

DOI:10.16136/j.joel.2022.01.0563

基于注意力机制的特征点匹配网络的 SLAM 方法

沮晨洋¹, 刘凤连^{1*}, 汪日伟²

(1. 天津理工大学 计算机视觉与系统教育部重点实验室和天津市智能计算及软件新技术重点实验室, 天津 300384; 2. 温州理工学院, 浙江 温州 325035)

摘要:已有的实时定位与地图构建(simultaneous localization and mapping, SLAM)方案采用的特征点匹配方法普遍会受视角变化的影响使得特征点的匹配比较困难,进而干扰到特征点匹配的精度,最终影响到三维(three-dimensional, 3D)点云地图构建以及相机运动位姿估计的精度。为此,本文提出一种基于注意力机制的特征点匹配网络的 SLAM 方法。相比于现有的 SLAM 方法,本文将 SLAM 中视觉里程计模块的特征点匹配的方法替换成了一个全新的、基于注意力机制的特征点匹配网络的特征点匹配方法,并和传统的特征点提取方法做了一个全新的特征点提取与匹配的组合,形成了一个新的视觉里程计,进而形成了一个新的 SLAM 方法。首先,通过传统的特征点提取算法进行特征点的提取,对提取的特征点及描述子向量进行编码,通过图注意力神经网络进行学习得到匹配描述子,根据匹配描述子创建得分矩阵,采用最优传输算法求解最优得分矩阵,计算得到最优匹配点对,到这里就完成了特征点提取与匹配的整个过程;基于匹配点对完成相机的定位、建图和回环检测。本文采用 KITTI 公开数据集进行实验,实验结果表明采用基于注意力机制特征点匹配网络的 SLAM 方案,在视角变化不稳定的情况下,相机运动轨迹误差和相机位姿估计误差的精度明显有所提升。

关键词:SLAM; 注意力机制; 特征点匹配; 神经网络**中图分类号:**TP311 **文献标识码:**A **文章编号:**1005-0086(2022)01-0014-09

A SLAM method for feature point matching network based on attention mechanism

ZU Chenyang¹, LIU Fenglian^{1*}, WANG Riwei²

(1. Key Laboratory on Computer Vision and Systems, Ministry of Education of China, Key Laboratory on Intelligence Computing and Novel Software Technology of the City of Tianjin, Tianjin University of Technology, Tianjin 300384, China; 2. Wenzhou University of Technology, Wenzhou, Zhejiang 325035, China)

Abstract: The current feature point matching method for simultaneous localization and mapping (SLAM) is generally affected by the change of perspective, which makes the matching of feature points difficult, which in turn deteriorates the accuracy of feature point matching, and ultimately influences the construction of three-dimensional (3D) point cloud maps and the estimation accuracy of camera motion pose. For this reason, this paper presents an attention based on feature point matching network for SLAM. The innovation of this article is that compared with the existing SLAM and we replaces the feature point matching method of the visual odometer module in SLAM with an attention based on feature point matching network for feature point matching. And we make a new combination of feature point extraction and matching with the traditional feature point extraction method to form a new visual odometer and a new SLAM. Firstly, we encode the extracted feature points and descriptor vectors and we learn through the graph attention neural network to obtain matching descriptors. Then we create a score matrix based on

* E-mail: lflian@tjut.edu.cn

收稿日期:2021-08-10 修订日期:2021-09-28

基金项目:国家自然科学基金(62020106004)资助项目

the matching descriptors and use the optimal transmission algorithm to solve the optimal score matrix. In the end we calculate the optimal matching point pair and complete camera positioning, mapping, and loop detection based on the optimal matching point pairs. The experimental results show that when the viewing angle is unstable, an attention based on feature point matching network for SLAM can significantly improve the accuracy of the camera's trajectory and the estimation accuracy of camera motion pose.

Key words: SLAM; attention mechanism; matching feature point; neural network

1 引言

在实时定位与地图构建(simultaneous localization and mapping, SLAM)^[1]中,提取到图像序列特征点后,基于不同的匹配算法计算出的特征点间的对应关系对于相机运动位姿的估计和三维(three-dimensional, 3D)点云的构建至关重要。但是视角的改变、光照变化以及遮挡等因素严重影响着对应点对的准确性。

