一种改进的即时定位与地图构建系统

孙云雷[1,2,3,4,5*]，吴清潇[1,2,4,5**]

[1]中国科学院沈阳自动化研究所，辽宁 沈阳 110016；
[2]中国科学院机器人与智能制造创新研究院，辽宁 沈阳 110016；
[3]中国科学院大学，北京 100049；
[4]中国科学院光电信息处理重点实验室，辽宁 沈阳 110016；
[5]辽宁省图像理解和计算机视觉重点实验室，辽宁 沈阳 110016

摘要 针对 ORB-SLAM2（Oriented FAST Rotated BRIEF SLAM2）系统中相机位姿求解精度不高，只能生成稀疏地图的问题，提出了一种在 ORB-SLAM2 系统框架上将稠密的直接法和原系统采用的稀疏特征法结合在一起求解相机位姿，并生成稠密地图的方法。该方法改进之处包括：在原系统使用的第三方图优化库 g2o（General Graph Optimization）中创建一条新的稠密约束一元边，将稠密直接法的光度误差约束加入到图优化库 g2o 中；跟踪相机时先通过稠密直接法计算相邻两帧图像之间相机的旋转变换，再利用改进后的图优化库 g2o 同时最小化特征法的重投影误差和直接法的光度误差，优化求解 6 DOF（Degree of Freedom）相机位姿；在 ORB-SLAM2 系统框架上添加稠密重建线程，将周围场景的重建结果实时反馈给用户。在 TUM RGB-D 和 ICL-NUIM 数据集上的测试结果表明，本文方法在一定程度上提高了 ORB-SLAM2 系统中相机位姿的求解精度，不仅可生成稀疏地图，还可重建更高精度的稠密地图。

关键词 图像处理；即时定位与地图构建；图优化；重投影误差；光度误差

中图分类号 TP391 **文献标识码** A **doi**：10.3788/LOP56.161013

An Improved Simultaneous Localization and Mapping System

Sun Yunlei[1,2,3,4,5*], Wu Qingxiao[1,2,4,5**]

[1] *Shenyang Institute of Automation, Chinese Academy of Sciences, Shenyang, Liaoning* 110016, *China*;
[2] *Institutes for Robotics and Intelligent Manufacturing, Chinese Academy of Sciences, Shenyang, Liaoning* 110016, *China*;
[3] *University of Chinese Academy of Sciences, Beijing* 100049, *China*;
[4] *Key Laboratory of Opto-Electronic Information Processing, Chinese Academy of Sciences, Shenyang, Liaoning* 110016, *China*;
[5] *Liaoning Provincial Key Laboratory of Image Understanding and Computer Vision, Shenyang, Liaoning* 110016, *China*

Abstract Camera pose estimation has a low accuracy and only generates a sparse map in the oriented FAST rotated BRIEF SLAM2 (ORB-SLAM2) system. To compute camera pose and generate a dense map, this study proposes a method that combines the dense direct method and sparse feature-based method adopted by the original ORB-SLAM2 system framework. This method mainly makes three improvements to the ORB-SLAM2 system. First, a new dense constraint unary edge is created in the third-party general graph optimization (g2o) library used in the original system; the photometric error constraint of the dense direct method is added to the g2o library. Second, the rotation transformation between two executive frames is calculated using the dense direct method; then, the improved g2o library is used to simultaneously minimize the re-projection error of the feature-based method and the

收稿日期：2019-02-17；**修回日期**：2019-03-12；**录用日期**：2019-03-27

基金项目：国家自然科学基金（U1713216）

* **E-mail**：sunyunlei@sia.cn；** **E-mail**：wuqingxiao@sia.cn

photometric error of the direct method to compute the 6 degree-of-freedom (DOF) camera pose. Third, a dense reconstruction thread is added in the ORB-SLAM2 system framework and the reconstruction result of the surrounding scene is reported to the user in real time. Experiments conducted on TUM RGB-D and ICL-NUIM datasets reveal that the proposed method considerably improves the accuracy of the camera pose estimation in the ORB-SLAM2 system, produces sparse maps, and reconstructs high-precision dense maps.

Key words image processing; simultaneous localization and mapping; general graph optimization; re-projection error; photometric error

OCIS codes 100.3010; 150.1135; 150.4065; 150.6910

1 引 言

近几年,随着机器人[1]、无人机、无人车、增强现实[2](AR)和虚拟现实(VR)等行业的发展,即时定位与地图构建(SLAM)技术备受关注。SLAM 技术可以实时定位传感器在未知环境中的位姿,并构建由传感器感知的环境的地图。根据系统输入的数据源,可将 SLAM 技术分为激光雷达 SLAM 和视觉 SLAM 等,其中视觉 SLAM 又包括单目、双目 SLAM 和 RGB-D SLAM(Read-Green-Blue-Dense SLAM)等。根据应用视觉 SLAM 技术的系统输出的地图类别,又可将 SLAM 技术又可以分为基于特征的稀疏 SLAM[如 PTAM[3](Parallel Tracking and Mapping for Small AR Workspaces),ORB-SLAM2(Oriented FAST Rotated BRIEF SLAM2)[4-5]],半稠密 SLAM[如 LSD-SLAM[6](Large-Scale Direct Monocular SLAM)]和稠密 SLAM(如 KinectFusion[7-8] 和 ElasticFusion[9-10])等。目前,基于特征的稀疏 SLAM 技术已经趋于成熟,稀疏 SLAM 技术通常只计算相机位姿,重建稀疏的地图;半稠密 SLAM 技术可以重建场景的半稠密地图;而稠密 SLAM 技术可以定位系统,同时重建场景的稠密地图。

PTAM 系统是第一个将稀疏 SLAM 技术应用在小场景 AR 中的系统,在该系统中,跟踪相机和重构地图分别在两个独立的线程中实现。ORB-SLAM2[4-5]是一个相对完备的 SLAM 系统,相比于 PTAM 系统,ORB-SLAM2 系统加入了闭环检测和重定位功能,ORB-SLAM2 采用第三方位置识别库 DBoW2[11](Bags of Binary Words for Fast Place Recognition in Image Sequences)检测闭环候选帧并重定位相机的位姿,当检测到闭环时,闭环线程先执行基础图优化,再执行全局 BA(Bundle Adjustment)[4]优化。

2011 年,微软公司的消费级 RGB-D 相机[12-13] Kinect 的问世,使得基于 RGB-D 相机的稠密 SLAM 技术受到研究人员的广泛关注。KinectFusion 是第一个基于 RGB-D 相机的实时稠密重建系统,该系统采用 TSDF[14](Truncated Signed Distance Function)模型融合 RGB-D 图像序列实时重建环境的稠密表面。TSDF 采用等大小的网格表示重建的空间,系统消耗的存储空间随着重建空间体积的扩展不断增大。为了解决 TSDF 模型消耗存储空间的问题,研究人员提出了改进算法[15-18]。Kintinuous 采用循环利用存储空间的方式扩展重建场景的范围[16]。ElasticFusion 系统采用 Surfel 面元模型重建稠密地图[17-18]。BundleFusion 采用传统 SFM(Structure From Motion)的方法,将每一帧和所有图像配准,该系统在重建小范围场景时效果较好,但需要两个高性能的 GPU(Graphics Processing Unit)才能达到实时的帧率[19]。

