·图形图像处理· 文章编号：1000-3428(2021)09-0197-06 文献标志码：A 中图分类号：TP391.41

基于动态遍历的分层特征网络视觉定位

蒋雪源，陈青梅，黄初华
（贵州大学 计算机科学与技术学院，贵阳 550025）

摘 要：采用分层特征网络估计查询图像的相机位姿，会出现检索失败和检索速度慢的问题。对分层特征网络进行分析，提出采用动态遍历与预聚类的视觉定位方法。依据场景地图进行图像预聚类，利用图像全局描述符获得候选帧集合并动态遍历查询图像，利用图像局部特征描述符进行特征点匹配，通过PnP算法估计查询图像的相机位姿，由此构建基于MobileNetV3的分层特征网络，以准确提取全局描述符与局部特征点。在典型数据集上与AS、CSL、DenseVLAD、NetVLAD等主流视觉定位方法的对比结果表明，该方法能够改善光照与季节变化场景下对候选帧的检索效率，提升位姿估计精度和候选帧检索速度。

关键词：视觉定位；分层特征网络；动态遍历；预聚类；位姿估计

开放科学（资源服务）标志码（OSID）：

中文引用格式：蒋雪源，陈青梅，黄初华.基于动态遍历的分层特征网络视觉定位[J].计算机工程，2021，47(9)：197-202.
英文引用格式：JIANG X Y，CHEN Q M，HUANG C H.Hierarchical feature network for visual localization based on dynamic traversal[J].Computer Engineering，2021，47(9)：197-202.

Hierarchical Feature Network for Visual Localization Based on Dynamic Traversal

JIANG Xueyuan，CHEN Qingmei，HUANG Chuhua
（College of Computer Science and Technology，Guizhou University，Guiyang 550025，China）

【Abstract】When used to estimate the camera pose of the query image，Hierarchical Feature Network（HFNet）is limited by frequent retrieval failures and the low retrieval speed. This paper analyzes HFNet and proposes a visual location method based on dynamic traversal and pre-clustering. According to the scene map，the image is pre-clustered. Then the global image descriptor is used to obtain the candidate frame set and dynamically traverse the query image，while the local feature descriptor is used to match the feature points. In addition，the camera pose of the query image is estimated by using the PnP algorithm. On this basis，an HFNet based on MobilenetV3 is constructed for the extraction of global descriptors and local feature points. Experimental results on typical data sets show that，compared with mainstream visual localization methods such as AS，CSL，DenseVLAD and NetVLAD，the proposed method can improve the retrieval efficiency of candidate frames in the cases of changing illumination conditions and seasons. It can also improve the accuracy of pose estimation and the speed of retrieving candidate frames.

【Key words】visual localization；Hierarchical Feature Network（HFNet）；dynamic traversal；pre-clustering；pose estimation
DOI：10. 19678/j. issn. 1000-3428. 0059101

0 概述

实现移动设备实时精确的定位是移动增强现实技术的基本要求[1]。由于GPS、北斗等方式难以穿透建筑物，移动设备在建筑密集区域的定位比较困难，而利用超宽带、红外、超声波等设备接收信号实现定位，需要在场景中部署设备[2]。以上两类方法在实际应用时均有不足，而视觉定位方法不存在上述问题，因此受到众多学者的关注[3]。

针对从场景图像中估计查询图像位姿的问题，SARLIN等[4]提出了层次定位方法。该方法利用卷积神经网络在场景图像中找到候选帧，候选帧聚类

基金项目：贵州省自然科学基金（黔科合基础[2019]1088）；贵州大学引进人才科研项目（贵大人基合字(2017)31号，贵大人基合字(2015)52号）；贵州省教育厅创新群体重大研究项目（黔教合KY字[2018]026）。
作者简介：蒋雪源（1993—），男，硕士研究生，主研方向为计算机视觉、深度学习；陈青梅，硕士研究生；黄初华，副教授、博士。
收稿日期：2020-07-30 **修回日期：**2020-09-03 **E-mail：**chhuang@gzu.edu.cn

