



(12) 发明专利

(10) 授权公告号 CN 116758131 B

(45) 授权公告日 2023. 11. 28

(21) 申请号 202311050584.9

CN 116342802 A, 2023.06.27

(22) 申请日 2023.08.21

CN 115953447 A, 2023.04.11

(65) 同一申请的已公布的文献号

CN 111402310 A, 2020.07.10

申请公布号 CN 116758131 A

CN 111311664 A, 2020.06.19

(43) 申请公布日 2023.09.15

CN 115953468 A, 2023.04.11

(73) 专利权人 之江实验室

CN 111340867 A, 2020.06.26

地址 311121 浙江省杭州市余杭区中泰街

US 2021312650 A1, 2021.10.07

道科创大道之江实验室

US 2022392083 A1, 2022.12.08

(72) 发明人 邱奇波 华炜 李碧清 高海明

US 2022084230 A1, 2022.03.17

(74) 专利代理机构 杭州华进联浙知识产权代理

孙业昊. 基于无监督学习的单目深度估计研究.《中国优秀硕士学位论文全文数据库信息科技辑》.2021, (第6期), 第1138-706页.

有限公司 33250

Clement Godard et al.. Digging Into

专利代理师 方道杰

Self-Supervised Monocular Depth

(51) Int. Cl.

Estimation.《Proceedings of the IEEE/CVF

G06T 7/55 (2017.01)

International Conference on Computer

G06N 3/0455 (2023.01)

Vision (ICCV)》.2019, 第3828-3838页. (续)

G06N 3/08 (2023.01)

审查员 何华兵

(56) 对比文件

CN 111311685 A, 2020.06.19

权利要求书4页 说明书13页 附图4页

(54) 发明名称

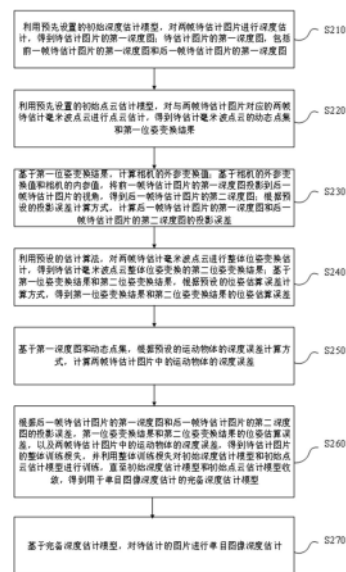
果的稳定性。

一种单目图像深度估计方法、装置和计算机设备

(57) 摘要

本申请涉及一种单目图像深度估计方法、装置和计算机设备。所述方法包括：得到待估计图片的第一深度图；得到待估计毫米波点云的动态点集和第一位姿变换结果；得到后一帧待估计图片的第二深度图；计算后一帧待估计图片的第一深度图和第二深度图的投影误差；得到待估计毫米波点云的第二位姿变换结果；得到第一位姿变换结果和第二位姿变换结果的位姿估算误差；计算两帧待估计图片中的运动物体的深度误差；根据投影误差、位姿估算误差以及深度误差，得到整体训练损失，并利用整体训练损失对初始模型进行训练，直至收敛，得到完备深度估计模型，对待估计的图片进行单目图像深度估计。利用完备深度估计模型可以保证实现图像的深度估计结

CN 116758131 B



[接上页]

**(56) 对比文件**

Akash Deep Singh et al..Depth  
Estimation From Camera Image and mmWave

Radar Point Cloud.《Proceedings of the  
IEEE/CVF Conference on Computer Vision  
and Pattern Recognition (CVPR)》.2023,第  
9275-9285页.

1. 一种单目图像深度估计方法,其特征在于,所述方法包括:

利用预先设置的初始深度估计模型,对两帧待估计图片进行深度估计,得到所述待估计图片的第一深度图;所述待估计图片的第一深度图,包括前一帧待估计图片的第一深度图和后一帧待估计图片的第一深度图;

利用预先设置的初始点云估计模型,对与两帧所述待估计图片对应的两帧待估计毫米波点云进行点云估计,得到所述待估计毫米波点云的动态点集和第一位姿变换结果;

基于所述第一位姿变换结果,计算相机的外参变换值;基于所述相机的外参变换值和所述相机的内参值,将所述前一帧待估计图片的第一深度图投影到所述后一帧待估计图片的视角,得到所述后一帧待估计图片的第二深度图;根据预设的投影误差计算方式,计算所述后一帧待估计图片的第一深度图和所述后一帧待估计图片的第二深度图的投影误差 $L_1$ ;

$L_1 = \frac{\partial}{2}(1 - \text{SSIM}(D_T, D_{T-1 \rightarrow T})) + (1 - \partial) \|D_T - D_{T-1 \rightarrow T}\|_1$ , 其中, $D_T$ 表示T时刻的待估计图片的第一深度图, $D_{T-1 \rightarrow T}$ 表示将T-1时刻的待估计图片的第一深度图投影到T时刻的待估计图片的视角,得到的T时刻的待估计图片的第二深度图,SSIM表示利用SSIM损失函数来计算T时刻的待估计图片的第一深度图和T时刻的待估计图片的第二深度图的投影误差的损失, $\partial$ 为预先设定的参数;

利用预设的估计算法,对两帧所述待估计毫米波点云进行整体位姿变换估计,得到所述待估计毫米波点云整体位姿变换的第二位姿变换结果;基于所述第一位姿变换结果和所述第二位姿变换结果,根据预设的位姿估算误差计算方式,得到所述第一位姿变换结果和所述第二位姿变换结果的位姿估算误差 $L_2$ ;  $L_2 = \|TR_1 - TR_2\|_2^2$ , 其中, $TR_1$ 表示所述第一位姿变换结果, $TR_2$ 表示所述第二位姿变换结果;

基于所述第一深度图和所述动态点集,根据预设的运动物体的深度误差计算方式,计算两帧所述待估计图片中的运动物体的深度误差 $L_3$ ;

$L_3 = \sum_{p \in RD_T, q \in RD_{T-1}} \text{Sim}(p, q) \cdot \pi(\cos(\text{Loc}_q, TR_{T-1 \rightarrow T}[:, 3])) \cdot \pi(D_T(p) < D_{T-1}(q))$ , 其中, $RD_{T-1}$ 表示T-1时刻的动态点集, $RD_T$ 表示T时刻的动态点集, $p, q$ 表示 $RD_{T-1}$ 和 $RD_T$ 中任意的点对, $\text{Sim}(p, q)$ 表示 $p, q$ 来自于同一个障碍物的概率, $\pi(\cos(\text{Loc}_q, TR_{T-1 \rightarrow T}[:, 3]))$ 为指示函数, $\text{Loc}_q$ 表示 $q$ 点的三维空间坐标, $TR_{T-1 \rightarrow T}$ 表示所述待估计毫米波点云从T-1时刻到T时刻的第一位姿变换结果, $\pi(D_T(p) < D_{T-1}(q))$ 为指示函数, $D_T(p)$ 表示 $p$ 在T时刻的待估计图像的第一深度图中的深度值, $D_{T-1}(q)$ 表示 $q$ 在T-1时刻的待估计图像的第一深度图中的深度值;

根据所述后一帧待估计图片的第一深度图和所述后一帧待估计图片的第二深度图的投影误差,所述第一位姿变换结果和所述第二位姿变换结果的位姿估算误差,以及两帧所述待估计图片中的所述运动物体的深度误差,得到所述待估计图片的整体训练损失 $L$ ,并利用所述整体训练损失对所述初始深度估计模型和所述初始点云估计模型进行训练,直至所述初始深度估计模型和所述初始点云估计模型收敛,得到用于单目图像深度估计的完备深度估计模型;

$L = L_1 + \beta \cdot L_2 + (1 - \beta) \cdot L_3$ , 其中, $\beta = 1 - \frac{ep}{\text{Max\_epoch}}$ ,  $ep$ 表示当前训练的轮次,Max\_epoch表示最大训练轮次;

基于所述完备深度估计模型,对待估计的图片进行单目图像深度估计。

2. 根据权利要求1所述的单目图像深度估计方法,其特征在于,在所述利用预先设置的初始深度估计模型,对两帧待估计图片进行深度估计,得到所述待估计图片的第一深度图之前,包括:

对待估计的两帧原始图片进行减均值除方差操作,生成两帧第一图片;

