



(21) 申请号 202311238629.5

(22) 申请日 2023.09.22

(65) 同一申请的已公布的文献号

申请公布号 CN 117197229 A

(43) 申请公布日 2023.12.08

(73) 专利权人 北京科技大学顺德创新学院

地址 528000 广东省佛山市顺德区大良致
慧路2号专利权人 航天科工集团智能科技研究院有
限公司

(72) 发明人 曾慧 梁溢友 江左 杨清港

(74) 专利代理机构 北京市广友专利事务所有限
责任公司 11237

专利代理师 张仲波 付忠林

(51) Int. Cl.

G06T 7/70 (2017.01)

G06T 5/90 (2024.01)

G06T 7/50 (2017.01)

G06N 3/0442 (2023.01)

G06N 3/0455 (2023.01)

G06N 3/0464 (2023.01)

G06N 3/048 (2023.01)

G06N 3/08 (2023.01)

(56) 对比文件

CN 110490928 A, 2019.11.22

WO 2022241874 A1, 2022.11.24

CN 108010081 A, 2018.05.08

CN 111275764 A, 2020.06.12

CN 111369608 A, 2020.07.03

CN 114663496 A, 2022.06.24

CN 114663509 A, 2022.06.24

CN 115661224 A, 2023.01.31

KR 102098687 B1, 2020.04.09

US 2021042937 A1, 2021.02.11

US 2021390723 A1, 2021.12.16

WO 2023165093 A1, 2023.09.07 (续)

审查员 吴海旋

权利要求书4页 说明书13页 附图6页

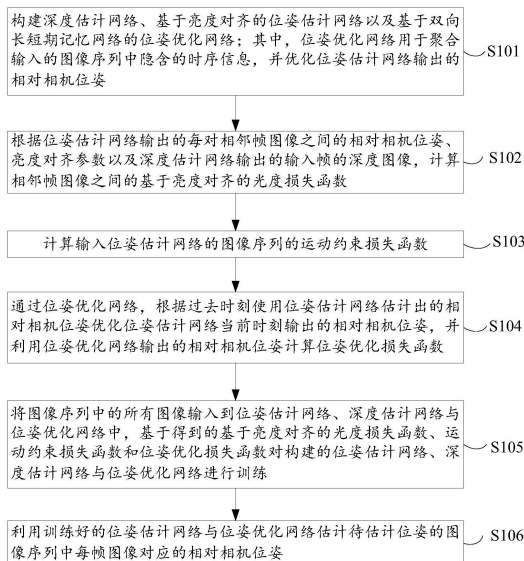
(54) 发明名称

一种基于亮度对齐的多阶段估计单目视觉
里程计方法

(57) 摘要

本发明提供一种基于亮度对齐的多阶段估计单目视觉里程计方法,属于计算机视觉技术领域。所述方法包括:构建深度估计网络、基于亮度对齐的位姿估计网络以及基于双向长短期记忆网络的位姿优化网络;计算相邻帧图像之间的基于亮度对齐的光度损失函数;计算输入位姿估计网络的图像序列的运动约束损失函数;通过位姿优化网络,根据过去时刻使用位姿估计网络估计出的相对相机位姿,并计算位姿优化损失函数;训练位姿估计网络、深度估计网络与位姿优化网络进行训练;利用训练好的位姿估计网络与位姿优化网络估计待估计位姿的图像序列中每帧图像对应的相对相机位姿。采用本发明,能够有效提高相

对相机位姿估计结果的精度与鲁棒性。



[接上页]

(56) 对比文件

岑仕杰;何元烈;陈小聪.结合注意力与无监督深度学习的单目深度估计.广东工业大学学报.2020,(04),全文.

苏健鹏;黄影平;赵柏淦;胡兴.基于深度卷

积神经网络的视觉里程计研究.光学仪器.2020,(04),全文.

李媛;彭晓东;周武根;李运;谢文明.小行星场景下基于循环卷积网络的位姿估计方法.传感器与微系统.2020,(08),全文.

1. 一种基于亮度对齐的多阶段估计单目视觉里程计方法,其特征在于,包括:

构建深度估计网络、基于亮度对齐的位姿估计网络以及基于双向长短期记忆网络的位姿优化网络;其中,位姿优化网络用于聚合输入的图像序列中隐含的时序信息,并优化位姿估计网络输出的相对相机位姿;

根据位姿估计网络输出的每对相邻帧图像之间的相对相机位姿、亮度对齐参数以及深度估计网络输出的输入帧的深度图像,计算相邻帧图像之间的基于亮度对齐的光度损失函数;

计算输入位姿估计网络的图像序列的运动约束损失函数;

通过位姿优化网络,根据过去时刻使用位姿估计网络估计出的相对相机位姿优化位姿估计网络当前时刻输出的相对相机位姿,并利用位姿优化网络输出的相对相机位姿计算位姿优化损失函数;

将图像序列中的所有图像输入到位姿估计网络、深度估计网络与位姿优化网络中,基于得到的基于亮度对齐的光度损失函数、运动约束损失函数和位姿优化损失函数对构建的位姿估计网络、深度估计网络与位姿优化网络进行训练;

利用训练好的位姿估计网络与位姿优化网络估计待估计位姿的图像序列中每帧图像对应的相对相机位姿;

其中,所述根据位姿估计网络输出的每对相邻帧图像之间的相对相机位姿、亮度对齐参数以及深度估计网络输出的输入帧的深度图像,计算相邻帧图像之间的基于亮度对齐的光度损失函数包括:

对于一对具有相对运动关系的图像对 I_t 和 I_{t+1} ,使用位姿估计网络估计对应的相对相机位姿和亮度对齐参数,使用深度估计网络估计输入帧的深度图像,使用亮度对齐参数对齐两幅图像的亮度信息,使用估计出的相对相机位姿与对应的深度图像,将亮度对齐后的第一帧图像 I_t 重投影至第二帧图像 I_{t+1} ,计算重投影图像与真实图像之间的基于亮度对齐的光度损失,获得基于亮度对齐的光度损失函数值;

其中,所述对于一对具有相对运动关系的图像对,使用位姿估计网络估计对应的相对相机位姿和亮度对齐参数,使用深度估计网络估计输入帧的深度图像,使用亮度对齐参数对齐两幅图像的亮度信息,使用估计出的相对相机位姿与对应的深度图像,将亮度对齐后的第一帧图像重投影至第二帧,计算重投影图像与真实图像之间的基于亮度对齐的光度损失,获得基于亮度对齐的光度损失函数值包括:

在当前 t 时刻,将具有相对运动关系的图像对 I_t 和 I_{t+1} 输入位姿估计网络,位姿估计网络输出相邻帧图像对 I_t 与 I_{t+1} 之间的6自由度相对相机位姿 P_{init}^t 和亮度对齐参数 a 和 b ;其中, P_{init}^t 表示为:

$$P_{init}^t = \text{Pose_BA}(I_t, I_{t+1})$$

其中, I_t 和 I_{t+1} 分别表示 t 时刻和 $t+1$ 时刻的图像, $\text{Pose_BA}()$ 为所述基于亮度对齐的位姿估计网络;

深度估计网络估计出第二帧图像 I_{t+1} 对应的深度图像 D_{t+1} ;

对于 P_{init}^t ,使用欧拉变换公式将其转换为 4×4 的位姿变换矩阵 $T_{init}^{I_t I_{t+1}}$:

$$T_{init}^{I_t, I_{t+1}} = Euler(P_{init}^t)$$

其中, $Euler()$ 表示欧拉变换公式;

使用亮度对齐参数 a 和 b 对输入图像 I_t 进行亮度仿射变换, 得到图像 $I_t^{a,b}$:

$$I_t^{a,b} = aI_t + b$$

对于图像 I_{t+1} 上的一个点 p_{t+1} , 其三维坐标由其深度 $D_{t+1}(p_{t+1})$ 还原; 其在图像 $I_t^{a,b}$ 上对应的投影点 $p_t^{a,b}$ 表示为:

$$p_t^{a,b} = K T_{init}^{I_t, I_{t+1}} D_{t+1}(p_{t+1}) K^{-1} p_{t+1}$$

其中, K 为摄像机内参数, D_{t+1} 为 $t+1$ 时刻的深度图像;

通过对图像 I_t 采样, 得到 $t+1$ 时刻图像 I_{t+1} 的重投影图像 I'_{t+1} :

$$I'_{t+1}(p_{t+1}) = I_t^{a,b}(p_t^{a,b})$$

确定重投影图像 I'_{t+1} 与真实图像 I_{t+1} 之间的基于亮度对齐的光度损失 $\hat{L}_p$, 表示为:

$$\begin{aligned} \hat{L}_p(I_{t+1}, I'_{t+1}) = & \frac{1}{|V|} \sum_{p \in V} (\mu_0 \|I_{t+1}(p) - I'_{t+1}(p)\|_1 \\ & + \frac{\mu_1}{2} (1 - SSIM(I_{t+1}(p), I'_{t+1}(p)))) \end{aligned}$$

其中, $SSIM(I_{t+1}, I'_{t+1})$ 表示真实图像 I_{t+1} 与重投影图像 I'_{t+1} 之间的结构相似性, μ_0 、 μ_1 为控制相应部分比例的超参数, $\|\cdot\|_1$ 表示 L1 范数;

其中, 所述运动约束损失函数包括: 非相邻帧光度损失函数以及位姿连续性损失函数;

所述计算输入位姿估计网络的图像序列的运动约束损失函数包括:

对于当前 t 时刻图像之前一段长度为 N 的图像序列中的非相邻帧图像对, 使用位姿估计网络估计其相对相机位姿及亮度对齐参数, 计算对应的基于亮度对齐的光度损失函数, 将序列中所有非相邻帧图像对的基于亮度对齐的光度损失函数分配不同权重相加获得图像序列的非相邻帧光度损失函数;

