一种基于射线模型的图像定位系统

邓 磊[1,2],陈宝华[1,2],黄思远[1,2],段岳圻[1,2],周 杰[1,2,3]
(1. 清华大学自动化系,北京 100084; 2. 清华信息科学与技术国家实验室,北京 100084;
3. 智能技术与系统国家重点实验室,北京 100084)

摘 要: 图像定位技术在现实中有广泛的应用,如导航、路径规划、虚拟现实等. 对于用户而言,只需拍摄一张图像即可实现定位. 本文提出了一种基于射线摄像机模型的图像定位系统,包含基于射线模型的三维重构算法和基于位姿图优化的图像定位方法. 提出的三维重构算法利用射线模型的内在几何性质,能够处理全景和鱼眼等广角摄像机模型,降低采集和重构的代价,提升重构的效果. 基于位姿图的定位方法融合了图像与点云匹配的信息和图像间相对位姿信息实施定位,得到更高的定位精度. 实验证明本文方法的有效性.

关键词: 图像定位; 射线摄像机模型; 三维重建; 位姿图优化; 全景相机
中图分类号: TP391 **文献标识码:** A **文章编号:** 0372-2112 (2017)01-0001-07
电子学报 URL: http://www.ejournal.org.cn **DOI:** 10.3969/j.issn.0372-2112.2017.01.001

A Ray-Based Image Localization System

DENG Lei[1,2], CHEN Bao-hua[1,2], HUANG Si-yuan[1,2], DUAN Yue-qi[1,2], ZHOU Jie[1,2,3]
(1. *Department of Automation, Tsinghua University, Beijing* 100084, *China*;
2. *Tsinghua National Laboratory for Information Science and Technology, Beijing* 100084, *China*;
3. *State Key Laboratory of Intelligent Technology and Systems, Beijing* 100084, *China*)

Abstract: Image-based localization has wide applications including auto navigation, path planning, virtual reality, etc. For the end user, a single image is required. In this paper, we present a ray-based image localization system including ray-based 3D reconstruction and pose graph optimization-based localization. By using the inherit geometry properties of ray-based camera model, the proposed reconstruction algorithm can deal with various wide FOV camera models (panorama and fisheye) to reduce the reconstruction cost and achieve a better reconstruction result. The proposed pose graph optimization localization framework can combine 2D-3D matches information and relative pose information for a better localization accuracy. Experiment result shows the effectiveness of the proposed system.

Key words: image-based localization; ray based camera model; 3D reconstruction; pose graph optimization; panorama

1 引言

近年来,基于图像的定位技术得到了广泛的研究. 该项技术可以用于机器人导航[1]、路径规划[2]、数字旅游、虚拟现实等. 能够适用于 GPS 无法工作的区域,如室内和地下. 相比基于传感器的定位技术,如蓝牙、WiFi 等,图像定位技术不依赖于专业设备,实施成本低. 对用户而言,仅需要拍摄一张图像便可以得到自身的位置和姿态,对于定位服务提供商,只需拍摄一组场景的图像即可.

目前基于图像定位的方法主要有两类,一种是基于图像检索的方法[3,4]. 此类方法寻找查询图像在数据库中的近邻图像,以其位置作为自身位置. 由于没有更充分的利用三维信息,其定位精度不优于库图像本身的位置精度和采样间隔. 另一种是基于三维重构结合图像-点云(2D-3D)匹配的方法[5,6]. 这类方法首先离线采集大量关于目标场景的平面图像,进行三维重构得到场景的三维特征点云[7,8]. 在线定位阶段,提取查询图像特征,并将其与三维特征点云进行 2D-3D 匹配,利用匹配结果估计目标摄像机的位姿. 相比图像检索方法,此类方法能够得到精度更高的定位结果. 但其中的三维重构算法只能用于平面摄像机. 受限于其较小的视场,通常需要对同一位置变换多个角度进行拍摄,得

收稿日期:2015-06-08;修回日期:2015-12-09;责任编辑:孙瑶
基金项目:国家自然科学基金(No. 61225008, No. 61373074, No. 61373090);国家 973 重点基础研究发展计划(No. 2014CB349304);教育部基金(No. 20120002110033)

到数量较大的图像集合进行三维重建,采集量大且重构代价较高.

随着拍摄设备和图像传感器技术的发展,全景、鱼眼等各类广角摄像机应用日益广泛,如图 1 所示. 相比传统的平面相机,这类相机具有更大的视场,覆盖相同场景信息所需的拍摄代价更低. 1 张全景图像等效于 5 ~8 张同光心不同朝向的平面图像. 特别适用于对场景完整性要求较高的应用,如本文所研究的图像定位问题,重构时应该尽可能将场景覆盖完整,以适应未知位姿的目标图像. 为利用广角摄像机的视场优势,本文提出了一种基于射线模型的图像定位系统,系统包含基于射线模型的三维重构算法,以及融合图像-点云匹配信息和图像间相对位姿信息的图像定位方法. 在基于射线模型的三维重构算法中,通过使用三维射线描述二维像素坐标,射线模型能够无畸变地表达多种摄像机模型(全景、鱼眼、平面),并将各种摄像机模型内在的几何约束引入进来. 如使用全景图像的三维重构等效于引入了同光心约束的平面图像重构,因而重构效果好,且采集成本低,计算速度快. 图像定位中,提出的位姿图优化框架,通过融合图像-点云匹配(2D-3D)信息和图像间相对位姿信息,改善定位效果. 相比传统的基于图像检索的方法和基于图像-点云匹配(2D-3D)的定位方法,本文方法定位精度更高.

