

DOI:10.11918/202007031

动态环境下融合边缘信息的稠密视觉里程计算法

周 凯¹, 罗 元¹, 张 毅², 李晋宏²

(1. 光电信息传感与技术重点实验室(重庆邮电大学), 重庆 400065;

2. 重庆邮电大学 信息无障碍与服务机器人工程技术研究中心, 重庆 400065)

摘 要: 针对传统的视觉里程计算法在动态环境下存在位姿估计精度不高且鲁棒性较差的问题, 提出一种融合边缘信息的稠密视觉里程计算法。首先, 使用深度信息计算像素点的空间坐标, 并采用 K -means 算法进行场景聚类。分别基于光度信息与边缘信息的聚类构建出光度及几何一致性误差与边缘对齐误差, 两者结合并进行正则化后得到数据融合的残差模型。将平均背景深度引入到残差模型中, 用以扩大动、静部分残差差距而有利于正确的运动分割。然后, 根据聚类残差分布的普遍特征, 构建运动似然的非参数统计模型, 通过动态阈值进行运动分割, 剔除动态物体并得到聚类权重。最后, 将加权聚类残差加入到位姿估计的非线性优化函数中, 以降低动态物体的影响, 提高位姿估计的精度。在 TUM 数据集上进行实验, 结果表明本文算法在静态环境下以及富有挑战性的高动态环境下都能取得较好的结果, 在动态环境下比现有算法具有更高的精度与鲁棒性。

关键词: 移动机器人; 视觉里程计; 边缘对齐; 动态环境; 运动分割

中图分类号: TP242.6

文献标志码: A

文章编号: 0367-6234(2021)02-0132-08

Dense visual odometry based on edge information fusion in dynamic environment

ZHOU Kai¹, LUO Yuan¹, ZHANG Yi², LI Jinhong²

(1. Key Laboratory of Optoelectronic Information Sensing and Technology (Chongqing University of Posts and Telecommunications), Chongqing 400065, China; 2. Engineering Research Center for Information Accessibility and Service Robots, Chongqing University of Posts and Telecommunications, Chongqing 400065, China)

Abstract: In view of the low robustness and poor accuracy of traditional visual odometry in dynamic environment, a dense visual odometry based on edge information fusion was proposed. First, the spatial coordinates of pixels based on depth information were calculated, and the K -means algorithm was adopted for scene clustering. Based on the clustering of photometric information and edge information, the photometric consistency error and edge alignment error were constructed respectively, and the residual model was obtained after fusion and regularization of the two errors. Next, the average background depth was introduced to the residual model so as to expand the residual difference between the dynamic and static parts, ensuring correct motion segmentation. Then, a non-parametric statistical model was constructed based on the general characteristics of the cluster residual distribution, and motion segmentation was performed through dynamic thresholds to eliminate dynamic objects and obtain clustering weights. Finally, the weighted cluster residuals were added to the nonlinear optimization function of pose estimation to reduce the effect of dynamic objects and improve the accuracy of pose estimation. Experiments on TUM dataset show that the proposed algorithm could achieve better results in both static and high dynamic environments, and it had higher accuracy and robustness than the existing algorithm in dynamic environment.

Keywords: mobile robot; visual odometry; edge alignment; dynamic environment; motion segmentation

视觉里程计(VO)是一种利用视觉传感器信息对机器人或其他载体进行准确的位姿估计的方法, 相比于使用激光传感器的方法, 视觉信息更加丰富, 可用于三维重建、物体识别、语义分割等。同时, 因其成本低、体积小、精度高等优点, 而广泛地应用于机器人自主导航、自动驾驶、增强现实等领域^[1-4]。

近年来, 许多精度高、效率快的视觉里程计方法被提出来, 可以将其大致分为3类。第1类是基于特征点法的视觉里程计^[4-6], 这些方法通过提取和匹配相邻图像的特征点, 建立重投影优化函数来估计帧间运动。第2类是基于直接法的视觉里程计^[7-9], 根据灰度不变假设, 建立光度误差函数来估计帧间运动。第3类是将直接法与特征点法融合的视觉里程计, Forster等^[10]提出一种半直接法视觉里程计(SVO), 通过对图像中的特征点块进行匹配来计算位姿, 加快了计算精度与速度。这些传统的算法都基

收稿日期: 2020-07-06

基金项目: 国家资金基金(61803058); 国家自然科学基金(51775076)

作者简介: 周 凯(1996—), 男, 硕士研究生

通信作者: 罗 元, luoyuan@cqupt.edu.cn

于静态环境假设,忽略了动态物体对运动估计的影响,从而导致计算结果在动态环境下有较大的误差.

在动态环境下,一般可采用随机采样一致性(RANSAC)或鲁棒 Huber 函数等概率方法将动态部分当作噪声剔除掉. 当动态物体在图像中占比较大时,则无法通过概率方法剔除,需要将动、静态物体区分开来,使用静态部分估算位姿,避免动态物体造成的误差. Besco 等^[11]提出的 DynaSLAM 算法采用深度学习结合几何的方法检测动态物体,该方法能准确剔除动态物体,具有较高的精度,但是该算法需要 GPU 加速,无法达到实时性与便携性. Li 等^[12]使用边缘点静态权重的方法判别动态点,该方法拥有较大的像素级噪声,容易导致动、静态点判断错误,且仅使用了部分图像信息,位姿估计精度较差. Jaimez 等^[13]和 Scona 等^[14]采用聚类的方法,通过构建背景分割优化函数来分离动、静态物体,然而该类方法在高动态环境下,不能有效地剔除动态对象,且无法进行准确的位姿估计.