通常,基于特征法的稀疏 SLAM 技术在相机定位精度、计算量和扩展性等方面表现较好,但是特征法在纹理特征较少的环境下容易跟踪失败。稠密配准的方法在运动模糊和低纹理特征的情况下更加稳健,此外稠密 SLAM 系统生成的稠密地图可以用作机器人导航、避障以及 AR 场景扫描等场合。结合稀疏和稠密 SLAM 技术各自的优势和不足,本文将两种方法组合在一起,通过在现有的图优化库 g2o[20](General Graph Optimization)中创建新的稠密约束边,同时最小化稀疏特征法的重投影误差和稠密直接法的光度误差,优化相机位姿,并新建稠密建图线程,采用 Surfel 面元模型融合 RGB-D 图像序列重建稠密地图,并将重建的结果实时反馈给用户。

2 系统概况

原始的 ORB-SLAM2 系统中有跟踪线程、局部建图线程和闭环线程。图 1 为改进后的系统概况图,图中粗线框框起来的三处为改进的部分。

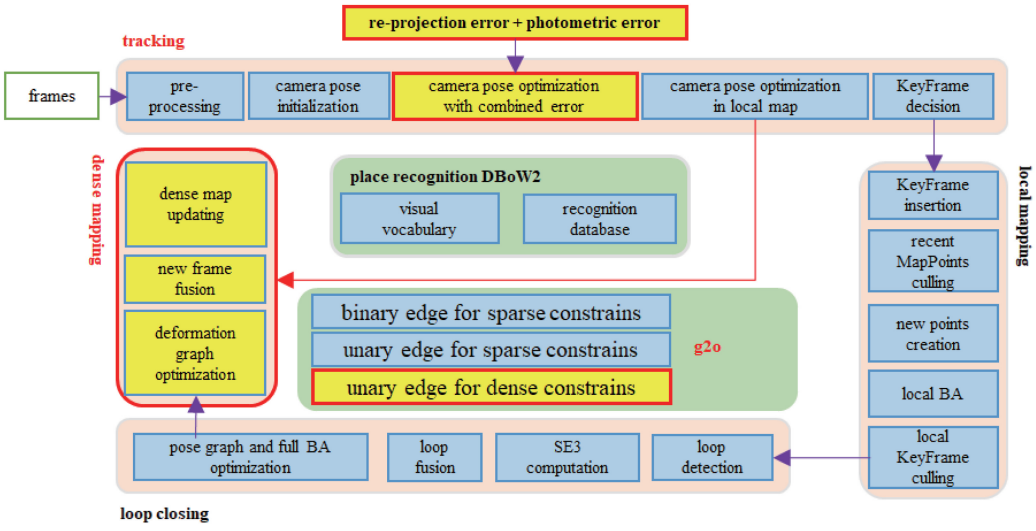

图 1 改进后的系统概括图

Fig. 1 Overview of improved system

ORB-SLAM2 系统中用来优化相机位姿的第三方图优化库 g2o 中，只提供了同时优化相机位姿和地图点位置坐标的稀疏约束两元边和只优化相机位姿的稀疏约束一元边，为了能同时最小化直接法的光度误差和特征法的重投影误差以优化相机位姿，在现有的图优化库 g2o 中新添加了稠密约束一元边，这也是本方法对原系统做的第一处改进。第二处改进是在原系统的跟踪线程中，初始化相机位姿之后，将稠密直接法应用到相机跟踪过程中，结合系统中原有的稀疏特征法，一起优化求解相机位姿。首先使用稠密直接法计算相邻两帧图像之间的相机旋转变换矩阵，再使用改进后的图优化库 g2o 同时最小化稀疏特征法的重投影误差和稠密直接法的光度误差来优化相机位姿。第三处改进是在原系统中创建了稠密建图线程，可以将相机周围场景的重建结果实时反馈给用户。局部建图线程和闭环检测线程沿用 ORB-SLAM2 系统中的方法。局部建图线程负责稀疏特征点的生成和选择以及关键帧的选择，并采用局部 BA[3] 算法优化局部地图中关键帧的相机位姿和地图点的位置坐标。闭环线程负责闭环检测以及检测到闭环后的基础图[4-5]优化和全局 BA[4] 优化。由于改进后的系统需要生成稠密地图，为了使系统可以实时工作，需要 GPU 加速计算。改进后的系统在联想笔记本电脑（CPU i7-7700HQ，GPU NVIDIA GeForce GTX 1060 MaxQ）上可以实时运行，实时估计相机位姿并实时生成稠密地图。虽然使用了 GPU 加速计算，由于生成稠密地图的计算量很大，所以改进后的系统与原 ORB-SLAM2 系统相比，运行速度还是会慢一些，经实验，改进后的系统可以在装有 960 及以上规格的 GPU 的电脑上运行并满足实时性。

2.1 相机位姿估计

在原 ORB-SLAM2 系统的跟踪线程中，初始化相机位姿之后，只使用第三方图优化库 g2o 最小化相邻两帧图像匹配点之间的重投影误差来优化相机位姿，而在本文提出的方法中，则是将稠密的直接法应用到求解相机位姿的过程中，使用改进后的图优化库 g2o 同时最小化稀疏特征法的重投影误差和稠密直接法的光度误差以优化求解相机位姿。在跟踪线程中使用固定速度模型[4]或参考帧模型[5]初始化相机位姿，将上一帧图像中的特征点对应的地图点向当前帧图像投影，在投影点周围寻找匹配点，如果找到的匹配点对数不够，则扩大搜索范围，并利用找到的匹配点在 RANSAC（Random Sample Consensus）框架下优化初始化的相机位姿，利用优化之后的位姿搜索更多的匹配点，这些匹配点可以用在后端优化中。构建相邻两帧图像的金字塔模型，在图像金字塔顶部的图像配准中，通过最小化稠密直接法的光度误差得到相邻两帧图像之间相机的旋转变换矩阵，在底层图像金字塔的图像配准中采用改进后的图优化库 g2o 同时最小化重投影误差和光度误差来优化相机位姿。将上一帧图像对应的地图点根据初始化的相机位姿向当前帧图像投影，在投影点周围寻找匹配点，利用匹配点之间的距离误差构建重投影误差代价方程，表达式为

$$E_{\text{sparse}}=\sum_{i\in S}\left\| \boldsymbol{u}_i-[\boldsymbol{K}\ \exp(\Delta\hat{\boldsymbol{\xi}})\exp(\hat{\boldsymbol{\xi}})\ \tilde{\boldsymbol{P}}_i]\right\|^2,\tag{1}$$

式中：E_{sparse}为采用稀疏的特征法优化相机位姿过程中产生的重投影误差；$\tilde{\boldsymbol{P}}_i$为上一帧图像中第i个特征点在世界坐标系下的齐次坐标；点$\boldsymbol{P}_i$在当前帧图像上匹配点的坐标为$\boldsymbol{u}_i$；$\exp(\hat{\boldsymbol{\xi}})$为李群表示的相机初始位姿；$\exp(\Delta\hat{\boldsymbol{\xi}})$为李群表示的相机位姿增量，$\boldsymbol{\xi}$为相机位姿的李代数形式，$\boldsymbol{\xi}=(\boldsymbol{v}_{3\times1},\boldsymbol{\omega}_{3\times1})^{\mathrm{T}}\in\mathbf{R}^6$，$\boldsymbol{v}_{3\times1}\in\mathbf{R}^3$，$\boldsymbol{\omega}_{3\times1}\in\mathrm{so}(3)$，so(3)表示用李代数的形式表示旋转，$\boldsymbol{v}_{3\times1}$表示相机的位移向量，$\boldsymbol{\omega}_{3\times1}$表示相机的旋转向量；$\boldsymbol{K}$是相机的内参矩阵；$S$表示匹配点集合；上标^表示将该向量转换为反对称矩阵的形式。并且

$$\hat{\boldsymbol{\xi}}=\begin{bmatrix}\hat{\boldsymbol{\omega}}_{3\times1} & \boldsymbol{v}_{3\times1}\\ \boldsymbol{0}_{1\times3} & 0\end{bmatrix}\in\mathbf{R}^{4\times4},\tag{2}$$