采用预设的缩放方法,将两帧所述第一图片缩放至预设的大小,得到缩放后的两帧所述待估计图片。

3. 根据权利要求1所述的单目图像深度估计方法,其特征在于,所述利用预先设置的初始深度估计模型,对两帧待估计图片进行深度估计,得到所述待估计图片的第一深度图,包括:

利用预先设置的所述初始深度估计模型的深度编码网络,获取所述待估计图片的深度特征;

利用预先设置的所述初始深度估计模型的深度解码网络,对获取到的所述深度特征进行深度估计相关的特征抽取,得到所述待估计图片的逆深度图;

对所述逆深度图进行取倒数处理,得到所述待估计图片的第一深度图。

4. 根据权利要求1所述的单目图像深度估计方法,其特征在于,所述利用预先设置的初始点云估计模型,对与两帧所述待估计图片对应的两帧待估计毫米波点云进行点云估计,得到所述待估计毫米波点云的动态点集和第一位姿变换结果,包括:

利用预先设置的所述初始点云估计模型的场景流预测网络,获取所述待估计毫米波点云的场景流;

根据预设的动态点的筛选条件,筛选出所述场景流中满足平移偏离量大于或等于平均平移偏移量一倍方差的动态点,得到所述待估计毫米波点云的动态点集;

利用预先设置的所述初始点云估计模型的位姿估计网络,获取所述待估计毫米波点云的一行六列矩阵;

利用预设的旋转公式,将所述一行六列矩阵转换为三行四列矩阵,得到所述待估计的毫米波点云的所述第一位姿变换结果。

5. 根据权利要求1所述的单目图像深度估计方法,其特征在于,所述基于所述第一位姿变换结果,计算相机的外参变换值;基于所述相机的外参变换值和所述相机的内参值,将所述前一帧待估计图片的第一深度图投影到所述后一帧待估计图片的视角,得到所述后一帧待估计图片的第二深度图,包括:

基于所述第一位姿变换结果,以及预设的毫米波雷达到所述相机的外参值,得到两帧所述待估计图片所对应的所述相机的外参变换值;

基于所述相机的外参变换值和所述相机的内参值,将所述前一帧待估计图片的第一深度图投影到所述后一帧待估计图片的视角,基于投影结果,得到所述后一帧待估计图片的第二深度图。

6. 根据权利要求1所述的单目图像深度估计方法,其特征在于,所述利用预设的估计算法,对两帧所述待估计毫米波点云进行整体位姿变换估计,得到所述毫米波点云整体位姿变换的第二位姿变换结果,包括:

利用ICP算法,对两帧所述待估计毫米波点云进行整体位姿变换估计,得到所述待估计

毫米波点云整体位姿变换的第二位姿变换结果。

7. 一种单目图像深度估计装置,其特征在于,所述装置包括:

深度估计模块,用于利用预先设置的初始深度估计模型,对两帧待估计图片进行深度估计,得到所述待估计图片的第一深度图;所述待估计图片的第一深度图,包括前一帧待估计图片的第一深度图和后一帧待估计图片的第一深度图;

点云估计模块,用于利用预先设置的初始点云估计模型,对与两帧所述待估计图片对应的两帧待估计毫米波点云进行点云估计,得到所述待估计毫米波点云的动态点集和第一位姿变换结果;

第一计算模块,用于基于所述第一位姿变换结果,计算相机的外参变换值;基于所述相机的外参变换值和所述相机的内参值,将所述前一帧待估计图片的第一深度图投影到所述后一帧待估计图片的视角,得到所述后一帧待估计图片的第二深度图;根据预设的投影误差计算方式,计算所述后一帧待估计图片的第一深度图和所述后一帧待估计图片的第二深度图的投影误差 $L_1$ ;

$$L_1 = \frac{\partial}{2} (1 - \text{SSIM}(D_T, D_{T-1 \rightarrow T})) + (1 - \partial) \|D_T - D_{T-1 \rightarrow T}\|_1,$$

其中, $D_T$ 表示T时刻的待估计图片的第一深度图, $D_{T-1 \rightarrow T}$ 表示将T-1时刻的待估计图片的第一深度图投影到T时刻的待估计图片的视角,得到的T时刻的待估计图片的第二深度图,SSIM表示利用SSIM损失函数来计算T时刻的待估计图片的第一深度图和T时刻的待估计图片的第二深度图的投影误差的损失, $\partial$ 为预先设定的参数;

第二计算模块,用于利用预设的估计算法,对两帧所述待估计毫米波点云进行整体位姿变换估计,得到所述待估计毫米波点云整体位姿变换的第二位姿变换结果;基于所述第一位姿变换结果和所述第二位姿变换结果,根据预设的位姿估算误差计算方式,得到所述第一位姿变换结果和所述第二位姿变换结果的位姿估算误差 $L_2$ ;

$$L_2 = \|TR_1 - TR_2\|_2^2,$$

其中, $TR_1$ 表示所述第一位姿变换结果, $TR_2$ 表示所述第二位姿变换结果;

第三计算模块,用于基于所述第一深度图和所述动态点集,根据预设的运动物体的深度误差计算方式,计算两帧所述待估计图片中的运动物体的深度误差 $L_3$ ;

$$L_3 = \sum_{p \in RD_T, q \in RD_{T-1}} \text{Sim}(p, q) \cdot \pi(\cos(\text{Loc}_q, TR_{T-1 \rightarrow T}[:, 3])) \cdot \pi(D_T(p) < D_{T-1}(q)),$$

其中, $RD_{T-1}$ 表示T-1时刻的动态点集, $RD_T$ 表示T时刻的动态点集, $p, q$ 表示 $RD_{T-1}$ 和 $RD_T$ 中任意的点对, $\text{Sim}(p, q)$ 表示 $p, q$ 来自于同一个障碍物的概率, $\pi(\cos(\text{Loc}_q, TR_{T-1 \rightarrow T}[:, 3]))$ 为指示函数, $\text{Loc}_q$ 表示q点的三维空间坐标, $TR_{T-1 \rightarrow T}$ 表示所述待估计毫米波点云从T-1时刻到T时刻的第一位姿变换结果, $\pi(D_T(p) < D_{T-1}(q))$ 为指示函数, $D_T(p)$ 表示p在T时刻的待估计图像的第一深度图中的深度值, $D_{T-1}(q)$ 表示q在T-1时刻的待估计图像的第一深度图中的深度值;

训练模块,用于根据所述后一帧待估计图片的第一深度图和所述后一帧待估计图片的第二深度图的投影误差,所述第一位姿变换结果和所述第二位姿变换结果的位姿估算误差,以及两帧所述待估计图片中的所述运动物体的深度误差,得到所述待估计图片的整体训练损失L,并利用所述整体训练损失对所述初始深度估计模型和所述初始点云估计模型进行训练,直至所述初始深度估计模型和所述初始点云估计模型收敛,得到用于单目图像深度估计的完备深度估计模型; $L = L_1 + \beta \cdot L_2 + (1 - \beta) \cdot L_3$ ,其中,

$\beta = 1 - \frac{ep}{\text{Max\_epoch}}$ , ep表示当前训练的轮次, Max\_epoch表示最大训练轮次;

以及估计模块,用于基于所述完备深度估计模型,对待估计的图片进行单目图像深度估计。

8.一种计算机设备,包括存储器和处理器,所述存储器存储有计算机程序,其特征在于,所述处理器执行所述计算机程序时实现权利要求1至权利要求6中任一项所述的单目图像深度估计方法的步骤。

## 一种单目图像深度估计方法、装置和计算机设备

### 技术领域

[0001] 本申请涉及图像处理技术领域,特别是涉及一种单目图像深度估计方法、装置和计算机设备。

### 背景技术

[0002] 三维环境感知是移动机器人和无人车领域的重要技术,而目前的三维环境感知主要是依赖昂贵的稠密激光雷达来获取环境精确的三维信息。相较于通过稠密的激光雷达来获取环境精确的三维信息,自监督单目图像深度估计通过相机采集到的图像来感知环境的深度,并且不依赖于额外的深度标记,具有一定的成本优势。