本文提出一种融合边缘信息的稠密视觉里程计算法,利用边缘信息对光照变化和运动模糊不敏感的特性,弥补直接法的缺点. 同时,在几何纹理较少

的环境下,加大直接法在计算中的占比,从而减小边缘信息缺失导致的误差. 本文的主要贡献如下:1) 将边缘对齐的 Distance Transform (DT) 误差引入残差模型中,增加算法在不同环境的鲁棒性;2) 引入平均背景深度,加大动态部分残差,以辅助运动分割;3) 将带有边缘信息的加权聚类残差加入到位姿估计的非线性优化函数中,提高位姿估计的精度. 实验结果表明,本文算法在动态环境下能准确剔除动态物体,并在位姿估算中具有较高的精度.

1 算法框架

本文算法的全过程见图 1. 首先,使用 K-means 聚类法将图像分为 N 个集群. 然后,将聚类后的边缘图与灰度图分别坐标变换到关键帧及其 DT 图上,构建基于聚类的残差模型. 接着,根据聚类残差建立基于 t 分布的非参数统计模型,并通过分段阈值函数分离出动态物体并得到残差权重. 最后,将加权残差加入到优化函数中进行位姿估算,并将得到的位姿进行跟踪质量检查,更新下次迭代的变换矩阵或关键帧.

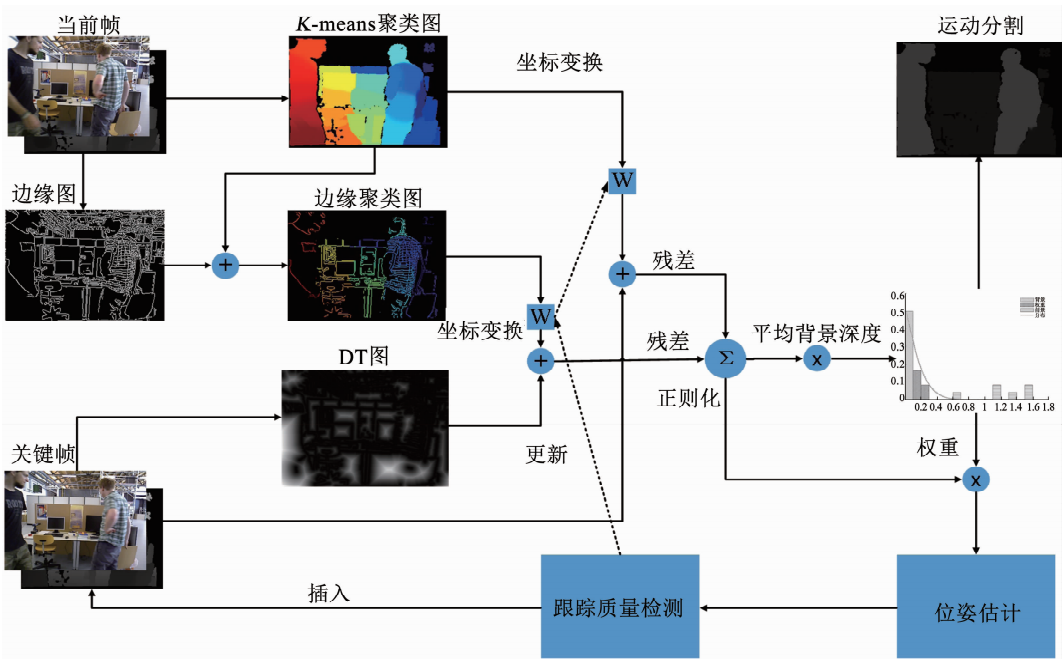


图 1 算法框架

Fig. 1 Algorithm framework

本文算法引入了边缘信息,对当前帧进行边缘提取与聚类后,计算出融合 DT 误差的残差模型,通过平均背景深度扩大动、静部分残差距离来辅助运动分割,并将边缘对齐作为约束加入到位姿估计中以获得更准确的估算结果.

2 融合边缘信息的视觉里程计

使用 RGB-D 传感器,每一个时间帧都可以获得一个灰度图 I_t 和一个深度图 Z_t . 深度信息用于计算场景聚类、参与构建残差模型并计算平均背景深度.

灰度信息用于提取边缘信息,并结合深度信息构建光度及几何一致性误差与 DT 误差.

2.1 相机运动模型

使用图像对应的三维点坐标 $P = (X, Y, Z)$ 和深度 $Z = Z_c(p)$ 可通过投影函数 π 计算出图像的像素坐标 $p = (x, y)$ 为

$$p = \pi(P) = \left(\frac{Xf_x}{Z} + c_x, \frac{Yf_y}{Z} + c_y \right), \quad (1)$$

式中 f_x 和 f_y 为焦距, c_x 和 c_y 为标准针孔相机模型定义的光学中心. 另外,为了将像素坐标 p 反投影为三维坐标点 P ,定义逆投影函数 π^{-1} 为

$$P = \pi^{-1}(p, Z) = \left(\frac{x - c_x}{f_x} Z, \frac{y - c_y}{f_y} Z, Z \right). \quad (2)$$

两个图像帧之间的相对运动可以表示为变换矩阵 $T \in SE(3)$,其中包含 3×3 的旋转矩阵 $R_{3 \times 3} \in SO(3)$ 和 3×1 的平移向量 $t_{3 \times 1} \in \mathbf{R}^3$. 通过上一帧的三维坐标点 P_{c-1} 以及两帧间的变换矩阵 T ,可以计算出当前帧的三维坐标点 P_c 为

$$P_c = T_{cc-1} P_{c-1} = \begin{bmatrix} R_{cc-1} & t_{cc-1} \\ 0 & 1 \end{bmatrix} P_{c-1}. \quad (3)$$

刚体运动有 6 个自由度,可使用李代数 $se(3)$ 中的 $\xi \in \mathbf{R}^6$ 来表示李群 $SE(3)$ 中的变换矩阵 $T \in \mathbf{R}^{4 \times 4}$. 定义坐标变换函数 ω 为

$$p_c = \omega(\xi_{cc-1}, p_{c-1}) = \pi(T_{cc-1} \pi^{-1}(p_{c-1}, Z_{c-1}(p_{c-1}))), \quad (4)$$

式中, ξ_{cc-1} 是 T_{cc-1} 的李代数, p_{c-1} 和 p_c 分别是相邻两帧对应的像素点坐标.

