



(12) 发明专利

(10) 授权公告号 CN 111462135 B

(45) 授权公告日 2023.04.21

(21) 申请号 202010246158.2

G06T 7/50 (2017.01)

(22) 申请日 2020.03.31

G06T 7/80 (2017.01)

G06T 7/90 (2017.01)

(65) 同一申请的已公布的文献号

申请公布号 CN 111462135 A

(56) 对比文件

CN 110097553 A, 2019.08.06

WO 2019169540 A1, 2019.09.12

(43) 申请公布日 2020.07.28

(73) 专利权人 华东理工大学

地址 200237 上海市徐汇区梅陇路130号

卞贤掌等. 基于语义分割的增强现实图像配准技术.《电子技术与软件工程》.2018, (第23期),

(72) 发明人 唐漾 钱锋 杜文莉 堵威

王廷银等. 基于北斗RDSS的核辐射监测应急通讯方法.《计算机系统应用》.2019, (第12期),

(74) 专利代理机构 上海专利商标事务所有限公司 31100

专利代理师 陆嘉

审查员 来文燕

(51) Int. Cl.

G06T 7/11 (2017.01)

G06N 3/0464 (2023.01)

G06N 3/08 (2023.01)

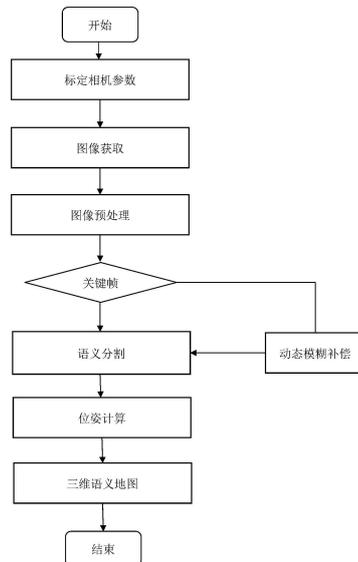
权利要求书3页 说明书11页 附图3页

(54) 发明名称

基于视觉SLAM与二维语义分割的语义建图方法

(57) 摘要

本发明涉及计算机视觉与深度学习的交叉融合领域,更具体的说,涉及一种基于视觉SLAM与二维语义分割的语义建图方法。本发明的方法,包括以下步骤:S1、标定相机参数,校正相机畸变;S2、获取图像帧序列;S3、图像预处理;S4、判断当前图像帧是否为关键帧,如果是,则转入步骤S6,如果不是,则转入步骤S5;S5、动态模糊补偿;S6、语义分割,针对图像帧进行ORB特征点的提取,利用掩膜区域卷积神经网络算法模型进行语义分割;S7、位姿计算,利用稀疏SLAM算法模型计算相机位姿;S8、将语义信息辅助稠密语义地图构建,实现全局点云地图的三维语义建图。本发明可以提升无人机语义建图系统性能,显著提升针对动态场景进行特征点的提取与匹配的鲁棒性。



1. 一种基于视觉SLAM与二维语义分割的语义建图方法,其特征在于,包括以下步骤:

S1、标定相机参数,校正相机畸变;

S2、获取图像帧序列,图像帧序列包括RGB图像和深度图像;

S3、图像预处理,采用针孔相机模型,得到RGB图像中每个像素点对应的真实三维空间点的坐标;

S4、判断当前图像帧是否为关键帧,如果是,则转入步骤S6,如果不是,则转入步骤S5;

S5、动态模糊补偿,计算得到当前图像帧的图像块质心作为语义特征点,作为ORB特征点的补充;

S6、语义分割,针对图像帧进行ORB特征点的提取,利用掩膜区域卷积神经网络算法模型进行语义分割,获取该帧图像每一个像素点的语义信息;

S7、位姿计算,利用稀疏SLAM算法模型计算相机位姿;

S8、将语义信息输入到稀疏SLAM算法模型,辅助稠密语义地图构建,完成关键帧的遍历,实现全局点云地图的三维语义建图。

2. 根据权利要求1所述的基于视觉SLAM与二维语义分割的语义建图方法,其特征在于,所述步骤S1中的校正相机畸变,进一步包括以下步骤:

S11、将相机坐标系的三维空间点 $P(X, Y, Z)$ ,投影到归一化图像平面形成该点的归一化坐标为 $[x, y]^T$ ;

S12、对归一化平面上的点 $[x, y]^T$ 进行径向畸变和切向畸变校正,通过以下公式实现:

$$\begin{cases} x_{\text{corrected}} = x(1 + k_1 r^2 + k_2 r^4 + k_3 r^6) + 2p_1 xy + p_2(r^2 + 2x^2) \\ y_{\text{corrected}} = y(1 + k_1 r^2 + k_2 r^4 + k_3 r^6) + p_1(r^2 + 2y^2) + 2p_2 xy \end{cases}$$

其中, $[x_{\text{corrected}}, y_{\text{corrected}}]^T$ 是校正后的点坐标, $p_1, p_2$ 为相机的切向畸变系数, $k_1, k_2, k_3$ 为相机的径向畸变系数, $r$ 为点 $P$ 离坐标系原点的距离;

S13、将校正后的点 $[x_{\text{corrected}}, y_{\text{corrected}}]^T$ 通过内参数矩阵,投影到像素平面得到其在图像上的正确位置 $[u, v]^T$ ,通过以下公式实现:

$$\begin{cases} u = f_x x_{\text{corrected}} + c_x \\ v = f_y y_{\text{corrected}} + c_y \end{cases}$$

其中, $f_x, f_y, c_x, c_y$ 为相机的内参数。

3. 根据权利要求2所述的基于视觉SLAM与二维语义分割的语义建图方法,其特征在于,所述步骤S3的图像预处理,进一步包括,像素点 $[u, v]^T$ 到真实三维空间点 $P(X, Y, Z)$ 的映射关系满足以下公式:

$$Z \begin{pmatrix} u \\ v \\ 1 \end{pmatrix} = \begin{pmatrix} f_x & 0 & c_x \\ 0 & f_y & c_y \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} X \\ Y \\ Z \end{pmatrix} = KP$$

其中, $K$ 称为相机内参数矩阵, $f_x, f_y, c_x, c_y$ 为相机的内参数, $P$ 为真实三维空间点坐标, $[u, v]^T$ 为像素点坐标。

4. 根据权利要求1所述的基于视觉SLAM与二维语义分割的语义建图方法,其特征在于,所述步骤S4的关键帧,使用稀疏SLAM算法模型进行筛选。

5. 根据权利要求1所述的基于视觉SLAM与二维语义分割的语义建图方法,其特征在于,所述步骤S5的图像块质心,通过以下步骤得到:

将该帧图像每一个物体标注为一个具体的类;

对于每一个分割出来的对象有对应标注区域,分割出来的图像称为图像块;

计算图像块的矩  $m_{pq} = \sum_{x,y \in B} x^p y^q I(x,y)$ ,  $p, q = \{0, 1\}$ ;

计算对应的质心C作为语义特征点,对ORB特征点进行补充,其中

$$C = \left( \frac{m_{10}}{m_{00}}, \frac{m_{01}}{m_{00}} \right)$$

6. 根据权利要求1所述的基于视觉SLAM与二维语义分割的语义建图方法,其特征在于,所述步骤S6,进一步包括:

每一个像素点的语义信息包括语义分类标签、包围框坐标以及该分类的置信分数;

基于语义分割结果,对于指定某种类别为动态物体的区域所提取的ORB特征点进行剔除。

7. 根据权利要求1所述的基于视觉SLAM与二维语义分割的语义建图方法,其特征在于,所述步骤S6的Mask R-CNN算法模型进行语义分割,进一步包括:

通过特征图金字塔网络提取输入图像不同层次上的特征;

通过区域生成网络提出感兴趣提案;

利用感兴趣区域排列进行提案区域对齐;

利用全卷积网络进行掩膜分割;

利用全连接层进行区域坐标确定和所述类别分类。

8. 根据权利要求1或权利要求4所述的基于视觉SLAM与二维语义分割的语义建图方法,其特征在于,所述稀疏SLAM算法模型,进一步包括,跟踪线程、局部建图线程、回环检测线程:

所述跟踪线程,通过寻找对局部地图特征进行匹配,利用纯运动光束平差法最小化重投影误差进行定位每帧图片的相机;

所述局部建图线程,通过执行局部光束平差法管理局部地图并优化,通过地图点维护关键帧之间的共视关系,通过局部光束平差法优化共视关键帧位姿和地图点;

所述回环检测线程,检测大的环并通过执行位姿图优化更正漂移误差,加速闭环匹配帧的筛选,并优化尺度,通过全局光束平差法优化本质图和地图点。

9. 根据权利要求1所述的基于视觉SLAM与二维语义分割的语义建图方法,其特征在于,所述稀疏SLAM算法模型,进一步包括全局光束平差法优化线程,在回环检测线程确认后触发,在位姿图优化之后,计算整个系统最优结构和运动结果。

10. 根据权利要求1所述的基于视觉SLAM与二维语义分割的语义建图方法,其特征在于,所述步骤S7的位姿计算,进一步包括:通过PnP求解初步计算相机位姿,利用后端位姿图优化计算相机位姿,构建相机位姿估计的最小化重投影误差:

$$\xi^* = \arg \min_{\xi} \frac{1}{2} \sum_{i=1}^n \left\| u_i - \frac{1}{s_i} K \exp(\xi^{\wedge}) P_i \right\|_2^2$$

其中,  $u_i$  为像素坐标,  $P_i$  为相机坐标,  $\hat{\xi}$  为相机位姿对应的李代数,  $s_i$  为特征点深度,  $K$  为相机内参数矩阵。

## 基于视觉SLAM与二维语义分割的语义建图方法

### 技术领域

[0001] 本发明涉及计算机视觉与深度学习的交叉融合领域,更具体的说,涉及一种基于视觉SLAM与二维语义分割的语义建图方法。

### 背景技术

[0002] 无人机一般由智能决策、环境感知、运动控制三个模块构成,其中环境感知是一切的基础。

[0003] 无人机感知周围环境,需要一套稳定、性能强大的传感器系统来充当“眼睛”,同时需要相应的算法和强有力的处理单元“读懂物体”。

[0004] 无人机的环境感知模块中,视觉传感器是不可或缺的一部分,视觉传感器可以是摄像头,相较于激光雷达、毫米波雷达,摄像头的分辨率更高,能获取足够的环境细节,例如可以描述物体的外观和形状、读取标识等。

[0005] 尽管全球定位系统(Global Positioning System,GPS)有助于定位过程,但是由于高大树木、建筑、隧道等造成的干扰会使得GPS定位不可靠,因此视觉传感器不能被GPS系统所取代。

[0006] 定位与建图(Simultaneous Localization and Mapping,SLAM)是指载有特定传感器的主体在没有先验信息的情况下,通过计算特定传感器获取图像帧来估计自身运动的轨迹,并建立周围环境的地图,其广泛应用于机器人、无人机、自动驾驶、增强现实、虚拟现实等应用中。

[0007] SLAM可以划分为激光SLAM和视觉SLAM两类。

[0008] 由于起步早,激光SLAM在理论技术和工程应用上都较为成熟,但是激光SLAM在机器人的应用上有一个致命的缺点,就是激光雷达智能感知的结构信息是二维信息,信息量较少,造成丢失了大量的环境信息。同时其高昂的成本、庞大的体积以及缺少语义信息使其在一些特定的应用场景中受限。

[0009] 视觉SLAM的感知信息源为相机图像。

[0010] 根据相机类型,可将视觉SLAM分为三种:单目、双目以及深度SLAM。类似于激光雷达,深度相机可以通过采集点云来直接计算到障碍物的距离。深度相机结构简单,易于安装操作,而且成本低、使用场景广泛。

[0011] 随着深度学习的兴起,视觉SLAM在近几年也取得了长足的进步。

[0012] 大部分的视觉SLAM方案都是特征点或像素级别,为了完成一个特定的任务,或者与周围环境进行智能化的交互,无人机需要获取语义信息。

[0013] 视觉SLAM系统能够选择有用信息,剔除无效信息。

[0014] 随着深度学习的发展,许多成熟的目标检测和语义分割的方法为精确的语义建图提供了条件。语义地图有利于提高无人机的自主性和鲁棒性,完成更复杂的任务,从路径规划转化为任务规划。

[0015] 随着硬件计算能力的提高,以及算法结构的优化,深度学习取得了越来越瞩目的

成就。

[0016] 在计算机视觉领域取得了巨大的飞跃,就RGB图像分割来看,可以大体分为目标检测和语义分割。

[0017] 在前期主要是目标检测框架的提出,实现了越来越精准的目标检测。

[0018] 主流的深度学习目标检测框架主要是基于CNN(Convolutional Neural Networks,卷积神经网络)的,其中较为高效的有YOLO(You Only Look Once,你只用看一次)系列和R-CNN(Region-CNN,区域卷积神经网络)系列。

[0019] 三维图像中的目标感知技术越来越成熟,三维理解的需求也越来越紧迫。由于点云的不规则性,大多数研究者会将点转化为规则的体素或者网格模型,利用深度神经网络进行预测。

