 35% 以上,标准差较 ORB-SLAM2 均降低 15% 以上。

表 3~5 更进一步表明了本算法在高动态序列中平移部分的贡献大于旋转部分。而在低动态序列中,DS-SLAM 在均方根误差与标准差上表现更好,但实际上,这 3 个算法的结果非常接近。综上所述,本算法在高动态序列下相较于 ORB-SLAM2 与 DS-SLAM 有显著提升,但在低动态序列中提升效果则不太显著。

与 ORB-SLAM2 和 DS-SLAM 相比,本算法可以显著提高动态场景的定位精度。对于高动态场景,本算法相较于 ORB-SLAM2 提高 95% 以上的精度,相较于 DS-SLAM 提高了 30% 以上。

2.4 定性分析

为了进一步评估系统,选取 fr3/w/xyz 与 fr3/w/half 两个序列进行比较。如图 6 和 7 所示,虚线为相机的真实轨迹,实线表示本算法的估计轨迹,点横线表示 DS-SLAM 算法的估计轨迹,点线表示 ORB-SLAM2 算

表 3 相对位姿误差平移部分对比结果
Table 3 Comparison results for the translational part of the relative pose error

序列	ORB-SLAM2		DS-SLAM		本算法	
	RMSE	SD	RMSE	SD	RMSE	SD
fr3/s/static	0.012 103	0.005 840	0.008 949	0.004 296	0.010 603	0.005 056
fr3/w/static	0.486 938	0.341 487	0.013 437	0.008 197	0.004 914	0.002 927
fr3/w/half	1.342 842	0.779 283	0.048 477	0.026 282	0.013 843	0.007 477
fr3/w/rpy	1.130 769	0.618 580	0.060 520	0.042 995	0.019 688	0.012 912
fr3/w/xyz	1.063 109	0.607 480	0.031 376	0.019 606	0.012 251	0.007 088

表 4 相对位姿误差旋转部分对比结果
Table 4 Comparison results for the rotational part of the relative pose error

序列	ORB-SLAM2		DS-SLAM		本算法	
	RMSE	SD	RMSE	SD	RMSE	SD
fr3/s/static	0.357 612	0.154 609	0.319 432	0.143 304	0.339 546	0.149 898
fr3/w/static	8.783 950	6.138 754	0.336 851	0.179 199	0.163 276	0.089 555
fr3/w/half	29.877 778	17.363 695	1.083 718	0.555 097	0.419 055	0.221 003
fr3/w/rpy	20.939 665	12.482 105	1.321 652	0.947 359	0.487 972	0.295 657
fr3/w/xyz	20.158 381	11.361 161	0.765 591	0.506 155	0.392 613	0.275 395

表 5 相对位姿误差的改进结果
Table 5 Improved results of RPE

序列	平移部分				旋转部分			
	ORB-SLAM2		DS-SLAM		ORB-SLAM2		DS-SLAM	
	RMSE	SD	RMSE	SD	RMSE	SD	RMSE	SD
fr3/s/static	12.39	13.42	-18.48	-17.69	5.05	3.05	-6.30	-4.60
fr3/w/static	98.99	99.14	63.43	64.29	98.14	98.54	51.53	50.02
fr3/w/half	98.97	99.04	71.44	71.55	98.60	98.73	61.33	60.19
fr3/w/rpy	98.26	97.91	67.47	69.97	97.67	97.63	63.08	68.79
fr3/w/xyz	98.85	98.83	60.95	63.85	98.05	97.58	48.72	45.59