计算机工程 www.ecice06.com

后与查询图像进行特征点匹配，建立特征点的2D-3D对应关系，使用PnP算法估计相机位姿。之后，SARLIN等[5]在层次定位方法的基础上提出了分层特征网络（Hierarchical Feature Network，HFNet），进一步改善了层次定位的性能。该方法联合估计图像的全局描述符与局部特征点，使层次定位方法具有出色的鲁棒性。HFNet将基于SIFT[6]的局部特征描述符替换为基于学习的局部特征描述符，引入了多任务知识蒸馏方法，使层次定位方法对计算资源的需求进一步减小。但是HFNet估计全局描述符的效果仍需提升，这导致层次定位方法查找候选帧的失败率较高同时，同时层次定位方法的聚类步骤计算量也比较大。

本文在HFNet的基础上，提出结合动态遍历与预聚类的视觉定位方法。利用动态遍历方式搜索候选帧，对粗略检索步骤的候选帧数量做动态调整。在此基础上，根据场景地图进行图像预聚类，引入压缩-激励模块和h-swish激活函数改进分层特征网络。

1 相关工作

目前视觉定位技术主要分为基于图像检索、基于场景地图、基于学习3类方法[5]。基于图像检索的方法通过在场景图像中检索查询图像，从场景图像返回与查询图像最相似的图像，近似得到查询图像的位姿[7-8]。该类方法可以应用到大型数据集中，但只能估计查询图像的近似位姿。基于场景地图的方法[9-10]通过直接匹配查询图像与场景地图建立查询图像特征点的2D-3D对应关系，从而得到查询图像的相机位姿。在大型场景中，由于场景地图规模庞大，该方法可靠性不高。基于学习的方法通过神经网络学习方法获得相机位姿，如KENDALL等[11]提出的PoseNet，通过训练室内与室外场景图像和图像对应的相机位姿，预测出查询图像的相机位姿。该方法在大规模变化场景下具有出色的鲁棒性，但定位精度不能达到增强现实应用的要求。

作为一种结合图像检索与场景地图的视觉定位方法，层次定位方法可在大型数据集中估计查询图像的精确位姿，但其需要提取场景图像的特征点构建场景地图。首先在场景图像中根据全局描述符对查询图像进行粗略检索，得到由场景图像组成的候选帧集合；然后将有共同特征点的候选帧聚类为一组，查询图像与各组候选帧进行局部特征点匹配，匹配的特征点可以从场景地图得到三维坐标，构建特征点的2D-3D对应关系；最后使用PnP算法估计查询图像的相机位姿。针对层次定位中全局描述符与局部特征点的提取，HFNet进行联合估计。该方法以MobileNetV2[12]为编码器，以NetVLAD[8]和SuperPoin[13]为解码器。其中，NetVLAD解码器输出图像的全局描述符，SuperPoint解码器输出图像的局部特征点。然而，HFNet存在全局描述符检索率低的问题，且聚类步骤的速度仍需提升。本文通过分析相机位姿恢复失败的查询图像，发现大部分恢复失败的图像并没有搜索出正确的候选帧。如图1所示，Query.jpg为查询图像，其余图像为粗略检索得到的候选帧。可以看出，查询图像与候选帧并不在同一地点，在候选帧数量有限的情况下会出现粗略检索候选帧失败的现象，导致恢复相机位姿失败。增加候选帧的数量可以缓解这个问题，但会增加聚类与局部特征点匹配的图像数量，导致运算量成倍增加。