[0020] 直接对点云空间进行语义分割需要消耗极大的计算资源,空间点之间的相互关系被削弱。

[0021] 2017年提出的PointNet(点网)是第一个可以直接处理原始三维点云的深度神经网络。

[0022] 现有大部分视觉SLAM系统采用的稠密建图方法,缺少语义信息,无法完成智能化的需求。

[0023] 视觉SLAM算法有一个典型假设是场景固定,其中一些动态物体的出现不仅影响相机位姿的估计而且在地图中留下残影,影响地图质量。

[0024] 相机在高速运动情况下所捕捉的照片容易模糊,极大的影响了特征点的提取与匹配。

## 发明内容

[0025] 本发明的目的是提供一种基于视觉SLAM与二维语义分割的语义建图方法,解决高速运动的动态物体影响建立地图质量的技术问题。

[0026] 为了实现上述目的,本发明提供了一种基于视觉SLAM与二维语义分割的语义建图方法,包括以下步骤:

[0027] S1、标定相机参数,校正相机畸变;

[0028] S2、获取图像帧序列,图像帧序列包括RGB图像和深度图像;

[0029] S3、图像预处理,采用针孔相机模型,得到RGB图像中每个像素点对应的真实三维空间点的坐标;

[0030] S4、判断当前图像帧是否为关键帧,如果是,则转入步骤S6,如果不是,则转入步骤S5;

[0031] S5、动态模糊补偿,计算得到当前图像帧的图像块质心作为语义特征点,作为ORB特征点的补充;

[0032] S6、语义分割,针对图像帧进行ORB特征点的提取,利用掩膜区域卷积神经网络算法模型进行语义分割,获取该帧图像每一个像素点的语义信息;

[0033] S7、位姿计算,利用稀疏SLAM算法模型计算相机位姿;

[0034] S8、将语义信息输入到稀疏SLAM算法模型,辅助稠密语义地图构建,完成关键帧的遍历,实现全局点云地图的三维语义建图。

[0035] 在一实施例中,所述步骤S1中的校正相机畸变,进一步包括以下步骤:

[0036] S11、将相机坐标系的三维空间点 $P(X, Y, Z)$ ,投影到归一化图像平面形成该点的归一化坐标为 $[x, y]^T$ ;

[0037] S12、对归一化平面上的点 $[x, y]^T$ 进行径向畸变和切向畸变校正,通过以下公式实现:

$$[0038] \begin{cases} x_{\text{corrected}} = x(1 + k_1 r^2 + k_2 r^4 + k_3 r^6) + 2p_1 xy + p_2(r^2 + 2x^2) \\ y_{\text{corrected}} = y(1 + k_1 r^2 + k_2 r^4 + k_3 r^6) + p_1(r^2 + 2y^2) + 2p_2 xy \end{cases}$$

[0039] 其中, $[x_{\text{corrected}}, y_{\text{corrected}}]^T$ 是校正后的点坐标, $p_1, p_2$ 为相机的切向畸变系数, $k_1, k_2, k_3$ 为相机的径向畸变系数, $r$ 为点 $P$ 离坐标系原点的距离;

[0040] S13、将校正后的点 $[x_{\text{corrected}}, y_{\text{corrected}}]^T$ 通过内参数矩阵,投影到像素平面得到其在图像上的正确位置 $[u, v]^T$ ,通过以下公式实现:

$$[0041] \begin{cases} u = f_x x_{\text{corrected}} + c_x \\ v = f_y y_{\text{corrected}} + c_y \end{cases}$$

[0042] 其中, $f_x, f_y, c_x, c_y$ 为相机的内参数。

[0043] 在一实施例中,所述步骤S3的图像预处理,进一步包括,像素点 $[u, v]^T$ 到真实三维空间点 $P(X, Y, Z)$ 的映射关系满足以下公式:

$$[0044] Z \begin{pmatrix} u \\ v \\ 1 \end{pmatrix} = \begin{pmatrix} f_x & 0 & c_x \\ 0 & f_y & c_y \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} X \\ Y \\ Z \end{pmatrix} = KP$$

[0045] 其中, $K$ 称为内参数矩阵, $f_x, f_y, c_x, c_y$ 为相机的内参数, $P$ 为真实三维空间点坐标, $[u, v]^T$ 为像素点坐标。

[0046] 在一实施例中,所述步骤S4的关键帧,使用稀疏SLAM算法模型进行筛选。

[0047] 在一实施例中,所述步骤S5的图像块质心,通过以下步骤得到:

[0048] 将该帧图像每一个物体标注为一个具体的类;

[0049] 对于每一个分割出来的对象有对应标注区域,分割出来的图像称为图像块;

[0050] 计算图像块的矩 $m_{pq} = \sum_{x, y \in B} x^p y^q I(x, y)$ ,  $p, q = \{0, 1\}$ ;

[0051] 计算对应的质心 $C$ 作为语义特征点,对ORB特征点进行补充,其中

$$[0052] C = \begin{pmatrix} \frac{m_{10}}{m_{00}} & \frac{m_{01}}{m_{00}} \end{pmatrix}。$$

[0053] 在一实施例中,所述步骤S6,进一步包括:

[0054] 每一个像素点的语义信息包括语义分类标签、包围框坐标以及该分类的置信分数;

[0055] 基于语义分割结果,对于指定某种类别为动态物体的区域所提取的ORB特征点进行剔除。

[0056] 在一实施例中,所述步骤S6的Mask R-CNN算法模型进行语义分割,进一步包括:

[0057] 通过特征图金字塔网络提取输入图像不同层次上的特征;

- [0058] 通过区域生成网络提出感兴趣提案；
- [0059] 利用感兴趣区域排列进行提案区域对齐；
- [0060] 利用全卷积网络进行掩膜分割；
- [0061] 利用全连接层进行区域坐标确定和所述类别分类。
- [0062] 在一实施例中，所述稀疏SLAM算法模型，进一步包括，跟踪线程、局部建图线程、回环检测线程：
- [0063] 所述跟踪线程，通过寻找对局部地图特征进行匹配，利用纯运动光束平差法最小化重投影误差进行定位每帧图片的相机；
- [0064] 所述局部建图线程，通过执行局部光束平差法管理局部地图并优化，通过地图点维护关键帧之间的共视关系，通过局部光束平差法优化共视关键帧位姿和地图点；
- [0065] 所述回环检测线程，检测大的环并通过执行位姿图优化更正漂移误差，加速闭环匹配帧的筛选，并优化尺度，通过全局光束平差法优化本质图和地图点。
- [0066] 在一实施例中，所述稀疏SLAM算法模型，进一步包括全局光束平差法优化线程，在回环检测线程确认后触发，在位姿图优化之后，计算整个系统最优结构和运动结果。
- [0067] 在一实施例中，所述步骤S7的位姿计算，进一步包括：通过PnP求解初步计算相机位姿，利用后端位姿图优化计算相机位姿，构建相机位姿估计的最小化重投影误差：

$$[0068] \quad \xi^* = \arg \min_{\xi} \frac{1}{2} \sum_{i=1}^n \left\| u_i - \frac{1}{s_i} K \exp(\xi^{\wedge}) P_i \right\|_2^2$$