较早期的视觉SLAM系统是由DAVISON等^[2]在2007年提出的Mono-SLAM,该系统通过单目相机构建稀疏的点云地图,但在复杂场景下光照的变化会严重影响系统的鲁棒性;KLEIN等^[3]利用FAST角点构建FAST-SLAM,由于其高效的特征点检测能力使得SLAM算法的运行速度获得有效提升,但只能应用于类似室内这样小环境的范围。ENGEL等^[4]提出了基于直接法的LSD-SLAM,该方法能够适用于大尺度的场景并且对于相机的旋转角度较大的问题还具有较高的鲁棒性;RUBLEE等^[5]提出了基于ORB特征提取方法的ORB-SLAM系统,该方法对于相机的运动视角变化的问题具有较高的鲁棒性^[6],并且系统的实时性非常高,但对于外界光照的影响不够鲁棒。现在随着深度学习的兴起,越来越多的学者会将深度学习神经网络^[7]应用于SLAM系统中来解决相关问题。KEISUKE等^[8]提出的CNN-SLAM系统在深度预测方面融合了CNN,相比于传统SLAM方法提高了点云地图的精度,并且对于低纹理的问题系统更加鲁棒。KENDALL等^[9]提出的基于PoseNet的SLAM系统使用深度学习的方法解决了复杂的图像外回归问题,也使系统更具有鲁棒性。UMMENHOFER等^[10]提出的基于端到端训练的卷积神经网络SLAM系统,该系统体现了在深度估计方面的优势,提高了建图的精度。

除了与深度学习的结合,还有很多针对应用场景的SLAM性能的改进,如GAN等^[11]提出了一种提高SLAM性能的动态检测方法,增强在动态环境中的SLAM系统性能。SLAM中的每一个重要模块都承担了非常重要的责任,目前也有很

多学者对已有模块的算法进行相应优化,LUO等^[12]提出的一种基于堆叠分类自动编码器的回环检测算法,相比于已有回环检测算法提高了精度和鲁棒性。在现实环境中总会存在很多外界因素影响SLAM系统的性能,便产生了很多针对不同环境因素对SLAM进行的改进,为了提升基于特征点的双目视觉定位算法在低光照环境下定位的准确性,李红莉等^[13]提出了一种基于在线估计的视觉同步定位与地图构建低光照图像增强算法,并且通过实验证明在低光照的环境下,本文提出的视觉SLAM系统,具有更好的定位准确性及鲁棒性。

通过对早期的SLAM方案和现在很多基于深度学习^[14]的SLAM方案的了解,从而发现复杂环境中各种因素会对SLAM系统产生影响,尤其是环境中的光照变化、相机视角变化、低纹理等因素,这些因素往往会影响到系统的鲁棒性,包括点云地图的准确性。传统方法的实时性很高,现在融合了深度学习的SLAM方案中,将深度学习的方法与特征点匹配、深度估计、位姿估计模块结合得较多^[15],最终能够得到相较于传统方法更加精确的点云地图。在已有的视觉SLAM方法会在室内机器人、室内外导航等方面进行实际应用,其中使用TUM公开数据集进行实验,在室内和室外进行实时定位,形成相机的运动轨迹,但现有方法的定位精度还不够高,在相机的轨迹和运动误差方面还应该再进一步进行优化,以达到更高的精度,那么针对现在存在的问题,本文也提出了一种基于注意力机制的特征点匹配网络^[16]的SLAM算法系统,该系统的视觉里程计模块结合了深度学习注意力机制^[17]的特征点匹配网络的特征点匹配方法,与传统的特征点提取方法形成一个全新的组合,来提高系统的鲁棒性、相机运动轨迹和运动的精度,这就是提出这个方法的意义所在。

2 基于注意力机制的匹配网络的SLAM算法分析

本文提出基于注意力机制特征匹配网络的SLAM方法示意图如图1所示。本文的创新点体现

在视觉里程计模块中传统的特征点提取算法与基于注意力机制的特征点匹配网络的特征点匹配算法的全新组合。在特征点匹配部分采用了图注意力神经网络,将特征点的位置坐标和描述子进行耦合,而后将耦合后的特征点通过注意力图神经网络进行更新,更新后得到的向量即为特征点的匹配描述子,再利用得到的匹配描述子计算出得分矩阵,通过最优传输算法进行求解,从而得出最优的特征点匹配,而后通过建图模块进行点云地图的构建。

本文的创新点就是特征点提取算法和基于注意力机制特征匹配网络的匹配算法的新组合。传统的特征点提取算法提取速度快提取效率高,基于注意力机制特征匹配网络的匹配算法匹配精度高,因此方法利用模拟人眼比较两幅图每个点的特异性,找出相同点的思想,并利用注意力机制以及图神经网络进行实现,实验证明这个全新组合能够大大提高相机位姿估计的精度以及相机运动轨迹的精度。