图1 相机位姿估计失败的查询图像与候选帧

Fig.1 The query images and candidate frames where camera pose estimation fails

2 本文方法

2.1 改进的层次定位算法

本文提出了基于动态遍历与预聚类的层次定位方法，算法框架如图2所示，其中上半部分为场景图像预处理，构建场景地图与聚类信息；下半部分为相机位姿估计。

计算机工程 www.ecice.com

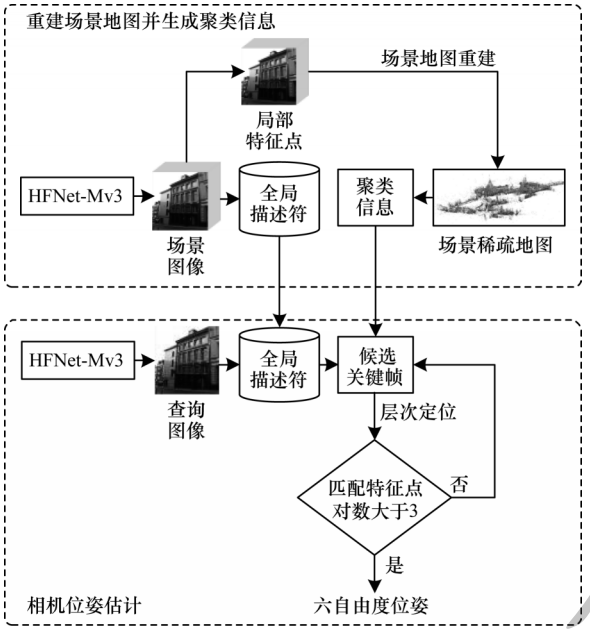

图2　本文方法框架

Fig.2　Framework of the proposed method

本文方法主要步骤如下：

1）利用基于MobileNetV3[14]的分层特征网络（简称HFNet-Mv3）提取全局描述符与局部特征点，全局描述符用于查询图像与场景图像的粗略检索，局部特征点用于构建场景稀疏地图Q。

2）通过场景地图Q对场景图像进行预聚类。层次定位方法先在场景图像中找到候选帧，再将这些候选帧按照是否为同一地点进行聚类，将有共同特征点的候选帧分为一组。笔者发现参与聚类步骤的图像均为场景图像，利用场景地图对场景图像进行预聚类处理，使得上述步骤可以直接对候选帧分组，降低了聚类步骤的运算量，且聚类结果不变。以场景图像L为例，预聚类步骤如下：记图像L的特征点为集合F，$F=\{\alpha_1,\alpha_2,\cdots,\alpha_n\}$，$\alpha$为集合$F$中的特征点，记$I=P(\alpha_k)$为拥有特征点$\alpha_k$的场景图像，$\alpha_k\in F$。遍历特征点集合$F$，将拥有这些特征点的场景图像记为图像$L$的聚类队列$M$，即$M=\{I\mid I=P(\alpha_k),\alpha_k\in F\}$。查找所有场景图像的聚类队列并保存，就能得到场景图像的聚类信息。

3）动态遍历搜索候选帧。如图3所示，查询图像与场景图像进行全局描述符匹配，得到m张场景图像组成的候选帧，这些候选帧选取前n张进行层次定位。聚类信息包含每个场景图像拥有共同特征点的图像序列，与候选帧匹配可直接完成聚类步骤。算法对参与层次定位的候选帧数量进行动态调整，若前n张候选帧层次定位失败，视为选中的候选帧没有查询图像所在的地点，则动态增加候选帧数量，候选帧向后选取n张图片继续进行层次定位。以n张图片为动态查询窗口，直到计算出查询图像的相机位姿。候选帧数量对粗略检索的速度影响不大，对聚类和局部特征点匹配的速度影响很大。动态遍历搜索候选帧的方法使聚类与局部特征点匹配的候选帧数量保持不变，对粗略检索步骤的候选帧数量进行动态调整，从而以较小的计算量代价缓解了查询图像定位失败的问题。