式中：$\hat{\boldsymbol{\omega}}_{3\times1}$为向量$\boldsymbol{\omega}_{3\times1}$的反对称矩阵。

在利用稠密直接法优化相邻两帧图像之间相机的旋转变换矩阵时，先为待配准的两幅图像建立4层图像金字塔模型，按照由粗到精[4]，由顶到底的配准方式求解相机位姿。对于图像金字塔顶部的两层图像，将上一帧图像上的像素点根据对应的深度信息逆投影到上一帧图像的相机坐标系下，将上一帧图像相机坐标系下的点，根据初始的相邻两帧图像之间相机的旋转变换矩阵，转换到当前帧图像的相机坐标系下，再根据当前帧图像相机内参矩阵，将相机坐标系下的点投影到当前帧图像上，根据灰度不变假设，利用两帧图像对应点之间的灰度值误差构建光度误差代价方程。利用最小化光度误差代价方程求解两帧图像之间相机的旋转变换。

从图像金字塔模型的第3层图像开始，通过图像配准构建求解当前帧相机位姿的代价方程。光度误差代价方程和当前帧相机位姿的代价方程分别为

$$E_{\text{pho\_rotation}}=\sum_{i\in\Omega}\left\|\boldsymbol{I}_m(\boldsymbol{u}_i)-\boldsymbol{I}_r\left[\boldsymbol{K}\exp(\Delta\hat{\boldsymbol{\xi}})\exp(\hat{\boldsymbol{\zeta}})D(\boldsymbol{x})\boldsymbol{K}^{-1}\tilde{\boldsymbol{x}}_i\right]\right\|^2,\tag{3}$$

$$E_{\text{pho\_pose}}=\sum_{i\in\Omega}\left\|\boldsymbol{I}_m(\boldsymbol{u}_i)-\boldsymbol{I}_r\left[\boldsymbol{K}\exp(\Delta\hat{\boldsymbol{\xi}})\exp(\hat{\boldsymbol{\xi}})D(\boldsymbol{x})\boldsymbol{K}^{-1}\tilde{\boldsymbol{x}}_i\right]\right\|^2,\tag{4}$$

式中：$E_{\text{pho_rotation}}$为优化相邻两帧图像之间的相机旋转变换过程中产生的光度误差；$E_{\text{pho_pose}}$为使用直接法优化当前帧相机位姿产生的光度误差；Ω为两幅图像中配准点的集合；$\boldsymbol{I}(\boldsymbol{u})$为图像上坐标点$\boldsymbol{u}$处的灰度值；$D(x)$为图像$r$上的像素点$\boldsymbol{u}$对应的深度；$\exp(\Delta\hat{\boldsymbol{\zeta}})$为旋转变换增量的李群表示形式；$\tilde{\boldsymbol{x}}_i$为图像上像素点的齐次坐标；$\exp(\hat{\boldsymbol{\zeta}})$为图像$r$和图像$m$之间的旋转变换的李群表示形式，$\boldsymbol{\zeta}=[\boldsymbol{\rho}_{3\times1},\boldsymbol{\phi}_{3\times1}]^{\mathrm{T}}\in\mathbf{R}^6$，$\boldsymbol{\rho}_{3\times1}\in\mathbf{R}^3$，$\boldsymbol{\phi}_{3\times1}\in\mathrm{so}(3)$，$\boldsymbol{\rho}_{3\times1}$表示相机平移向量，$\boldsymbol{\phi}_{3\times1}$表示相机旋转向量。并且

$$\hat{\boldsymbol{\zeta}}=\begin{bmatrix}\hat{\boldsymbol{\phi}}_{3\times1} & \boldsymbol{\rho}_{3\times1}\\ \boldsymbol{0}_{1\times3} & 0\end{bmatrix}\in\mathbf{R}^{4\times4}。\tag{5}$$

利用固定速度模型[4]或者参考帧模型初始化的相机位姿，分别将上一帧图像对应的世界坐标系下的三维点云和特征点，投影到当前帧图像的相机坐标系下，根据相机内参将相机坐标系下的点投影到当前帧图像上。对于得到的一组稀疏特征点的匹配点，构建对应的重投影误差代价方程，同时为了保证两幅图像对应区域外观不变，根据灰度不变原理构建相邻两帧图像之间的光度误差代价方程。采用改进后的图优化库g2o，同时最小化重投影误差和光度误差代价方程以优化相机位姿。

在优化的过程中总是将上一层图像配准得到的结果作为下一层图像配准的初始值，将第2层图像配准求解的相邻图像之间的相机旋转变换与上一帧图像相机位姿的乘积作为第3层图像配准中相机位姿的初始值，将第4层图像配准求得的相机位姿作为当前帧相机位姿。这种在像素较少的金字塔顶层图像中用于求解两帧图像之间的旋转变换，在像素较多的底层金字塔图像中用于求解6自由度的相机位姿的方法，使求得的相机位姿更稳健，算法的收敛速度也更快[21]。计算相邻两帧图像之间的相机旋转变换用到的光度误差代价方程如(3)式所示，计算当前帧相机位姿用到的重投影误差代价方程和光度误差代价方程分别如(1)式和(4)式所示。

如图2所示，$\boldsymbol{T}_{mr}^{(1)}$和$\boldsymbol{T}_{mr}^{(2)}$为金字塔顶部两层图像通过稠密直接法配准得到的图像$m$和图像$r$之间的旋转变换矩阵，将第1层图像配准的计算结果作为第2层图像配准的初始值；$\boldsymbol{T}_{wm}\boldsymbol{T}_{rm}^{(2)}$为图像$m$和图像$r$之间的旋转矩阵与图像$m$相机位姿的乘积，并将这个乘积作为第3层图像配准求解图像r相机位姿的初始值；将第3层图像配准求得的相机位姿作为第4层图像配准的初始值，并将第4层图像配准得到的相机位姿$\boldsymbol{T}_{wr}^{(2)}$作为图像$r$的相机位姿。

现有的图优化库g2o中只提供了如图3所示的同时优化相机位姿和地图点位置坐标的稀疏约束两元边和只优化相机位姿的稀疏约束一元边。为了能在图优化库g2o中同时最小化稀疏特征法的重投影