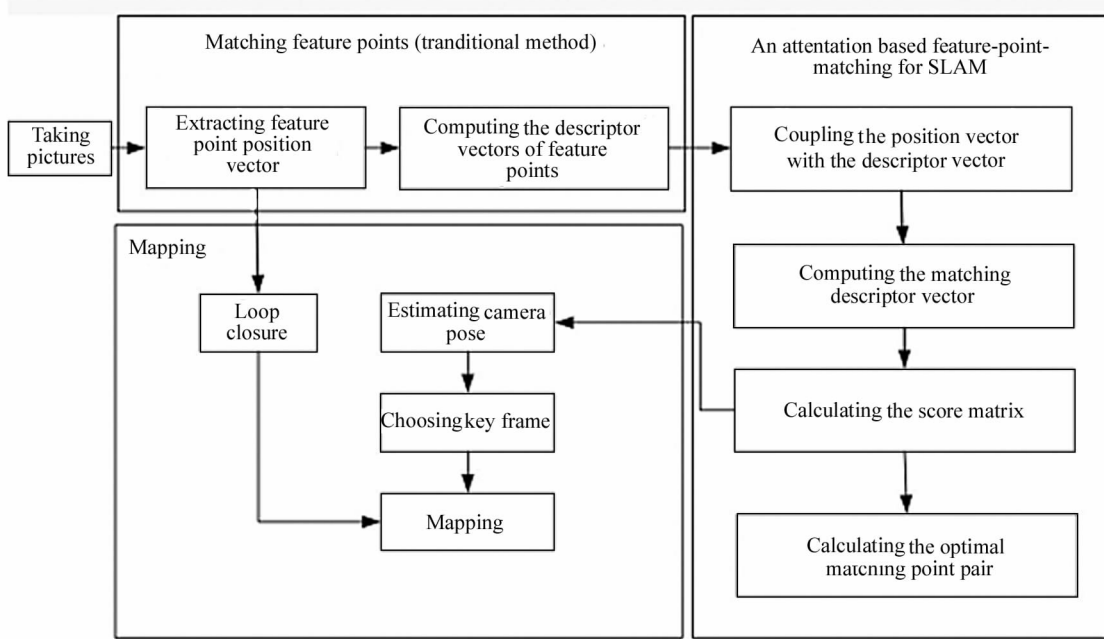


图1 基于注意力机制特征匹配网络的SLAM方法示意图

Fig. 1 Schematic diagram of a network SLAM method based on attention feature matching

2.1 特征点提取

图像的特征点可以简单地理解为图像中比较显著的点,如轮廓点,较暗区域中的亮点,较亮区域中的暗点等.本文特征点提取算法借鉴了ORB算法中FAST角点提取的思想,本文重新定义了新的提取特征点规则。

首先,从相机采集到的图像帧中提取候选特征点,对于图像中的每一个像素 p ,设置像素 p 的灰度值为 H_p ,并且设定一个阈值为这个像素的灰度值的30%,也就是 $30\% H_p$ 。然后以每一个像素为圆心,取半径为3的圆上的16个像素点为提取特征点,如图2所示。那么判断像素点 p 是否为特征点的条件如下,设 H_{p_i} 为圆上的某一个像素点 p_i 的灰度值:

$$\begin{cases} H_{p5} > H_p + 30\% H_p / H_{p5} < H_p - 30\% H_p \\ H_{p13} > H_p + 30\% H_p / H_{p13} < H_p - 30\% H_p \end{cases}, (1)$$

$$\begin{cases} H_{p2} > H_p + 30\% H_p / H_{p2} < H_p - 30\% H_p \\ H_{p6} > H_p + 30\% H_p / H_{p6} < H_p - 30\% H_p \\ H_{p10} > H_p + 30\% H_p / H_{p10} < H_p - 30\% H_p \\ H_{p14} > H_p + 30\% H_p / H_{p14} < H_p - 30\% H_p \end{cases}, (2)$$

$$H_{p_i} > H_p + 30\% H_p \text{ 或 } H_{p_i} < H_p - 30\% H_p, \\ i = 1 \dots 16. (3)$$

满足式(1),并同时满足式(2)中的任意3个像素点和式(3)中的任意13个像素点,那么就选定像素点 p 为特征点。按照上述方法,就能提取出每幅图像帧中的特征点,从而也就得到了特征点的相应二维坐标,采用以上方法选定特征点的速度会快很多。已有的快速排除一大部分非特征点的点的规则:选取1、5、9、13四个位置的灰度值,如果是一个特征点,那么上述4个像素点中至少有3个应该和点

相同。如果都不满足,那么不可能是一个特征点。针对筛选特征点的规则,本文规定新的筛选方法:以像素点 p 为圆心,以 3 个像素为半径形成一个圆,选取这个圆经过的 16 个像素点,如果圆上经过的像素点在圆内的面灰度值超过整个像素点的 30%,那么就选定为其中一个像素点,规定 p 点正上方在圆上的像素为第一个像素,顺时针选取,若圆上类似像素点 2 和 3 之间的像素点也满足要求,那么就比较这 3 个像素点哪个在圆内的灰度值更小,舍弃最小的,留下两个较大的即可满足一共选取 16 个圆上像素点进行计算的要求,如图 2 所示。

