文章编号:1671-6833(2021)01-0056-07

基于多任务学习的初始图像对选取方法

刘宇翔, 张茂军, 颜 深, 李京蓓, 彭 杨

(国防科技大学 系统工程学院,湖南 长沙 410073)

摘 要: 初始图像对选取是增量式从运动中恢复结构的一个关键环节,但传统方法中存在计算效率低、对特殊场景不鲁棒的问题。因此,提出基于多任务学习的初始图像对选取网络以提高该过程的效率,并针对某些特殊场景容易出现初始图像对位于场景边缘的问题,提出结合场景连接图的初始对选取策略。该策略首先构建图像间的拓扑结构,通过图像间连接的疏密程度判断初始图像对是否处于场景中心,从而避免初始图像对处于场景边缘导致重建不完整的问题。对比传统 SfM(structure from motion)中的初始图像对选取方法,结果表明:所提出的方法在多种不同场景中的选取速度提升 5 倍以上;同时,提出的结合场景图的选取策略可使得特殊场景中重建的空间点数量增加 10 倍,且重投影误差下降 0.05 px,显著提升了在特殊场景中初始图像对选取的鲁棒性,证明了所提方法的有效性,在提高了效率的同时,能够很好地保证特殊场景重建的完整性和稳定性。

关键词: 增量式 SfM; 初始图像对选取; 多任务学习; 场景连接图

中图分类号:TP391.4 **文献标志码**: A **doi**:10.13705/j.issn.1671-6833.2021.01.009

0 引言

近年来随着无人机与高清相机的广泛应用,基于图像的大规模三维场景重建技术得到了广泛关注,其主要运用多视图几何原理,通过不同视点所拍摄的图像计算出相机姿态和场景三维结构。从运动中恢复结构(structure from motion,SfM)是基于图像的三维重建中一个关键环节,主要完成相机位姿的估计和稀疏点云的重建,其中增量式(incremental)SfM 是目前最普遍、最稳定的 SfM 方法。增量式 SfM 首先需要选取一对图像作为起点进行两视图重建,最开始选取的这一对图像被称为初始图像对(initial image pair,InitIP),它对三维重建的最终效果影响巨大,整个初始图像对选取过程如图 1 所示。

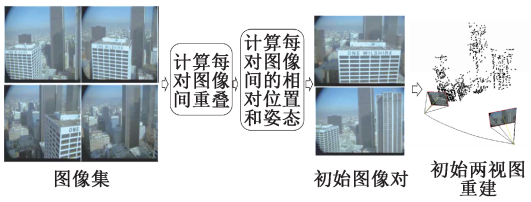


图 1 初始图像对选取流程图

Figure 1 Flow chart of initial image pair selection

2006 年,Beder 等[1]通过计算三维点所处空间区域的圆度,来衡量 InitIP 对于场景重建的稳定性,首次给出了增量式 SfM 中初始图像对的评价方法。之后,Haner 等[2]通过最小化每个相机到 InitIP 的距离来减小累积误差,该方法首次将 InitIP 与整个图像集之间的关系考虑到该选取过程之中。近年来,一些三维重建开源系统例如 Colmap、Alicevision 则使用 Schonberger 等[3]和 Moulon 等[4]提出的利用多尺度网格划分图像来计算特征点得分的方法,将匹配点的分布进行量化,一定程度上提高了 InitIP 的鲁棒性。但缺点在于面对大规模数据时,计算每幅图像多个尺度的特征分数也会额外增加计算开销。综上,现有的 InitIP 选取方法主要遵循两大原则:第一,足够多的匹配点;第二,两幅图像要具有足够的相对运动以保证不退化为单应模型。以上方法均需要进行大量的特征匹配导致效率较低,另一方面,特征点检测与匹配中的误差也会导致相机相对位置的估计不准确。

InitIP 的选取需要建立大量图像间的连接关

收稿日期:2020-08-16;**修订日期**:2020-11-17

基金项目:国家自然科学基金资助项目(62071478)

通信作者:张茂军(1971—),男,湖北黄梅人,国防科技大学教授,博士,博士生导师,主要从事虚拟现实与计算摄影方面的研究,E-mail:mjzhang@nudt.edu.cn。