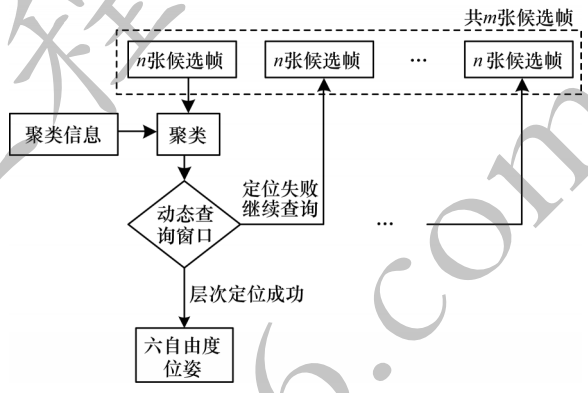

图3　动态查询流程

Fig.3　Procedure of dynamic query

2.2　基于MobileNetV3的分层特征网络

本文算法使用基于MobileNetV3的HFNet-Mv3提取查询图像的全局描述符与局部特征参与层次定位过程。MobileNetV3使用神经网络搜索算法构建全局的网络结构，利用NetAdapt[15]算法对每层卷积核数量进行优化。在网络结构上引入了压缩-激励模块[16]（Squeeze-and-Excite，SE），网络根据损失函数学习特征权重，增大有效的特征图权重，减小无效或效果小的特征图权重，从而达到更好的训练效果。

HFNet使用ReLU激活函数，swish激活函数在深层模型上的效果优于ReLU，但计算量大。MobileNetV3引入了h-swish激活函数，在降低计算量的同时对swish激活函数进行了拟合，其定义如式（1）所示：

$$\text{h-swish}[x]=x\cdot\frac{\text{ReLU6}(x+3)}{6} \tag{1}$$

为提升分层特征网络的性能，本文以MobileNetV3-Large[14]作为分层特征网络的编码器，以NetVLAD和SuperPoin作为解码器，提出了改进的分层特征网络HFNet-Mv3，其网络结构如图4所示。

计算机工程 www.ecice06.com

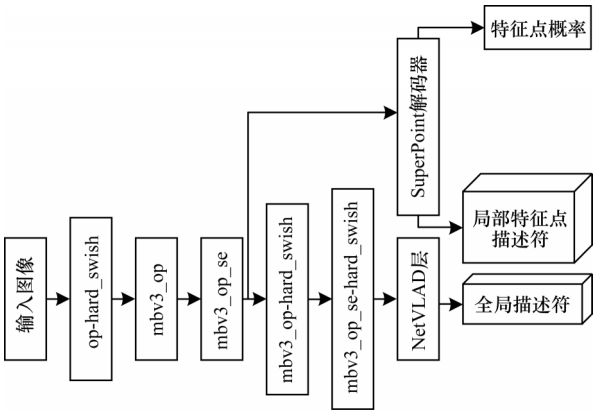

图4 HFNet-Mv3的网络结构

Fig.4 Network structure of HFNet-Mv3

HFNet-Mv3的编码器由op-hard_swish、mbv3_op、mbv3_op_se等6个模块组成，其中：mbv3_op_se与mbv3_op_se-hard_swish引入了压缩-激励模块；op-hard_swish、mbv3_op-hard_swish和mbv3_op_se-hard_swish使用了h-swish激活函数。除op-hard_swish外编码器以深度可分离卷积[17]作为卷积核的基本单位。

SuperPoint解码器需要更高的分辨率来保留特征，因此，在mbv3_op_se层之后，即在HFNet-Mv3的第7层引入了SuperPoint解码器，用来输出图像像素的特征点概率和局部特征描述符。由于NetVLAD解码器需要更多的语义信息用于输出图像的全局描述符，因此在HFNet-Mv3的最末端引入了NetVLAD解码器。

损失函数如式(2)所示：

$$L=\mathrm{e}^{-\omega_1}\left\|d_s^g-d_{t_1}^g\right\|_2^2+\mathrm{e}^{-\omega_2}\left\|d_s^l-d_{t_2}^l\right\|_2^2+2\mathrm{e}^{-\omega_3}\mathrm{CrossEntropy}(p_s,p_{t_3})+\sum_i\omega_i \tag{2}$$