现在有了特征点的坐标和灰度值,设特征点的坐标为 (a, b) ,以每一个特征点为中心设定一个 9×9 的图像块,在提取完特征点得到相应的坐标之后,还需要计算该像素点的描述子向量,设描述子向量为 128 维。将这个以特征点为中心点的 9×9 的图像块平均分成 9×9 个正方形,取每一个小正方形的中心点坐标(除了特征点所在方块的坐标不选),并将这些坐标存入数组,然后从数组中随机选两对点 (x, y) 和 (p, q) ,设点对的像素值为 $H_{p(i,j)}$,描述子向量设为 128 维向量,每一维向量代表着像素差,这 128 维由 0 和 1 组成,见式(4):

$$\begin{cases} H_{p(a-x, b+y)} > H_{p(a-p, b+q)}, 1 \\ H_{p(a-x, b+y)} < H_{p(a-p, b+q)}, 0 \end{cases} \quad (4)$$

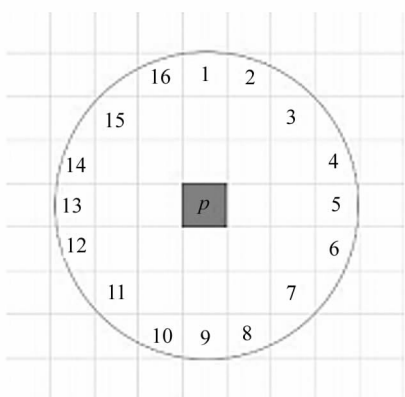


图2 提取特征点

Fig. 2 Extraction of feature points

2.2 匹配描述子

传统的特征点匹配算法中,粗匹配算法有:暴力匹配、FLANN 快速最近邻匹配。去粗取精算法有:Lower's 算法为了进一步筛选匹配点,来获取优秀的匹配点,这就是所谓的“去粗取精”,一般会采用 Lower's 算法来进一步获取优秀匹配点;RANSAC 算法设定阈值,抽样误差大于阈值的点就可认为不

是最佳匹配点,从而进一步获取优秀匹配点;LMEDS 算法不需要阈值,但是内点数小于 50% 会被判定不是最佳匹配点,从而进一步获取优秀匹配点;基于深度学习的特征匹配算法像 MatchNet,由一个深度卷积网络组成,网络拆解为并行的特征网络和测度网络,分两个阶段使用特征网络和测度网络。首先为所有补丁生成特征编码,然后将这些特征配对并推送它们通过测度网络获得分数,通过分数来获取特征的匹配。本文基于前面传统方法获取的特征点,将基于注意力机制特征匹配网络的深度学习算法用于特征点匹配,而这里的网络是图神经网络,图神经网络在特征点匹配的优势更为突出。

前面提取出了每一幅图像中的特征点,如果要两幅图像的特征点进行匹配,要先求出每个特征点的匹配描述子,首先将每个特征点的位置坐标和描述子向量进行耦合,目的是在后面在使用注意力机制的时候能够更多的考虑到每个特征点的外观特征,这里特征点的位置坐标 p_i 为 2 维向量,描述子向量 d_i 为 128 维,使用多层感知机全连接神经网络将两者进行相加成一个 128 维的向量 x_i ,设定中间 Hidden layer 为 3 层,分别为 32 维、64 维、128 维。选取 Leaky Relu 函数为激活函数:

$$f(x_i) = \max\{ax_i, x_i\}. \quad (5)$$

将两张用来匹配的图像提取的特征点放在同一张图中作为图的节点,节点与节点之间连接的边有两种:一种是图像自身节点之间连接的边,为 self-edge(self-attention);另一种是与另一帧图像连接的边,为 cross-edge(cross-attention),为了将两帧图像的特征点进行匹配,首先要利用基于注意力机制的图神经网络来将特征点进行更新,图注意力神经网络和图注意力层如图 3 所示。

右图为图注意力层,左图为图注意力网络结构。这里认为特征点之间全都是相邻的,并且设定图注意力层为 9 层,当层数 l 为奇数,更新由 self-edge 连接的节点,当层数 l 为偶数时,更新由 cross-edge 连接的节点,其中 $m_{e \rightarrow i}$ 为特征点 i 与相邻节点的聚合, $\parallel$ 为两个矩阵的拼接操作(concatenate),严格按照式 6 进行节点更新:

$$x_i^{l+1} = x_i^l + MLP(x_i^l \parallel m_{e \rightarrow i}), i = 1 \cdots 9. \quad (6)$$

利用注意力机制节点更新示意图如图 4 所示,设待更新的节点为 x_i , x_i 的相邻节点为 x_j ,每个节点为 128 维向量,每一维是特征点的位置向量和描述子向量耦合后的结果。

其中输入待更新节点和相邻节点,都是 128 维

向量,输出的更新后的节点也为 128 维向量, x_i 和 x_j 是相邻节点,其中通过线性映射得到 query、key、value, query 是 x_i 的线性映射得到的, key 和 value 是 x_j