系,传统的做法是特征点提取与匹配[5]。同时,选取 InitIP 还涉及两幅图像相对空间位置的计算,因此它是一个包含多任务、多输出的问题。近年来,多任务学习[6]作为深度学习中的一个分支,能够高效地在多个相关联任务中进行学习训练、共享特征,从而得到广泛应用。因此,本文借鉴多任务学习的思想,提出基于多任务学习的初始图像对选取网络,以提高选取 InitIP 的效率。为了避免 InitIP 位于场景的稀疏区域而导致重建场景不完整的问题,进一步提出了一种结合场景连接图的选取策略,以提高重建稳定性与完整性。

1　初始图像对选取网络

多任务学习针对不同但具有相关性的任务,同时对两个或两个以上任务进行学习,在一定程度上共享学习到的知识,以提升各自的性能。因此本文提出使用多任务学习网络同时预测图像相似性和相机的位置姿态,进而加速在大规模场景下选取 InitIP 的整个过程。

1.1　多任务网络框架

首先选定两个特定的网络对应两个子任务,整个多任务网络框架如图 2 所示。上方蓝色分支为 PoseNet,用来预测图像间的相对位移与旋转,下方红色分支为 MatchNet,输出每一对图像间的相似度,然后联合相似度、相对位移与旋转进行图像对的整体评分,从而选出得分最高的 InitIP。

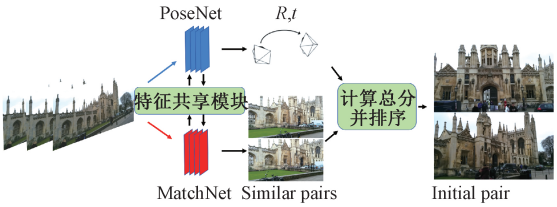


图 2　多任务初始对选取网络图

Figure 2　Multi-task initial pair selection network

其中,MatchNet 采用 Shen 等[7]提出的图像检索网络作为主体结构,如图 3 所示,训练时网络的一组输入为 3 幅图像:参考图像、正例图像、负例图像。正例图像即与参考图像相似的图像,负例图像也就是与参考图像无关的图像。

特征编码网络使用被广泛采用的卷积网络作为基础网络。该网络将卷积神经网络作为特征编码器,将输入图像编码成一个高维空间中的特征向量,使得包含相似场景或物体的图片经过编码后形成的向量在高维空间中尽可能接近,不包含相似场景的图像尽可能远离。因此,本文用归一化后的特征向量之间的 $L2$ 距离来度量相似度,如

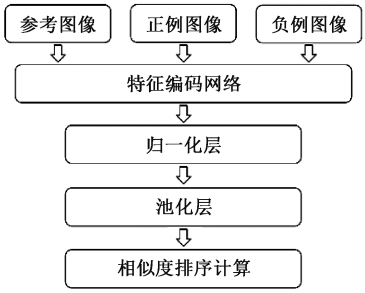


图 3　图像相似性检测网络结构图

Figure 3　Structure of image similarity detection network

式(1)所示:

$$d(f(\boldsymbol{I}_i),f(\boldsymbol{I}_j)) = \left\| \frac{f(\boldsymbol{I}_i)}{\|f(\boldsymbol{I}_i)\|} - \frac{f(\boldsymbol{I}_j)}{\|f(\boldsymbol{I}_j)\|} \right\|_2 。 \quad (1)$$

式中:$f(\cdot)$ 为深度神经网络;$\boldsymbol{I}_i$、$\boldsymbol{I}_j$ 为需要进行相似度量的两幅图像。

另一个分支则为相机位置姿态估计网络 PoseNet,使用 Kendall 等[8]提出的 PoseNet 对输入的图像进行六自由度的位置和姿态估计。该网络以 GoogLeNet[9]作为基础,将原有的 3 个 softmax 分类器修改为输出两个向量的仿射变换回归器,如图 4 红色标注框所示,$\boldsymbol{t}=(x,y,z)$ 表示位置的三维向量,$\boldsymbol{R}=(w,a,b,c)$ 表示相机旋转的四元数。图 4 整体为一个修改后的卷积模块。