[0003] 但是,直接从图像中估计深度是一个不适定的问题,并且,采用主流的深度学习方法不能仅通过相机采集到的图像来精准的估计出图像的绝对深度。因此,现有技术通过引入额外的廉价模态的数据,来增强图像深度的估计能力。具体地,可以通过引入GPS(Global Positioning System,全球定位系统)坐标、IMU(Inertial Measurement Unit,惯性测量单元)数据或稀疏激光雷达数据的方式来约束深度估计。但是,通过引入GPS坐标、IMU数据或稀疏激光雷达数据的方式来约束深度估计,需要去除图像中运动物体,并且依赖静态环境假设,使得存在运动物体的相邻两帧单目图像的深度估计结果抖动很大,无法保证图像的深度估计结果的稳定性。

[0004] 对于现有技术存在无法保证图像的深度估计结果的稳定性的问题,仍没有得到解决。

### 发明内容

[0005] 基于此,有必要针对上述技术问题,提供一种单目图像深度估计方法、装置和计算机设备。

[0006] 第一方面,本申请提供了一种单目图像深度估计方法。所述方法包括以下步骤:

[0007] 利用预先设置的初始深度估计模型,对两帧待估计图片进行深度估计,得到所述待估计图片的第一深度图;所述待估计图片的第一深度图,包括前一帧待估计图片的第一深度图和后一帧待估计图片的第一深度图;

[0008] 利用预先设置的初始点云估计模型,对与两帧所述待估计图片对应的两帧待估计毫米波点云进行点云估计,得到所述待估计毫米波点云的动态点集和第一位姿变换结果;

[0009] 基于所述第一位姿变换结果,计算相机的外参变换值;基于所述相机的外参变换值和所述相机的内参值,将所述前一帧待估计图片的第一深度图投影到所述后一帧待估计图片的视角,得到所述后一帧待估计图片的第二深度图;根据预设的投影误差计算方式,计算所述后一帧待估计图片的第一深度图和所述后一帧待估计图片的第二深度图的投影误差;

[0010] 利用预设的估计算法,对两帧所述待估计毫米波点云进行整体位姿变换估计,得到所述待估计毫米波点云整体位姿变换的第二位姿变换结果;基于所述第一位姿变换结果

和所述第二位姿变换结果,根据预设的位姿估算误差计算方式,得到所述第一位姿变换结果和所述第二位姿变换结果的位姿估算误差;

[0011] 基于所述第一深度图和所述动态点集,根据预设的运动物体的深度误差计算方式,计算两帧所述待估计图片中的运动物体的深度误差;

[0012] 根据所述后一帧待估计图片的第一深度图和所述后一帧待估计图片的第二深度图的投影误差,所述第一位姿变换结果和所述第二位姿变换结果的位姿估算误差,以及两帧所述待估计图片中的所述运动物体的深度误差,得到所述待估计图片的整体训练损失,并利用所述整体训练损失对所述初始深度估计模型和所述初始点云估计模型进行训练,直至所述初始深度估计模型和所述初始点云估计模型收敛,得到用于单目图像深度估计的完备深度估计模型;

[0013] 基于所述完备深度估计模型,对待估计的图片进行单目图像深度估计。

[0014] 在其中一个实施例中,在所述利用预先设置的初始深度估计模型,对两帧待估计图片进行深度估计,得到所述待估计图片的第一深度图之前,包括以下步骤:

[0015] 对待估计的两帧原始图片进行减均值除方差操作,生成两帧第一图片;

[0016] 采用预设的缩放方法,将两帧所述第一图片缩放至预设的大小,得到缩放后的两帧所述待估计图片。

[0017] 在其中一个实施例中,所述利用预先设置的初始深度估计模型,对两帧待估计图片进行深度估计,得到所述待估计图片的第一深度图,包括以下步骤:

[0018] 利用预先设置的所述初始深度估计模型的深度编码网络,获取所述待估计图片的深度特征;

[0019] 利用预先设置的所述初始深度估计模型的深度解码网络,对获取到的所述深度特征进行深度估计相关的特征抽取,得到所述待估计图片的逆深度图;

[0020] 对所述逆深度图进行取倒数处理,得到所述待估计图片的第一深度图。

[0021] 在其中一个实施例中,所述利用预先设置的初始点云估计模型,对与两帧所述待估计图片对应的两帧待估计毫米波点云进行点云估计,得到所述待估计毫米波点云的动态点集和第一位姿变换结果,包括以下步骤:

[0022] 利用预先设置的所述初始点云估计模型的场景流预测网络,获取所述待估计毫米波点云的场景流;

[0023] 根据预设的动态点的筛选条件,筛选出所述场景流中满足平移偏离量大于或等于平均平移偏移量一倍方差的动态点,得到所述待估计毫米波点云的动态点集;

[0024] 利用预先设置的所述初始点云估计模型的位姿估计网络,获取所述待估计毫米波点云的一行六列矩阵;

[0025] 利用预设的旋转公式,将所述一行六列矩阵转换为三行四列矩阵,得到所述待估计的毫米波点云的所述第一位姿变换结果。

[0026] 在其中一个实施例中,所述基于所述第一位姿变换结果,计算相机的外参变换值;基于所述相机的外参变换值和所述相机的内参值,将所述前一帧待估计图片的第一深度图投影到所述后一帧待估计图片的视角,得到所述后一帧待估计图片的第二深度图,包括以下步骤:

[0027] 基于所述第一位姿变换结果,以及预设的毫米波雷达到所述相机的外参值,得到

两帧所述待估计图片所对应的相机的外参变换值；

[0028] 基于所述相机的外参变换值和所述相机的内参值，将所述前一帧待估计图片的第一深度图投影到所述后一帧待估计图片的视角，基于投影结果，得到所述后一帧待估计图片的第二深度图。

[0029] 在其中一个实施例中，所述后一帧待估及图片的第一深度图和所述后一帧待估及图片的第二深度图的投影误差 $L_1$ 的计算公式为：

$$[0030] \quad L_1 = \frac{\theta}{2} (1 - \text{SSIM}(D_T, D_{T-1 \rightarrow T})) + (1 - \theta) \|D_T - D_{T-1 \rightarrow T}\|_1$$

[0031] 其中， $D_T$ 表示T时刻的待估计图片的第一深度图， $D_{T-1 \rightarrow T}$ 表示将T-1时刻的待估计图片的第一深度图投影到T时刻的待估计图片的视角，得到的T时刻的待估计图片的第二深度图，SSIM表示投影误差的损失， $\theta$ 为预先设定的参数。

[0032] 在其中一个实施例中，所述利用预设的估计算法，对两帧所述待估计毫米波点云进行整体位姿变换估计，得到所述毫米波点云整体位姿变换的第二位姿变换结果，包括：

[0033] 利用ICP算法，对两帧所述待估计毫米波点云进行整体位姿变换估计，得到所述待估计毫米波点云整体位姿变换的第二位姿变换结果。

[0034] 在其中一个实施例中，所述第一位姿变换结果和所述第二位姿变换结果的位姿估算误差 $L_2$ 的计算公式为：

$$[0035] \quad L_2 = \|TR_1 - TR_2\|_2^2$$

[0036] 其中， $TR_1$ 表示所述第一位姿变换结果， $TR_2$ 表示所述第二位姿变换结果。

[0037] 在其中一个实施例中，两帧所述待估计图片中的运动物体的深度误差 $L_3$ 的计算公式为：

$$[0038] \quad L_3 = \sum_{p \in RD_T, q \in RD_{T-1}} \text{Sim}(p, q) \cdot \pi(\cos(\text{Loc}_q, TR_{T-1 \rightarrow T}[:, 3])) \cdot \pi(D_T(p) < D_{T-1}(q))$$

[0039] 其中， $RD_{T-1}$ 表示T-1时刻的动态点集， $RD_T$ 表示T时刻的动态点集， $p, q$ 表示 $RD_{T-1}$ 和 $RD_T$ 中任意的点对， $\text{Sim}(p, q)$ 表示 $p, q$ 来自于同一个障碍物的概率， $\pi(\cos(\text{Loc}_q, TR_{T-1 \rightarrow T}[:, 3]))$ 为指示函数， $\text{Loc}_q$ 表示 $q$ 的三维空间坐标， $TR_{T-1 \rightarrow T}$ 表示所述待估计毫米波点云从T-1时刻到T时刻的第一位姿变换结果， $\pi(D_T(p) < D_{T-1}(q))$ 为指示函数， $D_T(p)$ 表示 $p$ 在T时刻的待估计图像的第一深度图中的深度值， $D_{T-1}(q)$ 表示 $q$ 在T-1时刻的待估计图像的第一深度图中的深度值。