将位姿估计网络估计的图像序列中每组相邻图像对的相对相机位姿转化为位姿变换矩阵, 将所有位姿变换矩阵相乘得到图像序列首尾帧图像之间的相对位姿变换矩阵, 使用所述位姿估计网络估计首尾真图像之间的相对相机位姿并转换为相对位姿变换矩阵, 计算这两种方式得到的图像序列首尾帧图像之间的相对位姿变换矩阵之间的 L1 范数获得图像序列的位姿连续性损失函数;

其中, 所述对于当前 t 时刻图像之前一段长度为 N 的图像序列中的非相邻帧图像对, 使用位姿估计网络估计其相对相机位姿及亮度对齐参数, 计算对应的基于亮度对齐的光度损失函数, 将序列中所有非相邻帧图像对的基于亮度对齐的光度损失函数分配不同权重相加获得图像序列的非相邻帧光度损失函数包括:

对于当前 t 时刻图像之前一段长度为 N 的图像序列中的非相邻帧图像对, 使用位姿估计网络估计图像序列中图像时间间隔为 2 至 N 的图像对之间的相对相机位姿及亮度对齐参数,

计算对应的基于亮度对齐的光度损失函数值；

根据图像对之间间隔对此非相邻帧图像对之间的基于亮度对齐的光度损失分配分配不同权重,获得整段图像序列的非相邻帧光度损失 L_{na} ：

$$L_{na} = \sum_{i=0}^{N-2} \sum_{j=i+2}^N \mu \hat{L}_p(I_j, I'_{i \rightarrow j})$$

其中, $I'_{i \rightarrow j}$ 表示由第*i*时刻的图像重投影至第*j*时刻的重投影图像, $\hat{L}_p(I_j, I'_{i \rightarrow j})$ 为重投影图像 $I'_{i \rightarrow j}$ 与真实图像 I_j 之间的基于亮度对齐的光度损失, μ 为人工设置的超参数,其值与输入的两帧图像之间间隔成负相关：

$$\mu = 10^{i-j}。$$

2.根据权利要求1所述的基于亮度对齐的多阶段估计单目视觉里程计方法,其特征在在于,所述位姿估计网络采用编码器-解码器结构；

所述位姿优化网络采用双向长短期记忆网络后连接全连接层构成；其中，

所述双向长短期记忆网络,用于聚合图像序列中隐含的时序信息；

所述全连接层,用于将聚合后的时序信息回归出一个6自由度的相对相机位姿。

3.根据权利要求1所述的基于亮度对齐的多阶段估计单目视觉里程计方法,其特征在在于,所述将位姿估计网络估计的图像序列中每组相邻图像对的相对相机位姿转化为位姿变换矩阵,将所有位姿变换矩阵相乘得到图像序列首尾帧图像之间的相对位姿变换矩阵,使用所述位姿估计网络估计首尾真图像之间的相对相机位姿并转换为相对位姿变换矩阵,计算这两种方式得到的图像序列首尾帧图像之间的相对位姿变换矩阵之间的L1范数获得图像序列的位姿连续性损失函数包括：

在一段长度为M的图像序列中,每组相邻图像对 I_i 和 I_{i+1} 之间存在一个位姿变换矩阵 $T^{I_i, I_{i+1}}$ ；其中, $i=0,1,\dots,M-2$, $T^{I_i, I_{i+1}}$ 由相邻图像对 I_i 和 I_{i+1} 的相对相机位姿转化而成；

在相机坐标系下,将所有位姿变换矩阵 $T^{I_i, I_{i+1}}$ 连续左乘得到图像序列首尾帧图像之间的相对位姿变换矩阵 $T^{I_0, I_{M-1}}$ ：

$$T^{I_0, I_{M-1}} = T^{I_{M-2}, I_{M-1}} T^{I_{M-3}, I_{M-2}} \dots T^{I_0, I_1}$$

同时,直接使用位姿估计网络估计首尾两帧图像之间相机运动产生的相对相机位姿,并使用欧拉变换将其转换为相对位姿变换矩阵 $\hat{T}^{I_0, I_{M-1}}$ ；

根据得到的 $T^{I_0, I_{M-1}}$ 与 $\hat{T}^{I_0, I_{M-1}}$ 之间的L1范数,得到长度为M的图像序列的位姿连续性损失 $L_{pc, M}$ ：

$$L_{pc, M} = ||T^{I_0, I_{M-1}} - \hat{T}^{I_0, I_{M-1}}||_1$$

对于当前*t*时刻图像之前一段长度为N的图像序列,将这段图像序列中采样间隔为2, 3, ..., N-2的每个子序列的位姿连续性损失相加获得总的位姿连续性损失：

$$L_{pc} = \sum_{M=2}^{N-2} L_{pc, M}。$$

4.根据权利要求1所述的基于亮度对齐的多阶段估计单目视觉里程计方法,其特征在在于,所述通过位姿优化网络,根据过去时刻使用位姿估计网络估计出的相对相机位姿优化

位姿估计网络当前时刻输出的相对相机位姿,并利用位姿优化网络输出的相对相机位姿计算位姿优化损失函数包括:

将当前 t 时刻之前的 $n-1$ 时刻位姿估计网络估计出的相对相机位姿 $P_{out}^{t-n+1}, \dots, P_{out}^{t-1}$ 与当前 t 时刻位姿估计网络估计出的相对相机位姿 P_{init}^t 输入位姿优化网络,位姿优化网络输出一个具有更高精度的6自由度相对相机位姿 P_{out}^t :

$$P_{out}^t = Pose_Optmization(P_{out}^{t-n+1}, P_{out}^{t-n+2}, \dots, P_{out}^{t-1}, P_{init}^t)$$

其中, $t > n-1$, $Pose_Optmization()$ 为位姿优化网络, P_{out}^t 包括3自由度的相对旋转和3自由度的相对位移;

使用此相对相机位姿进行图像重投影操作,并计算相应的基于亮度对齐的光度损失函数,作为训练位姿优化网络的损失函数 L_{p0} 。

一种基于亮度对齐的多阶段估计单目视觉里程计方法

技术领域

[0001] 本发明涉及计算机视觉技术领域,特别是指一种基于亮度对齐的多阶段估计单目视觉里程计方法。

背景技术

[0002] 视觉里程计是视觉同时定位与建图技术的重要组成部分,主要用于相机自我运动估计,其被广泛应用于广泛应用于机器人导航、自动驾驶和增强现实等多个领域。根据采用的传感器种类及其数量的不同,视觉里程计可分为单目视觉里程计、双目视觉里程计和基于视觉-惯性的里程计算法。其中,单目视觉里程计算法仅通过单目摄像机即可实现,具有成本低廉、易于部署等特性,在近近年来受到了研究者的广泛关注。

[0003] 传统的视觉里程计方法分为基于特征的方法和直接法。基于特征的方法提取并匹配两个连续输入帧的特征,根据匹配特征之间的几何关系估计摄像机的自我运动。直接法则利用灰度不变性假设,直接跟踪帧间的像素运动,通过最小化光度误差,进而估计出相机的自我运动。

[0004] 随着深度学习技术的发展,深度神经网络在计算机视觉领域得到了广泛应用。研究人员已经采用深度学习的方式实现视觉里程计方法,基于深度学习的视觉里程计方法包含基于有监督学习的视觉里程计算法和基于自监督学习的视觉里程计算法(简称:自监督视觉里计算法)。自监督视觉里计算法不需要相机的真实相对位姿作为监督信号,具有更加广泛的应用场景。

[0005] 许多现有的自监督视觉里程计算法在相对相机位姿的估计精度还相对较差,且在算法鲁棒性方面仍存在局限性,特别是在环境光照变化场景中。此外,在相机自我运动估计过程中,许多现有方法对于输入图像序列中隐含的时序信息未得到充分利用。

发明内容

[0006] 本发明实施例提供了基于亮度对齐的多阶段估计单目视觉里程计方法,能够有效提高相对相机位姿估计结果的精度与鲁棒性。所述技术方案如下:

[0007] 一方面,提供了一种基于亮度对齐的多阶段估计单目视觉里程计方法,该方法应用于电子设备,该方法包括:

[0008] 构建深度估计网络、基于亮度对齐的位姿估计网络以及基于双向长短期记忆网络的位姿优化网络;其中,位姿优化网络用于聚合输入的图像序列中隐含的时序信息,并优化位姿估计网络输出的相对相机位姿;

[0009] 根据位姿估计网络输出的每对相邻帧图像之间的相对相机位姿、亮度对齐参数以及深度估计网络输出的输入帧的深度图像,计算相邻帧图像之间的基于亮度对齐的光度损失函数;

[0010] 计算输入位姿估计网络的图像序列的运动约束损失函数;

[0011] 通过位姿优化网络,根据过去时刻使用位姿估计网络估计出的相对相机位姿优化

位姿估计网络当前时刻输出的相对相机位姿,并利用位姿优化网络输出的相对相机位姿计算位姿优化损失函数;

[0012] 将图像序列中的所有图像输入到位姿估计网络、深度估计网络与位姿优化网络中,基于得到的基于亮度对齐的光度损失函数、运动约束损失函数和位姿优化损失函数对构建的位姿估计网络、深度估计网络与位姿优化网络进行训练;

[0013] 利用训练好的位姿估计网络与位姿优化网络估计待估计位姿的图像序列中每帧图像对应的相对相机位姿。

[0014] 进一步地,所述位姿估计网络采用编码器-解码器结构;

[0015] 所述位姿优化网络采用双向长短期记忆网络后连接全连接层构成;其中,

[0016] 所述双向长短期记忆网络,用于聚合图像序列中隐含的时序信息;