2.2 场景聚类

为了获得稠密的运动分割,有的方法直接采用像素点进行权重计算^[12],逐点分割未考虑到像素点之间的空间关系,不可避免地会产生噪声,导致精度下降. 本文参考 Jaimez 等^[13,15] 的方法采用 K-means

聚类法,利用深度信息和逆投影函数 π^{-1} 得到三维空间点坐标,通过计算每一个三维空间点距离集群中心点的距离,来确定该空间点属于哪一个集群,属于同一个集群的点一般为同一个刚体. 使用集群计算能够避免产生像素级噪声,简化运动分割问题.

2.3 残差模型

2.3.1 光度及几何误差

为了识别环境中的动、静态部分,需要计算每个集群的残差大小,然后进行运动分割. 首先,将当前帧 F_c 完成聚类后坐标变换到关键帧 F_k . 然后,根据光度及几何一致性构建误差模型:

$$r_z^i(\xi) = Z_k(\omega(\xi_{kc}, p_c^i)) - |T_{kc} \pi^{-1}(p_c^i, Z_c(p_c^i))|_Z, \quad (5)$$

$$r_l^i(\xi) = I_k(\omega(\xi_{kc}, Z_c(p_c^i))) - I_c(p_c^i). \quad (6)$$

式中: r_l^i 和 r_z^i 分别表示点 i 对应的光度误差与深度误差, $|\cdot|_Z$ 表示三维坐标点的 Z 坐标值. ξ_{kc} 为当前帧到关键帧坐标变换的李代数,由匀速运动假设 $\xi_{cc-1} = \xi_{c-1c-2}$ 可得 $\xi_{kc} = \xi_{kc-1} \xi_{c-2c-1}$,这个假设值一般接近于相机运动的最终估计值,从而可减少收敛时间并避免陷入局部最优解.

2.3.2 边缘对齐误差

Schenk 等^[16] 对比了多种边缘检测算法的性能,本文借鉴他们的方法对当前帧 F_c 进行边缘提取. 然后,将边缘点按照集群分类,并将边缘坐标变换到关键帧的 DT 图以计算边缘对齐误差. 基于边缘对齐的误差模型为

$$r_E^i(\xi) = DT_k(\omega(\xi_{kc}, p_c^i)), \quad (7)$$

式中, r_E^i 表示边缘点 i 对应的 DT 误差, DT_k 表示关键帧的 DT 误差函数. 根据关键帧提取的边缘,可以很快速地计算出 DT 图^[17],如图 2 所示, (a) 为原始灰度图, (b) 为边缘检测图, (c) 为由边缘得到的 DT 图,其中距离边缘远的像素呈白色.



(a) 灰度图



(b) 边缘图



(c) DT图

图 2 DT 计算

Fig. 2 Calculation of distance transform (DT)

2.3.3 数据融合误差

直接法基于灰度不变假设,所以导致 r_l^i 受光照变化影响较大,且对于运动模糊图像的处理也会产

生较大误差,而边缘信息对于光照变化以及图像模糊不敏感. 因此,通过引入基于边缘对齐的误差模型 r_E^i ,来提高算法的鲁棒性. 同时,在纹理单一的环境

中,直接法可以避免边缘过少导致算法精度下降的问题^[18].

将边缘对齐误差引入到光度及几何误差中构建融合边缘信息的残差模型,能够计算出更准确的残差.但是,背景的遮挡区域往往会导致残差模型具有较高的值,类似于 Jaimez 等^[13]的方法,为降低遮挡的影响,定义残差模型为

$$r_n(\xi) = \frac{\sum_{i=1}^{S_n-O_n} (\alpha_l r_l^i + r_z^i / \bar{Z}_n)}{S_n - O_n} + \frac{\sum_{i=1}^{E_n-O_n} \alpha_E r_E^i}{E_n - O_n}. \quad (8)$$

式中: r_n 表示第 n 个集群的残差值, S_n 表示第 n 个集群中所有深度有效的像素个数, O_n 表示第 n 个集群中遮挡的像素个数, E_n 表示第 n 个集群中所有深度有效的边缘点, $\bar{Z}_n$ 为第 n 个集群的平均深度, α_l 与 α_E 分别表示光度误差与DT误差在残差模型中所占权重.为了加强算法应对各种环境的鲁棒性,构建权重公式为

$$\alpha_E = \frac{\eta v_c \sum_{n=1}^N E_n}{\sum_{n=1}^N S_n} \alpha_l. \quad (9)$$

式中: v_c 为相机运动速度, N 为集群总数量, η 为由 α_l 到 α_E 的尺度系数.在边缘信息丰富(边缘点占比高)或相机运动速度较快时(容易产生模糊图像), α_E 有较大值,相反, α_E 较小时边缘信息影响较小.

使用原始集群残差往往准确率不高,运动分割效果较差.因此,参考 Jaimez 等^[13]的方法进行残差正则化,对连续的集群偏向于相似的运动分割,建立最小化模型来求取正则化残差公式为

$$E(r_n) = \lambda_R \sum_{m=1}^N G_{mn} (r_n - r_m)^2 + \lambda_T (r_n - r_{n,c-1})^2. \quad (10)$$

式中, G_{mn} 为聚类连通函数,为1表示 m 与 n 集群连通; $r_{n,c-1}$ 表示上一帧中第 n 个集群的残差值,因为在时间上,物体都是趋于保持移动或静止的.通过求解最小问题来得到最佳 r_n ,正则化后的残差模型与动态物体有较高的相关性.