[0040] 在其中一个实施例中，所述待估计图片的整体训练损失公式 $L$ 为：

$$[0041] \quad L = L_1 + \beta \cdot L_2 + (1 - \beta) \cdot L_3$$

[0042] 其中， $\beta = 1 - \frac{ep}{\text{Max\_epoch}}$ ， $ep$ 表示当前训练的轮次， $\text{Max\_epoch}$ 表示最大训练轮次。

[0043] 第二方面，本申请还提供了一种单目图像深度估计装置。所述装置包括：

[0044] 深度估计模块，用于利用预先设置的初始深度估计模型，对两帧待估计图片进行深度估计，得到所述待估计图片的第一深度图；所述待估计图片的第一深度图，包括前一帧待估计图片的第一深度图和后一帧待估计图片的第一深度图；

[0045] 点云估计模块，用于利用预先设置的初始点云估计模型，对与两帧所述待估计图

片对应的两帧待估计毫米波点云进行点云估计,得到所述待估计毫米波点云的动态点集和第一位姿变换结果;

[0046] 第一计算模块,用于基于所述第一位姿变换结果,计算相机的外参变换值;基于所述相机的外参变换值和所述相机的内参值,将所述前一帧待估计图片的第一深度图投影到所述后一帧待估计图片的视角,得到所述后一帧待估计图片的第二深度图;根据预设的投影误差计算方式,计算所述后一帧待估计图片的第一深度图和所述后一帧待估计图片的第二深度图的投影误差;

[0047] 第二计算模块,用于利用预设的估计算法,对两帧所述待估计毫米波点云进行整体位姿变换估计,得到所述待估计毫米波点云整体位姿变换的第二位姿变换结果;基于所述第一位姿变换结果和所述第二位姿变换结果,根据预设的位姿估算误差计算方式,得到所述第一位姿变换结果和所述第二位姿变换结果的位姿估算误差;

[0048] 第三计算模块,用于基于所述第一深度图和所述动态点集,根据预设的运动物体的深度误差计算方式,计算两帧所述待估计图片中的运动物体的深度误差;

[0049] 训练模块,用于根据所述后一帧待估计图片的第一深度图和所述后一帧待估计图片的第二深度图的投影误差,所述第一位姿变换结果和所述第二位姿变换结果的位姿估算误差,以及两帧所述待估计图片中的所述运动物体的深度误差,得到所述待估计图片的整体训练损失,并利用所述整体训练损失对所述初始深度估计模型和所述初始点云估计模型进行训练,直至所述初始深度估计模型和所述初始点云估计模型收敛,得到用于单目图像深度估计的完备深度估计模型;

[0050] 以及估计模块,用于基于所述完备深度估计模型,对待估计的图片进行单目图像深度估计。

[0051] 第三方面,本申请还提供了一种计算机设备。所述计算机设备包括存储器和处理器,所述存储器存储有计算机程序,所述处理器执行所述计算机程序时实现上述第一方面所述的单目图像深度估计方法。

[0052] 上述单目图像深度估计方法、装置和计算机设备,其通过预设的算法对后一帧待估计图片的第二深度图进行计算,得到后一帧待估计图片的第二深度图,并计算后一帧待估计图片的第一深度图和后一帧待估计图片的第二深度图的投影误差。通过预设的算法对两帧待估计毫米波点云进行位姿变换计算,得到第二位姿变换结果,并计算第一位姿变换结果和第二位姿变换结果的位姿估算误差,最后计算两帧待估计图片中的运动物体的深度误差。进而将投影误差、位姿估算误差和运动物体的深度误差反馈至整体训练损失,用于对初始深度估计模型进行训练,采用这种方式能够考虑到图像中的运动物体,得到准确的完备深度估计模型,进而保证利用完备深度估计模型实现图像的深度估计结果的稳定性,解决了现有技术存在无法保证图像的深度估计结果的稳定性的问题。

[0053] 本申请的一个或多个实施例的细节在以下附图和描述中提出,以使本申请的其他特征、目的和优点更加简明易懂。

## 附图说明

[0054] 此处所说明的附图用来提供对本申请的进一步理解,构成本申请的一部分,本申请的示意性实施例及其说明用于解释本申请,并不构成对本申请的不当限定。在附图中:

- [0055] 图1为本申请一实施例提供的单目图像深度估计方法的终端的硬件结构框图；
- [0056] 图2为本申请一实施例提供的单目图像深度估计方法的流程图；
- [0057] 图3为本申请一优选实施例提供的单目图像深度估计方法的流程图；
- [0058] 图4为本申请一实施例提供的单目图像深度估计装置的结构框图。

### 具体实施方式

[0059] 为更清楚地理解本申请的目的、技术方案和优点，下面结合附图和实施例，对本申请进行了描述和说明。

[0060] 除另作定义外，本申请所涉及的技术术语或者科学术语应具有本申请所属技术领域具备一般技能的人所理解的一般含义。在本申请中的“一”、“一个”、“一种”、“该”、“这些”等类似的词并不表示数量上的限制，它们可以是单数或者复数。在本申请中所涉及的术语“包括”、“包含”、“具有”及其任何变体，其目的是涵盖不排他的包含；例如，包含一系列步骤或模块(单元)的过程、方法和系统、产品或设备并未限定于列出的步骤或模块(单元)，而可包括未列出的步骤或模块(单元)，或者可包括这些过程、方法、产品或设备固有的其他步骤或模块(单元)。在本申请中所涉及的“连接”、“相连”、“耦接”等类似的词语并不限于物理的或机械连接，而可以包括电气连接，无论是直接连接还是间接连接。在本申请中所涉及的“多个”是指两个或两个以上。“和/或”描述关联对象的关联关系，表示可以存在三种关系，例如，“A和/或B”可以表示：单独存在A，同时存在A和B，单独存在B这三种情况。通常情况下，字符“/”表示前后关联的对象是一种“或”的关系。在本申请中所涉及的术语“第一”、“第二”、“第三”等，只是对相似对象进行区分，并不代表针对对象的特定排序。

[0061] 在本实施例中提供的方法实施例可以在终端、计算机或者类似的运算装置中执行。比如在终端上运行，图1是本实施例的单目图像深度估计方法的终端的硬件结构框图。如图1所示，终端可以包括一个或多个(图1中仅示出一个)处理器102和用于存储数据的存储器104，其中，处理器102可以包括但不限于微处理器MCU或可编程逻辑器件FPGA等的处理装置。上述终端还可以包括用于通信功能的传输设备106以及输入输出设备108。本领域普通技术人员可以理解，图1所示的结构仅为示意，其并不对上述终端的结构造成限制。例如，终端还可以包括比图1中所示更多或者更少的组件，或者具有与图1所示出的不同配置。

[0062] 存储器104可用于存储计算机程序，例如，应用软件的软件程序以及模块，如本申请实施例提供的单目图像深度估计方法对应的计算机程序，处理器102通过运行存储在存储器104内的计算机程序，从而执行各种功能应用以及数据处理，即实现上述的方法。存储器104可包括高速随机存储器，还可包括非易失性存储器，如一个或者多个磁性存储装置、闪存、或者其他非易失性固态存储器。在一些实例中，存储器104可进一步包括相对于处理器102远程设置的存储器，这些远程存储器可以通过网络连接至终端。上述网络的实例包括但不限于互联网、企业内部网、局域网、移动通信网及其组合。

[0063] 传输设备106用于经由一个网络接收或者发送数据。上述的网络包括终端的通信供应商提供的无线网络。在一个实例中，传输设备106包括一个网络适配器(Network Interface Controller, 简称为NIC)，其可通过基站与其他网络设备相连从而可与互联网进行通讯。在一个实例中，传输设备106可以为射频(Radio Frequency, 简称为RF)模块，其用于通过无线方式与互联网进行通讯。

[0064] 在本实施例中提供了一种单目图像深度估计方法,图2是本实施例的单目图像深度估计方法的流程图,如图2所示,该流程包括如下步骤:

[0065] 步骤S210,利用预先设置的初始深度估计模型,对两帧待估计图片进行深度估计,得到待估计图片的第一深度图;待估计图片的第一深度图,包括前一帧待估计图片的第一深度图和后一帧待估计图片的第一深度图。