[0069] 其中， $u_i$ 为像素坐标， $P_i$ 为相机坐标， $\xi^{\wedge}$ 为相机位姿对应的李代数， $s_i$ 为特征点深度， $K$ 为相机内参数矩阵。

[0070] 本发明提供了一种基于视觉SLAM与二维语义分割的语义建图方法，基于ORB特征点的Mask R-CNN算法模型和稀疏SLAM算法模型，建立剔除动态物体的稠密语义地图，利用帧间信息以及图像帧上的语义信息来提升无人机语义建图系统性能，提升针对动态场景进行特征点的提取与匹配的鲁棒性。

## 附图说明

- [0071] 本发明上述的以及其他的特征、性质和优势将通过下面结合附图和实施例的描述而变的更加明显，在附图中相同的附图标记始终表示相同的特征，其中：
- [0072] 图1揭示了根据本发明一实施例的方法流程图；
- [0073] 图2揭示了根据本发明一实施例的相机标定用标定板；
- [0074] 图3a揭示了根据本发明一实施例的针孔相机的小孔成像模型图；
- [0075] 图3b揭示了根据本发明一实施例的针孔相机的相似三角形原理图；
- [0076] 图4揭示了根据本发明一实施例的Mask RCNN的系统流程图。

## 具体实施方式

[0077] 为了使本发明的目的、技术方案及优点更加清楚明白，以下结合附图及实施例，对本发明进行进一步详细说明。应当理解，此处所描述的具体实施例仅用以解释发明，并不用于限定发明。

[0078] 语义地图的概念,指的是一种包含丰富语义信息的地图,表示了对环境中 空间几何关系和存在的物体种类位置等语义信息的抽象。语义地图既包含环境 空间信息又包含环境语义信息的地图,这样移动机器人可以像人一样,既知道 环境中物体也知道物体是什么。

[0079] 针对现有技术中存在的问题与不足,本发明提出基于视觉SLAM与二维语义分割的语义建图系统,使用基于ORB (Oriented FAST and Rotated BRIEF, 快速导向与简要旋转) 特征点进行语义分割,结合稀疏SLAM算法模型,实现 定位的同时完成语义建图。

[0080] 图1揭示了根据本发明一实施例的基于视觉SLAM与二维语义分割的语义建图方法流程图,在图1所示的实施例中,本发明提出的基于视觉SLAM 与二维语义分割的语义建图方法,具体步骤如下:

[0081] S1、标定相机参数,校正相机畸变;

[0082] S2、获取图像帧序列,图像帧序列包括RGB图像和深度图像;

[0083] S3、图像预处理,采用针孔相机模型,得到RGB图像中每个像素点对应的真实三维空间点的坐标;

[0084] S4、判断当前图像帧是否为关键帧,如果是,则转入步骤S6,如果不是, 则转入步骤S5;

[0085] S5、动态模糊补偿,计算得到当前图像帧的图像块质心作为语义特征点, 作为ORB特征点的补充;

[0086] S6、语义分割,针对图像帧进行ORB特征点的提取,利用Mask R-CNN 算法模型进行语义分割,获取该帧图像每一个像素点的语义信息;

[0087] S7、位姿计算,利用稀疏SLAM算法模型计算相机位姿;

[0088] S8、将语义信息输入到稀疏SLAM算法模型,辅助稠密语义地图构建, 完成关键帧的遍历,实现全局点云地图的三维语义建图。

[0089] 下面对每一步骤进行详细的说明介绍。

[0090] 步骤S1:标定相机参数、校正相机畸变。

[0091] 在图像测量过程以及机器视觉应用中,为确定空间物体表面某点的三维几何位置与其在图像中对应点之间的相互关系,必须建立相机成像的几何模型, 这些几何模型参数就是相机参数。

[0092] 畸变系数属于其中一种相机参数,对应与相机畸变现象。在大多数条件下, 这些相机参数必须通过实验与计算才能得到,这个求解参数的过程就称之为相机标定(或摄像机标定)。

[0093] 相机畸变包括径向畸变和切向畸变。

[0094] 所述径向畸变由透镜形状引起。

[0095] 更具体的说,在针孔模型中,一条直线投影到像素平面上还是一条直线。

[0096] 可是,在实际拍摄的照片中,摄像机的透镜往往使得真实环境中的一条直线在图片中变成了曲线,这种畸变称为径向畸变。

[0097] 所述切向畸变,在相机的组装过程中由于不能使得透镜和成像面严格平行 而形成。

[0098] 由于光线投射导致实际对象物体跟投影到2D平面的图像不一致,这种不一致性

是稳定的,可以通过对相机标定,计算出畸变参数来实现对后续图像的畸变校正。

[0099] 对于径向畸变,用和距中心距离有关的二次及高次多项式函数进行纠正:

$$[0100] \quad \begin{cases} x_{\text{corrected}} = x(1 + k_1 r^2 + k_2 r^4 + k_3 r^6) \\ y_{\text{corrected}} = y(1 + k_1 r^2 + k_2 r^4 + k_3 r^6) \end{cases}$$

[0101] 其中,  $[x, y]^T$  是未纠正的点的坐标,  $[x_{\text{corrected}}, y_{\text{corrected}}]^T$  为纠正后的点坐标,  $k_1, k_2, k_3$  为相机的径向畸变系数,  $r$  为点P离坐标系原点的距离。

[0102] 对于切向畸变可以使用另外的两个参数  $p_1, p_2$  来进行纠正:

$$[0103] \quad \begin{cases} x_{\text{corrected}} = x + 2p_1 xy + p_2(r^2 + 2x^2) \\ y_{\text{corrected}} = y + p_1(r^2 + 2y^2) + 2p_2 xy \end{cases}$$

[0104] 其中,  $[x, y]^T$  是未纠正的点的坐标,  $[x_{\text{corrected}}, y_{\text{corrected}}]^T$  为纠正后的点坐标,  $p_1, p_2$  为相机的切向畸变系数,  $r$  为点P离坐标系原点的距离。