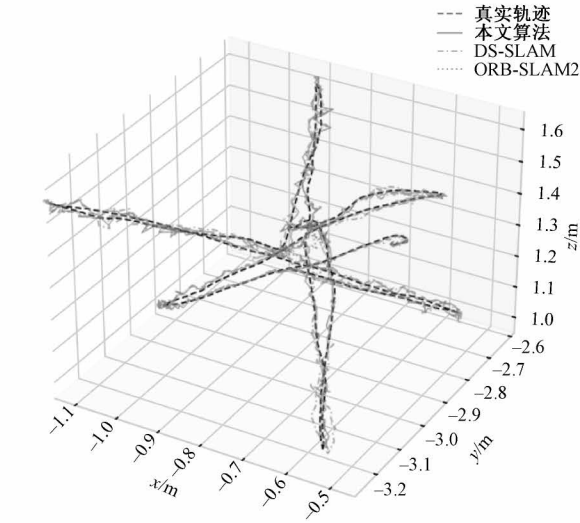


图 6 在 fr3/w/xyz 序列下的轨迹图
Fig. 6 Trajectory plot under fr3/w/xyz sequence

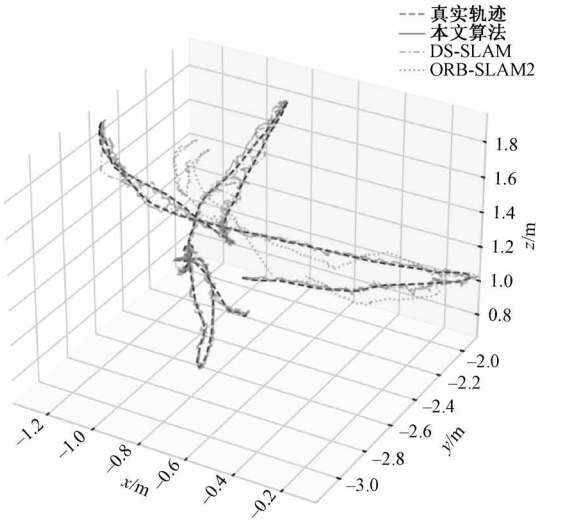


图 7 在 fr3/w/half 序列下的轨迹图
Fig. 7 Trajectory plot under fr3/w/half sequence

法的估计轨迹。在高动态序列下,ORB-SLAM2 算法的轨迹漂移较大,而 DS-SLAM 与本算法能够保持轨迹的稳定。在图 6 的下部与图 7 的上部,可以看到 DS-SLAM 出现了较大的漂移,这是由于光照变化和运动模糊导致 DS-SLAM 的运动检测算法失效所致。本算法克服了光照与运动模糊的影响并取得良好效果。

图 8~9 为 DS-SLAM 与本算法在 fr3/w/xyz 与 fr3/w/half 序列下的绝对轨迹误差。DS-SLAM 在 20 s 左右误差的峰值在 0.14,而本算法在同时刻的误差峰值只有 0.01。本算法与真实轨迹的误差更小,在 DS-SLAM 误差最大的时刻,本算法对应的误差较小,说明本算法提高了高动态序列下的精度与鲁棒性。

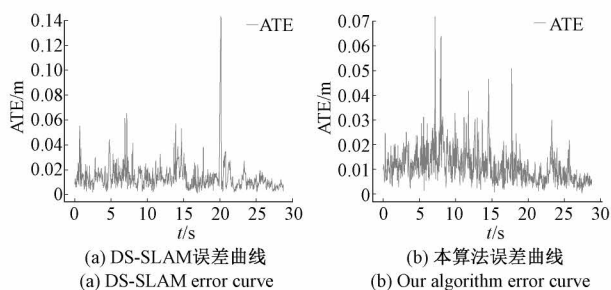


图 8 在 fr3/w/xyz 序列下 DS-SLAM 与本算法的绝对轨迹误差曲线

Fig. 8 Absolute trajectory error profile of DS-SLAM with this paper's algorithm at fr3/w/xyz sequence

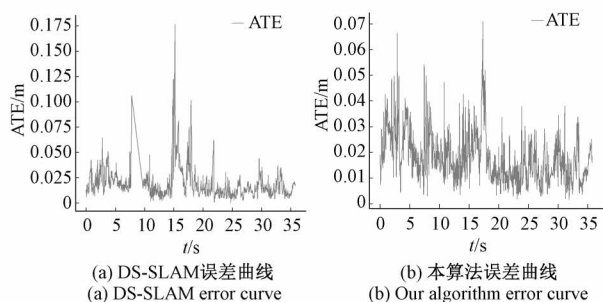


图 9 在 fr3/w/half 序列下 DS-SLAM 与本算法的绝对轨迹误差曲线

Fig. 9 Absolute trajectory error profile of DS-SLAM with this paper's algorithm at fr3/w/half sequence

2.5 复杂环境下的掩码提取实验

为验证算法在高动态场景中出现运动模糊和光照变化时的定位与建图效果,本文设计了一个出现运动模糊和光照变化的环境,并使用手持深度相机进行了实验。相机型号为 Intel RealSense Camera D415,RGB 图分辨率为 $1\,920 \times 1\,080$,深度视场为 $65^\circ \times 40^\circ$,深度范围为 0.16~10 m,深度图的分辨率为 $1\,280 \times 720$ 。

图 10 展示了一组具有代表性的运动模糊场景下的实验结果。其中,图 10(a)和(b)为输入的 RGB 图像,图 10(c)为其对应的动态掩码。当环境中存在如图 10(b)所示运动模糊时,本算法仍可精确地将动态与静态区域区分开来,如图 10(c)所示。