[0066] 在本步骤中,上述初始深度估计模型,可以包括深度编码网络和深度解码网络。其中,深度编码网络,可以用于对待估计的图片进行深度特征的提取。深度解码网络,可以用于对获取到的深度特征进行深度估计相关的特征抽取。具体地,深度解码网络,可以预先设定多层卷积层,并且每层卷积层后面设置一层上采样层,利用多层卷积层和上采样层实现对获取到的深度特征进行深度估计相关的特征抽取。上述两帧待估计图片,可以是按照采样时间顺序划分的前一帧待估计图片和后一帧待估计图片。相应的,上述待估计图片的第一深度图,包括前一帧待估计图片的第一深度图和后一帧待估计图片的第一深度图。

[0067] 上述利用预先设置的初始深度估计模型,对两帧待估计图片进行深度估计,得到待估计图片的第一深度图,可以是利用预先设置的初始深度估计模型的深度编码网络,获取待估计图片的深度特征,进而利用预先设置的初始深度估计模型的深度解码网络,对获取到的深度特征进行深度估计相关的特征抽取,得到待估计图片的逆深度图,最后,对逆深度图进行取倒数处理,得到待估计图片的第一深度图。其中,上述待估计图片的逆深度图,为待估计图片的深度图的倒数,可以通过对逆深度图进行取倒数处理,得到待估计图片的第一深度图。需要说明的是,待估计图片的第一深度图的尺寸与初始深度估计模型的深度编码网络中的下采样倍数和初始深度估计模型的深度解码网络中的上采样倍数有关。若初始深度估计模型的深度编码网络中的下采样倍数等于初始深度估计模型的深度解码网络中的上采样倍数,则待估计图片的第一深度图的尺寸与待估计图片的尺寸相同。若待估计图片的第一深度图的尺寸大于待估计图片的尺寸,可以对待估计图片的第一深度图进行缩放,使得待估计图片的第一深度图的尺寸与待估计图片的尺寸一致。采用本步骤的方法,可以实现对待估计图片的第一深度图的获取,便于后续利用待估计图片的第一深度图进行投影误差的计算。

[0068] 步骤S220,利用预先设置的初始点云估计模型,对与两帧待估计图片对应的两帧待估计毫米波点云进行点云估计,得到待估计毫米波点云的动态点集和第一位姿变换结果。

[0069] 其中,上述初始点云估计模型,可以包括场景流预测网络和位姿估计网络。其中,场景流预测网络,可以包括特征提取层,用于获取待估计毫米波点云的场景流。位姿估计网络,可以复用场景流预测网络的特征提取层,并将初始点云估计模型的场景流预测网络从上采样卷积层开始的所有网络层替换为N个线性层,并且最后一次输出为一行六列的矩阵。其中,N的数量以及每层线性层的输入输出可以预设。位姿估计网络,可以用于估计两帧待估计毫米波点云之间的位姿变换结果。上述利用预先设置的初始点云估计模型,对与两帧待估计图片对应的两帧待估计毫米波点云进行点云估计,得到待估计毫米波点云的动态点集和第一位姿变换结果,可以是利用预先设置的初始点云估计模型的场景流预测网络,获取待估计毫米波点云的场景流。进而,根据预设的动态点的筛选条件,筛选出场景流中满足平移偏离量大于或等于平均平移偏移量一倍方差的动态点,得到待估计毫米波点云的动态

点集。可以利用预先设置的初始点云估计模型的位姿估计网络,获取待估计毫米波点云的一行六列矩阵,进而,利用预设的旋转公式,将一行六列矩阵转换为三行四列矩阵,得到待估计的毫米波点云的第一位姿变换结果。上述预设的旋转公式,可以是罗德里格斯旋转公式。其中,待估计毫米波点云的一行六列矩阵,可以是矩阵的六列分别表示Pitch,Yaw,Roll三个角度值和X,Y,Z三个坐标轴所表示的方向上的平移量。其中,Pitch是指俯仰角,就是抬头或低头,和大地水平面的夹角,也可以解释为沿自身坐标系(X轴向前的坐标系)的Y轴旋转的角度;Yaw是指偏航角,是沿世界坐标系的Z轴旋转的角度;Roll是指翻滚角,就是左倾或右倾,和大地水平面的夹角,也可以解释为沿自身坐标系(X轴向前的坐标系)的X轴旋转的角度。例如,上述一行六列矩阵A可以表示为:

$$[0070] \quad A = [\text{Pitch Yaw Roll X Y Z}]$$

[0071] 上述三行四列矩阵,三行可以是指Pitch,Yaw,Roll三个角度值,四列可以是指X,Y,Z三个坐标轴所表示的方向上的平移量,以及径向速度V的值。例如,上述三行四列矩阵B可以表示为:

$$[0072] \quad B = \begin{bmatrix} X_1 & Y_1 & Z_1 & V_1 \\ X_2 & Y_2 & Z_2 & V_2 \\ X_3 & Y_3 & Z_3 & V_3 \end{bmatrix}$$

[0073] 上述第一位姿变换结果,可以是将一行六列矩阵转换为三行四列矩阵的变换矩阵。通过本步骤,可以实现对待估计毫米波点云的动态点集和第一位姿变换结果的获取,便于后续利用动态点集和第一位姿变换结果进行位姿估算误差和运动物体的深度误差的计算。

[0074] 步骤S230,基于第一位姿变换结果,计算相机的外参变换值;基于相机的外参变换值和相机的内参值,将前一帧待估计图片的第一深度图投影到后一帧待估计图片的视角,得到后一帧待估计图片的第二深度图;根据预设的投影误差计算方式,计算后一帧待估计图片的第一深度图和后一帧待估计图片的第二深度图的投影误差。

[0075] 上述基于第一位姿变换结果计算相机的外参变换值,可以是基于第一位姿变换结果,以及预设的毫米波雷达到相机的外参值,得到两帧待估计图片所对应的相机的外参变换值。例如,T-1时刻的待估计图片到T时刻的待估计图片所对应的相机的外参变换值 $TC_{T-1 \rightarrow T}$ 的具体的计算公式为:

$$[0076] \quad TC_{T-1 \rightarrow T} = T_{R \rightarrow C}^{-1} TR_{T-1 \rightarrow T} T_{R \rightarrow C}$$

[0077] 其中, $T_{R \rightarrow C}$ 表示毫米波雷达到相机的外参值, $TR_{T-1 \rightarrow T}$ 表示T-1时刻的待估计毫米波点云到T时刻的待估计毫米波点云的第一位姿变换结果。

[0078] 上述得到后一帧待估计图片的第二深度图,可以是基于相机的外参变换值和相机的内参值,将前一帧待估计图片的第一深度图投影到后一帧待估计图片的视角,基于投影结果,得到后一帧待估计图片的第二深度图。例如,可以是基于T-1时刻的待估计图片到T时刻的待估计图片所对应的相机的外参变换值 $TC_{T-1 \rightarrow T}$ 和相机的内参值,将T-1时刻的待估计图片的第一深度图 $D_{T-1}$ 投影到T时刻的待估计图片的视角,基于投影结果,得到T时刻的待估计图片的第二深度图 $D_{T-1 \rightarrow T}$ 。T时刻的待估计图片的第二深度图 $D_{T-1 \rightarrow T}$ 的计算公式如下:

$$[0079] \quad D_{T-1 \rightarrow T} = D_{T-1} T C_{T-1 \rightarrow T}^{-1}$$

[0080] 上述根据预设的投影误差计算方式,计算后一帧待估计图片的第一深度图和后一帧待估计图片的第二深度图的投影误差,可以是利用SSIM(Structure Similarity Index Measure,结构衡量指标)损失函数来计算后一帧待估计图片的第一深度图和后一帧待估计图片的第二深度图的投影误差的损失。例如,利用SSIM损失函数来计算T时刻的待估计图片的第一深度图 $D_T$ 和T时刻的待估计图片的第二深度图 $D_{T-1 \rightarrow T}$ 的投影误差 $L_1$ 的计算公式为:

$$[0081] \quad L_1 = \frac{\theta}{2} (1 - \text{SSIM}(D_T, D_{T-1 \rightarrow T})) + (1 - \theta) \|D_T - D_{T-1 \rightarrow T}\|_1$$

[0082] 其中, $\theta$ 为预先设定的参数。

[0083] 步骤S240,利用预设的估计算法,对两帧待估计毫米波点云进行整体位姿变换估计,得到待估计毫米波点云整体位姿变换的第二位姿变换结果;基于第一位姿变换结果和第二位姿变换结果,根据预设的位姿估算误差计算方式,得到第一位姿变换结果和第二位姿变换结果的位姿估算误差。