其中：ω_i为优化变量，用于避免手动调节各个损失函数的权重[18]；d_s^g为HFNet-Mv3的全局描述符输出；$d_{t_1}^g$为教师网络的全局描述符输出；d_s^l为HFNet-Mv3的局部特征描述符输出；$d_{t_2}^l$为教师网络的局部特征描述符输出；p_s为HFNet-Mv3的局部特征点输出；p_{t_3}为教师网络的局部特征点输出，局部特征点采用交叉熵计算损失。

3 实验结果与分析

3.1 实验环境及数据集

基于TensorFlow1.14.0和pytorch0.4.1框架进行实验，训练实验的硬件配置如下：IntelXeon®CPUE5-2620v42.10 GHz，NVIDIATITANXp12 GB；定位实验的硬件配置如下：IntelCorei58300H2.30 GHz，NVIDIAGTX10606 GB。

采用GoogleLandmarks和Berkeley DeepDrive数据集训练HFNet-Mv3。Google Landmarks数据集包括世界各地的城市场景，选取其中的185 000张图片进行训练；BerkeleyDeepDrive数据集由黄昏与夜间的道路场景组成，选取其中的37 681张图片参与训练。所有图片都预处理为480像素×640像素的灰度图。采用SuperPoint与NetVLAD作为HFNet-Mv3的教师网络，分别估计GoogleLandmarks数据集和Berkeley DeepDrive数据集的局部特征点、局部特征描述符和全局描述符作为图像标签。对HFNet-Mv3进行多任务知识蒸馏训练[19]，batch_size为16，通道乘数[12]为0.75，总迭代次数为85 000次，初始学习率设置为10^{-3}，随迭代进行，微调学习率。

在AachenDay-Night和CMUSeasons数据集上评估HFNet-Mv3对HFNet的提升效果，每个数据集由稀疏的场景地图和图像组成。AachenDay-Night数据集包含了4 328张来自Aachen旧城区白天的场景图像、824张白天查询图像和98张夜间的查询图像。CMUSeasons数据集记录了不同季节的图像，由17个子集组成，包含了市区与郊区的7 159张场景图像和75 335张查询图像，由于季节、光照和天气的变化导致环境条件的变化，该数据集具有挑战性。由数据集作者提供的场景地图并不适用于分层特征网络定位，因此，本文利用HFNet-Mv3在场景图像中提取特征点，利用COLMAP[20]软件对特征点进行匹配与三角测量，得到该数据集的场景地图。

3.2 定位结果及分析

选用通用的基准数据集AachenDay-Night和CMUSeasons进行测试，并将测试结果与ActiveSearch(AS)[21]、CityScaleLocalization(CSL)[9]、DenseVLAD[22]、NetVLAD、DIFL+FCL[23]和HFNet作为定位精度基准进行对比。AS和CSL是基于场景地图的视觉定位方法，DenseVLAD、NetVLAD和DIFL+FCL是基于图像检索的视觉定位方法，本文方法使用HFNet-Mv3提取图像的全局描述符与局部特征点，使用基于预聚类与动态遍历的视觉定位方法估计相机位姿。

通过相机位姿的估计值与真实值之间的偏差来评估算法的精度。将定位结果上传到数据集作者的网站(https://www.visuallocalization.net/submission/)，由数据集作者对算法的位姿精度进行评估。数据集作者设定3个位姿精度区间，即高精度(0.25 m，2°)、中精度(0.5 m，5°)和低精度(5 m，10°)，通过比较在不同精度条件下的召回率来比较算法的精度。