图 10 动态模糊下的掩码提取

Fig. 10 Mask extraction under dynamic blur

图 11 展示了一组在光照变化场景下的实验结果,图 11(a)和(b)为输入的 RGB 图像,图 11(c)为其对应的动态掩码。其中,图 11(a)为光照较为充足条件下的 RGB 图像,图 11(b)为光照较为缺乏条件下的 RGB 图像。可以看到,当环境中出现如图 11(a)和(b)所示的光照变化时,本文算法尽管存在些许噪音,但仍可精确的对动态与静态区域进行分割,如图 11(c)所示。



图 11 光照变化下的掩码提取

Fig. 11 Mask extraction under illumination changes

由于本文优化了动态掩码的提取策略,因此相较于 DS-SLAM 使用 RANSAC 算法计算离群点从而得到动态特征点,本文算法并不需要静态区域占据主要区域或感兴趣区域的假设。经验证,本文算法可以实现在动态模糊区域占图像大部分区域或场景发生剧烈光照变化的情况下准确地提取动态区域。

3 结 论

本文提出了一种基于 ORB-SLAM2 与 FMT 的动态 SLAM 算法。主要贡献如下:

1)通过改进的 FMT 对图像进行配准以实现运动补偿,使用帧间差分获得运动掩膜,降低了运动模糊与光照变化对运动检测的影响。

2)结合运动检测与语义分割,提出了一种动态特征点过滤方法,可以有效地剔除动态对象对位姿估计与建图的干扰。

3)通过 TUM 数据集的 5 个动态序列进行实验。在高动态序列中,本算法的绝对轨迹误差相比于 ORB-

SLAM2 平均减少了 95% 以上,相比于 DS-SLAM 平均减少了 30% 以上。这表明了本算法在动态环境下具有较高的准确性与鲁棒性。

未来考虑物体 SLAM 的思想,对动态对象也进行位姿估计与跟踪以增加约束条件,提高 SLAM 精度。

参考文献

- [1] LIU Y, MIURA J. RDS-SLAM: Real-time dynamic SLAM using semantic segmentation methods[J]. IEEE Access, 2021, 9: 23772-23785.
- [2] 冯明驰,刘景林,李成南,等. 一种多焦距动态立体视觉 SLAM[J]. 仪器仪表学报, 2021, 42(11): 200-209.
- FENG M CH, LIU J L, LI CH N, et al. A multi-focal length dynamic stereo vision SLAM[J]. Chinese Journal of Scientific Instrument, 2021, 42(11): 200-209.
- [3] MUR-ARTAL R, TARDÓS J D. ORB-SLAM2: An open-source SLAM system for monocular, stereo, and RGB-D cameras[J]. IEEE Transactions on Robotics, 2017, 33(5): 1255-1262.
- [4] ENDRES F, HESS J, STURM J, et al. 3D mapping with an RGB-D camera[J]. IEEE Transactions on Robotics, 2013, 30(1): 177-187.
- [5] 刘钰嵩,何丽,袁亮,等. 动态场景下基于光流的语义 RGBD-SLAM 算法[J]. 仪器仪表学报, 2022, 43(12): 139-148.
- LIU Y S, HE L, YUAN L, et al. Semantic RGBD-SLAM in dynamic scene based on optical flow[J]. Chinese Journal of Scientific Instrument, 2022, 43(12): 139-148.
- [6] KIM D H, KIM J H. Effective background model-based RGB-D dense visual odometry in a dynamic environment[J]. IEEE Transactions on Robotics, 2016, 32(6): 1565-1573.
- [7] SUN Y, LIU M, MENG M Q H. Improving RGB-D SLAM in dynamic environments: A motion removal approach[J]. Robotics and Autonomous Systems, 2017, 89: 110-122.
- [8] 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.
- [9] 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 incorporating improved SuperPoint network[J]. Chinese Journal of Scientific Instrument, 2021, 42(1): 116-126.
- [10] YU C, LIU Z, LIU X J, et al. DS-SLAM: A semantic visual SLAM towards dynamic environments[C]. International Conference on Intelligent Robots and Systems (IROS), 2018: 1168-1174.
- [11] BADRINARAYANAN V, KENDALL A, CIPOLLA R. Segnet: A deep convolutional encoder-decoder architecture for image segmentation[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017, 39(12): 2481-2495.
- [12] BESCOS B, FÁCIL J M, CIVERA J, et al. DynaSLAM: Tracking, mapping, and inpainting in dynamic scenes[J]. IEEE Robotics and Automation Letters, 2018, 3(4): 4076-4083.
- [13] HE K, GKIOXARI G, DOLLÁR P, et al. Mask R-CNN[C]. 16th IEEE International Conference on Computer Vision (ICCV), 2017: 2961-2969.
- [14] ZHONG F, WANG S, ZHANG Z, et al. Detect-SLAM: Making object detection and SLAM mutually beneficial[C]. IEEE Winter Conference on Applications of Computer Vision (WACV), 2018: 1001-1010.
- [15] BESCOS B, CAMPOS C, TARDÓS J D, et al. DynaSLAM II: Tightly-coupled multi-object tracking and SLAM[J]. IEEE Robotics and Automation Letters, 2021, 6(3): 5191-5198.
- [16] CHAPEL M N, BOUWMANS T. Moving objects detection with a moving camera: A comprehensive review[J]. Computer Science Review, 2020, 38: 100310.
- [17] XU Q, KUANG H, KNEIP L, et al. Rethinking the Fourier-Mellin transform: Multiple depths in the camera's view[J]. Remote Sensing, 2021, 13(5): 1000.
- [18] 余联想,郑明魁,欧文君等. 多传感器融合的移动机器人室外激光 SLAM 算法优化与系统实现[J]. 电子测量与仪器学报, 2023, 37(2): 48-55.