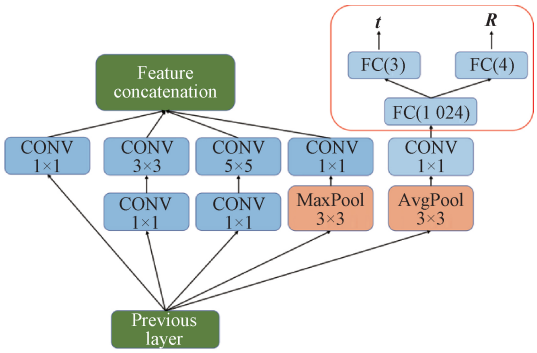


图 4　PoseNet 的相机位置姿态回归模块

Figure 4　Camera position attitude regression module of PoseNet

卷积层输出的特征图经过平均池化层改变尺寸,接着经过 1×1 的卷积层改变通道数,然后进入仿射变换回归器。在回归器中,特征向量先通过 1 024 维的全连接层,再分别经过维度为 3 和 4 的全连接层回归代表位置和旋转的两个向量,因此可以通过这个分支预测得到图像的位置和姿态,进而计算相对位移 T_{rel}、相对旋转 R_{rel}。通过文献[10]中所提方法,最终由四元数 R 转换为相对欧拉角度 R_{rel}。

至此,通过网络的两个分支分别得到了图像之间的特征向量距离和相对位移与旋转,然后,通

过设计的评分公式(2)进行评分和排序,从而选取最终的InitIP。

$$Score = T_{\mathrm{rel}}/d^2(R_{\mathrm{rel}} < 45°)。 \quad (2)$$

式中:$Score$ 表示InitIP的最终评分;d 表示式(1)中所计算的特征向量之间的距离,根据航拍三维重建中采集图像的重叠度至少为60%,相机相对夹角小于45°的原则[11],将3个轴上相对旋转的阈值设定为45°。

1.2 交叉连接网络

在确定了多任务网络框架后,就需要确定多任务网络中特征共享的方式。本文采用Fukuda等[12]所提出的交叉连接单元将PoseNet与MatchNet进行连接,实现不同任务之间特征的共享,交叉单元如图5所示。

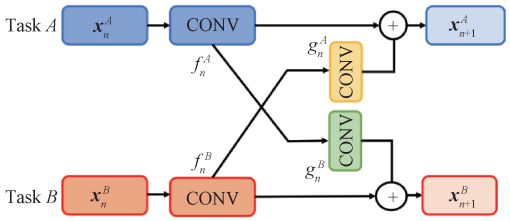


图5 交叉连接结构示意图[12]

Figure 5 Schematic diagram of cross-connection structure[12]

网络第 n 层的输入表示为 $\boldsymbol{x}_n^A$ 和 $\boldsymbol{x}_n^B$,将原任务中的卷积层表示为 f_n^A 和 f_n^B。假设交叉连接中的卷积层学习到变换 g_n^A 和 g_n^B,则两个任务经过交叉连接的输出 $\boldsymbol{x}_{n+1}^A$ 和 $\boldsymbol{x}_{n+1}^B$ 由式(3)计算得出[12]:

$$\begin{cases} \boldsymbol{x}_{n+1}^A = f_n^A(\boldsymbol{x}_n^A) + g_n^A(f_n^B(\boldsymbol{x}_n^B)); \\ \boldsymbol{x}_{n+1}^B = f_n^B(\boldsymbol{x}_n^B) + g_n^B(f_n^A(\boldsymbol{x}_n^A))。 \end{cases} \quad (3)$$

上面两个等式右边的第二项 $g_n^A(f_n^B(\boldsymbol{x}_n^B))$ 和 $g_n^B(f_n^A(\boldsymbol{x}_n^A))$ 表示另一个任务中有用的信息。如果将PoseNet与MatchNet所有卷积层都进行交叉连接会增加大量的参数量和计算时间,并且两个网络在基本结构上存在许多差异,所以并不适合把两个网络进行全部交叉连接。因此,本文采用了浅层交叉连接和深层交叉连接的两种思路,构建两种网络分别进行实验,以探索性能更好的交叉连接方式。如图6所示,矩形框标注区域为建立交叉连接的层级,其余层级为两个任务各自特征提取层,不参与特征的共享。

图6(a)中MatchNet的主干网络Resnet50的第2、3个残差块中的卷积层与PoseNet中Inception3中的卷积层进行交叉连接,输出的特征图尺寸分别为56×56、28×28。图6(b)中MatchNet的第4、5个残差块与PoseNet中的Inception4、Inception5层中的卷积层之间构建交叉连接模块,输出特征图的尺寸分别为14×14和7×7。