3.3 运行时间

通过记录算法在不同步骤的运行时间，比较本文算法与HFNet的运行速度。由于实验硬件配置差异的原因，实验结果与原文献数据会略有不同。

表1为视觉定位方法在不同距离与角度阈值下的召回率，从左到右依次为高精度、中精度和低精度的召回率，算法性能已在数据集网站公开(本文方法名称为Dynamic retrieval and pre-clustering for HFNet-Mv3)，其中DIFL+FCL的作者只在CMUSeasons数据集上做

计算机工程 www.ecice06.com

了测试。由表1可以看出,本文方法在不同距离与角度阈值下的召回率均大于其他算法,在Aachen夜间与CMU公园场景的精度上,相对HFNet提升5%以上。夜间图像粗略检索正确的候选帧比较困难,公园场景图像的季节、光照和植物的变化更明显,导致层次定位失败。对比HFNet,本文方法缓解了这个问题,因此,Aachen的夜间图像与CMU的公园场景图像估计相机位姿精度的提升效果更明显。

表1　在AachenDay-Night和CMUSeasons数据集上的定位精度对比

Table 1　Localization accuracy comparison on AachenDay-Night and CMUSeasons datasets　%

方法	AachenDay	AachenNight	CMUurban	CMUsuburban	CMUpark
AS方法	57.3/83.7/96.6	28.6/37.8/51.0	55.2/60.3/65.1	20.7/25.9/29.9	12.7/16.3/20.8
CSL方法	52.3/80.0/94.3	29.6/40.8/56.1	36.7/42.0/53.1	8.6/11.7/21.1	7.0/9.6/17.0
DenseVLAD方法	0.0/0.1/22.8	0.0/1.0/19.4	22.2/48.6/92.8	9.8/26.6/85.2	10.3/27.1/77.0
NetVLAD方法	0.0/0.2/18.9	0.0/0.0/14.3	17.4/40.3/93.2	7.6/21.0/80.5	5.6/15.7/65.8
DIFL+FCL方法	—	—	20.2/45.0/87.1	9.1/23.3/69.3	10.1/26.4/74.0
HFNet方法	76.2/85.4/91.9	62.2/73.5/81.6	90.3/93.0/96.1	71.7/78.0/86.8	49.3/56.6/71.3
本文方法	77.9/87.1/93.1	67.3/78.6/88.8	92.0/94.9/97.9	74.7/82.2/91.7	56.0/65.4/81.2

表2为视觉定位方法在不同步骤的运行时间,从左到右依次为特征提取、全局检索候选帧、聚类、局部特征点匹配、PnP求解相机位姿和总耗时,CMUslice2、CMUslice9、CMUslice25分别为CMUSeasons在市区、郊区与公园场景的子集。由表2可以看出,本文方法在特征提取与全局检索候选帧步骤的耗时与HFNet几乎一致。由于本文方法采用了预聚类策略,聚类步骤的耗时缩短为HFNet的10%。针对HFNet检索正确候选帧失败的查询图像,本文方法动态增加候选帧的数量,导致局部特征点匹配与PnP求解相机位姿的平均耗时增加。由于聚类步骤耗时的大幅缩短,本文方法总耗时与HFNet几乎一致。

表2　在AachenDay-Night和CMUSeasons数据集上的定位速度对比

Table 2　Localization speed comparison on AachenDay-Night and CMUSeasons datasets　ms

数据集	HFNet方法	本文方法
AachenDay	81/5/100/24/38/248	98/6/8/26/72/210
AachenNight	82/6/95/22/72/277	98/5/7/46/112/268
CMUslice2	86/2/66/15/96/265	97/2/4/15/122/240
CMUslice9	80/2/41/13/144/280	97/2/3/13/198/313
CMUslice25	81/2/33/8/137/261	98/2/7/7/156/270

3.4　动态遍历参数

本文提出的动态遍历方法有2个关键参数,即候选帧数量m和动态查询图像数量n。候选帧数量决定了分层特征网络获取正确候选帧的能力,随着候选帧数量的提升,定位精度的提升变得不明显,而速度明显降低。动态查询图像数量决定了聚类图像数量,太小会减少匹配的特征点数量,影响定位精度,太大会导致下一窗口得不到充足的聚类图像,并降低运行速度。