2.3.4 平均背景深度辅助运动分割

为了减少场景的错误分割与过度分割,提高运动分割的准确性,本文引入平均背景深度来辅助算法进行运动分割.动态物体一般都为独立的刚体且远离背景,通过扩大远离背景部分的集群残差,使得靠近背景的集群偏向于分割为静态部分,而远离背景的集群有更大几率被判定为动态部分.定义平均背景深度为

$$\bar{Z}_{b,c} = \lambda_D \bar{Z}_{b,c-1} + (1 - \lambda_D) \frac{\sum_{n=1}^{N-d_{c-1}} \bar{Z}_n}{N - d_{c-1}}. \quad (11)$$

式中, $\bar{Z}_{b,c}$ 表示当前帧平均背景深度, d_{c-1} 为上一帧所分割出的动态集群数量.采用上一帧 d_{c-1} 计算当前帧的平均深度原因有3点:1)相邻两帧图像的聚类与运动分割大致相同;2)可通过权重 λ_D 来降低两帧间运动分割不一致产生的误差;3)背景不会移动,只有当相机移动时,背景深度才会变化,通过引入加权 $\bar{Z}_{b,c-1}$ 可使背景深度缓慢变化,符合相机缓慢移动的事实.引入平均背景深度后的残差模型为

$$\delta_n = (1 + |\bar{Z}_n - \bar{Z}_{b,c}|) r_n. \quad (12)$$

式中, δ_n 为引入平均背景深度后的第 n 个集群的残差值, $|\bar{Z}_n - \bar{Z}_{b,c}|$ 表示第 n 个集群与背景的距离.平均背景深度的引入使得距离背景较远的动态部分有更大的残差值,而距离背景较近的动态部分本身就有较大的残差,以此来扩大动、静部分残差差距而有利于正确的运动分割.

2.4 相机位姿估计

动态物体的独立运动会使得两帧之间坐标变换后的动态物体图像不对齐,而导致属于该物体的集群有较大的残差,而静态集群的残差在图像对齐的情况下往往很小或接近于0.图3是一个高度动态场景的统计残差直方图,根据聚类残差分布的普遍特征,参考文献[12,19]的方法建立一个基于t分布的非参数统计模型

$$w_n = \frac{v_0 + 1}{v_0 + ((\delta_n - \mu)/\sigma)^2}, \quad (13)$$

$$\sigma = 1.4826 \text{Median}\{|\delta_n - \mu|\}. \quad (14)$$

式中: v_0 为t分布的自由度,它决定了分布曲线的陡峭程度;平均值 μ 决定分布中心所处位置,将其设置为0表示残差小的静态集群占比较大; σ 为统计方差,通过中位数绝对偏差估算得到.

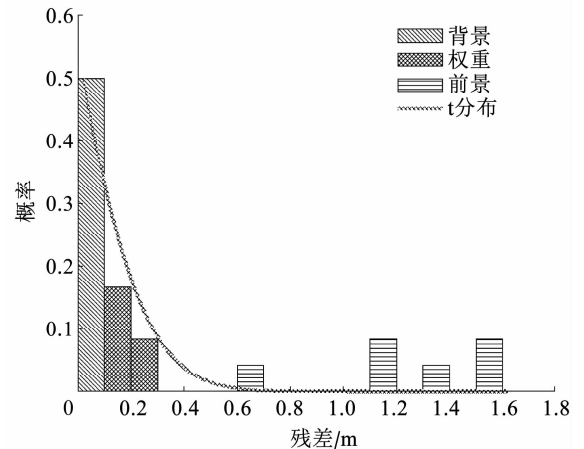


图3 残差直方图与t分布曲线

Fig. 3 Residual histogram and t distribution curve

如图3所示,非参数统计模型能够较好地拟合实际实验残差的直方图.由于统计模型的概率代表

了聚类的运动似然,因此可以指导场景进行运动分割.在各种复杂的场景下,采用固定阈值会使得分割效果不佳,参考 Zhou 等^[19]的方法使用自适应阈值构建场景分割模型,残差高于自适应阈值的集群可判定为动态部分,而低于阈值的集群可作为静态部分或赋权部分加入到位姿估算中,其公式为

$$\varphi_n = \begin{cases} 0, & \delta_n \geq 1/(\min(10, \max(3, \beta v_c \sigma))); \\ 1, & \delta_n < 1/10; \\ w_n, & \text{Others.} \end{cases} \quad (15)$$

式中: φ_n 代表第 n 个集群的权重,为 0 时表示该集群为动态集群,为 1 时表示为静态集群; v_c 为相机速度; σ 为统计方差; β 用于调整因变量到残差尺度的系数,通过公式 $10^3 \times \min(10^3, 10^{d_c-1})$ 计算得到.

计算出集群权重后,便可得到加权聚类残差,通过加权聚类残差来构建非线性优化函数,用以更准确地估计帧间运动.文献[13,19]的方法表明,采用 Cauchy M-estimator 模型进行优化能获得较好的结果.构建融合边缘信息的加权聚类残差的非线性优化函数

$$\xi = \underset{\xi}{\operatorname{argmin}} \left\{ \sum_{i=1}^M [F(\varphi_n^i r_z^i(\xi)) + F(\alpha_1 \varphi_n^i r_1^i(\xi)) + F(\alpha_E \varphi_n^i r_E^i(\xi))] \right\}, \quad (16)$$

$$F(r) = \frac{c^2}{2} \log \left(1 + \left(\frac{r}{c} \right)^2 \right). \quad (17)$$

式中: M 表示所有深度有效的点的数量, φ_n^i 表示属于第 n 个集群的点 i 的权重,参数 c 标记了 F 的转折点.通过最小化非线性优化函数求得最佳的帧间运动.