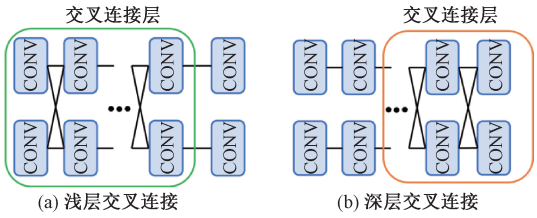


图6 两种交叉连接示意图

Figure 6 Diagram of two kinds of cross connection

1.3 联合损失函数

对整个网络框架的损失函数进行设计时,首先需要确定两个子网络各自的损失函数。对于相似性检测网络采用如下三元损失函数[7]:

$$L(\boldsymbol{a},\boldsymbol{p},\boldsymbol{n}) = [d^+ - \min(d^-, d^{-\prime}) + \alpha]_+。 \quad (4)$$

式中:$\boldsymbol{a}$,$\boldsymbol{p}$,$\boldsymbol{n}$ 分别代表参考图像、正例图像、负例图像;$d^+ = d(f(\boldsymbol{a}), f(\boldsymbol{p}))$,$d^- = d(f(\boldsymbol{a}), f(\boldsymbol{n}))$,$d^{-\prime} = d(f(\boldsymbol{n}), f(\boldsymbol{p}))$,$d$ 为式(1)所述的两幅图像之间的距离;α 为一个设定的边界参数使得两个距离的计算保持一定的界限。式(4)中 $[\sigma]_+ = \max(\sigma, 0)$,即当正例距离远大于负例距离时,该组数据的损失值是需要被抑制的。

对于另一个分支的PoseNet,Kendall[8]提出的整体损失函数如式(5)所示:

$$L_{\mathrm{pose}} = \| \hat{\boldsymbol{x}} - \boldsymbol{x} \|_2 + \beta \left\| \hat{\boldsymbol{q}} - \frac{\boldsymbol{q}}{\| \boldsymbol{q} \|} \right\|_2。 \quad (5)$$

式中:$\boldsymbol{x}$ 与 $\boldsymbol{q}$ 为通过传统SfM得到的位置姿态的参考值;$\hat{\boldsymbol{x}}$ 与 $\hat{\boldsymbol{q}}$ 为位置和姿态的预测值;β 为用来保持位置和方向误差的期望值近似相等的尺度因子。

训练多任务网络时,可能出现梯度主导[13]问题,导致无法收敛。主要原因是由于各任务输入数据的不平衡,以及反向传播中梯度数值相差过大。因此,本文采用一种为多个损失函数动态赋予权重的方法[14]设计联合损失函数,如式(6)所示:

$$L_{\mathrm{final}} = W_1 \cdot L_{\mathrm{similar}} + W_2 \cdot L_{\mathrm{pose}}。 \quad (6)$$

式中:L_{similar} 为式(4)所表示的相似性检测网络中的三元损失函数;L_{pose} 为式(5)表示的位置姿态回归损失函数;权重 W_1、W_2 分别为 $\frac{1}{2\sigma_1^2}$、$\frac{1}{2\sigma_2^2}$,通过噪声参数 σ_1、σ_2 动态调整两个损失函数之间的平衡[13]。

2　结合场景连接图的选取策略

传统的 InitIP 选取框架都是通过图像的外观特征和几何关系来进行筛选。然而,在某些实际场景中,这些方法选择的 InitIP 会出现与整个场景关联度较低,或者处于场景边缘的情况,导致场景重建不完整的问题。因此,利用多任务网络的中间输出建立一个场景连接图,在该连接图中选取处于场景稠密区域的 InitIP。