图1 使用的全景设备及图像(用于重构)和平面设备及图像(用于定位)

2 基于射线模型的三维重构

系统包括离线三维重构和在线图像定位两部分(如图 2 所示). 离线三维重构阶段,对一个场景采集一定数量的广角图像(如全景,鱼眼). 随后用提出的基于射线摄像机模型的三维重构算法对采集到的图像做三维重构得到场景的点云和广角图像的摄像机位姿,并建立 3D 点云特征的索引树来加速在线定位,三维重构部分介绍如下.

2.1 射线摄像机模型

定义摄像机局部坐标系为以摄像机光心为原点,摄像机光轴为 z 轴,图像坐标的两轴为 x,y 轴构成的右手坐标系(如图 3 所示). 摄像机矩阵为 $\boldsymbol{P}=[\boldsymbol{R}|\boldsymbol{t}]=\boldsymbol{R}[\boldsymbol{I}|-\boldsymbol{C}]$,其中 $\boldsymbol{R}$,$\boldsymbol{C}$ 是摄像机在世界坐标系下的旋转矩阵和光心坐标,即摄像机外参数. 世界坐标系下的 3D 点坐标用 $\boldsymbol{X}=(X,Y,Z)$ 表示,其在摄像机 $\boldsymbol{P}$ 的局部坐标系下的坐标为 $\boldsymbol{x}=\boldsymbol{P}[\boldsymbol{X};1]$.

在以摄像机光心为原点的局部摄像机坐标系中,通过光心的每条射线 $\boldsymbol{r}$ 可以用原点和单位球面上的另一个点 $\boldsymbol{x}=(x,y,z)$ 所定义. 该射线与图像坐标 $\boldsymbol{u}=(u,v)$ 通过映射函数一一对应. 映射函数定义为 $\boldsymbol{x}=f(\boldsymbol{u},\boldsymbol{K})$,$\boldsymbol{u}=f^{-1}(\boldsymbol{x},\boldsymbol{K})$,其中 $\boldsymbol{K}$ 为摄像机的内参. 对于不同的摄像机模型,其映射函数各不同. 全景、鱼眼、平面映射函数,分别由式(1) ~ (3)描述,如图 3 所示.

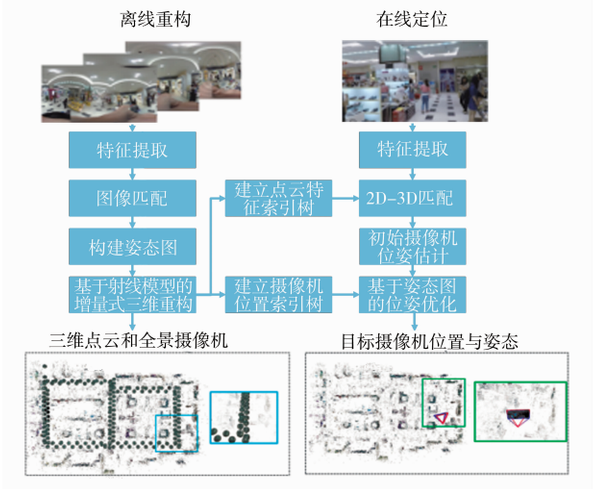


图2 系统流程图

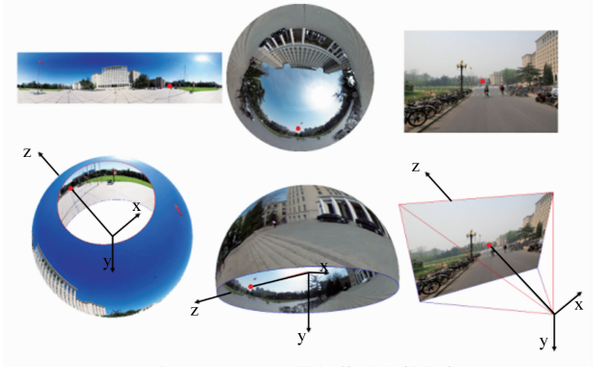


图3 全景、鱼眼、平面摄像机映射示意图

$$
\begin{aligned}
&pan=\frac{u-u_c}{f},\ tilt=\frac{v-v_c}{f},\boldsymbol{u}_c=(u_c,v_c),\\
&f(\boldsymbol{u},(f,\boldsymbol{u}_c))=(\cos(tilt)\sin(pan),\\
&-\sin(tilt),\cos(tilt)\cos(pan))\\
&u_1=\frac{u-u_c}{f},v_1=\frac{v-v_c}{f},\boldsymbol{u}_c=(u_c,v_c),
\end{aligned}
\tag{1}
$$

$$\phi=\arctan2(v_1,u_1),r_1=\sqrt{u_1^2+v_1^2},\theta=2\arctan\left(\frac{r_1}{2}\right),$$
$$f(\boldsymbol{u},(f,\boldsymbol{u}_c))=$$
$$(\cos(\phi)\sin(\theta),-\cos(\theta),\sin(\phi)\sin(\theta)) \tag{2}$$
$$pan=\arctan\left(\frac{u-u_c}{f}\right),tilt=\arctan\left(\frac{v-v_c}{f}\right),\boldsymbol{u}_c=(u_c,v_c),$$
$$f(\boldsymbol{u},(f,\boldsymbol{u}_c))=$$
$$(\cos(tilt)\sin(pan),-\sin(tilt),\cos(tilt)\cos(pan)) \tag{3}$$

式中 $\boldsymbol{u}_c$ 为摄像机主点坐标,f 为焦距.