[0105] 在相机使用前,通过标定相机的径向畸变系数和切向畸变系数,从二维的图像中获取三维信息,实现图像的畸变校正、对象测量、三维重建等。

[0106] 图2揭示了根据本发明一实施例的相机标定用标定板,将图2所示的标定板摆在相机可视的范围内,每拍一张照片,标定板换一个位置和朝向,检测出图像中的特征点,求出相机的内参数、外参数,进而得到畸变系数。

[0107] 优选的,使用MATLAB中的Camera Calibrator(相机校正)工具箱进行求解相机参数。

[0108] 对于相机坐标系中的点P(X,Y,Z),本发明的步骤S1通过5个畸变系数进行相机畸变校正,找到这个点在像素平面上的正确位置。

[0109] 相机畸变的校正步骤如下:

[0110] S11、将三维空间点投影到归一化图像平面。设它的归一化坐标为  $[x, y]^T$ 。

[0111] S12、对归一化平面上的点进行径向畸变和切向畸变纠正。

$$[0112] \quad \begin{cases} x_{\text{corrected}} = x(1 + k_1 r^2 + k_2 r^4 + k_3 r^6) + 2p_1 xy + p_2(r^2 + 2x^2) \\ y_{\text{corrected}} = y(1 + k_1 r^2 + k_2 r^4 + k_3 r^6) + p_1(r^2 + 2y^2) + 2p_2 xy \end{cases}$$

[0113] 其中,  $[x_{\text{corrected}}, y_{\text{corrected}}]^T$  是校正后的点坐标,  $p_1, p_2$  为相机的切向畸变系数,  $k_1, k_2, k_3$  为相机的径向畸变系数,  $r$  为点P离坐标系原点的距离。

[0114] S13、将纠正后的点  $[x_{\text{corrected}}, y_{\text{corrected}}]^T$  通过内参数矩阵投影到像素平面,得到该点在图像上的正确位置坐标  $[u, v]^T$ 。

$$[0115] \quad \begin{cases} u = f_x x_{\text{corrected}} + c_x \\ v = f_y y_{\text{corrected}} + c_y \end{cases}$$

[0116] 其中,  $f_x, f_y, c_x, c_y$  为相机的内参数。

[0117] 步骤S2、获取图像帧序列。

[0118] 利用Kinect相机获取RGB-D图像帧序列,图像帧序列包括RGB图像和深度图像。

[0119] 步骤S3、图像预处理

[0120] 在一实施例中,采用RGB-D相机作为主要传感器,同时获得RGB图像与深度图像,

采用针孔相机模型进行RGB图像的像素点到真实三维空间的映射。

[0121] 图3a揭示了根据本发明一实施例的针孔相机的小孔成像模型图,图3b 揭示了根据本发明一实施例的针孔相机的相似三角形原理图,如图3a和图 3b所示,建立相机坐标系 $0-x-y-z$ ,以相机的光心位置为坐标系原点 $0$ ,约定 箭头方向为正向。

[0122] 通过图3b所示的相似三角形的映射变换,在相机的成像平面上建立坐标系 $0'-x'-y'-z'$ ,约定箭头方向为正向。

[0123] 假设 $P$ 点坐标是 $[X,Y,Z]^T$ ,相机镜片的焦距为 $f$ ,焦距是相机光心到物理成 像平面的距离。

[0124] 点 $P$ 穿过光心投影到成像平面的点 $P'$ ,像素点 $P'$ 的坐标 $[u,v]^T$ 。

[0125] 根据相应的对应关系,映射关系对应着一个尺度的缩放以及平移的量,推 导可得:

$$[0126] \quad Z \begin{pmatrix} u \\ v \\ 1 \end{pmatrix} = \begin{pmatrix} f_x & 0 & c_x \\ 0 & f_y & c_y \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} X \\ Y \\ Z \end{pmatrix} = KP$$

[0127] 其中, $K$ 称为相机内参数矩阵,为固有参数,在步骤S1中已经进行标定, $f_x, f_y, c_x, c_y$ 为相机的内参数, $P$ 为真实三维空间点坐标, $[u,v]^T$ 为像素点坐标。

[0128] 步骤S4、判断是否是关键帧,如果是,则转入步骤S6,如果不是,则转 入步骤S5;

[0129] 如果采用每一帧图像来进行视觉SLAM和语义分割计算,计算量太大,因 此,选取其中质量高的作为关键帧。

[0130] 本发明中,使用基于ORB (Oriented FAST and Rotated BRIEF,快速导向 与简要旋转)特征点的稀疏SLAM算法模型来筛选关键帧。

[0131] 每一个关键帧,都包含一张RGB图像和一张深度图像。

[0132] 步骤S5、动态模糊

[0133] 由于每帧图像中均可能存在动态物体,每次执行语义建图任务时,指定某 几种目标为动态目标。在图像序列中,如果在该帧图像中识别出该动态目标, 则本发明在二维像素点到三维空间坐标转化时,对相应的点云进行剔除,防止 动态物体在地图中留下残影,影响建图质量。

[0134] 本发明的步骤S5中,如果该帧图像不是关键帧,因为运动模糊,ORB特 征点提取不足,在步骤S6的图像语义分割步骤之前,进行如下操作作为补充:

[0135] 该帧图像每一个物体标注为一个具体的类,对于每一个分割出来的对象有 对应标注区域,分割出来的图像称为图像块,计算图像块 $B$ 的矩 $m_{pq}$ :

$$[0136] \quad m_{pq} = \sum_{x,y \in B} x^p y^q I(x,y), \quad p,q = \{0,1\}$$

[0137] 质心位置 $C$ 为:

$$[0138] \quad C = \begin{pmatrix} \frac{m_{10}}{m_{00}} & \frac{m_{01}}{m_{00}} \\ \frac{m_{20}}{m_{00}} & \frac{m_{02}}{m_{00}} \end{pmatrix}$$

[0139] 该质心作为语义特征点,对ORB特征点的不足进行补充。

[0140] 针对ORB特征点损失严重的模糊图像进行语义特征点的补充,抑制跟踪 算法使用

属于动态对象的匹配,进而综合筛选关键帧,进行相机位姿的估计,防止建图算法将移动对象包括为3D地图的一部分。

[0141] 步骤S6、语义分割

[0142] 针对每一个图像帧进行ORB特征点的提取,利用Mask RCNN (Mask Region-CNN,掩膜区域卷积神经网络) 算法模型进行语义分割,获取该帧图像 每一个像素点的语义信息。