2.5 跟踪质量检测

本文采用帧到关键帧的跟踪方法,只有在更新关键帧的时候才会进行复杂的 DT 计算,相比于传统的帧对帧计算每一帧 DT 的方法^[20],大大提高了计算效率.另外,采用帧与关键帧的跟踪方法,可使非刚性物体的轻微运动更加容易区分,从而降低轻微运动导致的累计误差.参考 Schenk 等^[16]的方法,通过将先前跟踪过的 N 帧的边缘投影到当前帧,得到像素点重叠计数图,通过对相同重叠数构建直方图,来获取图像重叠情况,从而评估当前帧跟踪质量,公式为

$$\sum_{j=1}^N \gamma_j H(j) \leq \gamma_0 H(0). \quad (18)$$

式中, $H(j)$ 表示重叠数为 j 的点的数量, γ_j 表示重叠数为 j 的点在跟踪质量评估中所占的权重.当边缘重叠的加权和小于非重叠的数量时,说明跟踪质量较差,此时将上一跟踪较好的帧更新为关键帧.

3 实验与分析

本实验所用电脑配置为:CPU 为英特尔 i5-4210,主频为 1.7 GHz,内存为 4 G,不使用图形处理器 (GPU) 加速,在 Ubuntu 16.04 下进行实验.本文算法的运行基于每一时间帧的灰度与深度信息,因此主要适用于 RGB-D 相机.在慕尼黑工业大学提供的 TUM RGB-D 标准数据集^[21]上进行实验,每个序列包含 RGB 图像、深度图像以及真实轨迹等,该数据集已被广泛用于测试各种 SLAM 和 VO 算法的准确性和性能.为了方便算法评估,将数据集分为 2 类:1) 低动态与静态环境,此类数据集下没有移动物体或仅有少量移动物体;2) 高动态环境,人或其他移动物体在图像中占比较大.同时在这 2 类数据集下测试来评估算法的综合性能.

本文实验取 K -means 聚类集群数 N 为 24,并设置其他参数如下:1) 在残差模型建立部分,如图 4 所示,(a) 为原始图像;仅采用几何信息进行运动分割时,即 α_1 为 0,可见(b)中由于信息不足而导致分割不完全;(c)中融合光度信息并设置较高的 α_1 值导致光度信息占比较高而产生过度分割;(d)为融合光度信息并取合适的 α_1 所产生的较准确的分割结果;(e)中融合边缘信息并取较高的 η 值时,同样会导致过度分割;(f)为最准确的分割结果,此时设置光度误差权重 α_1 为 0.15,计算 DT 误差权重的尺度因子 η 为 0.35,来平衡几何、光度、边缘信息对残差模型的影响.2) 在残差正则化部分,参考文献[13]的方法,将式(10)中的参数设置为 $\lambda_R = 0.5, \lambda_T = 1.5$ 来设置空间正则化与时间正则化的权重.3) 将式(11)中的 λ_D 设为 0.7 来计算稳定且缓慢变化的平均背景深度.4) 在跟踪质量检测中,参考文献[16]的方法,将前 3 帧的边缘投影到当前帧进行质量评估,设置式(18)中 $\gamma_j = [1, 1, 1.25, 1.5]$,其中 $j \in [0, 3]$.以上参数为本次实验中获取的经验值,并非最优参数,但在实验中提供了良好的结果.

为了评估本文算法的精度,与现存的 3 种算法进行了比较:包括结合场景流的 VO-SF^[13]、基于边缘对齐的 REVO^[16]以及文献[19]的方法.以上算法都是基于 RGB-D 数据进行实验,其中 REVO 仅利用边缘信息进行位姿估计,且未处理动态目标,与其进行对比的原因在于评估算法在静态环境下的精度以及效率.VO-SF 与文献[19]的方法都是目前动态环境下较为先进的视觉里程计算法.

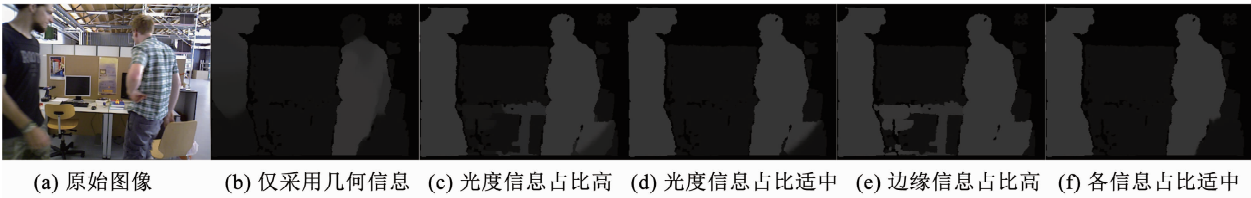


图 4 信息权重

Fig. 4 Schematic diagram of information weights

3.1 运动分割

运动分割用于剔除动态部分,使视觉里程计仅使用静态部分计算相机位姿.如若分割错误,部分动态物体未被检测而加入位姿计算,会导致相机与动态物体产生错误的相对运动而造成位姿估计误差.若过度分割,将部分静态物体剔除,会使得位姿估计中可用像素信息减少而增大位姿估计误差.因此,正确的运动分割对于动态环境下的视觉里程计起到至关重要的作用.

在多种动态环境下采用 VO-SF 算法与本文算法进行运动分割比较实验,结果见图 5. (a) 为原始 RGB 图, (b) 为有聚类信息的边缘图, (c) 为 VO-SF 算法运动分割结果, (d) 为本文算法运动分割结果,

深色为静态部分,浅色为动态部分. (1) 与 (2) 展示了动态物体在图像像素占比较大的情况,因引入了丰富的边缘信息则相比于 VO-SF 有更为准确的运动分割.引入的平均背景深度能辅助运动分割,从 (3) 与 (4) 可以看出,平均背景深度加大了远离背景的物体的残差,从而避免了 (3) 中动、静态部分分割错误并在 (4) 中准确地检测出了远离相机的移动物体. (5) 中椅子因靠近移动的人而被分割为动态物体,轻量的过度分割是合理的且不会影响整体的位姿估计. (6) 与 (7) 分别为有效深度过少的图像以及相机移动过快导致的模糊图像,这两种情况会因严重的信息丢失而导致较差的分割结果,但是引入边缘信息使得本文算法在类似情况下仍优于 VO-SF.