相较于传统的基于像素的平面模型,射线模型能够适应不同的摄像机类型(全景、鱼眼、平面),并将其统一起来. 在传统平面模型下的三维重构中的基本模块都可以扩展到此处定义的射线模型中,例如本征矩阵及相对位姿的估计,三角测量,摄像机位姿估计,非线性优化等.

几何距离,即重投影误差(Reprojection Error),用来度量在高斯噪声假设下的投影点和观测值之间的距离,对其最小化几乎是所有多视图几何方法所公认的黄金法则[9]. 传统像素坐标下的距离度量为

$$d_{\text{pix}}(\boldsymbol{P}_i,\boldsymbol{X}_j,\boldsymbol{u}_{ij})=\left\|\boldsymbol{u}_{ij}-f^{-1}\left(\frac{\boldsymbol{P}_i\boldsymbol{X}_j}{[\boldsymbol{P}_i\boldsymbol{X}_j]_3},\boldsymbol{K}_i\right)\right\| \tag{4}$$

其中 $\boldsymbol{P}_i=\boldsymbol{R}_i[I|-\boldsymbol{C}_i]$ 是第 i 个摄像机的投影矩阵,$\boldsymbol{X}_j=(X,Y,Z,1)$ 是第 j 个 3D 点的齐次坐标. $\boldsymbol{K}_i$ 为第 i 个摄像机的内参,f 为该摄像机的映射函数,$\boldsymbol{u}_{ij}=(u_{ij},v_{ij})$ 为观测值.

而在射线模型下,该距离度量为

$$d_{\text{ray}}(\boldsymbol{P}_i,\boldsymbol{X}_j,\boldsymbol{u}_{ij})=\angle\left(f(\boldsymbol{u}_{ij},\boldsymbol{K}_i),\frac{\boldsymbol{P}_i\boldsymbol{X}_j}{\|\boldsymbol{P}_i\boldsymbol{X}_j\|}\right) \tag{5}$$

其中 $\angle(\cdot)$ 为射线间的夹角. 其具有非线性的特点,优化难度大. 在误差较小的情况下,可以用弦距离来近似,降低了优化复杂度. 如

$$d_{\text{ch}}(\boldsymbol{P}_i,\boldsymbol{X}_j,\boldsymbol{x}_{ij})=\left\|f(\boldsymbol{u}_{ij},\boldsymbol{K}_i)-\frac{\boldsymbol{P}_i\boldsymbol{X}_j}{\|\boldsymbol{P}_i\boldsymbol{X}_j\|}\right\|=\left\|\boldsymbol{x}_{ij}-\frac{\boldsymbol{P}_i\boldsymbol{X}_j}{\|\boldsymbol{P}_i\boldsymbol{X}_j\|}\right\| \tag{6}$$

式中 $f(\boldsymbol{u}_{ij},\boldsymbol{K}_i)$ 为摄像机 i 对 3D 点 j 的观测射线. 为描述方便,以 $\boldsymbol{x}_{ij}$ 来表示.

在本征矩阵的估计中,会涉及到 Sampson 距离,它是几何距离的一阶近似. 同样,将其扩展到射线意义下

$$d_{\text{sampson}}(\boldsymbol{E},\boldsymbol{x}_0,\boldsymbol{x}_1)=\frac{(\boldsymbol{x}_1\boldsymbol{E}\boldsymbol{x}_0)^2}{\|\boldsymbol{E}\boldsymbol{x}_0\|^2+\|\boldsymbol{E}^{\mathrm{T}}\boldsymbol{x}_1\|^2} \tag{7}$$

其中 $\boldsymbol{x}_0,\boldsymbol{x}_1$ 为给定的两图像对同一 3D 点的观测射线. $\boldsymbol{E}$ 是本征矩阵,类似于基础矩阵,描述了两摄像机之间的相对位姿,关系为 $\boldsymbol{E}=[\boldsymbol{t}]_\times\boldsymbol{R}$,通过分解可以得到两摄像机的相对位姿[9].

对于重构的摄像机集合和 3D 点集合,其非线性优化目标函数为 $\min\limits_{\boldsymbol{P},\boldsymbol{X}}\sum_{i,j}d_{\text{ch}}(\boldsymbol{P}_i,\boldsymbol{X}_j,\boldsymbol{u}_{ij})^2$ 为非线性最小二乘. 非线性来自于式(6)中的 $\frac{\boldsymbol{P}_i\boldsymbol{X}_j}{\|\boldsymbol{P}_i\boldsymbol{X}_j\|}$. 本文使用 LM(Levenberg-Marquardt)算法[10]求取最优解. LM 算法是使用广泛的非线性最小二乘算法,它可以被视为一种梯度下降法和牛顿法相结合的迭代求解方法. 当接近极值点时,步长等于牛顿法步长以获得更快收敛速度,当远离极值点时,步长约等于梯度下降法的步长以保证收敛. 由于非线性最小二乘问题通常不能保证单个极小点,优化结果严重依赖于初值. 所以本文采用增量式结合去噪的方法不断为其提供有效的初值,如算法 1.