2.1　场景图构建方法

首先,给出所使用符号的定义,$I=\{\boldsymbol{I}_i\}$ 代表图像集合,多任务网络的中间输出为图像特征向量间的距离集合 D、相对位移集合 T、相对旋转集合 R。定义图像相似度集合为 S,任意两幅图像 $\boldsymbol{I}_i$ 和 $\boldsymbol{I}_j$ 相似度为 $s_{ij}\in S$,s_{ij} 由式(7)计算得到:

$$s_{ij}=\frac{1}{d_{ij}^2}。\tag{7}$$

式中:$d_{ij}\in D$,为式(1)所求两特征向量间的距离。

类似地,定义两幅图像的相对位移 $t_{ij}\in T$ 和相对旋转 $R_{ij}\in R$,其中 t_{ij}、R_{ij} 分别对应 1.1 节中的 T_{rel}、R_{rel}。然后,定义场景连接图为节点和边缘的集合 $G=(N,E)$,其中 N 代表连接图中节点集合,任意 $n_i\in N$ 对应 I 中一幅图像 $\boldsymbol{I}_i$,E 则代表边缘的集合,初始时为空。当两图像之间的相似度、相对位移、相对旋转均处于所设定的取值范围时,则为两个节点 n_i 与 n_j 连接一条边缘 $e_{ij}\in E$,边缘的权重被设置为相似度与相对位移的乘积,如式(8)所示:

$$w_{ij}=s_{ij}\cdot t_{ij}。\tag{8}$$

该权重可以综合度量两图像的特征相似程度和几何关系,因此边缘的数据结构可表示为 $e_{ij}=(n_i,n_j,w_{ij})$。建立场景连接图的完整算法具体步骤如下。

Step 0　初始化连接图。设定图像集合 I,图像间的相似度集合 S,图像间的相对姿态集合 R、T,为 I 中的每一幅图像 $\boldsymbol{I}_i$ 生成一个节点 n_i,组成节点集合 N,生成一个空的边缘集合 E。设置相似度阈值 s_0、相对位移阈值 t_0,相对旋转阈值 R_0;

Step 1　访问 N 中任一未被访问的节点 n_i,在 S、T、R 中查询 n_i 所对应图像 $\boldsymbol{I}_i$ 与其余图像 $\boldsymbol{I}_j$ 的相似度 s_{ij}、相对位移 t_{ij} 与相对旋转 R_{ij};

Step 2　当 $s_{ij}\geqslant s_0$,且 $t_{ij}\leqslant t_0$、$R_{ij}\leqslant R_0$ 时,为节点 n_i 与图像 $\boldsymbol{I}_j$ 所对应的节点 n_j 之间连接一条边缘 e_{ij},边缘权重为 w_{ij},然后将边缘 $e_{ij}=(n_i,n_j,w_{ij})$ 存储到集合 E 中;

Step 3　当 N 中所有节点都被访问,输出最终场景连接图 $G=(N,E)$,否则返回 Step 1。

2.2　基于场景图的选取方法

通过 2.1 节方法建立了场景连接图后,本节提出基于场景连接图的初始图像对评分方法,首先计算每个节的度 deg,然后通过两幅图像度之和 (deg_i+deg_j) 来衡量候选 InitIP 与剩余图像的关联程度,判断该图像对是否位于场景稠密区域。但是,在将度的数量纳入参考指标时,可能出现 InitIP 中一幅图像的度较大,而另一幅图像的度很小的情况,这样仍有可能导致重建结果精度不高,甚至无法重建出完整场景的问题。因此,本文将两幅图像度之间差的绝对值的指数函数 $e^{|deg_i-deg_j|}$ 定义为度平衡因子 b_{ij},以衡量两幅图像度的平衡程度,两幅图像度的差值越小,则平衡因子越小,最后计算的总评分也越高。

计算 n_i,n_j 两个节点所代表的图像的最终评分,令这两个节点的度的和为 m_{ij},度平衡因子为 b_{ij},则有式(9)、(10):

$$m_{ij}=deg_i+deg_j;\tag{9}$$

$$b_{ij}=e^{|deg_i-deg_j|}。\tag{10}$$

评分的总体公式如式(11)所示:

$$Gscore_{ij}=w_{ij}\frac{m_{ij}}{b_{ij}}。\tag{11}$$

式中:$Gscore_{ij}$ 代表基于场景连接图方法的评分;w_{ij} 由式(8)计算得出。

整个场景连接图的初始对选取算法具体步骤如下。

Step 0　初始化:以 2.1 节中建立的场景连接图、图像集合 I 为输入,遍历连接图 G 的节点集合 N 中每一个节点 n_i,计算每个节点的度 deg_i 并存储;

Step 1　访问 E 中任一未被访问的边缘 e_{ij},读取 e_{ij}所连接的两个节点 n_i 和 n_j,根据式(9)、式(10)计算两节点间的度的和 m_{ij}与度平衡因子 b_{ij};