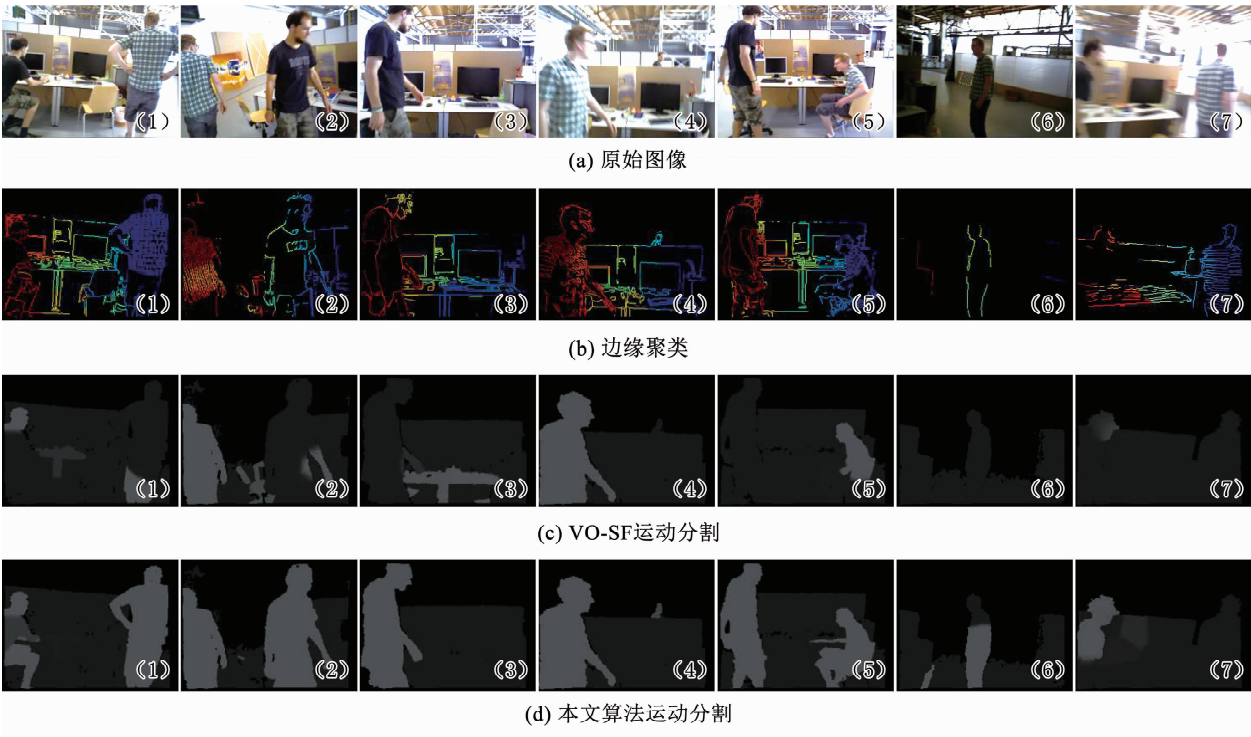


图 5 运动分割

Fig. 5 Motion segmentation results

3.2 评估指标

相对位姿误差 (RPE)^[21] 用于测量每秒的平均漂移量,使用 RPE 对本文算法以及另外 3 种算法进行评估,分别在静态与低动态环境和高动态环境下

进行了实验,文献[19]采用该文献部分原始数据.实验结果见表 1,在静态与低动态环境下,本文算法可能因场景过度分割而造成边缘信息减少,但光度误差能为位姿估计提供更多的约束,对比仅采用边

表 1 平移和旋转均方根误差
Tab. 1 RMSE of translation and rotation

环境	序列	平移 RMSE/(cm·s ⁻¹)				旋转 RMSE/((°)·s ⁻¹)			
		REVO	VO-SF	文献[19]方法	本文方法	REVO	VO-SF	文献[19]方法	本文方法
静态与低动态	fr1/rpy	3.568	4.554		3.920	2.998	2.371		2.506
静态与低动态	fr2/desk	1.876	2.389	2.490	1.327	0.681	1.005	1.056	0.650
静态与低动态	fr3/sitting_xyz	2.039	5.721	6.500	2.944	0.719	1.449	1.600	0.883
静态与低动态	fr3/sitting_static	1.320	2.438	2.570	1.437	0.367	0.711	0.732	0.472
高动态	fr3/walking_rpy	61.636	36.208	26.900	20.982	12.561	6.007	4.801	3.919
高动态	fr3/walking_static	22.149	9.547	4.600	5.354	3.847	1.615	0.956	1.019
高动态	fr3/walking_xyz	46.305	27.729	12.700	11.597	8.942	5.109	2.615	2.551
高动态	fr3/walking_halfsphere	41.266	33.442	27.360	19.176	8.155	6.681	5.618	4.612

缘信息的 REVO 算法也拥有较好的性能,而对比 VO-SF 与文献[19]的方法,本文算法明显有更优的性能. 在高动态环境下,本文算法采用帧与关键帧的跟踪方法避免非刚性物体造成的累积误差,并利用平均背景深度来辅助场景分割. 从表 1 中可看出本文算法在高动态环境下拥有更高的精度,相比 VO-SF 平均精度提升了 45%,相比文献[19]平均精度提升了 15%. 图 6 通过对 REVO、VO-SF 与本文算法在序列 fr3/walking_xyz 下进行了轨迹漂移比较,可以很直观地看出本文算法精度高于其他算法.