2.2　生成位姿图

当采集到足够的室内场景的全景图像后,首先提取图像 SIFT[11]特征并进行图像间的特征匹配. 然后基于匹配的特征点结合 RANSAC[12]鲁棒估计框架估计两两摄像机间的本征矩阵(Essential Matrix)并分解得到两两相对位姿,由于 RANSAC 框架的随机采样模型并计算其余样本对该模型支持度的特性,噪声匹配被过滤掉. 将这些过滤后的内点匹配组织起来,形成多条轨迹,每条轨迹对应于将被重构的一个 3D 点.

随后一个无向的位姿图(Pose Graph)将被建立(如图 4 所示),目的是将前面得到的多种信息进行有效组织,为三维重构提供输入. 位姿图定义为 $G=(NP,NX,EP,EX)$,其中 NP 为摄像机节点,NX 为 3D 点节点,EP 为摄像机-摄像机连接边,边上附有摄像机 i 和 k 之间的相对位置位姿属性,包括相对旋转 $\boldsymbol{R}_{ik}$ 和相对平移方向 $\boldsymbol{C}_{ik}$. EX 为摄像机-3D 点连接边,边上附有该摄像机观测到的特征点射线坐标 $\boldsymbol{x}_{ij}$. 根据位姿图可以定义可视性函数 visX($\boldsymbol{X}_j,\boldsymbol{Ps}$) 和 visP($\boldsymbol{P}_i,\boldsymbol{Xs}$). 其中 visX($\boldsymbol{X}_j,\boldsymbol{Ps}$) = $\{i:(i,j)\in EX,i\in\boldsymbol{Ps}\}$ 意为给定 3D 点 $\boldsymbol{X}_j$ 和摄像机集合 $\boldsymbol{Ps}$ 的条件下,返回 $\boldsymbol{Ps}$ 中观测到 $\boldsymbol{X}_j$ 的摄像机集合. visP($\boldsymbol{P}_i,\boldsymbol{Xs}$) = $\{j:(i,j)\in EX,i\in\boldsymbol{Xs}\}$ 意为给定 3D 点集合 $\boldsymbol{Xs}$ 和摄像机 $\boldsymbol{P}_i$ 的条件下,返回 $\boldsymbol{Xs}$ 中被 $\boldsymbol{P}_i$ 观测到的 3D 点集合. 他们共同用于描述摄像机集合与 3D 点集合之间的可视性关系.

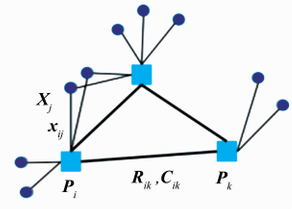


图4 用于三维重建的位姿图

2.3　射线模型下的增量式三维重构算法

根据输入的位姿图 $G=(NP,NX,EP,EX)$,本文提出了一种基于射线模型的增长式三维重构算法,用于

同时恢复场景 3D 点云和摄像机的位置. 由于引入了射线模型,这些摄像机模型可以是全景,鱼眼或者平面. 具体步骤如算法 1 所示. 首先选择相对位姿估计质量较高的一对摄像机作为初始种子,相对位姿质量高指的是两摄像机之间的特征匹配较多且相对位姿差异较大以保证最佳的初始重构 3D 点的数量和质量. 然后利用三角测量和可视化函数寻找新的 3D 点,再次利用新的 3D 点寻找更多的摄像机,不断迭代,直至寻找不到更多的摄像机或者 3D 点. 算法期间不断实施非线性优化,用于减小三维重构的误差. 同时使用质量评价函数剔除质量不高的摄像机和 3D 点. 算法中的距离度量、三角测量、摄像机位姿估计,非线性优化、质量评价函数等模块都是针对射线模型所改进的,相比传统的仅适用平面图像的重构算法[7,8],具有更广泛的普适性.

算法 1 基于射线模型增长式三维重构算法

输入:位姿图 $G=(NP,NX,EP,EX)$

输出:3D 点云 $\boldsymbol{Xs}=\{\boldsymbol{X}_1,\boldsymbol{X}_2,\cdots,\boldsymbol{X}_n\}$ 和摄像机位姿 $\boldsymbol{Ps}=\{\boldsymbol{P}_1,\boldsymbol{P}_2,\cdots,\boldsymbol{P}_m\}$

Define:$\boldsymbol{Ps}$ 为当前已重构得到位姿的摄像机集合,$\boldsymbol{Xs}$ 为当前已重构得到坐标的 3D 点集合. 新增的 3D 点数量为 n_X,新增的摄像机数量为 n_P.

Initialization:从位姿图中选择相对位姿估计质量较高的图像对对作为初始种子 $\boldsymbol{Ps}=\{\boldsymbol{P}_i,\boldsymbol{P}_k\}$. $\boldsymbol{Xs}=\{\}$,$n_X=0$,$n_P=|\boldsymbol{Ps}|$.