[0017] 所述全连接层,用于将聚合后的时序信息回归出一个6自由度的相对相机位姿。

[0018] 进一步地,所述根据位姿估计网络输出的每对相邻帧图像之间的相对相机位姿、亮度对齐参数以及深度估计网络输出的输入帧的深度图像,计算相邻帧图像之间的基于亮度对齐的光度损失函数包括:

[0019] 对于一对具有相对运动关系的图像对 I_t 和 I_{t+1} ,使用位姿估计网络估计对应的相对相机位姿和亮度对齐参数,使用深度估计网络估计输入帧的深度图像,使用亮度对齐参数对齐两幅图像的亮度信息,使用估计出的相对相机位姿与对应的深度图像,将亮度对齐后的第一帧图像 I_t 重投影至第二帧图像 I_{t+1} ,计算重投影图像与真实图像之间的基于亮度对齐的光度损失,获得基于亮度对齐的光度损失函数值。

[0020] 进一步地,所述对于一对具有相对运动关系的图像对,使用位姿估计网络估计对应的相对相机位姿和亮度对齐参数,使用深度估计网络估计输入帧的深度图像,使用亮度对齐参数对齐两幅图像的亮度信息,使用估计出的相对相机位姿与对应的深度图像,将亮度对齐后的第一帧图像重投影至第二帧,计算重投影图像与真实图像之间的基于亮度对齐的光度损失,获得基于亮度对齐的光度损失函数值包括:

[0021] 在当前 t 时刻,将具有相对运动关系的图像对 I_t 和 I_{t+1} 输入位姿估计网络,位姿估计网络输出相邻帧图像对 I_t 与 I_{t+1} 之间的6自由度相对相机位姿 P_{init}^t 和亮度对齐参数 a 和 b ;

其中, P_{init}^t 表示为:

[0022] $P_{init}^t = Pose_BA(I_t, I_{t+1})$

[0023] 其中, I_t 和 I_{t+1} 分别表示 t 时刻和 $t+1$ 时刻的图像, $Pose_BA()$ 为所述基于亮度对齐的位姿估计网络;

[0024] 深度估计网络估计出第二帧图像 I_{t+1} 对应的深度图像 D_{t+1} ;

[0025] 对于 P_{init}^t ,使用欧拉变换公式将其转换为 4×4 的位姿变换矩阵 $T_{init}^{I_t, I_{t+1}}$:

[0026] $T_{init}^{I_t, I_{t+1}} = Euler(P_{init}^t)$

[0027] 其中, $Euler()$ 表示欧拉变换公式;

[0028] 使用亮度对齐参数 a 和 b 对输入图像 I_t 进行亮度仿射变换,得到图像 $I_t^{a,b}$:

[0029] $I_t^{a,b} = aI_t + b$

[0030] 对于图像 I_{t+1} 上的一个点 p_{t+1} ,其三维坐标由其深度 $D_{t+1}(p_{t+1})$ 还原;其在图像 $I_t^{a,b}$ 上对应的投影点 $p_t^{a,b}$ 表示为:

$$[0031] \quad p_t^{a,b} = K T_{init}^{I_t, I_{t+1}} D_{t+1}(p_{t+1}) K^{-1} p_{t+1}$$

[0032] 其中, K 为摄像机内参数, D_{t+1} 为 $t+1$ 时刻的深度图像;

[0033] 通过对图像 I_t 采样,得到 $t+1$ 时刻图像 I_{t+1} 的重投影图像 I'_{t+1} :

$$[0034] \quad I'_{t+1}(p_{t+1}) = I_t^{a,b}(p_t^{a,b})$$

[0035] 确定重投影图像 I'_{t+1} 与真实图像 I_{t+1} 之间的基于亮度对齐的光度损失 $\hat{L}_p$,表示为:

$$[0036] \quad \hat{L}_p(I_{t+1}, I'_{t+1}) = \frac{1}{|V|} \sum_{p \in V} (\mu_0 \|I_{t+1}(p) - I'_{t+1}(p)\|_1 + \frac{\mu_1}{2} (1 - SSIM(I_{t+1}(p), I'_{t+1}(p))))$$

[0037] 其中, $SSIM(I_{t+1}, I'_{t+1})$ 表示真实图像 I_{t+1} 与重投影图像 I'_{t+1} 之间的结构相似性, μ_0, μ_1 为控制相应部分比例的超参数, $\|\cdot\|_1$ 表示L1范数。

[0038] 进一步地,所述运动约束损失函数包括:非相邻帧光度损失函数以及位姿连续性损失函数;

[0039] 所述计算输入位姿估计网络的图像序列的运动约束损失函数包括:

[0040] 对于当前 t 时刻图像之前一段长度为 N 的图像序列中的非相邻帧图像对,使用位姿估计网络估计其相对相机位姿及亮度对齐参数,计算对应的基于亮度对齐的光度损失函数,将序列中所有非相邻帧图像对的基于亮度对齐的光度损失函数分配不同权重相加获得图像序列的非相邻帧光度损失函数;

[0041] 将位姿估计网络估计的图像序列中每组相邻图像对的相对相机位姿转化为位姿变换矩阵,将所有位姿变换矩阵相乘得到图像序列首尾帧图像之间的相对位姿变换矩阵,使用所述位姿估计网络估计首尾真图像之间的相对相机位姿并转换为相对位姿变换矩阵,计算这两种方式得到的图像序列首尾帧图像之间的相对位姿变换矩阵之间的L1范数获得图像序列的位姿连续性损失函数。

[0042] 进一步地,所述对于当前 t 时刻图像之前一段长度为 N 的图像序列中的非相邻帧图像对,使用位姿估计网络估计其相对相机位姿及亮度对齐参数,计算对应的基于亮度对齐的光度损失函数,将序列中所有非相邻帧图像对的基于亮度对齐的光度损失函数分配不同权重相加获得图像序列的非相邻帧光度损失函数包括:

[0043] 对于当前 t 时刻图像之前一段长度为 N 的图像序列中的非相邻帧图像对,使用位姿估计网络估计图像序列中图像时间间隔为2至 N 的图像对之间的相对相机位姿及亮度对齐参数,计算对应的基于亮度对齐的光度损失函数值;

[0044] 根据图像对之间间隔对此非相邻帧图像对之间的基于亮度对齐的光度损失分配不同权重,获得整段图像序列的非相邻帧光度损失 L_{na} :

$$[0045] \quad L_{na} = \sum_{i=0}^{N-2} \sum_{j=i+2}^N \mu \hat{L}_p(I_j, I'_{i \rightarrow j})$$

[0046] 其中, $I'_{i \rightarrow j}$ 表示由第 i 时刻的图像重投影至第 j 时刻的重投影图像, $\hat{L}_p(I_j, I'_{i \rightarrow j})$ 为重投影图像 $I'_{i \rightarrow j}$ 与真实图像 I_j 之间的基于亮度对齐的光度损失, μ 为人工设置的超参数, 其值与输入的两帧图像之间间隔成负相关:

[0047] $\mu = 10^{i-j}$ 。

[0048] 进一步地, 所述将位姿估计网络估计的图像序列中每组相邻图像对的相对相机位姿转化为位姿变换矩阵, 将所有位姿变换矩阵相乘得到图像序列首尾帧图像之间的相对位姿变换矩阵, 使用所述位姿估计网络估计首尾真图像之间的相对相机位姿并转换为相对位姿变换矩阵, 计算这两种方式得到的图像序列首尾帧图像之间的相对位姿变换矩阵之间的 L1 范数获得图像序列的位姿连续性损失函数包括:

[0049] 在一段长度为 M 的图像序列中, 每组相邻图像对 I_i 和 I_{i+1} 之间存在一个位姿变换矩阵 $T^{I_i, I_{i+1}}$; 其中, $i = 0, 1, \dots, M-2$, $T^{I_i, I_{i+1}}$ 由相邻图像对 I_i 和 I_{i+1} 的相对相机位姿转化而成;

[0050] 在相机坐标系下, 将所有位姿变换矩阵 $T^{I_i, I_{i+1}}$ 连续左乘得到图像序列首尾帧图像之间的相对位姿变换矩阵 $T^{I_0, I_{M-1}}$:

[0051] $T^{I_0, I_{M-1}} = T^{I_{M-2}, I_{M-1}} T^{I_{M-3}, I_{M-2}} \dots T^{I_0, I_1}$

[0052] 同时, 直接使用位姿估计网络估计首尾两帧图像之间相机运动产生的相对相机位姿, 并使用欧拉变换将其转换为相对位姿变换矩阵 $\hat{T}^{I_0, I_{M-1}}$;

[0053] 根据得到的 $T^{I_0, I_{M-1}}$ 与 $\hat{T}^{I_0, I_{M-1}}$ 之间的 L1 范数, 得到长度为 M 的图像序列的位姿连续性损失 $L_{pc, M}$:

[0054] $L_{pc, M} = ||T^{I_0, I_{M-1}} - \hat{T}^{I_0, I_{M-1}}||_1$

[0055] 对于当前 t 时刻图像之前一段长度为 N 的图像序列, 将这段图像序列中采样间隔为 $2, 3, \dots, N-2$ 的每个子序列的位姿连续性损失相加获得总的位姿连续性损失:

[0056] $L_{pc} = \sum_{M=2}^{N-2} L_{pc, M}$ 。

[0057] 进一步地, 所述通过位姿优化网络, 根据过去时刻使用位姿估计网络估计出的相对相机位姿优化位姿估计网络当前时刻输出的相对相机位姿, 并利用位姿优化网络输出的相对相机位姿计算位姿优化损失函数包括:

[0058] 将当前 t 时刻之前的 $n-1$ 时刻位姿估计网络估计出的相对相机位姿 $P_{out}^{t-n+1}, \dots, P_{out}^{t-1}$ 与当前 t 时刻位姿估计网络估计出的相对相机位姿 P_{init}^t 输入位姿优化网络, 位姿优化网络输出一个具有更高精度的 6 自由度相对相机位姿 P_{out}^t :

[0059] $P_{out}^t = \text{Pose_Optimization}(P_{out}^{t-n+1}, P_{out}^{t-n+2}, \dots, P_{out}^{t-1}, P_{init}^t)$

[0060] 其中, $t > n-1$, $\text{Pose_Optimization}()$ 为位姿优化网络, P_{out}^t 包括 3 自由度的相对旋转和 3 自由度的相对位移;

[0061] 使用此相对相机位姿进行图像重投影操作, 并计算相应的基于亮度对齐的光度损失函数, 作为训练位姿优化网络的损失函数 L_{p0} 。

[0062] 一方面, 提供了一种电子设备, 所述电子设备包括处理器和存储器, 所述存储器中存储有至少一条指令, 所述至少一条指令由所述处理器加载并执行以实现上述基于亮度对

齐的多阶段估计单目视觉里程计方法。

[0063] 一方面,提供了一种计算机可读存储介质,所述存储介质中存储有至少一条指令,所述至少一条指令由处理器加载并执行以实现上述基于亮度对齐的多阶段估计单目视觉里程计方法。

[0064] 本发明实施例提供的技术方案带来的有益效果至少包括:

[0065] 本发明实施例中,构建深度估计网络、基于亮度对齐的位姿估计网络以及基于双向长短期记忆网络的位姿优化网络;其中,位姿优化网络用于聚合输入的图像序列中隐含的时序信息,并优化位姿估计网络输出的相对相机位姿;根据位姿估计网络输出的每对相邻帧图像之间的相对相机位姿、亮度对齐参数以及深度估计网络输出的输入帧的深度图像,计算相邻帧图像之间的基于亮度对齐的光度损失函数;计算输入位姿估计网络的图像序列的运动约束损失函数;通过位姿优化网络,根据过去时刻使用位姿估计网络估计出的相对相机位姿优化位姿估计网络当前时刻输出的相对相机位姿,并利用位姿优化网络输出的相对相机位姿计算位姿优化损失函数;将图像序列中的所有图像输入到位姿估计网络、深度估计网络与位姿优化网络中,基于得到的基于亮度对齐的光度损失函数、运动约束损失函数和位姿优化损失函数对构建的位姿估计网络、深度估计网络与位姿优化网络进行训练;利用训练好的位姿估计网络与位姿优化网络估计待估计位姿的图像序列中每帧图像对应的相对相机位姿。这样,能够对齐输入图像的亮度信息,并采用多阶段位姿估计策略能够更加有效运用图像序列中隐含的时序信息,从而有效提高相对相机位姿估计结果的精度与鲁棒性。

附图说明

[0066] 为了更清楚地说明本发明实施例中的技术方案,下面将对实施例描述中所需要使用的附图作简单地介绍,显而易见地,下面描述中的附图仅仅是本发明的一些实施例,对于本领域普通技术人员来讲,在不付出创造性劳动的前提下,还可以根据这些附图获得其他的附图。

[0067] 图1为本发明实施例提供的基于亮度对齐的多阶段估计单目视觉里程计方法的流程示意图;

[0068] 图2为本发明实施例提供的总体流程框架示意图;

[0069] 图3为本发明实施例提供的位姿估计网络结构示意图;

[0070] 图4为本发明实施例提供的位姿优化网络结构示意图;

[0071] 图5为本发明实施例提供的初始位姿估计流程示意图;

[0072] 图6为本发明实施例提供的位姿优化流程示意图;

[0073] 图7(a)为本发明实施例提供的方法在KITTI里程计数据集中序列09上估计的轨迹示意图;

[0074] 图7(b)为本发明实施例提供的方法在KITTI里程计数据集中序列10上估计的轨迹示意图;

[0075] 图8为本发明实施例提供的亮度对齐实验结果示意图;

[0076] 图9(a)为本发明实施例提供的方法在KITTI里程计数据集中序列11上估计的轨迹示意图;

[0077] 图9(b)为本发明实施例提供的方法在KITTI里程计数据集中序列15上估计的轨迹示意图;

[0078] 图9(c)为本发明实施例提供的方法在KITTI里程计数据集中序列16上估计的轨迹示意图;

[0079] 图9(d)为本发明实施例提供的方法在KITTI里程计数据集中序列17上估计的轨迹示意图;

[0080] 图10是本发明实施例提供的一种电子设备的结构示意图。

具体实施方式

[0081] 为使本发明的目的、技术方案和优点更加清楚,下面将结合附图对本发明实施方式作进一步地详细描述。

[0082] 如图1-图2所示,本发明实施例提供了一种基于亮度对齐的多阶段估计单目视觉里程计方法,该方法可以由电子设备实现,该电子设备可以是终端或服务器,该方法包括:

[0083] S101,构建深度估计网络 (Depth Net)、基于亮度对齐的位姿估计网络 (Pose-BA Net) 以及基于双向长短期记忆网络的位姿优化网络 (Pose-OPT Net);

[0084] 本实施例中,深度估计网络用于输出输入帧的深度图像。深度估计网络选择ResNet50结构作为编码器,以类似于DispNet解码器的多层反卷积结构作为解码器,并通过跳跃链接结构与编码器连接,输出层激活函数为Sigmoid。

[0085] 本实施例中,所述位姿估计网络,用于输出输入的每组图像对之间的相对相机位姿以及亮度对齐参数;如图3所示,所述位姿估计网络采用编码器-解码器结构,其网络结构如表1所示:

[0086] 表1Pose-BA Net网络结构

	module	layer	inplane	kernal	outplane
[0087]	特征编码器	Resblock	6		64
		Resblock	64		128
		Resblock	128		256
		Resblock	256		512
	亮度对齐解码器-a	Conv	512	1	256
		Conv	256	3	256
		Conv	256	1	1
	亮度对齐解码器-b	Conv	512	1	256
		Conv	256	3	256
		Conv	256	1	1
	位姿解码器	Conv	512	1	256
		Conv	256	3	256
		Conv	256	1	6

[0088] 如图3所示,位姿估计网络由ResNet18编码器 (ResBlock是构成ResNet18的基础模块) 以及三个并行的三层卷积结构组成。三层卷积层中的前两层卷积层以ReLU (Rectification Linear Unit,整流线性单元) 为激活函数,最后一层卷积层为纯卷积层,无激活函数;位姿估计网络的输入经过ResNet18后,再经过三个并行的三层卷积层,输出一个6自由度的相对相机位姿 (包括:3自由度的相对旋转和3自由度的相对位移) 以及两个亮

度对齐参数。

[0089] 本实施例中,位姿估计网络与深度估计网络的输入为 832×256 的RGB图像。

[0090] 本实施例中,使用多阶段估计的方法进行相对相机位姿估计,首先使用位姿估计网络对相对相机位姿进行初步估计,进而使用位姿优化网络优化位姿估计网络的输出结果。因此,位姿优化网络不仅可以用于聚合输入的图像序列中隐含的时序信息,还可以优化位姿估计网络输出的相对相机位姿。所述位姿优化网络采用双向长短期记忆网络后连接全连接层构成,其基本结构如图4中虚线框所示;其中,

[0091] 所述双向长短期记忆网络,用于聚合图像序列(即:相机运动过程)中隐含的时序信息;

[0092] 所述全连接层,用于将聚合后的时序信息回归出一个6自由度的相对相机位姿。

[0093] S102,根据位姿估计网络输出的每对相邻帧图像之间的相对相机位姿、亮度对齐参数以及深度估计网络输出的输入帧的深度图像,计算相邻帧图像之间的基于亮度对齐的光度损失函数;

[0094] 本实施例中,对于一对具有相对运动关系的图像对 I_t 和 I_{t+1} ,使用位姿估计网络估计对应的相对相机位姿和亮度对齐参数,使用深度估计网络估计输入帧的深度图像,使用亮度对齐参数对齐两幅图像的亮度信息,使用估计出的相对相机位姿与对应的深度图像,将亮度对齐后的第一帧图像 I_t 重投影至第二帧图像 I_{t+1} ,计算重投影图像与真实图像之间的基于亮度对齐的光度损失,获得基于亮度对齐的光度损失函数值,如图5所示,具体可以包括以下步骤:

[0095] A1,在当前 t 时刻,将具有相对运动关系的图像对 I_t 和 I_{t+1} 输入位姿估计网络,位姿估计网络输出相邻帧图像对 I_t 与 I_{t+1} 之间的6自由度相对相机位姿 P_{init}^t 和亮度对齐参数 a 和 b ;其中, P_{init}^t 表示为:

$$[0096] \quad P_{init}^t = \text{Pose_BA}(I_t, I_{t+1})$$

[0097] 其中, I_t 和 I_{t+1} 分别表示 t 时刻和 $t+1$ 时刻的图像, $\text{Pose_BA}()$ 为所述基于亮度对齐的位姿估计网络;

[0098] 值得注意的是,这里的相对相机位姿 P_{init}^t 虽然没有位姿优化网络输出的相对相机位姿精度高,但是也已经是较为可信的当前时刻相对相机位姿估计值。

[0099] A2,深度估计网络估计出第二帧图像 I_{t+1} 对应的深度图像 D_{t+1} ;

[0100] A3,对于 P_{init}^t ,使用欧拉变换公式将其转换为 4×4 的位姿变换矩阵 $T_{init}^{I_t, I_{t+1}}$:

$$[0101] \quad T_{init}^{I_t, I_{t+1}} = \text{Euler}(P_{init}^t)$$

[0102] 其中, $\text{Euler}()$ 表示欧拉变换公式;