的线性映射得到的, W_1 、 W_2 、 W_3 、 b 每一层是固定的,是所有节点的共享参数,但不同层 W 就不一样:

$$q_i = W_1 x_i + b_1, \quad (7)$$

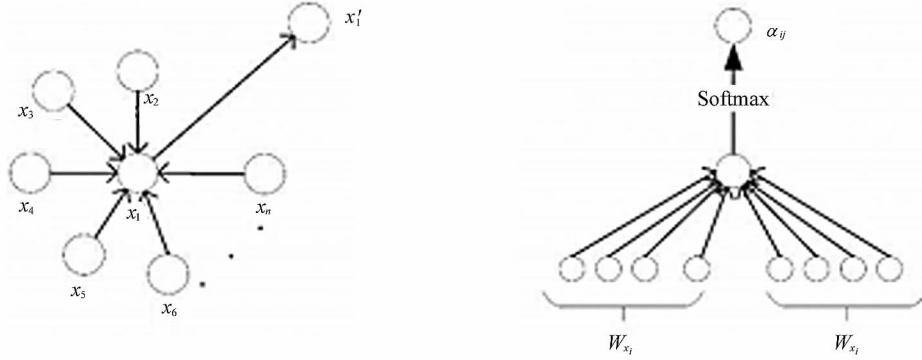


图 3 图注意力神经网络和图注意力层

Fig. 3 Graph attention networks and graph attention layer

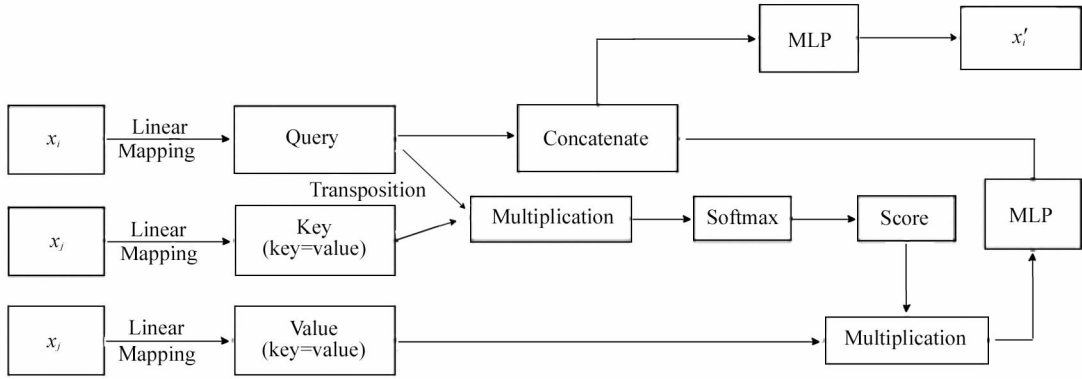


图 4 节点更新示意图

Fig. 4 Update node schematic

$$\begin{bmatrix} k_j \\ v_j \end{bmatrix} = \begin{bmatrix} W_2 \\ W_3 \end{bmatrix} x_j + \begin{bmatrix} b_2 \\ b_3 \end{bmatrix}. \quad (8)$$

为了更好地分配权重,需要将当前中心节点与其所有邻居计算出的相关度进行统一的归一化处理,具体形式为 Softmax 归一化,有了 query、key 和 value,就可以利用 Softmax 函数求注意力权重,也就是 x_i 和 x_j 的相似度:

$$\alpha_{ij} = \text{Softmax}(q_i^T k_j), \quad (9)$$

求得了注意力权重,便可以求得 $m_{e \rightarrow i}$:

$$m_{e \rightarrow i} = \sum \alpha_{ij} v_j, \quad (10)$$

更新后得到的节点即为匹配描述子:

$$f_i = W_{x_i} + b = W[x_i + \text{MLP}(x_i \parallel \sum \text{Softmax}(q_i^T k_j) v_j)] + b. \quad (11)$$

2.3 特征点的最优匹配

通过注意力机制和图神经网络求得每个特征点的匹配描述子 f_i ,设两帧图像为 P 和 Q ,每帧图像的

特征点的匹配描述子表示为 f_i^P 和 f_i^Q ,将 P 中的 N 个特征点与 Q 中的 M 个特征点做匹配,即将两个匹配描述子做内积,即可得到得分矩阵 $score_{ij}$:

$$score_{ij} = \langle f_i^P, f_j^Q \rangle, \forall (i, j) \in \{P, Q\}. \quad (12)$$

构建一个由 0 和 1 组成的分配矩阵 D ,利用最优传输算法求解 $\sum_{i,j} score_{ij} D_{ij}$,即可得到分配矩阵 D 。可能 P 中有特征点无法与 Q 中的特征点进行匹配,将 $score_{ij}$ 扩充一行和一列,这样就保证了 P, Q 的特征点的线性映射, $score$ 矩阵如图 5 所示。

实验环境中利用基于注意力机制特征匹配网络的特征点分布和匹配示意图如图 6 所示。

至此,视觉里程计模块中的特征提取与匹配的工作就完成了,这样一种全新的组合的目的就是减少因特征匹配产生的误差和误匹配,从而提高匹配准确度,进而提高相机的运动轨迹精度以及相机位姿估计精度。