Step 2　读取 e_{ij} 中的 w_{ij},结合 b_{ij} 与 m_{ij},根据式(11)计算该对节点的 $Gscore_{ij}$;

Step 3　当 E 中所有边缘 e_{ij} 都被访问,对 $Gscore_{ij}$ 排序,取得分最高的节点对 (n_i,n_j) 所对应的图像对 $(\boldsymbol{I}_i,\boldsymbol{I}_j)$ 组成初始图像对 $I_{\text{pair}}=(\boldsymbol{I}_i,\boldsymbol{I}_j)$(InitIP)并输出,否则返回 Step 1。

总体来说,本方法更倾向于选择靠近场景中心的 InitIP,旨在解决特殊场景中传统方法与多任务方法所选的 InitIP 容易陷入局部最优,而导致

的重建精度低、不完整等问题,相较于传统方法提升了计算效率的同时对特殊场景的鲁棒性更好,适用范围更广。

3 实验与结果分析

3.1 实验环境及数据

本文所有实验均在配备 Intel i76700 K 处理器和单个 NVIDIA GTX 1080 Ti 图形显卡的实验机上进行。采用的深度学习框架为 TensorFlow,学习率的更新策略为每迭代 10 000 步,将学习率调整为当前学习率的 0.9 倍直至完成训练。稀疏重建对比的传统方法为 Alicevision[6]。

实验中采用由香港科技大学的计算机科学与工程系建立的公开数据集 GL3D[7] 作为训练数据集。其中包含了 90 630 张涉及 378 个不同场景的高分辨率图像。在测试时,采用 Cambridge Landmarks Dataset 室外数据集[15],这是一个大型的城市数据集,包含来自剑桥大学周围的多个不同建筑场景。

3.2 多任务网络的实验

在测试数据集的 5 个场景的数据集上进行 InitIP 选取实验,图 7 为一个场景的 InitIP 示意图,图 7(a)为 Alicevision 所选取的;图 7(b)、图 7(c)分别为浅层交叉网络与深层交叉网络的选取结果。从外观上看,交叉连接的两个网络所选择的 InitIP 也基本符合特征相似与空间位移的原则。

(a) Alicevision选取

(b) 浅层交叉网络选取

(c) 深层交叉网络选取

图 7 3 种方法所选 InitIP 对比图

Figure 7 Comparison of InitIP selected by three methods

进一步定量对比传统方法与多任务方法所选 InitIP 作为起点进行稀疏重建时的表现,如表 1 所示。定量结果显示,两种交叉连接方式相较于 Alicevision 速度上都有较大提升,对比选取时间,所提出的多任务方法在多种不同场景中的选取速度提升 5 倍以上。在相同的测试场景下,深层交叉连接网络因为在更深的层级中嵌入了通道数更多的卷积层,导致模型参数量上升,从而使得整体的推理时间略长于浅层连接。但深层交叉连接的网络的实验结果,在 5 个重建场景中的最终误差都最低,其中 2、3 场景中的重投影误差有明显降低。综合来看,深层交叉网络性能要高于浅层交叉。

表 1 稀疏重建结果定量对比表

Table 1 Quantitative comparison of sparse reconstruction results

场景	图像数量	对比方法	t/s	重投影误差/px
1	112	Alicevision	207	1.092 34
		浅层交叉	26	1.096 97
		深层交叉	27	1.091 03
2	278	Alicevision	537	1.138 86
		浅层交叉	54	1.130 06
		深层交叉	58	1.108 86
3	350	Alicevision	988	1.101 37
		浅层交叉	81	1.099 73
		深层交叉	87	1.088 37
4	895	Alicevision	2 068	1.447 96
		浅层交叉	147	1.448 78
		深层交叉	159	1.440 32
5	1 237	Alicevision	4 923	1.868 16
		浅层交叉	323	1.867 17
		深层交叉	343	1.865 16

3.3 场景图方法的实验

进一步对 IVRTC 比赛中无人机采集的、更大范围的场景数据[16]进行基于场景图方法的实验,该数据集包含 498 张分辨率为 5 472×3 648 像素的航拍图像,多任务网络采用精度更高的深层交叉网络,所建立场景连接图如图 8 所示。