[0103] A4,使用亮度对齐参数 a 和 b 对输入图像 I_t 进行亮度仿射变换,得到图像 $I_t^{a,b}$:

$$[0104] \quad I_t^{a,b} = aI_t + b$$

[0105] A5,对于图像 I_{t+1} 上的一个点 p_{t+1} ,其三维坐标由其深度 $D_{t+1}(p_{t+1})$ 还原;其在图像 $I_t^{a,b}$ 上对应的投影点 $p_t^{a,b}$ 表示为:

[0106] $p_t^{a,b} = KT_{init}^{I_t, I_{t+1}} D_{t+1}(p_{t+1}) K^{-1} p_{t+1}$

[0107] 其中, K 为摄相机内参数, D_{t+1} 为 $t+1$ 时刻的深度图像;

[0108] $A6$, 通过对图像 I_t 采样, 得到 $t+1$ 时刻图像 I_{t+1} 的重投影图像 I'_{t+1} :

[0109] $I'_{t+1}(p_{t+1}) = I_t^{a,b}(p_t^{a,b})$

[0110] $A7$, 确定重投影图像 I'_{t+1} 与真实图像 I_{t+1} 之间的基于亮度对齐的光度损失 $\hat{L}_p$, 表示为:

[0111]
$$\hat{L}_p(I_{t+1}, I'_{t+1}) = \frac{1}{|V|} \sum_{p \in V} (\mu_0 \|I_{t+1}(p) - I'_{t+1}(p)\|_1 + \frac{\mu_1}{2} (1 - SSIM(I_{t+1}(p), I'_{t+1}(p))))$$

[0112] 其中, $SSIM(I_{t+1}, I'_{t+1})$ 表示真实图像 I_{t+1} 与重投影图像 I'_{t+1} 之间的结构相似性, μ_0, μ_1 为控制相应部分比例的超参数, $\|\cdot\|_1$ 表示 $L1$ 范数。

[0113] $S103$, 计算输入位姿估计网络的图像序列的运动约束损失函数;

[0114] 本实施例中, 所述运动约束损失函数包括: 非相邻帧光度损失函数以及位姿连续性损失函数;

[0115] 所述计算输入位姿估计网络的图像序列的运动约束损失函数, 具体可以包括以下包括:

[0116] $B1$, 对于当前 t 时刻图像之前一段长度为 N 的图像序列中的非相邻帧图像对, 使用位姿估计网络估计其相对相机位姿及亮度对齐参数, 计算对应的基于亮度对齐的光度损失函数, 将序列中所有非相邻帧图像对的基于亮度对齐的光度损失函数分配不同权重相加获得图像序列的非相邻帧光度损失函数; , 具体可以包括以下包括:

[0117] $B11$, 对于当前 t 时刻图像之前一段长度为 N 的图像序列中的非相邻帧图像对, 使用位姿估计网络估计图像序列中图像时间间隔为 2 至 N 的图像对之间的相对相机位姿及亮度对齐参数, 计算对应的基于亮度对齐的光度损失函数值;

[0118] $B12$, 根据图像对之间间隔对此非相邻帧图像对之间的基于亮度对齐的光度损失分配不同权重, 获得整段图像序列的非相邻帧光度损失 L_{na} :

[0119]
$$L_{na} = \sum_{i=0}^{N-2} \sum_{j=i+2}^N \mu \hat{L}_p(I_j, I'_{i \rightarrow j})$$

[0120] 其中, $I'_{i \rightarrow j}$ 表示由第 i 时刻的图像重投影至第 j 时刻的重投影图像, $\hat{L}_p(I_j, I'_{i \rightarrow j})$ 为重投影图像 $I'_{i \rightarrow j}$ 与真实图像 I_j 之间的基于亮度对齐的光度损失, μ 为人工设置的超参数, 其值与输入的两帧图像之间间隔成负相关:

[0121] $\mu = 10^{i-j}$ 。

[0122] $B2$, 将位姿估计网络估计的图像序列中每组相邻图像对的相对相机位姿转化为位姿变换矩阵, 将所有位姿变换矩阵相乘得到图像序列首尾帧图像之间的相对位姿变换矩阵, 使用所述位姿估计网络估计首尾真图像之间的相对相机位姿并转换为相对位姿变换矩阵, 计算这两种方式得到的图像序列首尾帧图像之间的相对位姿变换矩阵之间的 $L1$ 范数获

得图像序列的位姿连续性损失函数,具体可以包括以下包括:

[0123] B21,在相机的运动过程中,连续的几个相对相机位姿之间存在相关性。在一段长度为M的图像序列中,每组相邻图像对 I_i 和 I_{i+1} 之间存在一个位姿变换矩阵 $T^{I_i I_{i+1}}$;其中, $i=0,1,\dots,M-2$, $T^{I_i I_{i+1}}$ 由相邻图像对 I_i 和 I_{i+1} 的相对相机位姿转化而成;

[0124] B22,在相机坐标系下,将所有位姿变换矩阵 $T^{I_i I_{i+1}}$ 连续左乘得到图像序列首尾帧图像之间的相对位姿变换矩阵 $T^{I_0 I_{M-1}}$:

$$[0125] \quad T^{I_0 I_{M-1}} = T^{I_{M-2} I_{M-1}} T^{I_{M-3} I_{M-2}} \dots T^{I_0 I_1}$$

[0126] 同时,直接使用位姿估计网络估计首尾两帧图像之间相机运动产生的相对相机位姿,并使用欧拉变换将其转换为相对位姿变换矩阵 $\hat{T}^{I_0 I_{M-1}}$;

[0127] B23,根据得到的 $T^{I_0 I_{M-1}}$ 与 $\hat{T}^{I_0 I_{M-1}}$ 之间的L1范数,得到长度为M的图像序列的位姿连续性损失 $L_{pc,M}$:

$$[0128] \quad L_{pc,M} = ||T^{I_0 I_{M-1}} - \hat{T}^{I_0 I_{M-1}}||_1$$

[0129] B24,对于当前t时刻图像之前一段长度为N的图像序列,将这段图像序列中采样间隔为2,3, $\dots$,N-2的每个子序列的位姿连续性损失相加获得总的位姿连续性损失:

$$[0130] \quad L_{pc} = \sum_{M=2}^{N-2} L_{pc,M}。$$

[0131] S104,通过位姿优化网络,根据过去时刻使用位姿估计网络估计出的相对相机位姿优化位姿估计网络当前时刻输出的相对相机位姿,并利用位姿优化网络输出的相对相机位姿计算位姿优化损失函数;

[0132] 本实施例中,如图6所示,将当前t时刻之前的n-1时刻位姿估计网络估计出的相对相机位姿 P_{out}^{t-n+1} , $\dots$, P_{out}^{t-1} 与当前t时刻位姿估计网络估计出的相对相机位姿 P_{init}^t 输入位姿优化网络,位姿优化网络输出一个具有更高精度的6自由度相对相机位姿 P_{out}^t :

$$[0133] \quad P_{out}^t = \text{Pose_Optimization}(P_{out}^{t-n+1}, P_{out}^{t-n+2}, \dots, P_{out}^{t-1}, P_{init}^t)$$

[0134] 其中, $t > n-1$,Pose_Optimization()为位姿优化网络, P_{out}^t 包括3自由度的相对旋转和3自由度的相对位移;

[0135] 使用此相对相机位姿进行图像重投影操作,即:使用 P_{out}^t 、 D_{t+1} 将 I_t 重投影至 I'_{t+1} ,并计算相应的基于亮度对齐的光度损失函数,作为训练位姿优化网络的损失函数 L_{p0} 。

[0136] S105,将图像序列中的所有图像输入到位姿估计网络、深度估计网络与位姿优化网络中,基于得到的基于亮度对齐的光度损失函数、运动约束损失函数和位姿优化损失函数对构建的位姿估计网络、深度估计网络与位姿优化网络进行训练;

[0137] 本实施例中,除去上述提到的损失函数外,还使用了深度平滑损失函数 L_s 以及几何一致性约束损失函数 L_{gc} 对深度估计网络进行训练。此外,为了减少或消除训练过程中由于动态场景对于光度损失函数的影响,本实施例中,使用自发现掩膜M对光度损失函数进行逐像素加权,因此,本实施例中,使用的总体损失函数L为:

$$[0138] \quad L = M(\alpha \hat{L}_p + \beta_1 L_{na} + \beta_2 L_{pc} + \gamma L_{p0}) + \delta L_s + \varepsilon L_{gc}$$

[0139] 其中, α 为控制基于亮度对齐的光度损失函数比例的超参数, β_1 为控制非相邻帧光度损失函数比例的超参数, β_2 为控制位姿连续性损失函数比例的超参数, γ 为控制位姿优化损失函数比例的超参数, δ 为深度平滑损失函数比例的超参数, ϵ 为控制几何一致性损失函数比例的超参数。

[0140] 本实施例中, 利用得到的总体损失函数训练位姿估计网络、深度估计网络与位姿优化网络。

[0141] 需要说明的是, 深度平滑损失函数、几何一致性约束损失函数以及自发现掩膜 M 均为其他文献使用方法, 其相关信息如下:

[0142] 在训练深度估计网络时, 基于亮度对齐的光度损失函数在低纹理场景和深度不均匀区域中具有局限性, 对网络的约束不足, 对此引入了深度平滑度损失作为自监督视觉里程计算法中的额外约束。深度平滑度损失可确保估计的深度图显示出与相应输入图像相似的纹理和边缘特征。深度平滑度损失 L_s 可以表示为:

$$[0143] \quad L_s = \sum_p ((e^{-\nabla I_t(p)} \cdot \nabla D_t(p))^2)$$

[0144] 其中, ∇ 表示图像沿空间方向的一阶导数, D_t 为深度估计网络输出的 t 时刻输入图像 I_t 对应的深度图像。

[0145] 在自监督视觉里程计方法中, 由于缺乏尺度约束, 深度估计网络同一序列中不同帧的深度估计值会有尺度差异。这种差异性会导致相机自我运动估计不准确。为了解决这个问题, 本实施例中引入了几何一致性约束损失函数 L_{GC} 对不同帧之间深度估计进行一致性约束:

$$[0146] \quad L_{GC} = \frac{1}{|V|} \sum_{p \in V} \frac{|D_{t+1}^t(p) - D'_{t+1}(p)|}{D_{t+1}^t(p) + D'_{t+1}(p)}$$

[0147] 其中, V 表示在图像重投影过程中有效投影的像素点的集合, D_{t+1}^t 是使用视图合成方法通过 D_t 合成的 $t+1$ 时刻的深度图像, D'_{t+1} 是通过 D_{t+1} 插值的深度图。

[0148] 动态场景会违背自监督算法中图像重投影过程中的几何假设, 本实施例中, 使用一个自发现掩膜 M 来处理场景中的运动物体与遮挡:

$$[0149] \quad M = 1 - \frac{|D_{t+1}^t - D'_{t+1}|}{D_{t+1}^t + D'_{t+1}}$$

[0150] S106, 利用训练好的位姿估计网络与位姿优化网络估计待估计位姿的图像序列中每帧图像对应的相对相机位姿。

[0151] 本实施例中, 在训练时, 将一个批次的视频图像序列中的所有图像输入到位姿估计网络、深度估计网络与位姿优化网络中, 对三个网络同时进行训练。训练完成后, 利用训练好的位姿估计网络与位姿优化网络估计待估计位姿的图像序列中每帧图像对应的相机位姿, 深度估计网络不参与此过程。

[0152] 本发明实施例所述的基于亮度对齐的多阶段估计单目视觉里程计方法, 至少具有以下优点:

[0153] (1) 自监督视觉里程计算法基于灰度不变假设, 然而在实际应用中, 由于环境光照

变化或摄像机在移动过程中的曝光调整,可能会导致相邻帧的亮度不一致,从而导致算法的性能受限甚至算法失灵。对此,本实施例设计了一种基于亮度对其的位姿估计网络,位姿估计网络在训练过程中考虑到了输入图像亮度不一致的情况,在对待估计位姿的图像序列进行相对相机位姿估计的过程中能够隐式地对齐输入图像的亮度信息,这样,使得位姿估计网络不仅能估计相对相机位姿,还能调整两幅输入图像的亮度,确保了输入的图像能够满足灰度不变假设,进而能够提高算法在具有明显亮度变化场景下的鲁棒性并提高算法精度;

[0154] (2) 为了更大程度地利用摄像机运动过程中拍摄的图像序列中的隐含的时间信息,本实施例中提出了一种基于多阶段估计策略的相对相机位姿估计方法。在这种方法中,位姿估计网络输出当前时刻相对相机位姿的近似估计,随后位姿优化网络通过整合时间信息进行微调,以获得更高精度的相对相机位姿估计结果,本实施例适用于用于自监督单目视觉里程计;

[0155] (3) 针对摄像机运动过程的内在特性引入了两个新的损失函数,旨在提高运动估计的一致性和连续性。第一个损失函数是非相邻帧光度损失,其核心是计算非相邻帧之间的光度误差,该损失函数有助于在采样间隔不同的图像对中对相对相机位姿进行尺度一致的估计。第二个损失函数被命名为位姿连续性损失,其目的是建立连续相对相机位置之间的相关性,以减小算法运行过程中产生的累积误差。

[0156] 为了验证本发明实施例提供的基于亮度对齐的多阶段估计单目视觉里程计方法的有效性,使用KITTI里程计数据集中提供的评估指标测试其性能:

[0157] (1) 相对位移均方误差 (Rel.trans.): 一个序列中全部长度为100、200、……、800米的子序列的平均位移RMSE (Root Mean Square Error), 以%度量, 即每100米偏差的米数, 数值越小越好。

[0158] (2) 相对旋转均方误差 (Rel.rot.): 一个序列中全部长度为100、200、……、800米的子序列的平均旋转RMSE, 以deg/m度量, 数值越小越好。

[0159] 本实施例中, 应用了KITTI里程计数据集中00-07这八个序列作为训练集与验证集训练位姿估计网络、深度估计网络与位姿优化网络, 并用09-10这两个序列来测试所述方法的性能。

[0160] KITTI里程计数据集是车载相机等设备采集的城市中公路环境的双目图像, 雷达点以及实际轨迹。

[0161] 在实施过程中, 执行S101-S106的操作, 其中, 光度误差损失函数的超参数的参数 $\alpha=1, \beta_1=0.25, \beta_2=0.25, \gamma=0.2, \delta=0.1, \delta=0.5$; 光度损失函数中 $\mu_0=0.85, \mu_1=0.15$ 。网络的训练过程中, 初始学习率为 3×10^{-4} , 并随着训练的进行逐渐减小, 采用Adam优化器进行150次迭代, 每轮迭代的批量大小为8, 运动约束损失函数中 $N=4$, 位姿优化网络的输入位姿数量 $n=5$ 。

[0162] 为了验证本实施例所述方法的性能, 本实施例中, 选择了近几年基于深度学习的自监督的单目视觉里程计方法进行了对比, 实验结果如表2所示。本实施例生成的轨迹如图7(a)、图7(b)所示, 其中, Groundtruth为真实的轨迹, Ours为本实施例中估计出的轨迹, SC-SfMLearner是一种基础方法估计出的轨迹。

[0163] 表2本实施例的方法与其他方法对比

Methods	Seq. 09			Seq. 10		
	t_{err} (%)	r_{err} (°/m)	ATE(m)	t_{err} (%)	r_{err} (°/m)	ATE(m)
SfMLearner	8.28	0.031	77.79	12.20	0.030	67.34
GeoNet	28.72	0.098	158.45	23.90	0.090	43.04
DepthVOFeat	11.93	0.039	52.12	12.82	0.035	24.7
Vid2depth	-	-	-	21.54	0.125	-
UnDeepVO	7.01	0.036	-	10.63	0.046	-
Wang et al.	9.88	0.034	-	12.24	0.052	-
[0164] Ranjan et al.	6.92	0.018	-	7.97	0.031	-
DeepMatchVO	9.91	0.038	27.08	12.18	0.059	24.44
PoseGraph	8.1	0.028	14.85	12.9	0.032	11.2
MonoDepth2	11.47	0.032	55.47	7.73	0.034	20.46
SC-SfMLearner	11.2	0.034	83.27	10.10	0.050	21.19
Jau et al.	8.64	0.007	-	11.72	0.010	-
SC-Depth	7.31	0.031	23.56	7.79	0.050	12.00
Liang et al.	8.13	0.030	-	6.76	0.024	-
Xiu et al.	6.297	0.0249	17.276	6.97	0.032	17.344
Ours	5.25	0.0034	13.475	6.19	0.0087	10.237

[0165] 为了验证本实施例所述的方法各部分的意义,本实施例中还进行了消融实验。首先验证了亮度对齐对于相对相机位姿估计精度提升的帮助。实验结果如图8所示。接着验证了为基础方法增加本方法提出的各项改进后的性能指标。表2展示了本实施例所提出的各模块对提高方法性能的意义;其中,“basic”表示不增加任何改进的基础方法;“+BA”表示在算法中加入位姿估计网络;“+MC”表示在算法中加入运动约束损失函数;“+PO”表示在算法中加入位姿优化网络。可以看出,在基础方法中引入本实施例提出的各种模块均能提高算法对于相对相机位姿的估计精度,算法性能随着各个部分的增加而逐渐上升,证明了本实施例的方法中各个部分的有效性。

[0166] 表3消融实验结果

Methods	Seq. 09			Seq. 10		
	t_{err} (%)	r_{err} (°/m)	ATE(m)	t_{err} (%)	r_{err} (°/m)	ATE(m)
Basic	15.16	0.0489	60.247	13.51	0.0401	20.134
[0167] Basic+PO	12.59	0.0320	48.239	11.01	0.0309	18.382
Basic+BA	9.20	0.0205	25.135	11.60	0.0289	15.367
Basic+MC	7.88	0.0215	22.347	7.152	0.0247	12.541
Basic+BA+MC	5.15	0.0039	14.788	6.29	0.0096	10.276
Basic+BA+MC+PO	5.25	0.0034	13.475	6.01	0.0087	10.227

[0168] 此外,本实施例还在KITTI数据集的部分其它序列(11-21序列,未提供真实相机位姿)上对本实施例提供的方法进行了附加的测试实验,由于这些序列没有提供真实的相机位姿,使用带闭环的ORB-SLAM2的结果作为参考。在这些序列中,选择了四个序列(11、15、16和17)进行附加实验。附加实验结果如图9(a)-(d)所示。图9(a)-(d)中虚线和实线分别表示ORB-SLAM和本实施例提出的方法得到的轨迹。

[0169] 图10是本发明实施例提供的一种电子设备600的结构示意图,该电子设备600可因配置或性能不同而产生比较大的差异,可以包括一个或一个以上处理器(central

processing units,CPU) 601和一个或一个以上的存储器602,其中,所述存储器602中存储有至少一条指令,所述至少一条指令由所述处理器601加载并执行以实现上述基于亮度对齐的多阶段估计单目视觉里程计方法。

[0170] 在示例性实施例中,还提供了一种计算机可读存储介质,例如包括指令的存储器,上述指令可由终端中的处理器执行以完成上述基于亮度对齐的多阶段估计单目视觉里程计方法。例如,所述计算机可读存储介质可以是ROM、随机存取存储器(RAM)、CD-ROM、磁带、软盘和光数据存储设备等。