	1	...	N	N+1
1				
...				
M				
M+1				

图 5 Score 矩阵

Fig. 5 Score matrix



图 6 特征点分布和匹配示意图(室内)

Fig. 6 Distribution and matching diagram of feature points (indoor)

2.4 构建点云地图

匹配完成之后,就要求解相机的位姿,实验使用单目相机,只知道图像序列中特征点的 2D 坐标,那么就利用两组 2D 坐标来计算相机位姿,也就是对极几何的方法, p_1 和 p_2 为像素坐标,采用八点法便可求出基础矩阵 F ,见式(12):

$$p_2^T F p_1 = 0. \quad (13)$$

采用 SVD 求解出相机的位姿,其位姿可能有四个解,需要将任意一个点去代入验证,验证这个点的深度 s 是否为正,如果是,则为正确解,否则不正确,其中 K 为相机内参,见式(13):

$$sp = K(PR + t). \quad (14)$$

构建 3D 点云地图最关键的一步就是关键帧的筛选,通过筛选可以减少 3D 地图点云拼接重复等错误和减少不必要的计算量,筛选关键帧的方式为:首先设定当前帧图像提取出来的特征点数量为 S_i ,上一帧关键帧提取出来的特征点数量为 S_{i-1} ,和上一帧关键帧进行特征匹配,匹配成功的点对数量为 Q_i ,设定一个阈值 100,当前帧的特征点数量必须大于

100;设定另一个阈值 0.8,代表着相机的运动系数,若小于 0.8,则证明运动尺度较大,可设为关键帧:

$$\begin{cases} S_i > 100 \\ \frac{Q_i}{S_{i-1}} < 0.8 \end{cases} \quad (15)$$

地图的构建是通过点云的变换得来的,点云的变换方式也就是相机的运动位姿 R, t ,设变换前点云的空间点为 P_i ,变换之后为 P_{i+1} ,设 T 为变换矩阵,并通过变换矩阵得到变换之后的点云信息:

$$T = \begin{bmatrix} R & t \\ 0 & 1 \end{bmatrix} \in R^{4 \times 4}, \quad (16)$$

$$P_{i+1} = P_i \times T. \quad (17)$$

因为前面已经得到了最优的特征点匹配,所以在检测是否产生回环的时候,将当前关键帧和历史每一个关键帧做特征点匹配,选取匹配点对成功数量最多的关键帧,这里一般会设定一个阈值,为当前特征点数量的 85%,如果匹配成功的特征点数量超过此阈值,则说明产生了回环,构建的部分点云地图如图 7 所示。

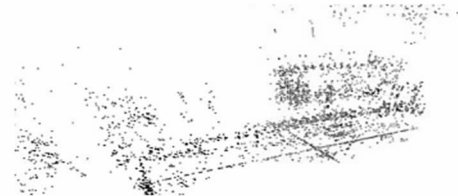


图 7 部分点云地图(TUM 数据集)

Fig. 7 Partial point cloud map(TUM)

3 实验结果及分析

本文从 SLAM 方法的精度方面进行评估。本文实验为了体现在室外环境中在视角变换情况下 SLAM 方案的精度,实验使用了 KITTI 官方数据集,在实验过程中不可避免会使用到两个精度指标:绝对轨迹误差(ATE)和相对位姿误差(RPE)。本文选取了两种传统的特征点匹配算法在 SLAM 中的应用和一种最新的基于深度学习的特征点匹配算法在 SLAM 中的应用:暴力匹配算法、FLANN 算法及 MatchNet 算法;先比较一下和本文 SLAM 方法特征点匹配及运动轨迹的效果图,分别是基于暴力匹配的配准方法(a)、基于 FLANN 匹配的配准方法(b)、基于 MatchNet^[15]的配准方法(c),以及本文的基于注意力特征点匹配网络的配准方法(d)的特征匹配效果图及四种配准方法的相机运动轨迹,如图 8 和图 9 所示。

绝对轨迹误差是估计位姿和真实位姿的直接差值,可以非常直观地反应算法精度和轨迹全局一致。相对位姿误差主要描述的是相隔固定时间差 Δ 两帧位姿差的精度(相比真实位姿)。首先定义一下公式标注:SLAM 算法估计的相机位姿 $P_1 \cdots P_N \in SE(3)$,真实的相机位姿 $Q_1 \cdots Q_N \in SE(3)$,这里的字母下标代表时间 t (或帧),假设算法估计的相机位姿和