[0143] 基于语义分割结果,如果识别出动态目标,对于指定某种类别为动态物体 的区域所提取的ORB特征点进行剔除。

[0144] 抑制视觉SLAM算法在建图过程中将移动对象包括为3D地图的一部分。

[0145] 本发明的步骤S6中,Mask RCNN算法模型采用COCO数据集进行训练。

[0146] COCO的全称是Common Objects in COntext,是微软团队提供的一个可以 用来进行图像识别的数据集,可以获得80个类别的分类信息。

[0147] 图4揭示了根据本发明一实施例的Mask RCNN的系统流程图,如图4 所示,基于Mask R-CNN算法模型实现图像帧的RGB图像语义分割,所述基 于Mask R-CNN算法模型的卷积神经网络框架,进行语义分割的步骤如下所 示:

[0148] 通过FPN (Feature Pyramid Networks,特征图金字塔网络) 提取输入图像 不同层次上的特征;

[0149] 通过RPN (Region Proposal Network,区域生成网络) 提出感兴趣提案;

[0150] 利用RoI Align (Region of Interest Align,感兴趣区域排列) 进行提案区域 对齐;

[0151] 利用FCN (Fully Convolutional Networks,全卷积网络) 进行掩膜分割;

[0152] 利用FC (Fully Connected Layers,全连接层) 进行区域坐标确定以及所 属类别分类。

[0153] 该帧图像经过Mask RCNN算法模型处理,生成像素级别的语义分类结果, 即每一个像素点的语义分类标签,同时输出包围框坐标以及该分类的置信分 数。

[0154] 本发明采用ORB特征点进行追踪、建图和位置识别任务,ORB特征点的 优点是具有旋转不变性和尺度不变性,并且能够迅速的提取特征和进行匹配, 能够满足实时操作的需求,能够在基于词袋的位置识别过程中,显示出良好的 精度。

[0155] S7位姿计算

[0156] 视觉里程计位姿的估计是对于相邻两帧图像而言的,不难理解,多个这样 的帧间位姿估计累积就是相机的运动轨迹。

[0157] 使用基于ORB (Oriented FAST and Rotated BRIEF,快速导向与简要旋转) 特征点的稀疏SLAM算法模型计算相机位姿。

[0158] 在提取图像帧特征点后,基于关键帧使用PnP进行相机位姿的估计。

[0159] PnP为Perspective-n-Point (n点透视) 的简称,是求解3D到2D点对的运 动的方法:即给出n个3D空间点及其投影位置时,如何求解相机的位姿。

[0160] 假设在时刻k,相机的位置为 $x_k$ ,相机输入数据为 $u_k$ , $w_k$ 为噪声,构建运 动方程:

[0161]  $x_k = f(x_{k-1}, u_k, w_k)$ 。

[0162] 在 $x_k$ 位置上观测到路标点 $y_j$ ,产生一系列观测数据 $z_{k,j}$ , $v_{k,j}$ 为观测噪声, 构建观测方程:

[0163]  $z_{k,j} = h(y_j, x_k, v_{k,j})$ 。

[0164] 本发明的步骤S7中,通过PnP问题求解可以初步计算相机位姿,进而利用后端位姿图优化进一步计算更为精确的相机位姿。

[0165] 本发明的步骤S7中,把相机位姿估计的PnP问题,构建成一个定义域李代数上的非线性最小二乘问题。

[0166] 更进一步的,本发明步骤S7的相机位姿估计,构建为一个BA(Bundle Adjustment, 光束平差法)问题,构建相机位姿估计的最小化重投影误差:

$$[0167] \quad \xi^* = \arg \min_{\xi} \frac{1}{2} \sum_{i=1}^n \left\| u_i - \frac{1}{s_i} K \exp(\xi^{\wedge}) P_i \right\|_2^2$$

[0168] 其中, $u_i$ 为像素坐标, $P_i$ 为相机坐标, $\xi^{\wedge}$ 为相机位姿对应的李代数, $s_i$ 为特征点深度, $K$ 为相机内参数矩阵, $n$ 为点的个数。

[0169] 更进一步的,本发明的步骤S7进一步包括,采用基于DBOW(Direct index Bag of words,词袋模型)嵌入式位置识别模型进行重定位,来防止跟踪失败、或已知地图场景重新初始化、回环检测等。

[0170] 本发明中采用稀疏SLAM算法模型进行关键帧的筛选和相机位姿计算,所述稀疏SLAM算法模型,是在ORB-SLAM2(Oriented FAST and Rotated BRIEF-Simultaneous Localization and Mapping 2,第二代快速导向与简要旋转的即时定位与地图构建)的基础上进行改进得到。

[0171] 所述SLAM算法模型,由4个平行线程组成,包括跟踪线程、局部建图线程、回环检测线程以及全局BA优化线程。

[0172] 更进一步的,全局BA优化线程,仅在回环检测线程确认后才执行。

[0173] 前三个线程为并行线程,定义分别如下:

[0174] 1) 跟踪线程。

[0175] 通过寻找对局部地图特征进行匹配,利用纯运动BA最小化重投影误差进行定位每帧图片的相机。

[0176] 优选的,采用恒速模型进行匹配。

[0177] 2) 局部建图线程。

[0178] 通过执行局部BA管理局部地图并优化,通过MapPoints(地图点)维护关键帧之间的共视关系,通过局部BA优化共视关键帧位姿和MapPoints。

[0179] 3) 回环检测线程。

[0180] 检测大的环并通过执行位姿图优化更正漂移误差,通过Bow加速闭环匹配帧的筛选,并通过Sim3优化尺度,通过全局BA优化Essential Graph(本质图)和MapPoints。所述Sim3变换就是相似变换。

[0181] 回环检测线程触发全局BA优化线程。

[0182] 全局BA线程,在位姿图优化之后,计算整个系统最优结构和运动结果。

[0183] 与现有技术的稠密SLAM算法模型相比,在本发明的稀疏SLAM算法模型,通过语义信息融合,在最终的建图过程中,增添了丰富的图像的语义分割信息。

[0184] 步骤S8、三维语义建图

[0185] 利用步骤S6的语义分割结果,结合步骤S7获取的帧间位姿信息以及图像帧像素点的真实三维坐标,将语义信息输入到稀疏SLAM算法模型,将该帧图像中语义包含的同种物体,以相同标注颜色投射到三维点云地图中,辅助稠密语义地图构建,完成关键帧的遍历,实现全局点云地图的三维语义建图。

[0186] 本发明的步骤S8进一步包括:

[0187] S81、将第一帧关键帧生成的三维空间像素投影到一个初始点云中;

[0188] S82、通过针孔模型计算得到的当前关键帧每一个像素点对应的三维空间坐标,生成一幅点云地图;

[0189] S83、计算得到当前关键帧与上一关键帧的位姿变化;