[0171] 本领域普通技术人员可以理解实现上述实施例的全部或部分步骤可以通过硬件来完成,也可以通过程序来指令相关的硬件完成,所述的程序可以存储于一种计算机可读存储介质中,上述提到的存储介质可以是只读存储器,磁盘或光盘等。

[0172] 以上所述仅为本发明的较佳实施例,并不用以限制本发明,凡在本发明的精神和原则之内,所作的任何修改、等同替换、改进等,均应包含在本发明的保护范围之内。

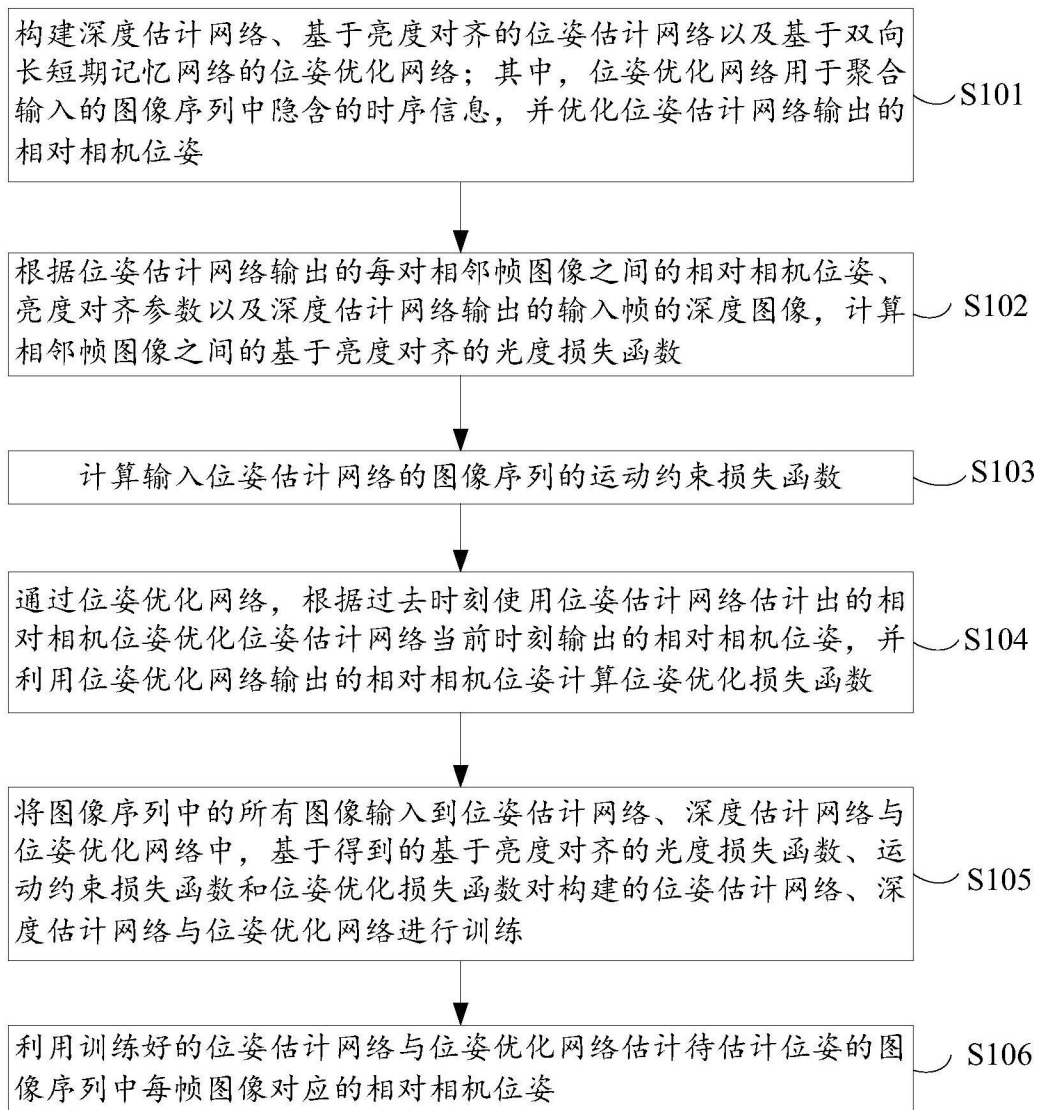


图1

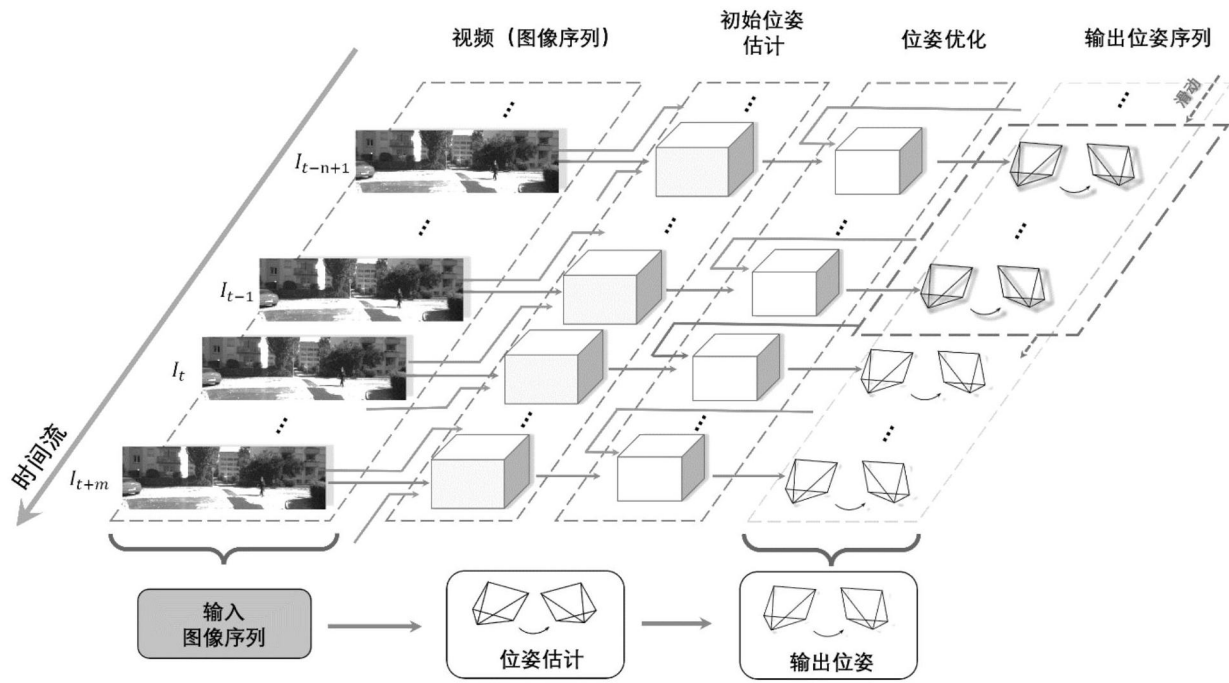


图2

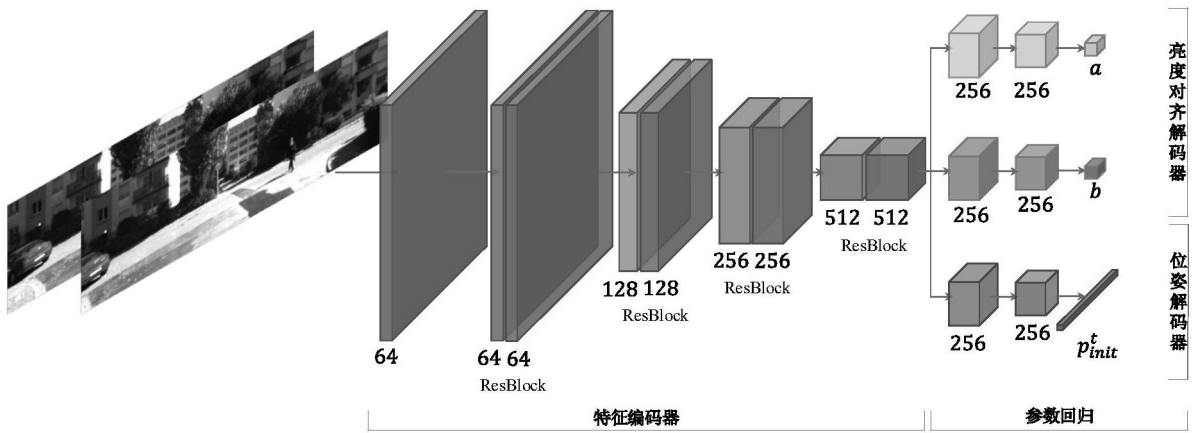


图3

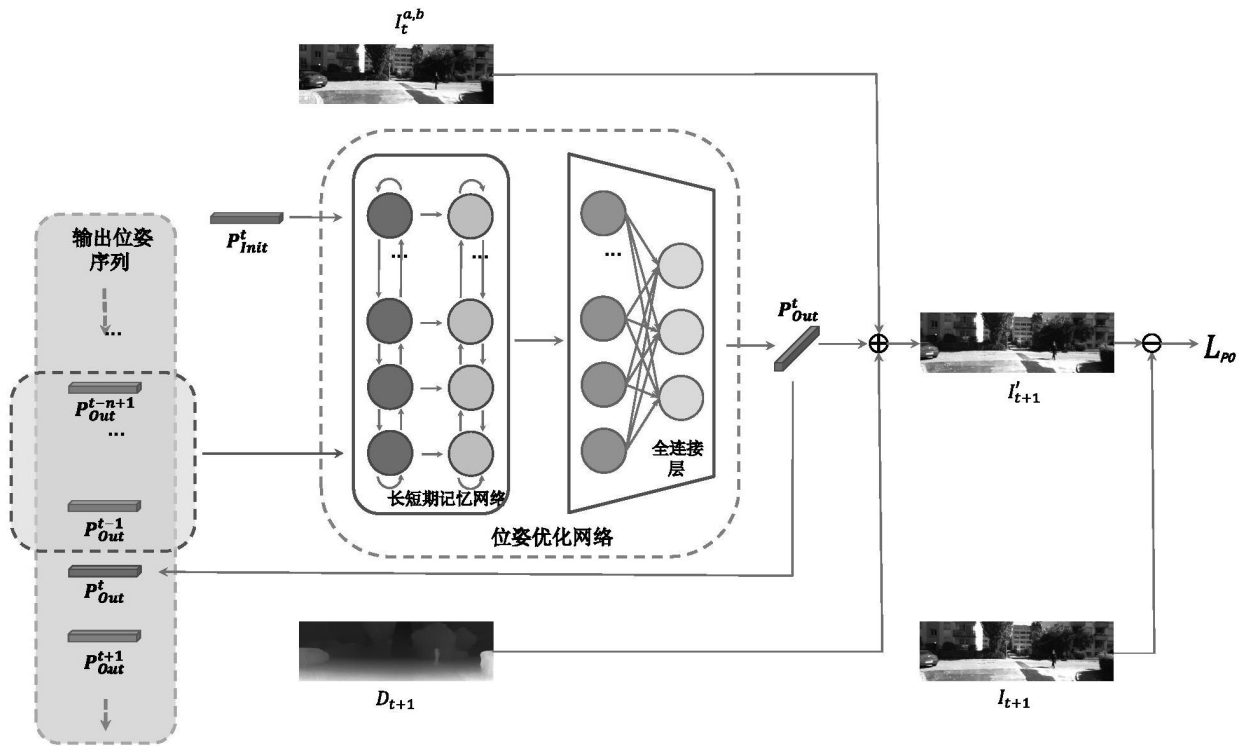


图4

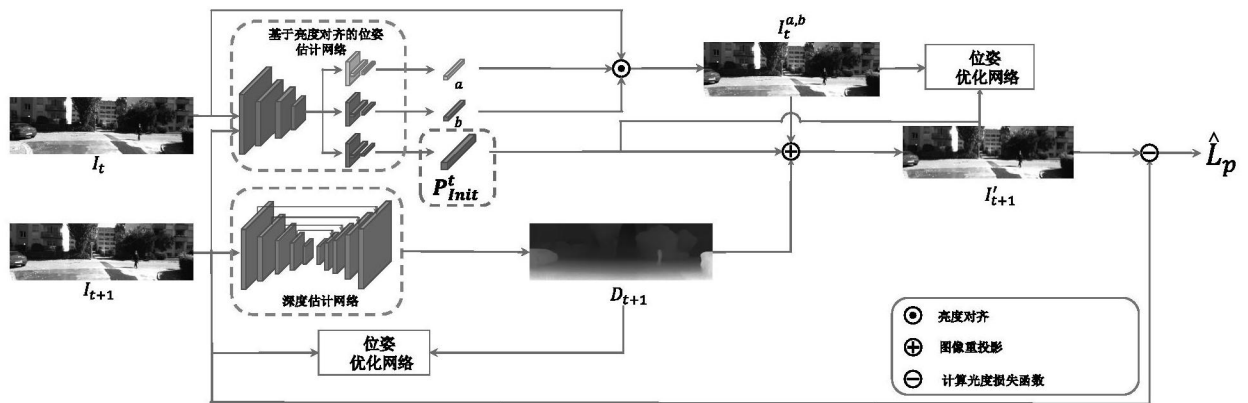