Loop

选择待重构的点集合. 其中每个点 $\boldsymbol{X}_j$ 被不少于 T_X 个已知位姿的摄像机所观测到,即$\{\boldsymbol{X}_j:\ |\text{visX}(\boldsymbol{X}_j,\boldsymbol{Ps})\ |>T_X,\boldsymbol{X}_j\notin\boldsymbol{Xs}\}$.

采用基于射线模型的三角测量方法估计候选集中的每一个 3D 点得到其 3D 坐标. 将质量较好的点集添加到当前重构集合 $\boldsymbol{Xs}$ 中. 即 $\boldsymbol{Xs}\leftarrow\boldsymbol{Xs}\cup\boldsymbol{Xs}_{\text{new}}$,$n_X\leftarrow|\boldsymbol{Xs}_{\text{new}}|$.

if $n_X>0$ then

非线性优化射线模型下的弦误差.

elseif $n_X=0$ and $n_P==0$ then

break

end

选择待重构的摄像机集合. 其中每个摄像机 $\boldsymbol{P}_i$ 观测到不少于 T_P 个的已知坐标的 3D 点. 即$\{\boldsymbol{P}_i:\ |\text{visP}(\boldsymbol{P}_i,\boldsymbol{Xs})\ |>T_P,\boldsymbol{P}_i\notin\boldsymbol{Ps}\}$.

利用射线模型估计每一个待重构的摄像机位姿. 将质量较好的摄像机集合添加到当前重构集合 $\boldsymbol{Ps}$ 中,即 $\boldsymbol{Ps}\leftarrow\boldsymbol{Ps}\cup\boldsymbol{Ps}_{\text{new}}$,$n_P\leftarrow|\boldsymbol{Ps}_{\text{new}}|$.

if $n_P>0$ then

非线性优化射线模型下的弦误差.

去除优化后质量较差的摄像机和 3D 点.

elseif $n_X=0$ and $n_P==0$ then

break

end

end loop

估计得到的 3D 点的的质量评价指标定义如下,利用得到的 3D 点坐标与参与重构的摄像机光心的连线间的最大夹角作为该 3D 点的质量评价指标. 角度小于一定阈值认为质量不良,该 3D 点被认为是噪声并舍弃. 同理,摄像机的质量评价指标用最大立体角定义. 含义是以摄像机光心为中心的单位球面上三个点所对应的球冠所占球面的比例,这些点是由摄像机光心和参与估计的 3D 点间的连线与该球面相交所产生的,选择其中的三个构成的最大立体角作为摄像机的质量评价指标. 质量低于一定阈值的摄像机将被舍弃.

三维重构完成后,三维点云中的每个点都附带若干特征,这些特征来自于观测到该 3D 点的图像. 在后续线定位环节中,需要建立查询图像的特征与该点云特征的匹配,以实现图像定位. 为了加速匹配过程,本文对点云特征建立 Kd-tree 索引树,以加快检索速度. 此外,由于在线定位环节需要检索查询图像的空间近邻,所以本文又对重构出的摄像机建立空间位置的 Kd-tree 索引树用来加速在线检索.

3 在线图像定位

在线图像定位的一般流程是:首先对查询图像提取特征,而后将这些特征与离线部分生成的 3D 点云的特征进行特征匹配(2D-3D 匹配),最后根据这些足够数量的 2D-3D 匹配,利用摄像机位姿估计算法[12,13]估计查询图像的位姿. 为得到精度更高的定位结果,本文在传统的 2D-3D 匹配[5]的基础上,增加使用查询图像与近邻图像间的相对位置关系,并利用图优化的方法整合这两种信息得到精度更高的查询图像的位姿. 在线定位算法的流程是:

① 实施查询图像与 3D 点云的特征匹配. 利用这些 2D-3D 匹配结果初步估计查询图像的空间位置.

② 获取与查询图像空间位置相近的重构库图像. 利用查询图像与近邻库图像进行特征匹配,并估计其相对位姿.

③ 使用图优化的方法融合前两步得到的 2D-3D 匹配和近邻图像的相对位姿以提升查询图像的位姿精度.

3.1 基于 2D-3D 匹配的初始定位

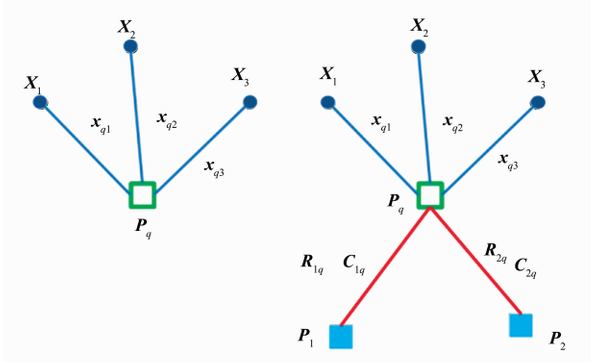
输入查询图像后,首先提取其图像特征并与 3D 点云进行特征匹配. 匹配方法类似于 2D 图像特征匹配的采用交叉验证近邻与次近邻的距离比. 当双向近邻与次近邻距离比均小于一定阈值时,即近邻相比次近邻更近时,认为成功得到一对候选匹配. 利用这些候选匹配做摄像机位姿估计[12,13]得到查询图像的初始位姿.