绝对轨迹误差 (ATE)^[21] 用于测量轨迹的整体质量,使用 ATE 图可以直观地看出估计轨迹与真实轨迹的区别,结果见图 7. 图中, (a) 为 VO-SF 的绝对

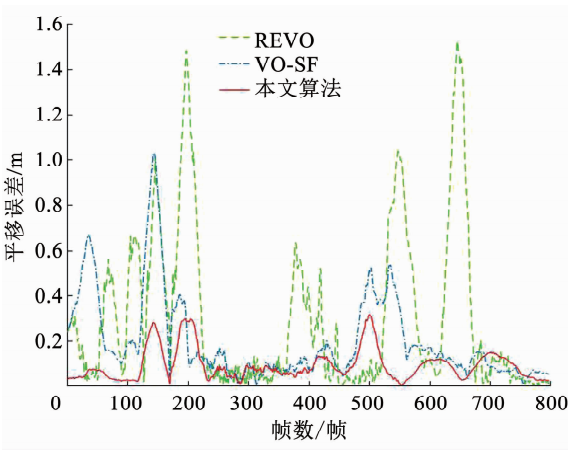


图 6 fr3/walking_xyz 下相对轨迹误差
Fig. 6 Relative trajectory error of fr3/walking_xyz

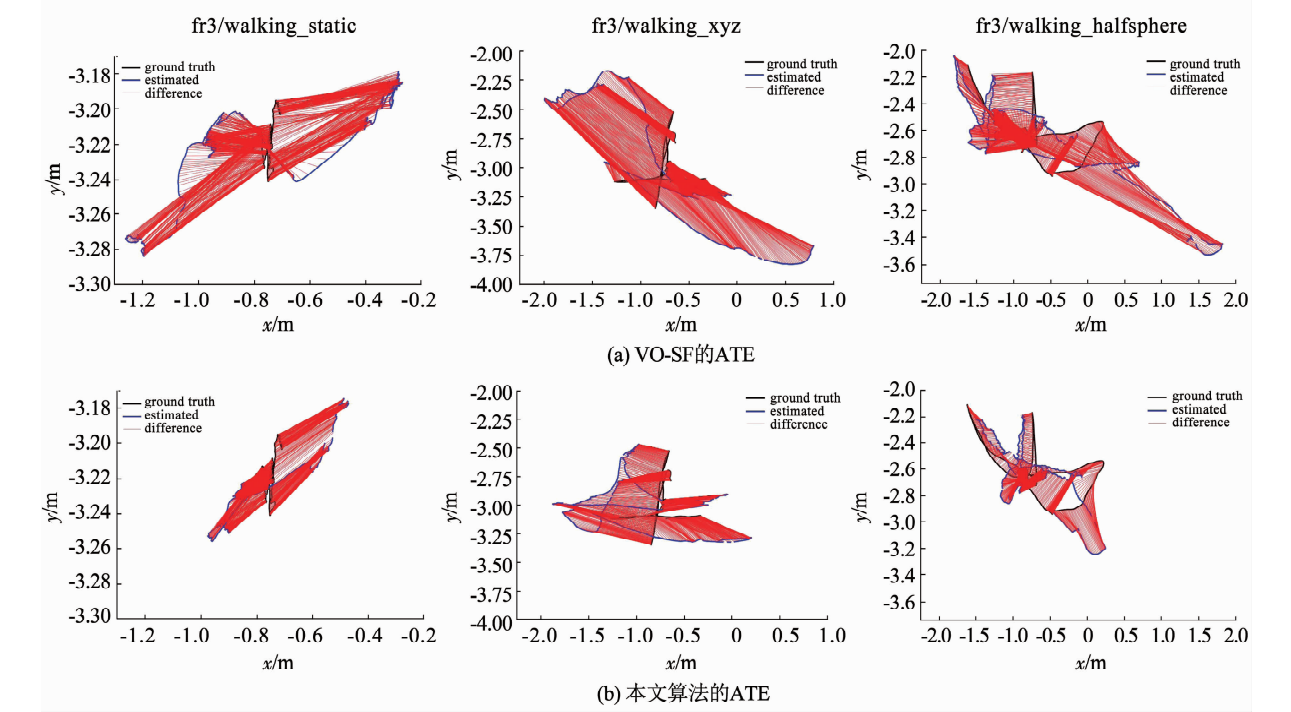


图 7 不同算法的绝对轨迹误差
Fig. 7 Absolute trajectory error of different algorithms

轨迹误差,(b)为本文算法的绝对轨迹误差.从图中可以看出,本文算法具有更小的轨迹误差,验证了本文算法改进的有效性.

为了比较运行时间,在同样的实验环境,本文算法采用与其他算法相同的分辨率为 320×240 的 QVGA 图像进行实验.结果表明,在不采用 GPU 加速的前提下,REVO 因仅采用边缘信息且未对动态物体进行处理,因此计算效率较高,平均每帧运行时间为 55 ms,VO-SF 平均每帧运行时间为 80 ms.因本文引入边缘信息,并在插入关键帧的时候进行了更多的处理,因此运行时间比其他算法要较长,平均每帧运行时间为 100 ms,基本满足实时性.

4 结 论

本文提出了一种融合边缘信息的稠密视觉里程计算法,将边缘信息加入到运动分割与位姿估计中,并引入平均背景深度辅助运动分割,然后利用边缘进行关键帧选择.边缘信息的融合使得算法在多种不同的环境下拥有更好的鲁棒性,同时对于几何纹理较少的环境,更多地采用光度结合深度信息来进行位姿估计,降低环境特征过少而导致的误差.在 TUM 公共数据集上的实验结果表明,本文算法在静态环境下以及富有挑战性的高动态环境下都能取得较好的结果,能够检测出距离相机较远的动态部分,在信息丢失、图像模糊等情况下也有更好的鲁棒性.