[0084] 在本步骤中,上述利用预设的估计算法,对两帧待估计毫米波点云进行整体位姿变换估计,得到待估计毫米波点云整体位姿变换的第二位姿变换结果,可以是利用ICP(Iterative Closest Point,迭代最近点点云匹配算法)算法,对两帧待估计毫米波点云进行整体位姿变换估计,得到待估计毫米波点云整体位姿变换的第二位姿变换结果。其中,根据预设的位姿估算误差计算方式,得到第一位姿变换结果和第二位姿变换结果的位姿估算误差,可以是根据预设的位姿估算误差的计算公式,来计算第一位姿变换结果和第二位姿变换结果的位姿估算误差。上述预设的第一位姿变换结果和第二位姿变换结果的位姿估算误差 $L_2$ 的计算公式为:

$$[0085] \quad L_2 = \|TR_1 - TR_2\|_2^2$$

[0086] 其中,其中, $TR_1$ 表示第一位姿变换结果, $TR_2$ 表示第二位姿变换结果。

[0087] 步骤S250,基于第一深度图和动态点集,根据预设的运动物体的深度误差计算方式,计算两帧待估计图片中的运动物体的深度误差。

[0088] 其中,上述根据预设的运动物体的深度误差计算方式,计算两帧待估计图片中的运动物体的深度误差,可以是根据两帧待估计图片所对应的两帧毫米波点云中的任意点对 $p$ 和 $q$ 来自同一障碍物的概率,点 $q$ 处所形成的三维向量与前后两帧毫米波点云整体平移方式是否一致,以及点对 $p$ 和 $q$ 所在的两帧毫米波点云所对应的两帧待估计图片的位置的待估计图像的第一深度图的深度值,来计算两帧待估计图片中的运动物体的深度误差。其中,两帧待估计图片中的运动物体的深度误差 $L_3$ 的计算公式为:

$$[0089] \quad L_3 = \sum_{p \in RD_T, q \in RD_{T-1}} \text{Sim}(p, q) \cdot \pi(\cos(\text{Loc}_q, TR_{T-1 \rightarrow T}[:, 3])) \cdot \pi(D_T(p) < D_{T-1}(q))$$

[0090] 其中, $RD_{T-1}$ 表示T-1时刻的动态点集, $RD_T$ 表示T时刻的动态点集, $p$ 、 $q$ 表示 $RD_{T-1}$ 和 $RD_T$ 中任意的点对, $\text{Sim}(p, q)$ 表示 $p$ 、 $q$ 来自于同一个障碍物的概率, $\text{Loc}_q$ 表示以毫米波为原点的 $q$ 点的三维空间坐标, $TR_{T-1 \rightarrow T}$ 表示待估计毫米波点云从T-1时刻到T时刻的第一位姿变换结果, $D_T(p)$ 表示 $p$ 在T时刻的待估计图像的第一深度图中的深度值, $D_{T-1}(q)$ 表示 $q$ 在T-1时刻的待估计图像的第一深度图中的深度值。

[0091] 其中,  $\text{Sim}(p, q)$  的计算公式如下:

$$[0092] \quad \text{Sim}(p, q) = \sum_{p \in \text{RD}_T, q \in \text{RD}_{T-1}} \frac{\text{Loc}_p \cdot \text{Loc}_q}{\|\text{Loc}_p\|^2 + \|\text{Loc}_q\|^2 - \text{Loc}_p \cdot \text{Loc}_q} \cdot \left( \frac{V_p}{V_p + V_q} \cdot \frac{V_q}{V_p + V_q} \right)$$

[0093] 其中,  $\text{Loc}_p$  表示  $p$  的三维空间坐标,  $V_p$  和  $V_q$  为点  $p$  和点  $q$  处的标量, 表示点  $p$  和点  $q$  处的径向速度。

[0094] 其中,  $\pi(\cos(\text{Loc}_q, \text{TR}_{T-1 \rightarrow T}[\cdot, 3]))$  为指示函数, 当点  $q$  处所形成的三维向量与前后两帧毫米波点云整体平移方向一致时, 指示函数的值为 1, 否则指示函数的值为 0。

[0095] 其中,  $\pi(D_T(p) < D_{T-1}(q))$  为指示函数, 表示, 当  $p$  在  $T$  时刻的待估计图像的第一深度图中的深度值小于  $q$  在  $T-1$  时刻的待估计图像的第一深度图中的深度值时, 指示函数的值为 1, 否则, 指示函数的值为 0。

[0096] 步骤 S260, 根据后一帧待估计图片的第一深度图和后一帧待估计图片的第二深度图的投影误差, 第一位姿变换结果和第二位姿变换结果的位姿估算误差, 以及两帧待估计图片中的运动物体的深度误差, 得到待估计图片的整体训练损失, 并利用整体训练损失对初始深度估计模型和初始点云估计模型进行训练, 直至初始深度估计模型和初始点云估计模型收敛, 得到用于单目图像深度估计的完备深度估计模型。

[0097] 在本步骤中, 上述根据后一帧待估计图片的第一深度图和后一帧待估计图片的第二深度图的投影误差, 第一位姿变换结果和第二位姿变换结果的位姿估算误差, 以及两帧待估计图片中的运动物体的深度误差, 得到待估计图片的整体训练损失  $L$  的计算公式为:

$$[0098] \quad L = L_1 + \beta \cdot L_2 + (1 - \beta) \cdot L_3$$

[0099] 其中,  $\beta = 1 - \frac{ep}{\text{Max\_epoch}}$ ,  $ep$  表示当前训练的轮次,  $\text{Max\_epoch}$  表示最大训练轮次。

[0100] 上述初始深度估计模型和初始点云估计模型的收敛条件, 可以是待估计图片的整体训练损失  $L$  达到预设阈值, 也可以是初始深度估计模型和初始点云估计模型的训练次数达到预设阈值条件。

[0101] 步骤 S270, 基于完备深度估计模型, 对待估计的图片进行单目图像深度估计。

[0102] 上述步骤 S210 至步骤 S270, 其通过利用预先设置的初始深度估计模型, 得到待估计图片的第一深度图, 利用预先设置的初始点云估计模型, 得到待估计毫米波点云的动态点集和第一位姿变换结果。然后利用预设的算法对后一帧待估计图片的第二深度图进行计算, 得到后一帧待估计图片的第二深度图, 并计算后一帧待估计图片的第一深度图和后一帧待估计图片的第二深度图的投影误差。通过预设的算法对两帧待估计毫米波点云进行位姿变换计算, 得到第二位姿变换结果, 并计算第一位姿变换结果和第二位姿变换结果的位姿估算误差, 最后计算两帧待估计图片中的运动物体的深度误差。进而将投影误差、位姿估算误差和运动物体的深度误差反馈至整体训练损失, 用于对初始深度估计模型进行训练, 采用这种方式能够考虑到图像中的运动物体, 得到准确的完备深度估计模型, 进而保证利用完备深度估计模型实现图像的深度估计结果的稳定性, 解决了现有技术存在无法保证图像的深度估计结果的稳定性的问题。

[0103] 在一个实施例中, 在步骤 S210, 利用预先设置的初始深度估计模型, 对两帧待估计

图片进行深度估计,得到待估计图片的第一深度图之前,包括以下步骤:

[0104] 步骤S202,对待估计的两帧原始图片进行减均值除方差操作,生成两帧第一图片。

[0105] 在本步骤中,上述对待估计的两帧原始图片进行减均值除方差操作,可以是按照通道对待估计的两帧原始图片进行减均值除方差操作。

[0106] 步骤S204,采用预设的缩放方法,将两帧第一图片缩放至预设的大小,得到缩放后的两帧待估计图片。

[0107] 其中,上述预设的缩放方法,可以是最临近缩放法、双线性缩放法和双立方缩放法等能够实现对图片进行缩放的方法。需要说明的是,对第一图片的缩放,可以是上述一种或多种方法,也可以是其他可用于图像缩放的方法,在此不做具体限定。

[0108] 上述步骤S202至步骤S204,通过对待估计的两帧原始图片进行减均值除方差操作,生成两帧第一图片,进而对两帧第一图片进行缩放,使得第一图片缩放至预设的大小,得到待估计图片,能够提高利用初始深度估计模型,对待估计图片进行深度估计的效率。