- YU L X, ZHENG M K, OU W J, et al. Optimization and system implementation of outdoor lidar SLAM algorithm for mobile robots based on multi-sensor fusion[J]. *Journal of Electronic Measurement and Instrumentation*, 2023, 37 (2): 48-55.
- [19] YI Z, ZHIGUO C, YANG X. Multi-spectral remote image registration based on SIFT [J]. *Electronics Letters*, 2008, 44(2): 1-2.
- [20] LOWE D G. Distinctive image features from scale-invariant keypoints[J]. *International Journal of Computer Vision*, 2004, 60: 91-110.
- [21] REDDY B S, CHATTERJI B N. An FFT-based technique for translation, rotation, and scale-invariant image registration [J]. *IEEE Transactions on Image Processing*, 1996, 5(8): 1266-1271.
- [22] BÜLOW H, BIRK A. Fast and robust photomapping with an unmanned aerial vehicle (UAV) [C]. *International Conference on Intelligent Robots and Systems (IROS)*, 2009: 3368-3373.
- [23] FAN M, LAI S, HUANG J, et al. Rethinking Bisenet for real-time semantic segmentation [C]. *Conference on Computer Vision and Pattern Recognition (CVPR)*, 2021: 9716-9725.
- [24] YU C, GAO C, WANG J, et al. Bisenet v2: Bilateral network with guided aggregation for real-time semantic segmentation [J]. *International Journal of Computer Vision*, 2021, 129: 3051-3068.

作者简介



栾添添, 2011 年、2018 年于哈尔滨工程大学分别获得工学学士学位、博士学位, 现为哈尔滨理工大学副教授, 主要研究方向为视觉 SLAM、路径规划。

E-mail: luantiantian@hrbust.edu.cn

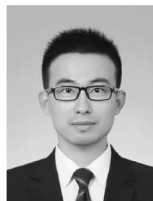
Luan Tiantian received her B.Sc. and Ph.D. degrees both from Harbin Engineering University in 2011 and 2018, respectively. She is currently an associate professor at Harbin University of Science and Technology. Her main research interests include visual SLAM, and path planning.



吕奉坤, 2022 年于南昌工程学院获得学士学位, 现于哈尔滨理工大学攻读硕士研究生, 主要研究方向为动态环境下的视觉 SLAM。

E-mail: 1446306727@qq.com

Lyu Fengkuan received his B.Sc. degree from Nanchang College of Engineering in 2022. He is currently pursuing his M.Sc. degree at Harbin Institute of Technology. His main research interest is visual SLAM in dynamic environments.



孙明晓(通信作者), 2010 年、2018 年于哈尔滨工程大学分别获得工学学士学位、博士学位, 现为哈尔滨理工大学副教授, 主要研究方向为视觉 SLAM、路径规划。

E-mail: sunmingxiao@hrbust.edu.cn

Sun Mingxiao (Corresponding author) received his B.Sc. and Ph.D. degrees both from Harbin Engineering University in 2010 and 2018, respectively. He is currently an associate professor at Harbin University of Science and Technology. His main research interests include visual SLAM, and path planning.