表3和表4为不同动态遍历参数下本文方法在AachenDay-Night数据集中的定位精度与速度。由于动态遍历参数不影响特征提取,因此特征提取步骤耗时相同。候选帧数量m设为80,定位精度最高,但速度降低幅度过大。由实验结果可得,候选帧数量m设为40,动态查询图像数量n设为10的情况下,定位精度与速度最优。

表3　不同参数情况下本文方法在AachenDay-Night数据集上的定位精度

Table 3　Localization accuracy of the proposed method on AachenDay-Night dataset under different parameters　%

参数	AachenDay	AachenNight
m=10,n=10	76.9/85.8/91.0	65.3/76.5/86.7
m=40,n=10	77.9/87.1/93.1	67.3/78.6/88.8
m=80,n=10	78.3/87.7/93.8	69.4/80.6/92.9
m=40,n=5	77.2/84.8/92.6	67.3/77.6/86.7
m=40,n=20	77.4/86.0/93.1	63.3/76.5/85.7

表4　不同参数情况下本文方法在AachenDay-Night数据集上的定位速度

Table 4　Localization speed of the proposed merthod on AachenDay-Night dataset under different parameters　ms

参数	AachenDay	AachenNight
m=10,n=10	98/5/8/26/51/188	98/6/7/47/91/249
m=40,n=10	98/6/8/26/72/210	98/5/7/46/112/268
m=80,n=10	98/7/9/28/123/265	98/5/7/54/243/407
m=40,n=5	98/5/4/17/76/200	98/6/4/45/151/304
m=40,n=20	98/6/18/42/91/255	98/6/15/64/114/297

4　结束语

本文针对层次定位方法粗略检索失败率高和聚类耗时长的问题,在分层特征网络中引入压缩-激励模块与h-swish激活函数,提出基于动态遍历与预聚

计算机工程 www.ecice06.com

类的视觉定位方法。在2个通用基准数据集上的实验结果表明,本文方法通过对粗略检索步骤的候选帧数量进行动态调整,有效降低了检索候选帧的失败率,与HFNet方法相比,高精度、中精度和低精度的召回率均有所提升,其中夜间与公园场景的精度提升达到5%以上。同时该方法采用了预聚类策略,使得运行速度与HFNet基本一致。本文方法适用于城市和郊区等室外场景,下一步将针对室内环境训练优化HFNet-Mv3,并对室内环境下的层次定位方法进行改进。

参考文献

[1] 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):855-868.
LIU H M,ZHANG G F,BAO H J. A survey of monocular simultaneous localization and mapping[J]. Journal of Computer-Aided Design & Computer Graphics,2016,28(6):855-868.(in Chinese)

[2] HAN G J,JIANG J F,ZHANG C Y,et al. A survey on mobile anchor node assisted localization in wireless sensor networks[J]. IEEE Communications Surveys & Tutorials,2016,18(3):2220-2243.

[3] 邓中亮,尹露,唐诗浩,等. 室内定位关键技术综述[J]. 导航定位与授时,2018,5(3):14-23.
DENG Z L,YIN L,TANG S H,et al. A survey of key technology for indoor positioning[J]. Navigation Positioning and Timing,2018,5(3):14-23.(in Chinese)

[4] SARLIN P E,DEBRAINE F,DYMCZYK M,et al. Leveraging deep visual descriptors for hierarchical efficient localization[C]//Proceedings of Conference on Robot Learning. Zurich,Switzerland:PMLR,2018:456-465.

[5] SARLIN P E,CADENAC R,SIEGWART R,et al. From coarse to fine:robust hierarchical localization at large scale[C]//Proceedings of IEEE/CVF Conference on Computer Vision and Pattern Recognition. Washingtong D. C.,USA:IEEE Press,2019:12708-12717.