[0190] S84、两幅点云地图通过位姿变换矩阵进行三维坐标点的叠加融合,生成一幅信息更多的点云地图;

[0191] S85、上述步骤不断迭代,当完成所有关键帧的遍历,实现全局点云地图的构建。

[0192] 下面结合具体的试验,对本发明的基于视觉SLAM与二维语义分割的无人机语义建图方法的试验结果作进一步详细说明。

[0193] 本次试验基于操作系统Ubuntu16.04和硬件显卡Nvidia Geforce GTX 1050,借助Tensorflow、OpenCV、g2o、Point Cloud Library等软件工具,以真实场景为实验条件,利用Kinect V1相机实拍数据。

[0194] 对于三维语义建图评估, $Q_1$ 代表正确检测物品个数, $Q_2$ 代表检测出物体但是分类错误以及实际有物体但是没有检测出的数量, $Q_3$ 代表没有物体但是检测出结果的数量, $P$ 代表三维物体正确检出率,计算方式如下所示:

$$[0195] \quad P = Q_1 / (Q_1 + Q_2 + Q_3)$$

[0196] 通过9次构建稠密语义建图进行实验记录,计算地图中的平均三维物体正确检出率为48.1086%,具体实验结果如下表所示:

[0197] 表1

实验序号	$Q_1$	$Q_2$	$Q_3$	$P_2$
1	46	50	2	0.469388
2	47	49	2	0.479592
[0198] 3	42	54	2	0.428571
4	44	52	2	0.44898
5	78	85	3	0.46988
6	73	90	3	0.439759
7	77	86	3	0.463855
8	79	84	3	0.475904
[0199] 9	170	87	3	0.653846
平均值	/	/	/	0.481086

[0200] 本发明提供一种基于视觉SLAM与二维语义分割的语义建图方法,基于 ORB特征点的Mask R-CNN算法模型和稀疏SLAM算法模型,建立剔除动态 物体的稠密语义地图,利用帧间信息以及图像帧上的语义信息来提升无人机语义建图系统性能,提升针对动态场景进行特征点的提取与匹配的鲁棒性。

[0201] 尽管为使解释简单化将上述方法图示并描述为一系列动作,但是应理解并 领会,这些方法不受动作的次序所限,因为根据一个或多个实施例,一些动作 可按不同次序发生和/或与来自本文中图示和描述或本文中未图示和描述但本 领域技术人员可以理解的其他动作并发地发生。

[0202] 如本申请和权利要求书中所示,除非上下文明确提示例外情形,“一”、“一 个”、“一种”和/或“该”等词并非特指单数,也可包括复数。一般说来,术语“包 括”与“包含”仅提示包括已明确标识的步骤和元素,而这些步骤和元素不构成 一个排它性的罗列,方法或者设备也可能包含其他的步骤或元素。

[0203] 上述实施例是提供给熟悉本领域内的人员来实现或使用本发明的,熟悉本 领域的人员可在不脱离本发明的发明思想的情况下,对上述实施例做出种种修 改或变化,因而本发明的保护范围并不被上述实施例所限,而应该是符合权利 要求书提到的创新性特征的最大范围。