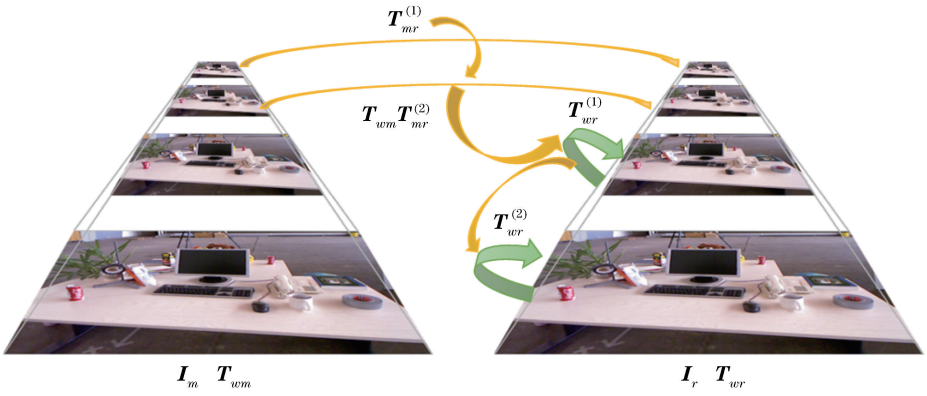

图 2 4 层图像金字塔模型和改进后的相机位姿求解步骤示意图

Fig. 2 Schematic of four-layer image pyramid model and improved camera pose estimation process

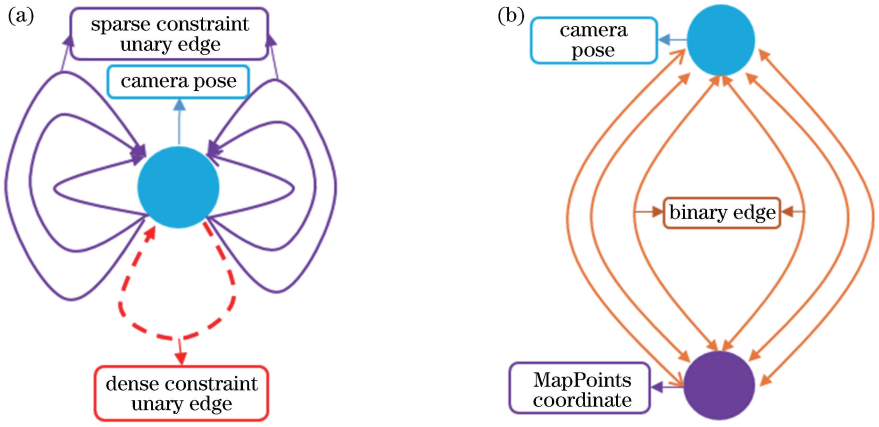

图 3 图优化库 g2o 中的一元边和两元边模型。(a)一元边;(b)两元边

Fig. 3 Unary edge and binary edge models in g2o library. (a) Unary edge model; (b) binary edge model

误差和稠密直接法的光度误差,优化求解相机位姿,在图优化库 g2o 中添加了稠密约束一元边,如图 3(a)中虚线表示的边。

2.2 相机位姿优化

系统中的相机位姿优化是通过图优化库 g2o 完成的。对于重投影误差,图优化库 g2o 为每一对匹配点创建一个稀疏约束一元边,如果也为稠密直接法中参与运算的每一个像素点创建一个稠密约束一元边,将大幅度增加计算机负担,降低整个系统的工作效率。为了将稠密的直接法和稀疏的特征法结合到一起求解相机位姿,又不至于使系统的工作效率降低,在图优化库 g2o 上创建了一条稠密约束一元边,将所有光度误差产生的约束初始化为图优化 g2o 中新添加的稠密约束一元边。

通过最小化重投影误差和光度误差优化相机位姿的问题本质上是一个非线性最小二乘问题。一个简单的最小二乘问题可以用 $\min \frac{1}{2}\| f(\boldsymbol{x})\|_2^2$ 来表示,$f(\boldsymbol{x})$ 为任意的一个非线性函数,$\boldsymbol{x}\in\mathbf{R}^n$,$f(\boldsymbol{x})\in\mathbf{R}^m$,表达式前面的 $\frac{1}{2}$ 是为了方便后期的求导加上的,$\| f(\boldsymbol{x})\|_2^2$ 表示 $f(\boldsymbol{x})$ 的二范数的平方。

高斯牛顿法[22]是求解非线性最小二乘问题常用的优化算法之一,该方法的核心思想是将目标函数一阶泰勒展开。非线性函数 $f(\boldsymbol{x})$ 的一阶泰勒展开后的形式为

$$f(\boldsymbol{x}+\Delta\boldsymbol{x})\approx f(\boldsymbol{x})+\boldsymbol{J}(\boldsymbol{x})\Delta\boldsymbol{x}, \tag{6}$$

式中:$\boldsymbol{J}(\boldsymbol{x})$ 是 $f(\boldsymbol{x})$ 关于 $\boldsymbol{x}$ 的导数,当 $\boldsymbol{x}$ 是一个向量时,$\boldsymbol{J}(\boldsymbol{x})$ 是一个 $m\times n$ 阶雅可比矩阵;$\Delta\boldsymbol{x}$ 为变量 $\boldsymbol{x}$ 的微小增量的向量形式。原来的非线性最小二乘问题可以转换为通过寻找目标微小增量 $\Delta\boldsymbol{x}^*$ 使得 $\| f(\boldsymbol{x}+\Delta\boldsymbol{x})\|_2^2$ 达到最小的线性最小二乘问题,表达式为

$$\Delta \boldsymbol{x}^{*}=\arg\min_{\Delta \boldsymbol{x}}\frac{1}{2}\left\|f(\boldsymbol{x})+\boldsymbol{J}(\boldsymbol{x})\Delta \boldsymbol{x}\right\|^{2}。\quad(7)$$

根据高斯牛顿法求解非线性最小二乘问题，下面分别介绍最小化重投影误差和光度误差求解相机位姿的数学公式推导过程。

2.2.1 重投影误差

首先将(1)式中的重投影误差代价方程中目标函数一阶泰勒展开，展开后的形式为

$$E_{\text{sparse}}\approx\sum_{i\in S}\left\|\left\{\boldsymbol{u}_{i}-\left[\boldsymbol{K}(\boldsymbol{I}+\Delta\boldsymbol{\xi})\exp(\hat{\boldsymbol{\xi}})\tilde{\boldsymbol{P}}_{i}\right]\right\}\right\|^{2}=\sum_{i\in S}\left\|\left[\boldsymbol{r}(\boldsymbol{\xi})-\boldsymbol{J}(\xi)\Delta\boldsymbol{\xi}\right]\right\|^{2},\quad(8)$$

式中：$\boldsymbol{I}$ 为单位矩阵。世界坐标系下的点 $\boldsymbol{P}_{i}$，在相机坐标系下对应点的坐标为 $\boldsymbol{q}=\Delta\boldsymbol{\xi}\exp(\hat{\boldsymbol{\xi}})\tilde{\boldsymbol{P}}_{i}$，点 $\boldsymbol{q}$ 在图像上投影点的像素坐标为 $\boldsymbol{u}=\boldsymbol{K}\boldsymbol{q}$，$\boldsymbol{r}(\boldsymbol{\xi})=\boldsymbol{u}_{i}-\boldsymbol{K}\exp(\hat{\boldsymbol{\xi}})\tilde{\boldsymbol{P}}_{i}$ 为余差。重投影误差代价方程中的目标函数对相机位姿增量的导数即雅可比矩阵表示为 $\boldsymbol{J}(\boldsymbol{\xi})=-\frac{\partial \boldsymbol{u}}{\partial \boldsymbol{q}}\frac{\partial \boldsymbol{q}}{\partial \Delta\boldsymbol{\xi}}$。根据相机内参，相机坐标系下的点 $\boldsymbol{q}=[x,y,z]^{\mathrm{T}}$ 投影到图像上得到的像素坐标表示为

$$\begin{bmatrix}u\\v\\1\end{bmatrix}=\frac{1}{z}\begin{bmatrix}f_{x}&0&c_{x}\\0&f_{y}&c_{y}\\0&0&1\end{bmatrix}\begin{bmatrix}x\\y\\z\end{bmatrix},\quad(9)$$