真实的相机位姿的各图像帧时间已经对齐准确,并且总帧数相同,其中, Δ 代表相隔的时间。单目相机具有尺度不确定性,所以需要计算一个从 SLAM 算法估计出来的相机位姿到真实的相机位姿的相似转换矩阵 $S \in Sim(3)$,那么第 i 帧的绝对轨迹误差定义如下:

$$F_i = Q_i^{-1} S P_i. \quad (18)$$



(a) Violent matching for SLAM



(b) FLANN for SLAM



(c) MatchNet for SLAM



(d) Method of this paper

图 8 4 种方法特征匹配效果图

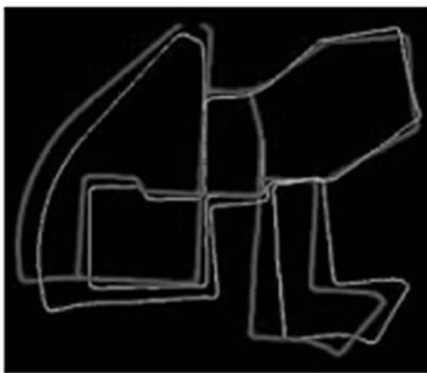
Fig. 8 Four methods of feature matching renderings



(a) Violent matching for SLAM



(b) FLANN for SLAM



(c) MatchNet for SLAM



(d) Method of this paper

图 9 基于 4 种配准方法的相机运动轨迹

Fig. 9 Four methods of camera motion track

在已知总帧数 n 和 Δ 的情况下,可以计算出 $m = n - \Delta$ 个绝对轨迹误差,然后可以用均方根误差 $RMSE$ 进行误差统计:

$$RMSE(F_{1:n}, \Delta) = \left(\frac{1}{m} \sum_{i=1}^m \| \text{trans}(F_i) \|^2 \right)^{\frac{1}{2}}, \quad (19)$$

同样,第 i 帧的相对位姿误差定义如下:

$$E_i = (Q_i^{-1} Q_{i+\Delta})^{-1} (P_i^{-1} P_{i+\Delta}). \quad (20)$$

同理,利用均方根误差 $RMSE$ 进行误差统计,需要注意的是,RPE 包含了两部分的误差:旋转误差和平移误差,为了能综合衡量算法表现,可以计算遍历所有 Δ 的 $RMSE$ 的平均值:

$$RMSE(E_{1:n}) = \frac{1}{n} \sum_{\Delta=1}^n RMSE(E_{1:n}, \Delta). \quad (21)$$

可以看到实验测试序列 00—05 这 6 个序列在 4 个配准方法的实验结果如表 1、表 2 和表 3 所示。

表 1 绝对轨迹误差对比(单位:m)

Tab. 1 Absolute trajectory error comparison(Unit:m)

seq.	Violent	FLANN	MatchNet	Ours
00	342.836 3	5.570 7	4.465 3	4.367 1
01	243.738 2	18.975 4	14.567 3	11.762 1
02	189.764 4	6.903 4	2.873 5	2.665 4
03	279.334 6	43.764 5	38.554 3	34.667 8
04	18.653 5	1.782 1	1.446 5	1.114 5
05	220.778 3	43.214 5	36.902 1	34.224 6

表 2 相对位姿误差对比(平移误差)

Tab. 2 Relative pose error comparison (motion error)

seq.	Violent	FLANN	MatchNet	Ours
00	0.2765	0.1145	0.1109	0.1097
01	0.5028	0.2796	0.1089	0.1012
02	0.9544	0.2019	0.1590	0.1344
03	0.1198	0.1009	0.0905	0.0897
04	0.8465	0.1795	0.0371	0.0310
05	1.2907	0.5069	0.9071	0.8902

表 3 相对位姿误差对比(旋转误差)

Tab. 3 Relative pose error comparison (rotation error)

seq.	Violent	FLANN	MatchNet	Ours
00	0.303 9	0.172 8	0.158 9	0.139 8
01	0.564 2	0.300 6	0.119 3	0.110 5
02	0.208 7	0.187 8	0.066 5	0.045 9
03	0.639 8	0.046 6	0.043 1	0.039 8
04	0.962 1	0.210 7	0.061 3	0.592 1
05	1.409 8	0.608 2	0.125 4	0.120 1

4 结 论

本文提出了基于注意力机制的特征匹配网络的 SLAM 方法,在特征点配准部分采用了基于注意力机制的图神经网络得到每个特征点的匹配描述子,而后将待匹配的点对的匹配描述子做内积,从而得到一个得分矩阵,利用最优传输算法计算出最优分配矩阵,从而得到最优的特征点匹配,而后通过匹配好的特征点对估计单目相机的运动位姿,根据运动位姿进行局部点云地图的建立以及回环检测。实验结果表明,相比于传统的配准方法与最新的基于深度学习的特征匹配方法,采用基于注意力机制特征点匹配网络的配准方法,不仅对视角不稳定的环境更鲁棒,而且在相机运动轨迹误差和相机位姿估计误差的精度明显有所提升。但是不足之处在于,求解分配矩阵时使用的最优传输算法的时间复杂度较高,直接影响了视觉 SLAM 的运行速度,在这一点不如 ORB-SLAM 等传统主流方法。那么未来的改进之处就主要着眼于运行速度的优化。