视觉里程计算法在各种环境下不可避免地产生误差,累积误差会导致估计位姿偏离真实位姿.因此,下一步工作将融合 IMU(惯性测量单元)数据,在动态环境下为位姿求解添加更多的约束项,用以降低位姿估计误差,并考虑加入回环检测以消除累积误差,进一步提高算法精度.

参考文献

- [1] 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. DOI: 10.1109/TRO.2016.2624754
- [2] BRESSON G, ALSAYED Z, YU Li, et al. Simultaneous localization and mapping: a survey of current trends in autonomous driving[J]. IEEE Transactions on Intelligent Vehicles, 2017, 2(3): 194. DOI: 10.1109/TIV.2017.2749181
- [3] 刘浩敏,章国锋,鲍虎军.基于单目视觉的同时定位与地图构建方法综述[J].计算机辅助设计与图形学学报,2016,28(6): 855
LIU Haomin, ZHANG Guofeng, BAO Hujun. A survey of monocular simultaneous localization and mapping[J]. Journal of Computer-Aided Design and Computer Graphics, 2016, 28(6): 855. DOI: 10.3969/j.issn.1003-9775.2016.06.001
- [4] KLEIN G, MURRAY D. Parallel tracking and mapping for small AR workspaces[C]//2007 6th IEEE and ACM International Symposium on Mixed and Augmented Reality. Nara: IEEE, 2007: 225. DOI: 10.1109/ISMAR.2007.4538852
- [5] MUR-ARTAL R, MONTIEL J M M, TARDÓS J D. ORB-SLAM: a

- versatile and accurate monocular SLAM system [J]. IEEE Transactions on Robotics, 2015, 31(5): 1147. DOI: 10.1109/TRO.2015.2463671
- [6] 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, 2016, 33(5): 1255. DOI: 10.1109/TRO.2017.2705103
- [7] ENGEL J, SCHÖPS T, CREMERS D. LSD-SLAM: large-scale direct monocular SLAM[C]//Computer Vision - ECCV 2014. Cham: Springer, 2014: 834. DOI: 10.1007/978-3-319-10605-2_54
- [8] ENGEL J, STÜCKLER J, CREMERS D. Large-scale direct SLAM with stereo cameras[C]//2015 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS). Hamburg: IEEE, 2015: 1935. DOI: 10.1109/IROS.2015.7353631
- [9] ENGEL J, KOLTUN V, CREMERS D. Direct sparse odometry[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2018, 40(3): 611. DOI: 10.1109/TPAMI.2017.2658577
- [10] FORSTER C, PIZZOLI M, SCARAMUZZA D. SVO: fast semi-direct monocular visual odometry[C]//2014 IEEE International Conference on Robotics and Automation (ICRA). Hong Kong: IEEE, 2014: 15. DOI: 10.1109/ICRA.2014.6906584
- [11] 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. DOI: 10.1109/LRA.2018.2860039
- [12] LI Shile, LEE D H. RGB-D SLAM in dynamic environments using static point weighting[J]. IEEE Robotics and Automation Letters, 2017, 2(4): 2263. DOI: 10.1109/LRA.2017.2724759
- [13] JAIMEZ M, KERL C, GONZALEZ-JIMENEZ J, et al. Fast odometry and scene flow from RGB-D cameras based on geometric clustering[C]//2017 IEEE International Conference on Robotics and Automation (ICRA). Singapore: IEEE, 2017: 3992. DOI: 10.1109/ICRA.2017.7989459
- [14] SCONA R, JAIMEZ M, PETILLOT Y R, et al. StaticFusion: background reconstruction for dense RGB-D SLAM in dynamic environments [C]//2018 IEEE International Conference on Robotics and Automation (ICRA). Brisbane: IEEE, 2018: 3849. DOI: 10.1109/ICRA.2018.8460681
- [15] JAIMEZ M, SOULAI M, STÜCKLER J, et al. Motion cooperation: smooth piece-wise rigid scene flow from RGB-D images[C]//2015 IEEE International Conference on 3D Vision. Lyon: IEEE, 2015: 64. DOI: 10.1109/3DV.2015.15
- [16] SCHENK F, FRAUNDORFER F. Robust edge-based visual odometry using machine-learned edges [C]//2017 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS). Vancouver: IEEE, 2017: 1297. DOI: 10.1109/IROS.2017.8202305
- [17] FELZENSZWALB P F, HUTTENLOCHER D P. Distance transforms of sampled functions[J]. Theory of Computing, 2012, 8:415. DOI: 10.4086/toc.2012.v008a019
- [18] WANG Xin, WEI Dong, ZHOU Mingcai, et al. Edge enhanced direct visual odometry [C]//British Machine Vision Conference. York: BMVA Press, 2016: 35. DOI: 10.5244/C.30.35
- [19] ZHOU Wugen, PENG Xiaodong, WANG Haijiao, et al. Nonparametric statistical and clustering based RGB-D dense visual odometry in a dynamic environment[J]. 3D Research, 2019, 10(2): 1. DOI: 10.1007/s13319-019-0220-4
- [20] KUSE M, SHEN Shaojie. Robust camera motion estimation using direct edge alignment and sub-gradient method [C]//2016 IEEE International Conference on Robotics and Automation (ICRA). Stockholm: IEEE, 2016: 573. DOI: 10.1109/ICRA.2016.7487181
- [21] STURM J, ENGELHARD N, ENDRES F, et al. A benchmark for the evaluation of RGB-D SLAM systems [C]//2012 IEEE/RSJ International Conference on Intelligent Robots and Systems, 2012: 573. DOI: 10.1109/IROS.2012.6385773