式中：$u=\frac{f_{x}}{z}x+c_{x}$；$v=\frac{f_{y}}{z}y+c_{y}$；$f_{x}=f/d_{u}$；$f_{y}=f/d_{v}$；d_{u} 和 d_{v} 为相机靶面上一个像素点横向方向和纵向方向的物理宽度，f 为相机焦距；c_{x} 和 c_{y} 分别是相机靶面横向和纵向长度的一半。

图像上 2D 像素点坐标对相机坐标系下的 3D 点坐标的偏导数为

$$\frac{\partial \boldsymbol{u}}{\partial \boldsymbol{q}}=\begin{bmatrix}\frac{\partial u}{\partial x}&\frac{\partial u}{\partial y}&\frac{\partial u}{\partial z}\\\frac{\partial v}{\partial x}&\frac{\partial v}{\partial y}&\frac{\partial v}{\partial z}\end{bmatrix}=\begin{bmatrix}\frac{f_{x}}{z}&0&-\frac{f_{x}x}{z^{2}}\\0&\frac{f_{y}}{z}&-\frac{f_{y}y}{z^{2}}\end{bmatrix}。\quad(10)$$

相机坐标系下的 3D 点坐标对当前帧相机位姿增量 $\Delta\boldsymbol{\xi}$ 的导数表示为

$$\frac{\partial \boldsymbol{q}}{\partial \Delta\boldsymbol{\xi}}=[\boldsymbol{I}-\hat{\boldsymbol{q}}]=\begin{bmatrix}1&0&0&0&-z&y\\0&1&0&z&0&x\\0&0&1&-y&x&0\end{bmatrix}。\quad(11)$$

根据链式求导法则，得到最终的雅可比矩阵为

$$\boldsymbol{J}(\boldsymbol{\xi})=-\frac{\partial \boldsymbol{u}}{\partial \Delta\boldsymbol{\xi}}=\frac{\partial \boldsymbol{u}}{\partial \boldsymbol{q}}\frac{\partial \boldsymbol{q}}{\partial \Delta\boldsymbol{\xi}}=-\begin{bmatrix}\frac{f_{x}}{z}&0&-\frac{f_{x}x}{z^{2}}&-\frac{f_{x}xy}{z^{2}}&f_{x}+\frac{f_{x}x^{2}}{z^{2}}&-\frac{f_{x}y}{z}\\0&\frac{f_{y}}{z}&-\frac{f_{y}y}{z^{2}}&-f_{y}-\frac{f_{y}y^{2}}{z^{2}}&\frac{f_{y}xy}{z^{2}}&\frac{f_{y}x}{z}\end{bmatrix}。\quad(12)$$

重投影误差方程可以转换为高斯牛顿方程，$\boldsymbol{J}(\boldsymbol{\xi})^{\mathrm{T}}\boldsymbol{J}(\boldsymbol{\xi})\Delta\boldsymbol{\xi}=\boldsymbol{J}(\boldsymbol{\xi})\boldsymbol{r}(\boldsymbol{\xi})$，通过迭代法求解高斯牛顿方程的步骤如下：

1) 给定李代数形式的初始相机位姿 $\boldsymbol{\xi}_{0}$；2)对于第 m 次迭代，求解当前的雅各比矩阵 $\boldsymbol{J}(\boldsymbol{\xi}_{0})$ 和余差 $\boldsymbol{r}(\boldsymbol{\xi}_{0})$；3)求解高斯牛顿方程：$\boldsymbol{J}(\boldsymbol{\xi}_{0})^{\mathrm{T}}\boldsymbol{J}(\boldsymbol{\xi}_{0})\Delta\boldsymbol{\xi}_{m}=\boldsymbol{J}(\boldsymbol{\xi}_{0})\boldsymbol{r}(\boldsymbol{\xi}_{0})$；4)若 $\Delta\boldsymbol{\xi}_{m}$ 足够小，则停止迭代。否则，令 $\boldsymbol{\xi}_{m+1}=\ln[\exp(\Delta\hat{\boldsymbol{\xi}}_{m})\exp(\hat{\boldsymbol{\xi}}_{m})]$，返回到第 2)步。

2.2.2 光度误差

首先将(4)式中的光度误差代价方程的目标函数一阶泰勒展开，展开后的形式为

$$E_{\text{pho\_pose}}\approx\sum_{i\in\Omega}\left\|\boldsymbol{I}_{m}(\boldsymbol{u}_{i})-\boldsymbol{I}_{r}\left[\boldsymbol{K}(\boldsymbol{I}+\Delta\boldsymbol{\xi})\exp(\hat{\boldsymbol{\xi}})D(\boldsymbol{x})\boldsymbol{K}^{-1}\tilde{\boldsymbol{x}}_{i}\right]\right\|^{2}=\sum_{i\in\Omega}\left\|\boldsymbol{r}(\boldsymbol{\xi})-\boldsymbol{J}(\boldsymbol{\xi})\Delta\boldsymbol{\xi}\right\|^{2}=\sum_{i\in\Omega}\left[\left\|\boldsymbol{r}(\boldsymbol{\xi})\right\|_{2}^{2}-2\boldsymbol{J}(\boldsymbol{\xi})^{\mathrm{T}}\boldsymbol{r}(\boldsymbol{\xi})\Delta\boldsymbol{\xi}+(\Delta\boldsymbol{\xi})^{\mathrm{T}}\boldsymbol{J}(\boldsymbol{\xi})^{\mathrm{T}}\boldsymbol{J}(\boldsymbol{\xi})(\Delta\boldsymbol{\xi})\right]=\sum_{i\in\Omega}\left[c_{i}-2\boldsymbol{b}_{i}^{\mathrm{T}}\Delta\boldsymbol{\xi}+(\Delta\boldsymbol{\xi})^{\mathrm{T}}\boldsymbol{H}_{i}(\Delta\boldsymbol{\xi})\right]=c-2\boldsymbol{b}^{\mathrm{T}}\Delta\boldsymbol{\xi}+(\Delta\boldsymbol{\xi})^{\mathrm{T}}\boldsymbol{H}(\Delta\boldsymbol{\xi}),\quad(13)$$

$$\boldsymbol{r}(\boldsymbol{\xi})=\boldsymbol{E}(\boldsymbol{\xi})=\boldsymbol{I}_{r}(\boldsymbol{u}_{i})-\boldsymbol{I}_{m}\left[\boldsymbol{K}\exp(\hat{\boldsymbol{\xi}})D(\boldsymbol{x})\boldsymbol{K}^{-1}\tilde{\boldsymbol{x}}_{i}\right],\quad(14)$$

$$c=\sum_{i\in\Omega}\left\|\boldsymbol{r}(\boldsymbol{\xi})\right\|_{2}^{2},\quad(15)$$

$$\boldsymbol{b}=\sum_{i\in\Omega}\boldsymbol{J}(\boldsymbol{\xi})^{\mathrm{T}}\boldsymbol{r}(\boldsymbol{\xi}),\quad(16)$$

$$\boldsymbol{H}=\sum_{i\in\Omega}\boldsymbol{J}(\boldsymbol{\xi})^{\mathrm{T}}\boldsymbol{J}(\boldsymbol{\xi}),\quad(17)$$