根据查询图像的初始位姿,接下来需要进一步估计其与近邻库图像间相对位姿,为将来做精确定位做准备. 通过空间位姿检索得到查询图像近邻的库图像,随后与每个近邻图像进行 2D-2D 的特征匹配,本征矩阵并分解得到查询图像与每个近邻库图像的相对位姿.

3.2　构建定位位姿图及优化

综合得到的2D-3D匹配以及与近邻库图像的相对位姿,建立关于查询图像的定位位姿图(如图5所示),描述如下:

定义关于查询图像 q 的位姿图 $G_q=(NP,NX,EP,EX)$,其中 NP 为摄像机节点,包含查询图像的摄像机 $\boldsymbol{P}_q$ 和其近邻摄像机 $\{\boldsymbol{P}_i,i=1,\cdots,k\}$,$NX$ 为3D点节点,对应于2D-3D匹配中的3D点. EP 为查询摄像机 $\boldsymbol{P}_q$ 与近邻摄像机 $\{\boldsymbol{P}_i,i=1,\cdots,k\}$ 的连接边,边上附有 i 和 q 之间的相对位姿,包括相对旋转 $\boldsymbol{R}_{iq}$ 和相对平移方向 $\boldsymbol{C}_{iq}$. EX 为查询摄像机 $\boldsymbol{P}_q$ 与3D点 $\boldsymbol{X}_j$ 的连接边,边上附有查询摄像机 $\boldsymbol{P}_q$ 上关于该3D的观测射线坐标 $\boldsymbol{x}_{qi}$.



(a) 2D-3D匹配　　(b) 融合近邻摄像机相对位姿

图5 定位的位姿图

基于该图构建优化目标函数

$$\min_{P_q}\frac{1}{n}\sum_{j=1,\cdots,n}d_{ch}(\boldsymbol{P}_q,\boldsymbol{X}_j,\boldsymbol{x}_{qj})+\frac{\lambda}{m}\sum_{i=1,\cdots,m}d_{\mathrm{rel}}(\boldsymbol{P}_i,\boldsymbol{P}_q,\boldsymbol{R}_{iq},\boldsymbol{C}_{iq}) \tag{8}$$

其中 $\boldsymbol{P}_q=\boldsymbol{R}_q[\boldsymbol{I}|-\boldsymbol{C}_q]$ 为待优化的查询图像的摄像机矩阵,$\boldsymbol{R}_q,\boldsymbol{C}_q$ 为该摄像机在世界坐标系下的旋转和平移. $\{(\boldsymbol{x}_{qj},\boldsymbol{X}_j),j=1,\cdots,n\}$ 为输入的2D-3D匹配集合. $\{(\boldsymbol{P}_i,\boldsymbol{R}_{iq},\boldsymbol{C}_{iq}),i=1,\cdots,m\}$ 为查询图像的近邻图像以及相应的相对位姿集合. λ 为两类代价的平衡因子. $d_{\mathrm{rel}}()$ 为相对位姿边上的代价函数,其定义如下

$$d_{\mathrm{rel}}(\boldsymbol{P}_i,\boldsymbol{P}_q,\boldsymbol{R}_{iq},\boldsymbol{C}_{iq})=d_R(\boldsymbol{R}_{iq},\boldsymbol{R}_q,\boldsymbol{R}_i^{\mathrm{T}})+\left\|\frac{\boldsymbol{C}_q-\boldsymbol{C}_i}{\|\boldsymbol{C}_q-\boldsymbol{C}_i\|}-\boldsymbol{R}_i\boldsymbol{C}_{iq}\right\| \tag{9}$$

其中相对位姿的代价函数包含两项,分别为旋转的代价和平移方向的代价,二者相互独立. 旋转的代价定义为理论旋转 $\boldsymbol{R}_q\boldsymbol{R}_i^{\mathrm{T}}$ 与观测旋转 $\boldsymbol{R}_{iq}$ 之间的相对欧拉角 $d_R(\boldsymbol{R}_1,\boldsymbol{R}_2)=\arccos\left(\frac{\mathrm{trace}(\boldsymbol{R}_1\boldsymbol{R}_2^{\mathrm{T}})-1}{2}\right)$. 平移方向的代价为观测出的平移方向 $\boldsymbol{R}_i\boldsymbol{C}_{iq}$ 与理论平移方向 $\frac{\boldsymbol{C}_q-\boldsymbol{C}_i}{\|\boldsymbol{C}_q-\boldsymbol{C}_i\|}$ 之间的弦距离.

由于式(8)所示目标函数的非线性,求解过程依赖良好的初值. 本文采用初始定位的结果作为初值,通过Levenberg-Marquardt[10]算法对其优化. 相比传统仅使用2D-3D匹配信息的定位方法[5],本文方法使用图优化的方法融合了2D-3D的匹配信息和图像间的相对位姿信息,提高了定位结果的精确度.