图5

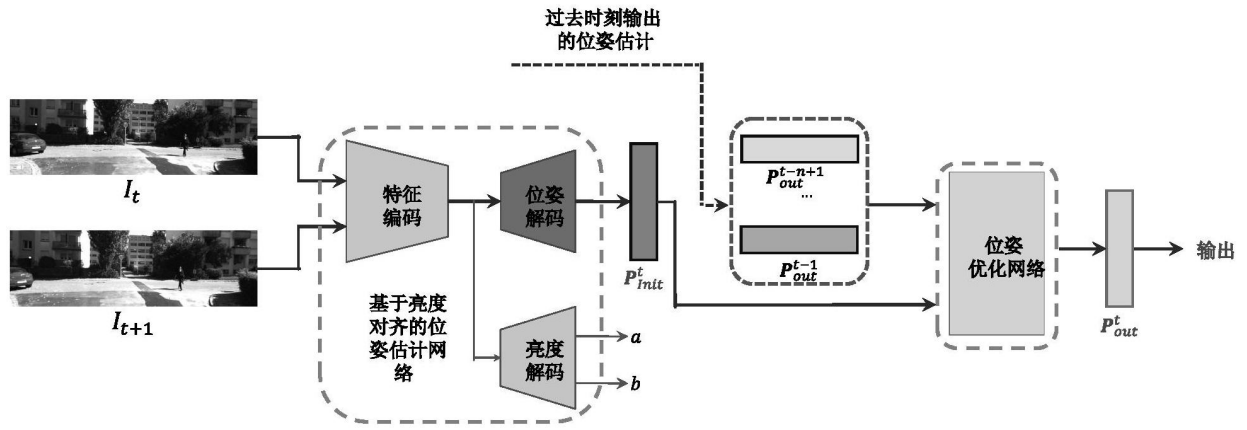


图6

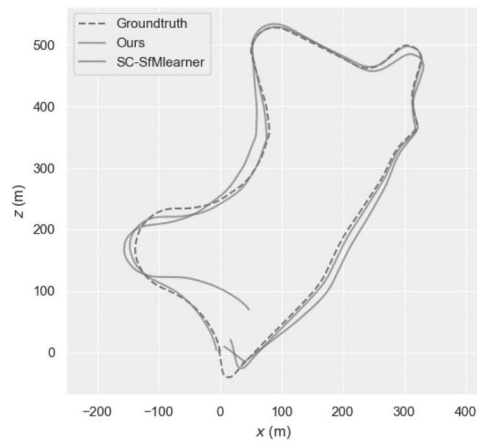


图7(a)

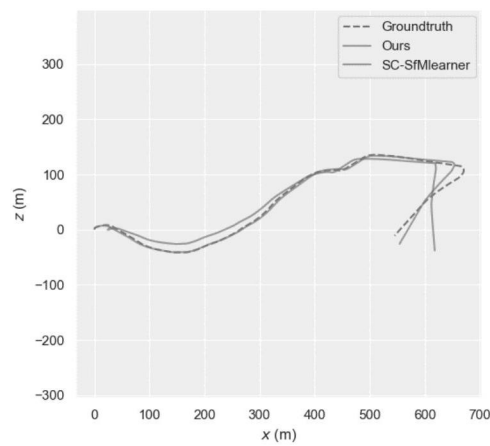


图7(b)



图8

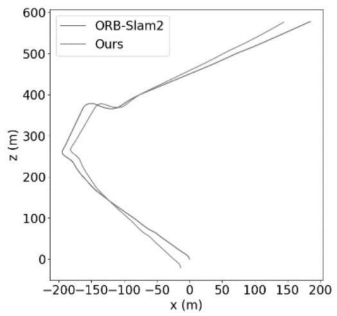


图9(a)

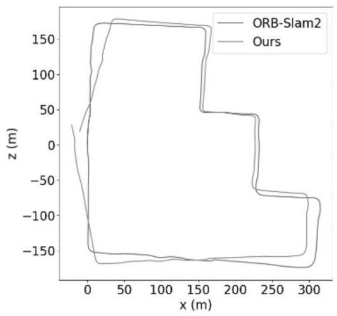


图9(b)

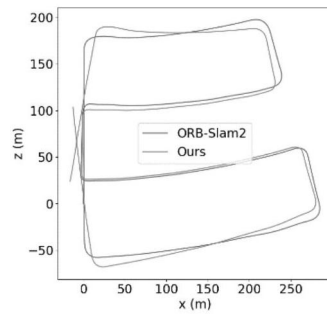


图9(c)

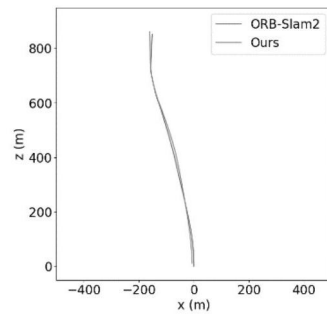


图9(d)

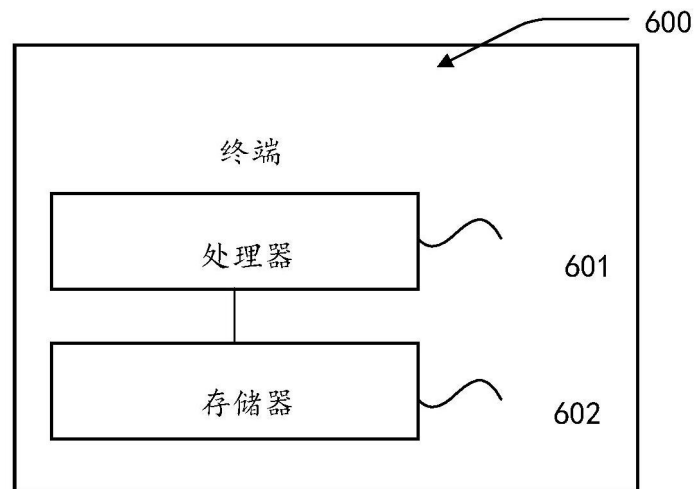


图10