式中：$\boldsymbol{I}_{r}$ 表示图像 r；$\boldsymbol{I}_{m}$ 表示图像 m。

在 GPU 上计算 $\boldsymbol{H}$ 和 $\boldsymbol{b}$，将计算结果传到 CPU，在 CPU 上分解矩阵 $\boldsymbol{H}$ 和 $\boldsymbol{b}$ 得到雅可比矩阵 $\boldsymbol{J}(\boldsymbol{\xi})$ 和余差 $\boldsymbol{r}(\boldsymbol{\xi})$。用分解出的 $\boldsymbol{J}(\boldsymbol{\xi})$ 和 $\boldsymbol{r}(\boldsymbol{\xi})$ 初始化改进图优化库 g2o，并在 g2o 中新加入的稠密约束一元边，构建稠密光度误差代价方程的高斯牛顿方程。最终在图优化库 g2o 中通过迭代法求解两个高斯牛顿方程，使总的误差方程达到最小，求解相机位姿。总的误差方程表示为

$$E_{\text{total}}=E_{\text{sparse}}+E_{\text{pho\_pose}}。\tag{18}$$

式中 E_{total} 为重投影误差和光度误差的组合误差。

2.3 稠密重建

在 ORB-SLAM2 系统框架下新创建了稠密重建线程，虽然在稠密建图线程中沿用了 ElasticFusion 系统中将 RGB-D 图像融合到 Surfel 面元模型中重建稠密模型的方法，但是在检测到闭环时采用了不同的方法更新重建的稠密模型中的点。在 ElasticFusion 系统中，检测到全局闭环之后，直接从重建的稠密模型表面按照时间顺序均匀地提取点构建变形图[23]，用环路两侧闭环帧图像之间产生的闭环约束优化变形图，将优化之后的变形图作用到重建的稠密模型表面，更新稠密模型上的点。而在改进后的 ORB-SLAM2 系统中，闭环线程检测到全局闭环之后，先优化基础图中所有结点上关键帧的相机位姿，将环路中积累的误差均匀到每一帧上，使得相机运动轨迹在闭环处闭合，然后再执行全局 BA 优化。在基础图上均匀地取点，构建新的变形图，在优化变形图时除了使用环路两侧闭环帧图像之间产生的闭环约束，还用到了优化基础图产生的约束，将经过两重约束优化的变形图作用到重建的稠密地图上更新地图中的点。

3 实验结果

3.1 相机运动轨迹评估

使用绝对轨迹误差(ATE)[24]的方法，对 DVO SLAM[25]，Kintinuous，ElasticFusion，ORB-SLAM2 和改进后的 ORB-SLAM2 系统在 TUM RGB-D 数据集的办公室场景 fr1 中的 9 个图像序列上测试得到的相机运动轨迹分别与对应的标准轨迹进行绝对误差计算，各个系统的绝对误差对比结果如表 1 所示。图 4 列出了改进后的 ORB-SLAM2 系统在 4 个图像序列上测试得到的相机运动轨迹与标准轨迹之间的绝对误差图。在图 4 中点线表示 ground truth，实线表示改进后的系统输出的相机运动轨迹，虚线表示上述两者之间对应点处的误差。

表 1 在 TUM RGB-D 标准数据集上的相机轨迹对比结果

Table 1 Comparison of camera trajectories on TUM RGB-D benchmark datasets

Sequence	Relative path /m				
	DVO SLAM	Kintinuous	ElasticFusion	ORB-SLAM2	Improved system
fr1/desk	0.022	0.142	0.022	0.016	**0.016**
fr1/plant	0.027	0.059	0.043	0.016	**0.014**
fr1/teddy	0.049	0.237	0.091	**0.035**	0.056
fr1/room	0.064	0.182	0.198	**0.059**	**0.056**
fr1/360	**0.074**	0.202	0.270	failed	**0.129**
fr1/desk2	0.035	0.140	0.058	0.031	**0.025**
fr1/floor	0.035	0.140	0.058	0.031	0.035
fr1/rpy	**0.022**	0.041	0.037	0.064	**0.023**
fr1/xyz	0.013	0.021	0.014	**0.009**	**0.009**

由表 1 可得，基于特征的方法包括 ORB-SLAM2 和改进后的系统的相机跟踪精度比 DVO SLAM、Kintinuous、ElasticFusion 的跟踪精度高。并且由于改进后的系统中使用了最小化组合的重投影误差和光度误差的方法优化相机位姿，相机的运动轨迹要比只优化重投影误差的 ORB-SLAM2 系统的跟踪精度高。DVO SLAM 在 fr1/360 和 fr1/rpy 序列上跟踪精度比其他系统高，是因为该系统在关键帧之间建立了额外的位姿约束并且采用位姿图优化提高相机位姿精度。由表 1 可得，高精度的相机位姿使改进后的 ORB-SLAM2 系统在大多数数据集上表现良好。因为在 TUM RGB-D 数据集中只有很少的闭环，跟踪线程中用最小化组合的重投影误差和光度误差优化相机位姿，局部建图线程中的局部 BA 进一步优化相机位姿使系统得到的相机位姿精度比只有遇到闭环时才执行位姿图优化的 DVO SLAM 系统得到的相机位姿精度高。

3.2 稠密模型精度评估

ICL-NUIM[26]数据集提供了一种评估稠密模型重建精度的方法，ICL-NUIM 数据集中包括了 4 个图像序列，lr_kt0，lr_kt1，lr_kt2，lr_kt3，4 个图像序列的重建效果如图 5 所示。紫色小块表示相机，蓝色小块表示提取的关键帧，绿色线表示 covisibility graph(无向由权图)中的共视关系[4-5]，粉色线表示检测到闭环之后纠正过的相机运动轨迹，红色点表示重建出的稀疏的地图点。在图 5 中不仅展示了改进后的系统重建的稠密地图，还可以看到输出的相机运动轨迹和稀疏地图，图 5 中，四周

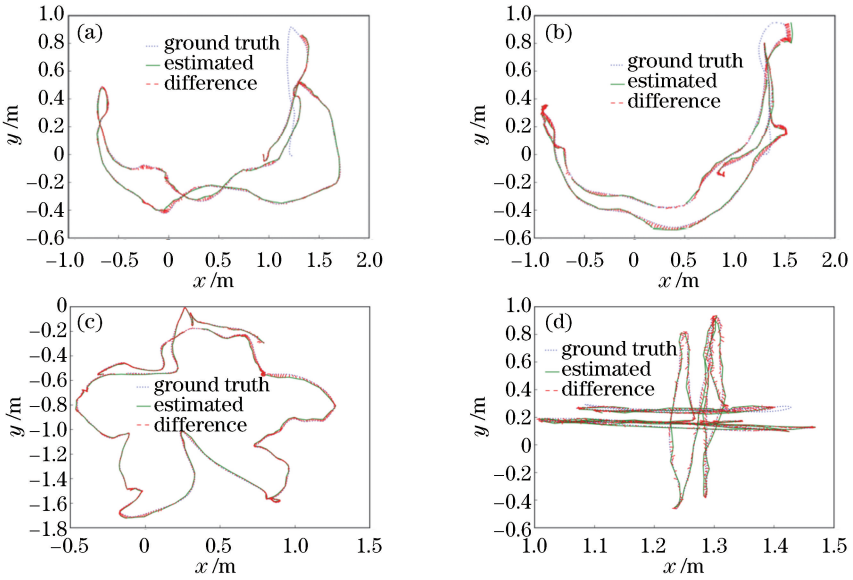

图 4 改进后的 ORB-SLAM2 系统在 4 个图像序列上的相机运动轨迹与标准轨迹对比图。(a) fr1/desk 图像序列；(b) fr1/desk2 图像序列；(c) fr1/plant 图像序列；(d) fr1/xyz 图像序列