4　实验

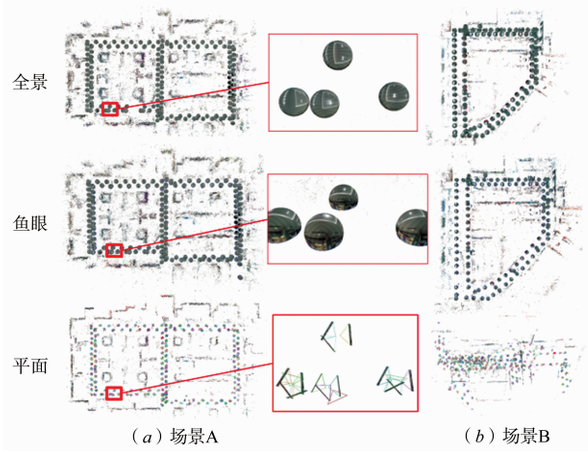
本文实验包含两部分,三维重构和图像定位. 选择A、B两个大型商场作为实验场景. A商场图像纹理丰富,B商场则存在大量的缺乏纹理的墙面和玻璃. 重构数据使用理光Theta全景相机(图1)拍摄的全景和鱼眼图像集合,图像分辨率分别为3584×1792,3259×3259. 鱼眼图像为中心朝上,垂直110度,水平360度. A、B场景对应的全景图像数量分别为197张和130张,分别覆盖约20m×40m和20m×30m的范围. 定位实验使用普通手机拍摄的平面图像,分辨率为1280×720. A、B场景中定位测试图像数量分别为211张和150张. 重构和定位实验图像位置的真实值由人工标定,其误差小于0.05m. 重构图像的采集方式为沿商场通道均匀采样,间隔大致为1~2m. 定位实验图像在不同位置随机朝向采样,间隔5~6m,且不同于重构图像的位置.

4.1　三维重构

此处将基于射线模型的三维重构方法与传统的基于平面图像的重构方法VisualSFM[8]进行比较. 基于射线模型的方法使用全景和鱼眼图像进行重构. 由于VisualSFM无法适应全景图像,我们在水平方向上将每张全景图像向间隔72度的5个方向投影,生成5张平面图像,每张图像的张角为90度,相互之间有一定的视场交叠,使得覆盖更为完整. A,B场景分别得到985张和650张平面图像. 基于这些平面图像使用VisualSFM进行三维重构.

重构效果分析:重构结果主要包括两部分,3D点云和摄像机位置和姿态. 重构效果如图6所示. 场景A中,由于VisualSFM使用可视范围较小平面图像进行重构,对估计摄像机位姿有影响. 实测表明,985张输入图像有888张图像对应的摄像机被估计出来,有97张图像未能估计出来,如图6左下方红框所示,正常应为每组包含5张相同光心的图像,但一些摄像机估计失败. 本文方法则将所有摄像机完整估计出来. VisualSFM和本文方法位置估计偏差分别为0.3m(平面)和0.07m(全景,鱼眼). 对于场景B,由于场景中存在大量缺少纹理的墙面和玻璃,如图7(a)所示,平面图像间特征点数较少,匹配困难,导致VisualSFM重构失败. 但全景和鱼眼图像视角广阔(360度),能够包含更丰富的信息,如

图 7(*b*)、(*c*)所示. 基于射线的重构方法可以将 130 个摄像机全部重构出来. 位置误差为 0.09m.



(*a*) 场景A (*b*) 场景B

图6 三维重构效果在数据集A、B上的对比，本文方法（全景，鱼眼），VisualSFM（平面）

(*a*) 平面图像 (*b*) 全景图像 (*c*) 鱼眼图像

图7 场景B中纹理信息较少的区域

计算量分析:重构过程中,图像匹配和增量式重构是主要的计算开销. 图像匹配的计算量由数据集内部图像两两匹配的次数描述. 重构的计算量由总的重构时间来描述,如表 1 所示.

表 1 重构效果和计算量对比分析

数据集	方法	重构误差(m)	图像匹配次数(1e4)	重构时间(s)
A	VisualSFM[8](平面)	0.3	48.4	690
	本文方法(全景,鱼眼)	0.07	1.93	150
B	VisualSFM[8](平面)	重构失败	21.1	400
	本文方法(全景,鱼眼)	0.09	0.84	80

两个数据集上的重构实验表明,基于射线模型的三维重构算法,由于采用了广角图像(全景,鱼眼),使用了更多的约束,摄像机位置估计更准确,信息更完整,重构效果明显优于使用平面图像的 VisualSFM 重构算法,且采集和计算代价上具有明显优势.

4.2 图像定位

图像定位是指根据输入图像估计摄像机位置(如图 8 所示). 本文在 2D-3D 匹配算法基础上添加位姿图优化方法实施摄像机位姿估计,对比的方法是基于图像检索的定位[3]和 2D-3D 匹配[5],误差统计结果如图 9 所示. 定位误差大于 5m 视为定位失败. 场景 A 和 B 中的定位测试图像数量分别为 211 张和 150 张,使用图像检索方法定位成功的图像数量为 180 和 145 张,2D-3D 匹配定位方法和本文方法定位成功的图像数量均为 186 和 148 张. 实验结果中可以看到,本文提出的定位方法平均定位误差(0.35m,0.21m)比图像检索的方法精度(1.19m,1.48m),2D-3D 匹配的方法精度(0.48m,0.42m)提高了至少 20% 以上.