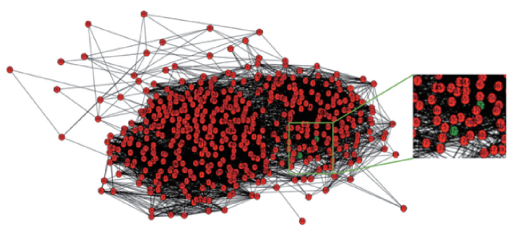


图 8 利用交叉网络中间输出建立的场景连接图

Figure 8 Scene connection graph established using the intermediate output of the cross-network

所选择的 InitIP 在场景连接图中使用绿色节点标注,可以看出其处于场景稠密区域,与 Alicevision 所选的 InitIP,以及交叉网络直接选取的 InitIP 对比如图 9 所示。

(a) Alicevision选取

(b)交叉网络+场景图方法选取

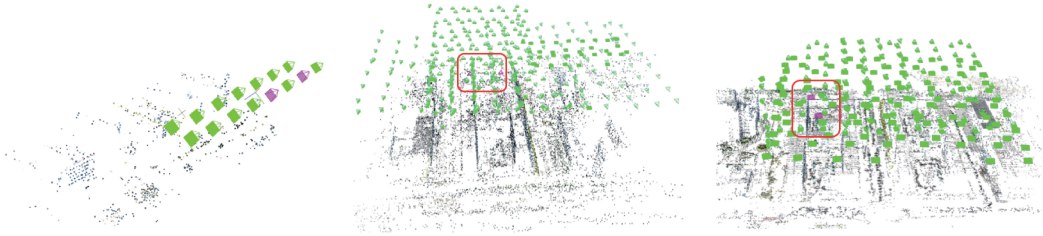
(c) 交叉网络直接选取

图 9　无人机场景 InitIP 对比图

Figure 9　Comparison of InitIP in aerial scenes

使用以上 3 组图像作为稀疏重建的起始点进行增量式 SfM,重建结果如图 10 所示,图中粉色方锥形代表的相机表示两幅初始图像。图 10(a)中 Alicevision 所选的 InitIP 仅引导重建出整个场景外围的一部分,只完成了少量图像的注册。图 10(b)为交叉网络+场景图方法所重建的场景,可见该方法所选 InitIP(由图中红色矩形框标出)能引导出完整的重建场景,其位置也更靠近场景中心,达到了所提方法的预期。图 10(c)中交叉网络直接选取的重建结果相对完整,但是 InitIP(由红色矩形框标出)距离场景稠密区域还有一定距离。进一步对比图 10(b)中的稀疏点云,该组实验结果仅重建出靠近 InitIP 中心区域的稀疏点云,在外侧树木部分还存在大量点云缺失的情况。

进一步定量地对比稀疏重建的结果,如表 2 所示,方法 A、B、C 分别为 Alicevision、交叉网络+场景图方法、交叉网络直接选取。



(a) Alicevision选取　　(b)交叉网络+场景图方法选取　　(c) 交叉网络直接选取

图 10　无人机场景稀疏重建对比图

Figure 10　Comparison of sparse reconstruction of aerial scenes

表 2　稀疏场景重建数据对比表

Table 2　Comparison of sparse reconstruction results

方法	注册相机数	空间点个数	重投影误差/px
A	15	17 607	1.182 42
B	445	198 906	1.139 76
C	240	87 799	1.153 93

可以看出,多任务方法与 Alicevision 对比,场景完整度均大幅提升,结合场景图的方法所选的 InitIP 则能够引导重建出最完整的稀疏场景,拥有最多的注册相机数、空间点数目,重建的空间点数量增加约 10 倍,重投影误差下降了约 0.05 px。综上所述,多任务学习结合场景连接图的选取策略能够高效地选取处于场景稠密区域的 InitIP,从而引导重建出更完整、精度更高的稀疏点云。

4　结论

根据初始图像对选取问题的特点,通过整合相似性检测和相机姿态回归两个子网络实现了一种基于多任务学习的 InitIP 选取网络,以提高增量式 SfM 中初始对选取过程的效率,并针对特殊重建场景提出结合场景连接图的选取策略,以提高重建的鲁棒性。实验结果证明所提方法在提高了效率的同时,能够很好地保证特殊场景重建的完整性和稳定性。

参考文献:

[1] BEDER C, STEFFEN R. Determining an initial image pair for fixing the scale of a 3d reconstruction from an image sequence[C]//Joint Pattern Recognition Symposium. Berlin: Springer, 2006: 657-666.