[0109] 下面通过优选实施例对本实施例进行描述和说明。

[0110] 图3是本申请一优选实施例提供的单目图像深度估计方法的流程图。如图3所示,该单目图像深度估计方法包括以下步骤:

[0111] 步骤S310,对待估计的两帧原始图片进行减均值除方差操作,生成两帧第一图片;

[0112] 步骤S320,采用预设的缩放方法,将两帧第一图片缩放至预设的大小,得到缩放后的两帧待估计图片;

[0113] 步骤S330,利用预先设置的初始深度估计模型,对两帧待估计图片进行深度估计,得到待估计图片的第一深度图;

[0114] 步骤S340,利用预先设置的初始点云估计模型,对与两帧待估计图片对应的两帧待估计毫米波点云进行点云估计,得到待估计毫米波点云的动态点集和第一位姿变换结果;

[0115] 步骤S350,基于第一位姿变换结果和相机的内参值,得到后一帧待估计图片的第二深度图,进而计算后一帧待估计图片的第一深度图和后一帧待估计图片的第二深度图的投影误差;

[0116] 步骤S360,对两帧待估计毫米波点云进行整体位姿变换估计,得到待估计毫米波点云整体位姿变换的第二位姿变换结果,进而基于第一位姿变换结果和第二位姿变换结果,计算第一位姿变换结果和第二位姿变换结果的位姿估算误差;

[0117] 步骤S370,基于第一深度图和动态点集,根据预设的运动物体的深度误差计算方式,计算两帧待估计图片中的运动物体的深度误差;

[0118] 步骤S380,基于投影误差、位姿估算误差和运动物体的深度误差,计算待估计图片的整体训练损失,并利用整体训练损失对初始深度估计模型和初始点云估计模型进行训练,直至初始深度估计模型和初始点云估计模型收敛,得到用于单目图像深度估计的完备深度估计模型;

[0119] 步骤S390,基于完备深度估计模型,对待估计的图片进行单目图像深度估计。

[0120] 上述步骤S310至步骤S390,其通过对待估计的两帧原始图片进行减均值除方差操作,生成两帧第一图片,进而采用预设的缩放方法,将两帧第一图片缩放至预设的大小,得到缩放后的两帧待估计图片。进而通过利用预先设置的初始深度估计模型,得到待估计图

片的第一深度图,利用预先设置的初始点云估计模型,得到待估计毫米波点云的动态点集和第一位姿变换结果。然后进而利用预设的算法对后一帧待估计图片的第二深度图进行计算,得到后一帧待估计图片的第二深度图,并计算后一帧待估计图片的第一深度图和后一帧待估计图片的第二深度图的投影误差。通过预设的算法对两帧待估计毫米波点云进行位姿变换计算,得到第二位姿变换结果,并计算第一位姿变换结果和第二位姿变换结果的位姿估算误差,最后计算两帧待估计图片中的运动物体的深度误差。进而将投影误差、位姿估算误差和运动物体的深度误差反馈至整体训练损失,用于对初始深度估计模型进行训练,采用这种方式能够考虑到图像中的运动物体,得到准确的完备深度估计模型,进而保证利用完备深度估计模型实现图像的深度估计结果的稳定性,解决了现有技术存在无法保证图像的深度估计结果的稳定性的问题。

[0121] 应该理解的是,虽然如上所述的各实施例所涉及的流程图中的各个步骤按照箭头的指示依次显示,但是这些步骤并不是必然按照箭头指示的顺序依次执行。除非本文中有明确的说明,这些步骤的执行并没有严格的顺序限制,这些步骤可以以其它的顺序执行。而且,如上所述的各实施例所涉及的流程图中的至少一部分步骤可以包括多个步骤或者多个阶段,这些步骤或者阶段并不必然是在同一时刻执行完成,而是可以在不同的时刻执行,这些步骤或者阶段的执行顺序也不必然是依次进行,而是可以与其它步骤或者其它步骤中的步骤或者阶段的至少一部分轮流或者交替地执行。

[0122] 基于同样的发明构思,在本实施例中还提供了单目图像深度估计装置,该装置用于实现上述实施例及优选实施方式,已经进行过说明的不再赘述。以下所使用的术语“模块”、“单元”、“子单元”等可以实现预定功能的软件和/或硬件的组合。尽管在以下实施例中描述的装置较佳地以软件来实现,但是硬件,或者软件和硬件的组合的实现也是可能并被构想的。

[0123] 在一个实施例中,图4是本申请一实施例提供的单目图像深度估计装置的结构框图,如图4所示,该单目图像深度估计装置,包括:

[0124] 深度估计模块41,用于利用预先设置的初始深度估计模型,对两帧待估计图片进行深度估计,得到待估计图片的第一深度图;第一深度图,包括前一帧的第一深度图和后一帧的第一深度图。

[0125] 点云估计模块42,用于利用预先设置的初始点云估计模型,对与两帧待估计图片对应的两帧待估计毫米波点云进行点云估计,得到待估计毫米波点云的动态点集和第一位姿变换结果。

[0126] 第一计算模块43,用于基于第一位姿变换结果,计算相机的外参变换值;基于相机的外参变换值和相机的内参值,将前一帧待估计图片的第一深度图投影到后一帧待估计图片的视角,得到后一帧待估计图片的第二深度图;根据预设的投影误差计算方式,计算后一帧待估计图片的第一深度图和后一帧待估计图片的第二深度图的投影误差。

[0127] 第二计算模块44,用于利用预设的估计算法,对两帧待估计毫米波点云进行整体位姿变换估计,得到待估计毫米波点云整体位姿变换的第二位姿变换结果;基于第一位姿变换结果和第二位姿变换结果,根据预设的位姿估算误差计算方式,得到第一位姿变换结果和第二位姿变换结果的位姿估算误差。

[0128] 第三计算模块45,用于基于第一深度图和动态点集,根据预设的运动物体的深度

误差计算方式,计算两帧待估计图片中的运动物体的深度误差。

[0129] 训练模块46,用于根据后一帧待估计图片的第一深度图和后一帧待估计图片的第二深度图的投影误差,第一位姿变换结果和第二位姿变换结果的位姿估算误差,以及两帧待估计图片中的运动物体的深度误差,得到待估计图片的整体训练损失,并利用整体训练损失对初始深度估计模型和初始点云估计模型进行训练,直至初始深度估计模型和初始点云估计模型收敛,得到用于单目图像深度估计的完备深度估计模型。

[0130] 以及估计模块47,用于基于完备深度估计模型,对待估计的图片进行单目图像深度估计。

[0131] 上述单目图像深度估计装置,其通过利用预先设置的初始深度估计模型,得到待估计图片的第一深度图,利用预先设置的初始点云估计模型,得到待估计毫米波点云的动态点集和第一位姿变换结果。然后利用预设的算法对后一帧待估计图片的第二深度图进行计算,得到后一帧待估计图片的第二深度图,并计算后一帧待估计图片的第一深度图和后一帧待估计图片的第二深度图的投影误差。通过预设的算法对两帧待估计毫米波点云进行位姿变换计算,得到第二位姿变换结果,并计算第一位姿变换结果和第二位姿变换结果的位姿估算误差,最后计算两帧待估计图片中的运动物体的深度误差。进而将投影误差、位姿估算误差和运动物体的深度误差反馈至整体训练损失,用于对初始深度估计模型进行训练,采用这种方式能够考虑到图像中的运动物体,得到准确的完备深度估计模型,进而保证利用完备深度估计模型实现图像的深度估计结果的稳定性,解决了现有技术存在无法保证图像的深度估计结果的稳定性的问题。

[0132] 需要说明的是,上述各个模块可以是功能模块也可以是程序模块,既可以通过软件来实现,也可以通过硬件来实现。对于通过硬件来实现的模块而言,上述各个模块可以位于同一处理器中;或者上述各个模块还可以按照任意组合的形式分别位于不同的处理器中。

[0133] 在一个实施例中,提供了一种计算机设备,包括存储器和处理器,存储器中存储有计算机程序,该处理器执行计算机程序时实现上述实施例中的任意一种单目图像深度估计方法。

[0134] 需要说明的是,本申请所涉及的用户信息(包括但不限于用户设备信息、用户个人信息等)和数据(包括但不限于用于分析的数据、存储的数据、展示的数据等),均为经用户授权或者经过各方充分授权的信息和数据。