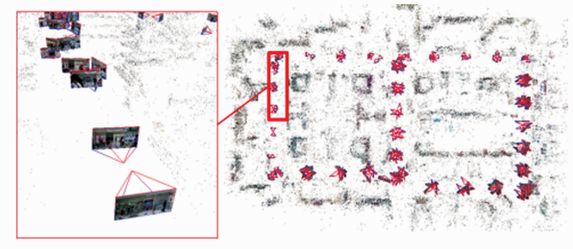


图8 本文方法对于场景A的定位结果示意图（186张查询图像）

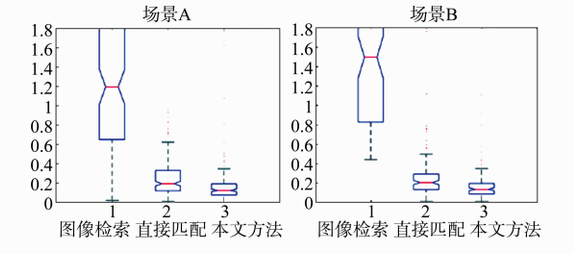


图9 场景A、B的定位误差boxplot图

5 总结与展望

图像定位具有广泛的应用前景. 为充分利用广角摄像机的优势,本文提出了一种基于射线模型三维重构的图像定位系统. 主要贡献包括:(1)提出的射线模型下的三维重构算法,能够适用于多种类型的摄像机(全景、鱼眼、平面),并充分利用其内在的几何性质,相比传统仅适用平面重构的方法,该方法重构效果好,采集成本低,计算速度快;(2)在图像定位方面,提出的基于位姿图优化的定位框架,融合了图像-点云匹配信息和图像间相对位姿信息,提高了图像定位的精度. 实验结果证明了本文方法的有效性. 未来我们将尝试在更具有挑战性的场景中实现定位,如特征较少的走廊、前景干扰较大的车站等.

参考文献

[1] 左敏,曾广平,涂序彦. 无人变电站智能机器人的视觉导航研究[J]. 电子学报,2012,39(10):2464－2468.
Zuo Min,Zeng Guang-ping,Tu Xu-yuan. Research on visual navigation of untended substation patrol robot[J]. Acta

Electronica Sinica, 2012, 39(10): 2464 – 2468. (in Chinese)

[2] 郝宗波,洪炳熔. 未知环境下基于传感器的移动机器人路径规划[J]. 电子学报,2006,34(5):953 – 956.
Hao Zong-bo, Hong Bing-rong. Sensor-based path planning for mobile robot in unknown environment[J]. Acta Electronica Sinica, 2006, 34(5): 953 – 956. (in Chinese)

[3] Schindler G, Brown M, Szeliski R. City-scale location recognition[A]. Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition[C]. Minneapolis, USA: IEEE Press, 2007. 1 – 7.

[4] Zamir A, Shah M. Accurate image localization based on google maps street view[A]. Lecture Notes in Computer Science[C]. Berlin Heidelberg: GER, 2010. 255 – 268.

[5] Sattler T, Leibe B, Kobbelt L. Fast image-based localization using direct 2D-to-3D matching[A]. Proceedings of IEEE International Conference on Computer Vision[C]. Barcelona, ESP: IEEE Press, 2011. 667 – 674.

[6] Li Yunpeng, Snavely N, Huttenlocher D, et al. Worldwide pose estimation using 3D point clouds[A]. Lecture Notes in Computer Science[C]. Berlin Heidelberg: GER, 2012. 15 – 29.

[7] Snavely N, Seitz S M, Szeliski R. Photo tourism: exploring photo collections in 3D[J]. ACM Transactions on Graphics, 2006, 25(3): 835 – 846.

[8] Wu ChangChang. Towards linear-time incremental structure from motion[A]. International Conference on 3D Vision [C]. Seattle, Washington, USA, 2013. 127 – 134.

[9] Hartley R, Zisserman A. Multiple View Geometry in Computer Vision[M]. England: Cambridge University Press, 2004. 1187 – 1865.

[10] Bill Triggs, Philip Mclauchlan, Richard Hartley, Andrew Fitzgibbon. Bundle adjustment-a modern synthesis[A]. Vision Algorithms: Theory and Practice, LNCS[C]. Berlin Heidelberg: Springer Verlag 2000. 153 – 177.

[11] Lowe D. Distinctive image features from scale-invariant keypoints[J]. International Journal of Computer Vision, 2004, 60(2): 91 – 110.

[12] Fisher M A, Bolles R C. Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography[J]. Communications of the ACM, 1981, 24(6): 726 – 740.

[13] Bujnak M, Kukelova Z, Pajdla T. A general solution to the P4P problem for camera with unknown focal length[A]. Proceedings of IEEE Conference on Computer Vision and Pattern Recognition[C]. Anchorage, USA: IEEE Press, 2008. 1 – 8.

[14] Kukelova Z, Bujnak M, Pajdla T. Real-time solution to the absolute pose problem with unknown radial distortion and focal length[A]. Proceedings of IEEE International Conference on Computer Vision[C]. Sydney, USA: IEEE Press, 2013. 2816 – 2823.

作者简介

邓　磊　男,1988 年 2 月生于山西侯马. 现为清华大学自动化系博士研究生. 主要研究方向为计算机视觉及模式识别.
E-mail: dengl09@ mails. tsinghua. edu. cn

陈宝华　男,1978 年 2 月出生于内蒙古兴安盟. 现为清华大学自动化系博士研究生. 主要研究方向为计算机视觉及模式识别.
E-mail: cbh10@ mails. tsinghua. edu. cn