[2] HANER S, HEYDEN A. Covariance propagation and next best view planning for 3D reconstruction[C]//European Conference on Computer Vision. Berlin: Springer, 2012:545-556.

[3] SCHONBERGER J L, FRAHM J M. Structure-from-motion revisited[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. New York: IEEE, 2016: 4104-4113.

[4] MOULON P, MONASSE P, MARLET R. Adaptive

structure from motion with a contrario model estimation [C]//Asian Conference on Computer Vision. Berlin: Springer, 2012:257-270.

[5] 张艺琨,唐雁,陈强. 基于多特征融合的三维模型检索[J].郑州大学学报(工学版), 2019, 40(1):1-6.

[6] RUDER S. An overview of multi-task learning in deep neural networks [EB/OL]. (2017-06-15) [2020-07-25].https://arxiv.org/abs/1706.05098.

[7] SHEN T W, LUO Z W, ZHOU L, et al. Matchable image retrieval by learning from surface reconstruction [C]//Asian Conference on Computer Vision. Berlin: Springer, 2018:415-431.

[8] KENDALL A, GRIMES M, CIPOLLA R. Posenet: a convolutional network for real-time 6-DOF camera relocalization[C]//Proceedings of the IEEE International Conference on Computer Vision. New York: IEEE, 2015:2938-2946.

[9] SZEGEDY C, LIU W, JIA Y, et al. Going deeper with convolutions[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. New York:IEEE, 2015:1-9.

[10] AKENINE-MOLLER T, HAINES E, HOFFMAN N. Real-time rendering[M]. 3rd ed. New York: A K Peters/CRC Press, 2019.

[11] 姜三. 无人机倾斜影像高效 SfM 重建关键技术研究[D]. 武汉:武汉大学, 2018.

[12] FUKUDA S, YOSHIHASHI R, KAWAKAMI R, et al. Cross-connected networks for multi-task learning of detection and segmentation [EB/OL]. (2018-05-15) [2020-07-25]. https://arxiv.org/abs/1805.05569.

[13] ZHANG Y, YANG Q. An overview of multi-task learning[J]. National science review, 2018, 5(1): 30-43.

[14] KENDALL A, GAL Y, CIPOLLA R. Multi-task learning using uncertainty to weigh losses for scene geometry and semantics[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. New York: IEEE, 2018:7482-7491.

[15] KENDALL A, CIPOLLA R. Modelling uncertainty in deep learning for camera relocalization [C]//2016 IEEE International Conference on Robotics and Automation (ICRA). New York: IEEE, 2016:4762-4769.

[16] IVRTC dataset [EB/OL]. (2019-10-31) [2020-07-25].http://www.ivrtc.org/? page_id=660.

Selecting Initial Image Pairs Based on Multi-task Learning

LIU Yuxiang, ZHANG Maojun, YAN Shen, LI Jingbei, PENG Yang

(School of Systems Engineering, National University of Defense Technology, Changsha 410073, China)

Abstract: The selection of the initial image pair was the key to the incremental structure from motion (SfM). However, traditional selection methods had some problems such as low computational efficiency and poor robustness in some special scenes. In this paper, an initial image pair selection network based on multi-task learning was proposed to improve the efficiency of selection, and a selection strategy combined with the scene connection graphs was proposed. The strategy firstly constructed the topological structure between the images, and then judged whether the initial image pair was in the center area of the scene based on the density of the connections between the images, so as to avoid the incomplete reconstruction in some special scenes due to the selected initial image pair being in the edge of the whole scene. Compared with traditional SfM (structure from motion) methods, the selecting speed of the proposed method in a variety of different scenes was increased by more than 5 times. At the same time, the proposed selection strategy combined with scene graphs could increase the number of reconstructed spatial points in special scenes by 10 times, and reduce the reprojection error by 0.05 px, which significantly improved the robustness of the initial image pair selection in special scenes. This proved the effectiveness of the proposed method. While improving the efficiency, it could ensure the completeness and stability of the reconstruction of special scenes.

Key words: incremental SfM; initial image pair selection; multi-task learning; scene graph