[6] LOWE D G. Distinctive image features from scale-invariant key points[J]. International Journal of Computer Vision,2004,60(2):91-110.

[7] KIM H J,DUNN E,FRAHM J M. Learned contextual feature reweighting for image Geo-localization[C]//Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. Washington D. C.,USA:IEEE Press,2017:3251-3260.

[8] ARANDJELOVIC R,GRONAT P,TORII A,et al. NetVLAD:CNN architecture for weakly supervised place recognition[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence,2018,40(6):1437-1451.

[9] SVARM L,ENQVIST O,KAHL F,et al. City-scale localization for cameras with known vertical direction[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence,2017,39(7):1455-1461.

[10] LIU L,LI H D,DAI Y C. Efficient global 2D-3D matching for camera localization in a large-scale 3D map[C]//Proceedings of IEEE International Conference on Computer Vision. Washington D. C.,USA:IEEE Press,2017:2391-2400.

[11] KENDALL A,GRIMES M,CIPOLLA R. PoseNet:a convolutional network for real-time 6-DOF camerare localization[C]//Proceedings of IEEE International Conference on Computer Vision. Washington D. C.,USA:IEEE Press,2015:2938-2946.

[12] SANDLER M,HOWARD A,ZHU M L,et al. MobileNetV2:inverted residuals and linear bottlenecks[C]//Proceedings of IEEE/CVF Conference on Computer Vision and Pattern Recognition. Washington D. C.,USA:IEEE Press,2018:4510-4520.

[13] DETONE D,MALISIEWICZ T,RABINOVICH A. SuperPoint:self-supervised interest point detection and description[C]//Proceedings of IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops. Washington D. C.,USA:IEEE Press,2018:337-350.

[14] HOWARD A,SANDLER M,CHEN B,et al. Searching for MobileNetV3[C]//Proceedings of IEEE/CVF International Conference on Computer Vision. Washington D. C.,USA:IEEE Press,2019:1314-1324.

[15] YANG T J,HOWARD A,CHEN B,et al. NetAdapt:platform-aware neural network adaptation for mobile applications[C]//Proceedings of European Conference on Computer Vision. Berlin,Germany:Springer,2018:285-300.

[16] HU J,SHEN L,ALBANIE S,et al. Squeeze-and-excitation networks[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence,2020,42(8):2011-2023.

[17] CHOLLET F. Xception:deep learning with depth wise separable convolutions[C]//Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. Washington D. C.,USA:IEEE Press,2017:1251-1258.

[18] CIPOLLA R,GAL Y,KENDALL A. Multi-task learning using uncertainty to weigh losses for scene geometry and semantics[C]//Proceedings of IEEE/CVF Conference on Computer Vision and Pattern Recognition. Washington D. C.,USA:IEEE Press,2018:7482-7491.

[19] FUKUI S,YU J,HASHIMOTO M. Distilling knowledge for non-neural networks[C]//Proceedings of NIPS'14. Washington D. C.,USA:IEEEPress,2019:1411-1416.

[20] SCHÖNBERGER J L,FRAHM J M. Structure-from-motionrevisited[C]//Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. Washington D. C.,USA:IEEE Press,2016:4104-4113.

[21] SATTLER T,LEIBE B,KOBBELT L. Efficient & effective prioritized matching for large-scale image-based localization[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence,2017,39(9):1744-1756.

[22] TORII A,ARANDJELOVIC R,SIVIC J,et al. 24/7 placere cognition by view synthesis[C]//Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. Washington D. C.,USA:IEEE Press,2015:1808-1817.

[23] HU H J,WANG H S,LIU Z,et al. Retrieval-based localization based on domain-invariant feature learning under changing environments[C]//Proceedings of IEEE/RSJ International Conference on Intelligent Robots and Systems. Washington D. C.,USA:IEEE Press,2019:3684-3689.

编辑 金胡考

计算机工程 www.ecice06.com