Fig. 4 Comparison between ground truths and real trajectories of cameras generated by improved ORB-SLAM2 system on four frame sequences. (a) fr1/desk frame sequence; (b) fr1/desk2 frame sequence; (c) fr1/plant frame sequence; (d) fr1/xyz frame sequence

为重建的场景的稠密模型，内部为相机的运动轨迹。表 2 列出了 4 个图像序列的具体情况。采用 ICL-NUIM 数据集中提供的 SurfReg 配准工具，在 ICL-NUIM 数据集上，分别将 Kintinuous 系统、ElasticFusion 系统和改进后的 ORB-SLAM2 系统生成的 3D 模型与数据集提供的标准 3D 模型配准，配准结果如表 3 所示。由结果可得，与另外两个系统相比，改进后的系统在 4 个图像序列上生成的稠密模型重建精度更高。

表 2 ICL-NUIM 数据集中 4 个图像序列的细节

Table 2 Details of four frame sequences on ICL-NUIM benchmark datasets

Sequence	Frame	Length /m	Withloop
lr_kt0	1510	6.54	No
lr_kt1	967	2.05	No
lr_kt2	882	8.43	No
lr_kt3	1242	11.32	Yes

表 3 在 ICL-NUIM 数据集上生成的 3D 模型对比结果

Table 3 Comparison of 3D models on ICL-NUIM benchmark datasets

System	3D model result /m			
	lr_kt0	lr_kt1	lr_kt2	lr_kt3
Kintinuous	0.087	0.470	0.162	0.205
ElasticFusion	0.058	0.505	0.185	0.219
Improved ORB-SLAM2	0.007	0.467	0.155	0.190

在 lr_kt0 序列中重建的场景比较小，由图 5(a)中相机运动轨迹可知，相机扫描的场景有重叠，每次相机返回到之前扫描过的区域时，ElasticFusion 系统和改进后的系统通过配准从非活动区域和活动区域投影得到的两帧点云，检测局部闭环，从检测到的闭环中生成闭环约束，利用闭环约束优化变形图，然后用优化后的变形图更新重建的稠密地图。因为 Kintinuous 系统中没有局部闭环检测机制，所以这两个系统重建的模型精度比 Kintinuous 系统的高。在 lr_kt1 序列中，相机在短距离内经过多次大角度旋转重建了较大的场景，由图 5(b)中相机运动轨迹可知，相机一直在扫描新的场景，没有重叠的视角，所以系统没有局部闭环和全局闭环修正，相机的累计误差较大，三个系统重建的模型的效果都比较差。虽然 lr_kt2 序列也重建了较大的场景，根据图 5(c)中相机运动轨迹可知，在扫描过程中相机运动平稳没有发生大角度的旋转，相机在运动过程中，相邻图像帧之间呈现的场景相似度比较高，系统有充足的时间进行相机位姿的优化和地图点的融合，使得重建场景的效果比在 lr_kt1 序列上重建的效果好。在图 5(d)中 lr_kt3 序列中有闭环存在，改进后的系统在检测到闭环时先优化基础图，使相机运动轨迹在闭环处闭合。在优化变形图时除了使用环路两侧闭环帧图像之间产生的闭环约束，还用到了优化基础图产生的约束，将经过两重约束优化的变形图作

用到重建的稠密地图上更新地图中的点，所以改进后的系统重建出的场景的表面更加平整光滑，如图6所示。在ElasticFusion系统中检测到全局闭环后直接用闭环约束优化变形图，没有执行基础图优化，所以在相机运动距离较大，轨迹中累积的误差较大时，ElasticFusion系统在闭环处的重建精度并不高。结合图5(d)和图6(c)可得，改进后的系统不仅可以生成稀疏的地图还可以生成更高精度的稠密地图。

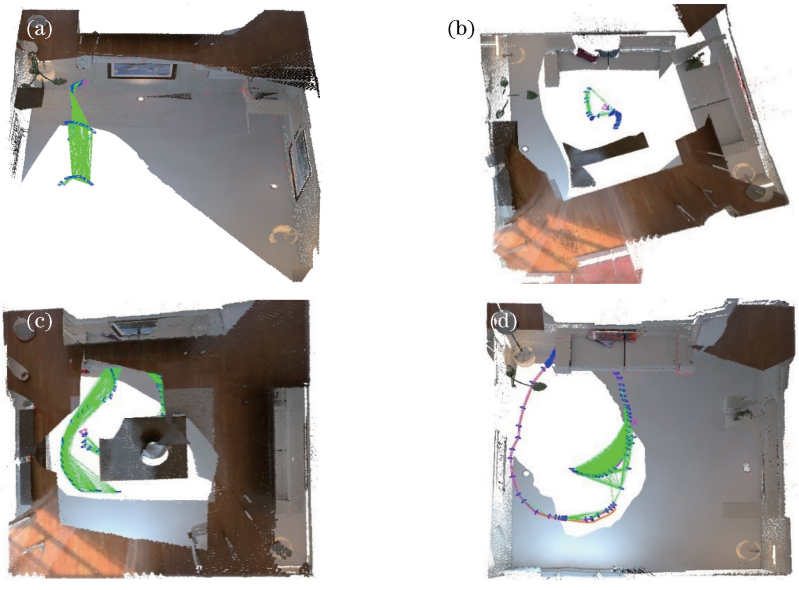

图5 改进后的ORB-SLAM2系统在ICL-NUIM数据集的4个图像序列上的重建结果和相机的运动轨迹。(a) lr_kt0；(b) lr_kt1；(c) lr_kt12；(d) lr_kt3

Fig. 5 Reconstruction results and camera trajectories on four frame sequences by improved ORB-SLAM2 system on ICL-NUIM benchmark datasets. (a) lr_kt0; (b) lr_kt1; (c) lr_kt12; (d) lr_kt3

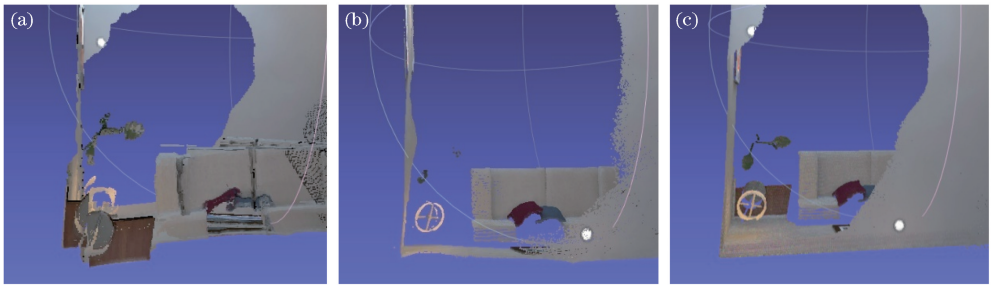

图6 Kintinuous，ElasticFusion和改进后的ORB-SLAM2系统在lr_kt3图像序列的闭环处重建结果对比。(a) Kintinuous系统；(b) ElasticFusion系统；(c)改进后的ORB-SLAM2系统