参考文献:

- [1] DURRANT-WHYTE H, BAILEY T, BAILEY T, et al. Simultaneous localization and mapping (SLAM): Part II[J]. IEEE Robotics & Automation Magazine, 2006, 13(3): 108-117.
- [2] DAVISON A J, REID I D, MOLTON N D, et al. Mono-SLAM: real-time single camera SLAM[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2007, 29(6): 1052-1067.
- [3] KLEIN G, MURRAY D. Parallel tracking and mapping for small AR workspaces[C]//6th IEEE and ACM International Symposium on Mixed and Augmented Reality, November 13-16, 2007, Nara, Japan. New York: IEEE, 2007: 225-234.
- [4] ENGEL J, SCHIPS T, CREMERS D. LSD-SLAM: Large-scale direct monocular SLAM[C]//13th European Conference on Computer Vision, September 6-12, 2014, Zurich, Switzerland. Berlin: Springer, 2014: 834-849.
- [5] RUBLEE E, RABAU D V, KONOLIGE K, et al. ORB: An efficient alternative to SIFT or SURF[C]//International Conference on Computer Vision, November 6-13, 2011, Barcelona, Spain. New York: IEEE, 2011: 564-2571.
- [6] RAUL M A. ORB-SLAM: a versatile and accurate monocular SLAM system[J]. IEEE Transactions on Robotics, 2017, 31(5): 1147-1163.

- [7] JIN J Y. Research on visual slam method based on feature points of convolutional neural network[D]. Wuhan: Hubei University of Technology, 2020.
金靖熠. 基于卷积神经网络特征点的视觉 SLAM 方法研究[D]. 武汉: 湖北工业大学, 2020.
- [8] KEISUKE T, FEDERICO T, IRO L, et al. CNN-SLAM: Real-time dense monocular SLAM with learned depth prediction[J]. 2017; 6565-6574.
- [9] KENDALL A, GRIMES M, CIPOLLA R. PoseNet: A convolutional network for real-time 6-dof camera relocalization[C]//IEEE International Conference on Computer Vision, February 7-13, 2015, Santiago, Chile. New York: IEEE, 2015; 2938-2946.
- [10] UMMENHOFER B, ZHOU H, UHRIG J, et al. DeMoN: Depth and Motion Network for Learning Monocular Stereo [C]//IEEE Conference on Computer Vision and Pattern Recognition, July 21-26, 2017, Honolulu, HI, USA. New York: IEEE, 2017; 5622-5631.
- [11] GAN Y, ZHANG J H, CHEN K Q, et al. A dynamic detection method to improve SLAM performance[J]. Optoelectronics Letters, 2021, 17(11): 693-698.
- [12] LUO Y, XIAO Y T, ZHANG Y, et al. Detection of loop closure in visual SLAM: a stacked assorted auto-encoder based approach [J]. Optoelectronics Letters, 2021, 17(6): 354-360.
- [13] LI H L, ZHANG Z B, XU X Y. Low-light image enhancement algorithm for visual SLAM based on online estimation [J]. Journal of Optoelectronics • Laser, 2021, 32(9): 945-952.
- 李红莉, 张智斌, 徐玄冀. 基于在线估计的视觉 SLAM 低光照图像增强算法 [J]. 光电子·激光, 2021, 32(9): 945-952.
- [14] LI S P, ZHANG T. A review of deep learning applications in visual SLAM[J]. Space control technology and application. 2019, 45(2): 1-10.
李少朋, 张涛. 深度学习在视觉 SLAM 中应用综述[J]. 空间控制技术与应用, 2019, 45(2): 1-10.
- [15] LI R N, XU G, GUO R, et al. Research on RGB-D slam loop detection based on convolutional neural network[J]. Journal of Heilongjiang Institute of Technology: General Edition, 2019, 19(11): 42-48.
李若楠, 许钢, 郭芮, 等. 基于卷积神经网络的 RGB-D SLAM 回环检测研究[J]. 黑龙江工业学院学报: 综合版, 2019, 19(11): 42-48.
- [16] SARLIN P E, DETONE D, MALISIEWICZ T, et al. SuperGlue: learning feature matching with graph neural networks[C]//IEEE/CVF Conference on Computer Vision and Pattern Recognition, June 14-19, 2020, Seattle, WA, USA. New York: IEEE, 2020: 4937-4946.
- [17] HAN X, LEUNG T, JIA Y, et al. MatchNet: Unifying feature and metric learning for patch-based matching[C]//IEEE Conference on Computer Vision and Pattern Recognition, June 7-12, 2015, Boston, MA, USA. New York: IEEE, 2015; 3279-3286.

作者简介:

刘凤连 (1965—), 男, 教授, 目前任教于天津理工大学计算机科学与工程学院, 主要从事计算机图像处理与视觉计算方面的研究。