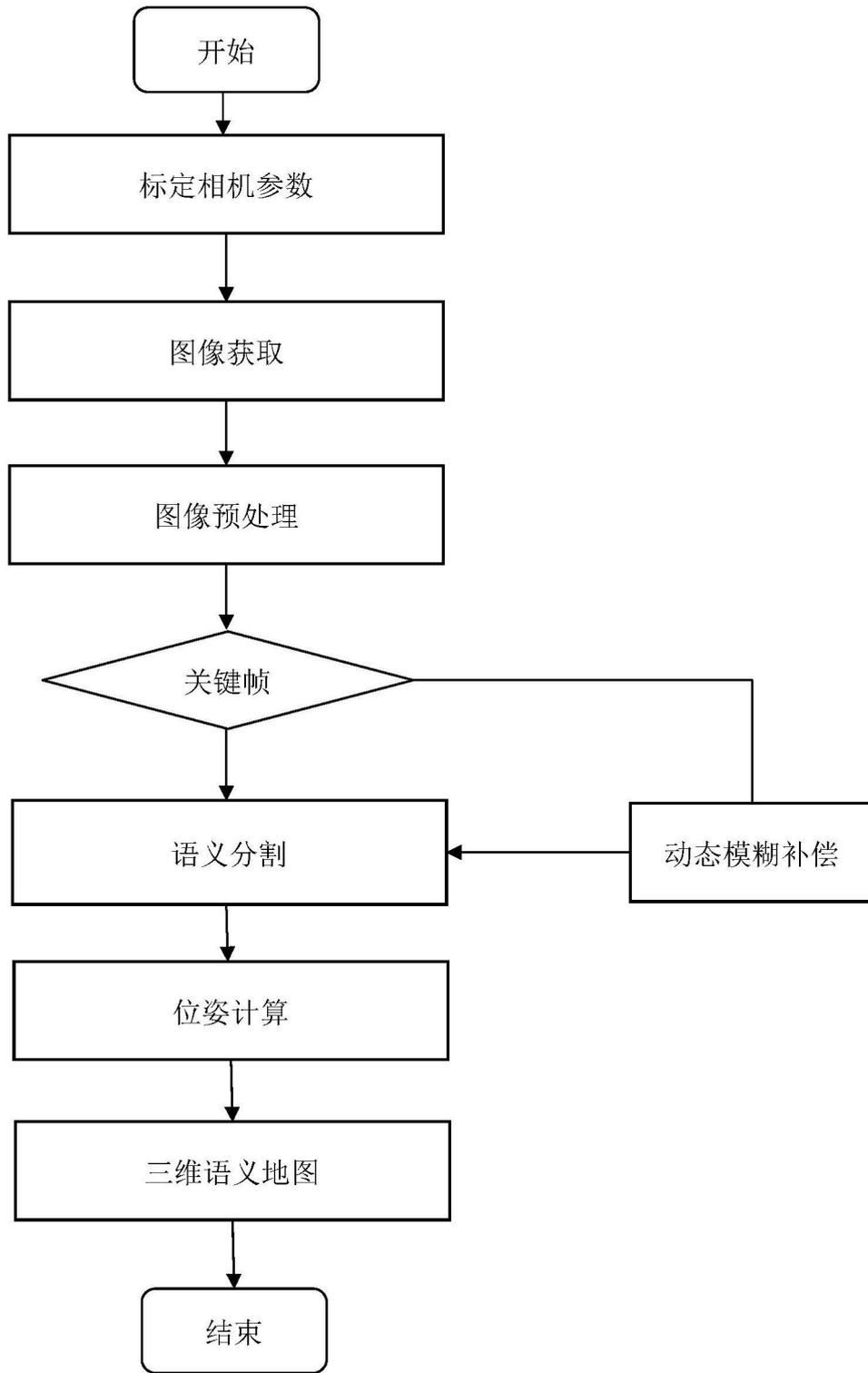


图1

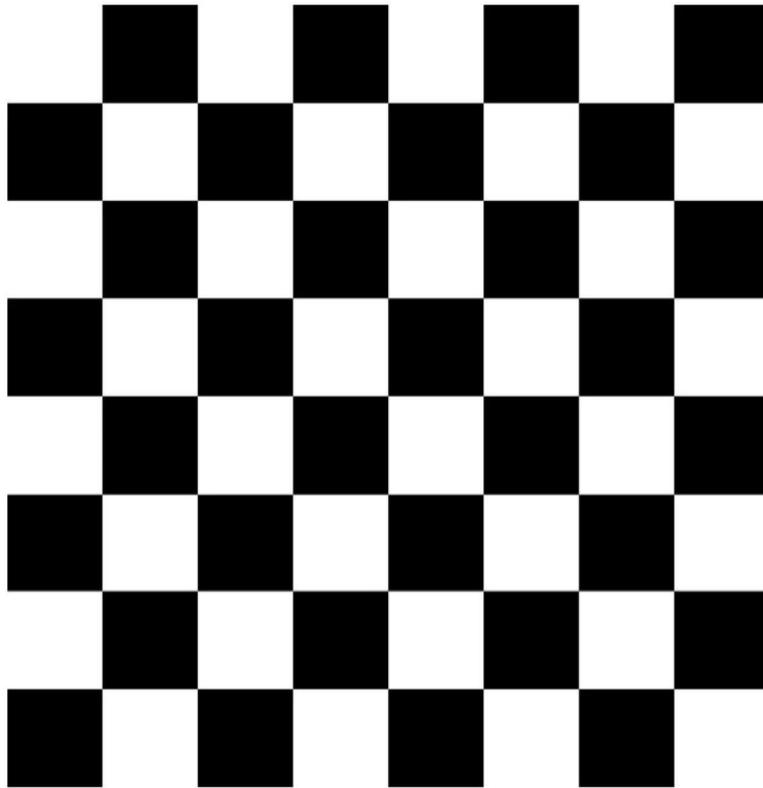


图2

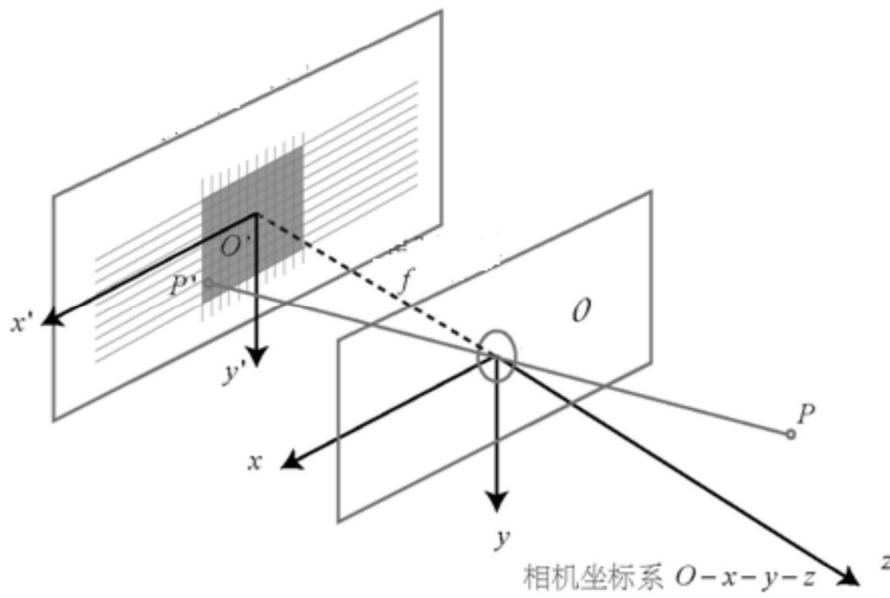


图3a

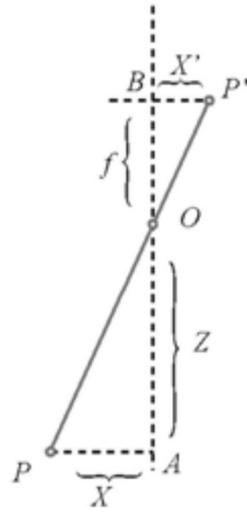


图3b

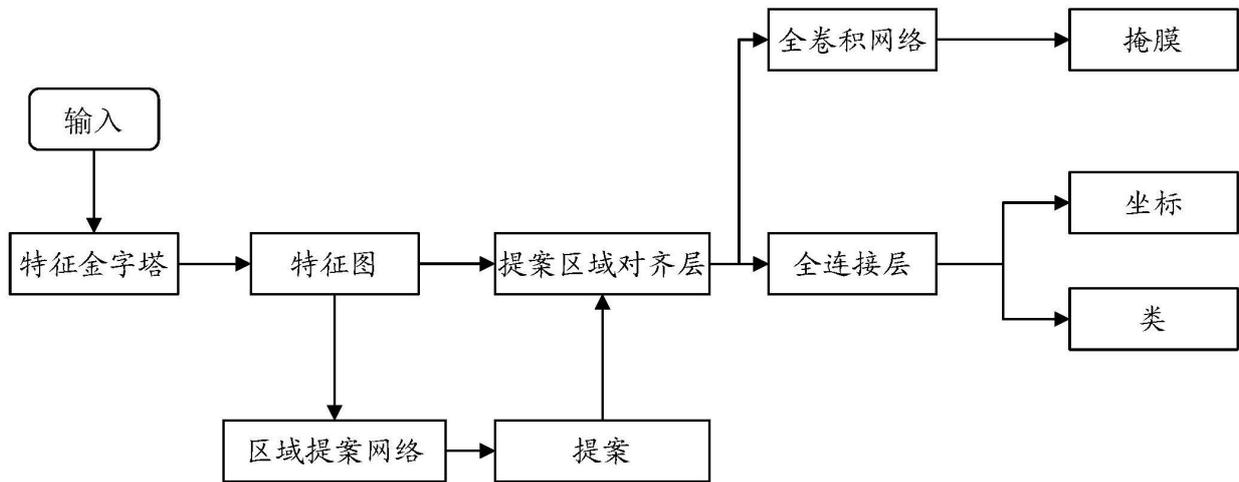


图4