Fig. 6 Comparison of reconstruction results of Kintinuous, ElasticFusion, and improved ORB-SLAM2 systems on closed-loop region of lr_kt3 frame sequence. (a) Kintinuous system; (b) ElasticFusion system; (c) improved ORB-SLAM2 system

4 结 论

在ORB-SLAM2系统框架下，将稠密直接法应用到跟踪线程中，结合系统中原有的稀疏特征法，利用加入稠密约束一元边的图优化库g2o联合优化直接法的光度误差和特征法的重投影误差代价方程求解相机位姿。在原系统中创建稠密重建线程，将输入的RGB-D图像融合到Surfel面元模型中生成稠密地图，并实时将重建结果反馈给用户。实验结果表明，所提方法在一定程度上可以提高相机的跟踪精度和稠密场景的重建精度，但是直接法的使用增加了系统的计算负担，为了不降低系统的效率，需要使用GPU加速，但是GPU的使用增加了系统的成本。为此，未来的研究方向将着眼于进一步提高相机位姿的求解精度和降低计算量。

参 考 文 献

[1] Liu T, Zong Q, Liu P H, *et al*. Generation and control of multi-robot formation based on structural

persistence[J]. Information and Control, 2018, 47(3): 314-323.
刘彤, 宗群, 刘朋浩, 等. 基于结构持久图和视觉定位的多机器人编队生成与控制[J]. 信息与控制, 2018, 47(3): 314-323.

[2] Zheng G Q, Zhou Z P. Improved augmented reality registration method based on VSLAM[J]. Laser & Optoelectronics Progress, 2019, 56(6): 061501.
郑国强, 周治平. 一种基于视觉即时定位与地图构建的改进增强现实注册方法[J]. 激光与光电子学进展, 2019, 56(6): 061501.

[3] 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, November 13-16, 2007, Nara, Japan. New York: IEEE, 2007: 10052991.

[4] 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): 1147-1163.

[5] 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.

[6] Engel J, Schöps T, Cremers D. LSD-SLAM: large-scale direct monocular SLAM[M]// Fleet D, Pajdla T, Schiele B, *et al*. Computer Vision-ECCV 2014. Lecture Notes in Computer Science. Cham: Springer, 2014, 8690: 834-849.

[7] Newcombe R A, Izadi S, Hilliges O, *et al*. Kinectfusion: real-time dense surface mapping and tracking[C]// 10th IEEE International Symposium on Mixed and Augmented Reality, ISMAR 2011, October 26-29, 2011, Basel, Switzerland. New York: IEEE, 2011: 127-136.

[8] Izadi S, Kim D Hilliges O, *et al*. KinectFusion: real-time 3D reconstruction and interaction using a moving depth camera[C]// Mixed and augmented reality (ISMAR), 2011 10th IEEE International Symposition, October 26-29, 2011, Basel, Switzerland. New York: IEEE, 2011: 127-136.

[9] Whelan T, Salas-Moreno R F, Glocker B, *et al*. ElasticFusion: real-time dense SLAM and light source estimation[J]. The International Journal of Robotics Research, 2016, 35(14): 1697-1716.

[10] Whelan T, Leutenegger S, Salas-Moreno R F, *et al*. ElasticFusion: dense SLAM without a pose graph[C]// 11th Robotics: Science and Systems XI, July 13-17, 2015, Rome, Italy. Germany: Schloss Dagstuhl, 2015.

[11] Galvez-López D, Tardos J D. Bags of binary words for fast place recognition in image sequences[J]. IEEE Transactions on Robotics, 2012, 28(5): 1188-1197.

[12] Tu S Q, Xue Y J, Liang Y, *et al*. Review on RGB-D image classification[J]. Laser & Optoelectronics Progress, 2016, 53(6): 060003.
涂淑琴, 薛月菊, 梁云, 等. RGB-D图像分类方法研究综述[J]. 激光与光电子学进展, 2016, 53(6): 060003.

[13] Jia T, Zhou Z X, Gao H H, *et al*. Depth measurement based on infrared coded structured light[J]. Infrared and Laser Engineering, 2015, 44(5): 1628-1632.
贾同, 周忠选, 高海红, 等. 基于红外编码结构光的深度测量方法[J]. 红外与激光工程, 2015, 44(5): 1628-1632.

[14] Curless B, Levoy M. A volumetric method for building complex models from range images[C]// The 23rd Annual Conference on Computer Graphics and Interactive Techniques-SIGGRAPH '96, August 4-9, 1996, New York, NY, USA. New York: ACM, 1996: 303-312.

[15] Nießner M, Zollhöfer M, Izadi S, *et al*. Real-time 3D reconstruction at scale using voxel hashing[J]. ACM Transactions on Graphics, 2013, 32(6): 169.

[16] Whelan T, Kaess M, Johannsson H, *et al*. Real-time large-scale dense RGB-D SLAM with volumetric fusion[J]. The International Journal of Robotics Research, 2015, 34(4/5): 598-626.

[17] Pfister H, Zwicker M, van Baar J, *et al*. Surfels: surface elements as rendering primitives[C]// The 27th Annual Conference on Computer Graphics and Interactive Techniques - SIGGRAPH '00, July 23-28, 2000, New York, NY, USA. New York: ACM, 2000: 335-342.

[18] Fu X Y, Zhu F, Wu Q X, *et al*. Real-time large-scale dense mapping with surfels[J]. Sensors, 2018, 18(5): 1493.

[19] Dai A, Nießner M, Zollhöfer M, *et al*. BundleFusion: real-time globally consistent 3D reconstruction using on-the-fly surface reintegration[J]. ACM Transactions on Graphics, 2017, 36(3): 24.

[20] Kümmerle R, Grisetti G, Strasdat H, *et al*. g2o: a general framework for graph optimization[C]// 2011

IEEE International Conference on Robotics and Automation, May 9-13, 2011, Shanghai, China. New York: IEEE, 2011: 3607-3613.

[21] Newcombe R A, Lovegrove S J, Davison A J. DTAM: dense tracking and mapping in real-time[C]// 2011 International Conference on Computer Vision, November 6-13, 2011, Barcelona, Spain. New York: IEEE, 2011: 2320-2327.

[22] Gao X, Zhang T, Liu Y, *et al*. 14 lectures on visual SLAM: from theory to practice [M]. Beijing: Publishing House of Electronics Industry, 2017: 111-114.
高翔, 张涛, 刘毅, 等. 视觉 SLAM 十四讲: 从理论到实践[M]. 北京: 电子工业出版社, 2017: 111-114.

[23] Sumner R W, Schmid J, Pauly M. Embedded deformation for shape manipulation [J]. ACM Transactions on Graphics, 2007, 26(3): 80.

[24] 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, October 7-12, 2012, Vilamoura, Portugal. New York: IEEE, 2012: 573-580.

[25] Kerl C, Sturm J, Cremers D. Dense visual SLAM for RGB-D cameras [C] // 2013 IEEE/RSJ International Conference on Intelligent Robots and Systems, November 3-7, 2013, Tokyo, Japan. New York: IEEE, 2013: 2100-2106.

[26] Handa A, Whelan T, McDonald J, *et al*. A benchmark for RGB-D visual odometry, 3D reconstruction and SLAM [C] // 2014 IEEE International Conference on Robotics and Automation (ICRA), May 31-June 7, 2014, Hong Kong, China. New York: IEEE, 2014: 1524-1531.