[0135] 本领域普通技术人员可以理解实现上述实施例方法中的全部或部分流程,是可以通过计算机程序来指令相关的硬件来完成,所述的计算机程序可存储于一非易失性计算机可读取存储介质中,该计算机程序在执行时,可包括如上述各方法的实施例的流程。其中,本申请所提供的各实施例中所使用的对存储器、数据库或其它介质的任何引用,均可包括非易失性和易失性存储器中的至少一种。非易失性存储器可包括只读存储器(Read-Only Memory,ROM)、磁带、软盘、闪存、光存储器、高密度嵌入式非易失性存储器、阻变存储器(ReRAM)、磁变存储器(Magnetoresistive Random Access Memory,MRAM)、铁电存储器(Ferroelectric Random Access Memory,FRAM)、相变存储器(Phase Change Memory,PCM)、石墨烯存储器等。易失性存储器可包括随机存取存储器(Random Access Memory,RAM)或外部高速缓冲存储器等。作为说明而非局限,RAM可以是多种形式,比如静态随机存

取存储器(Static Random AccessMemory,SRAM)或动态随机存取存储器(Dynamic Random Access Memory,DRAM)等。本申请所提供的各实施例中所涉及的数据库可包括关系型数据库和非关系型数据库中至少一种。非关系型数据库可包括基于区块链的分布式数据库等,不限于此。本申请所提供的各实施例中所涉及的处理器可为通用处理器、中央处理器、图形处理器、数字信号处理器、可编程逻辑器、基于量子计算的数据处理逻辑器等,不限于此。

[0136] 以上实施例的各技术特征可以进行任意的组合,为使描述简洁,未对上述实施例中的各个技术特征所有可能的组合都进行描述,然而,只要这些技术特征的组合不存在矛盾,都应当认为是本说明书记载的范围。

[0137] 以上所述实施例仅表达了本申请的几种实施方式,其描述较为具体和详细,但不能因此而理解为对本申请专利范围的限制。应当指出的是,对于本领域的普通技术人员来说,在不脱离本申请构思的前提下,还可以做出若干变形和改进,这些都属于本申请的保护范围。因此,本申请的保护范围应以所附权利要求为准。

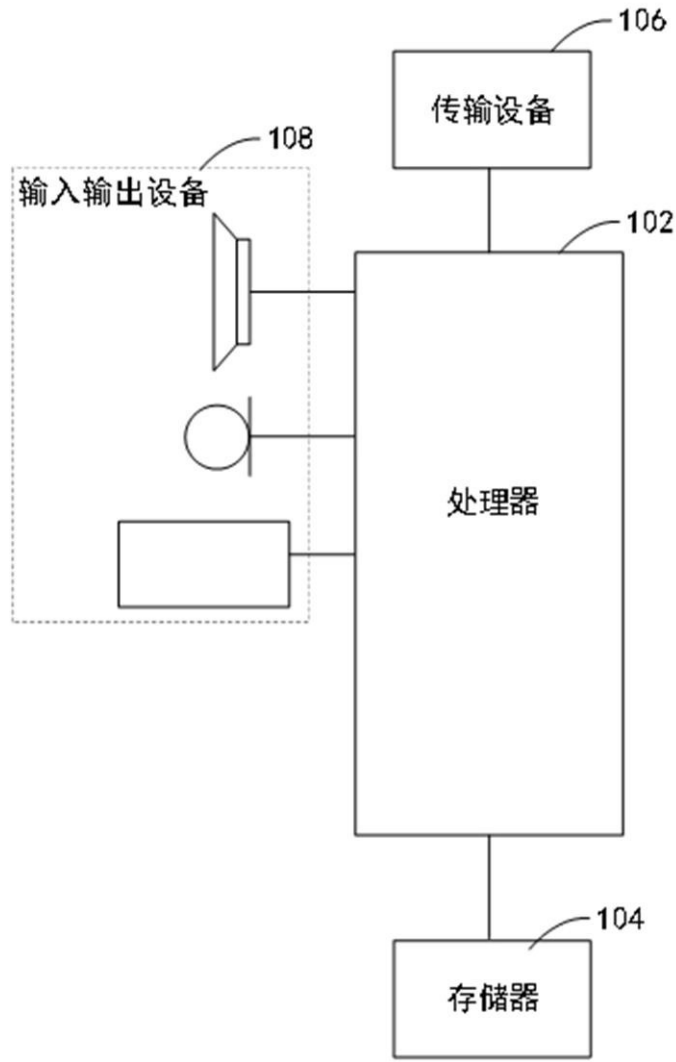


图 1

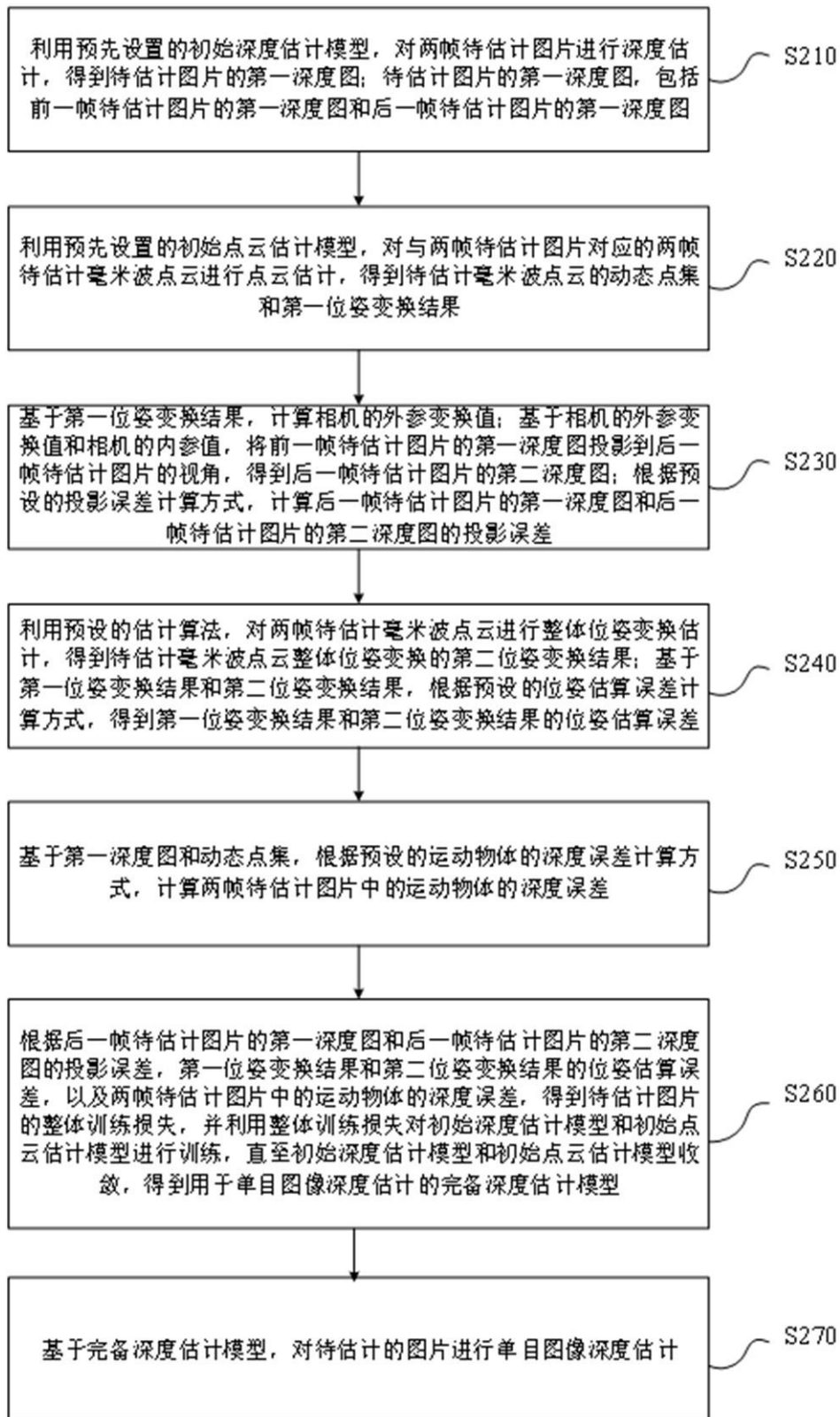


图 2

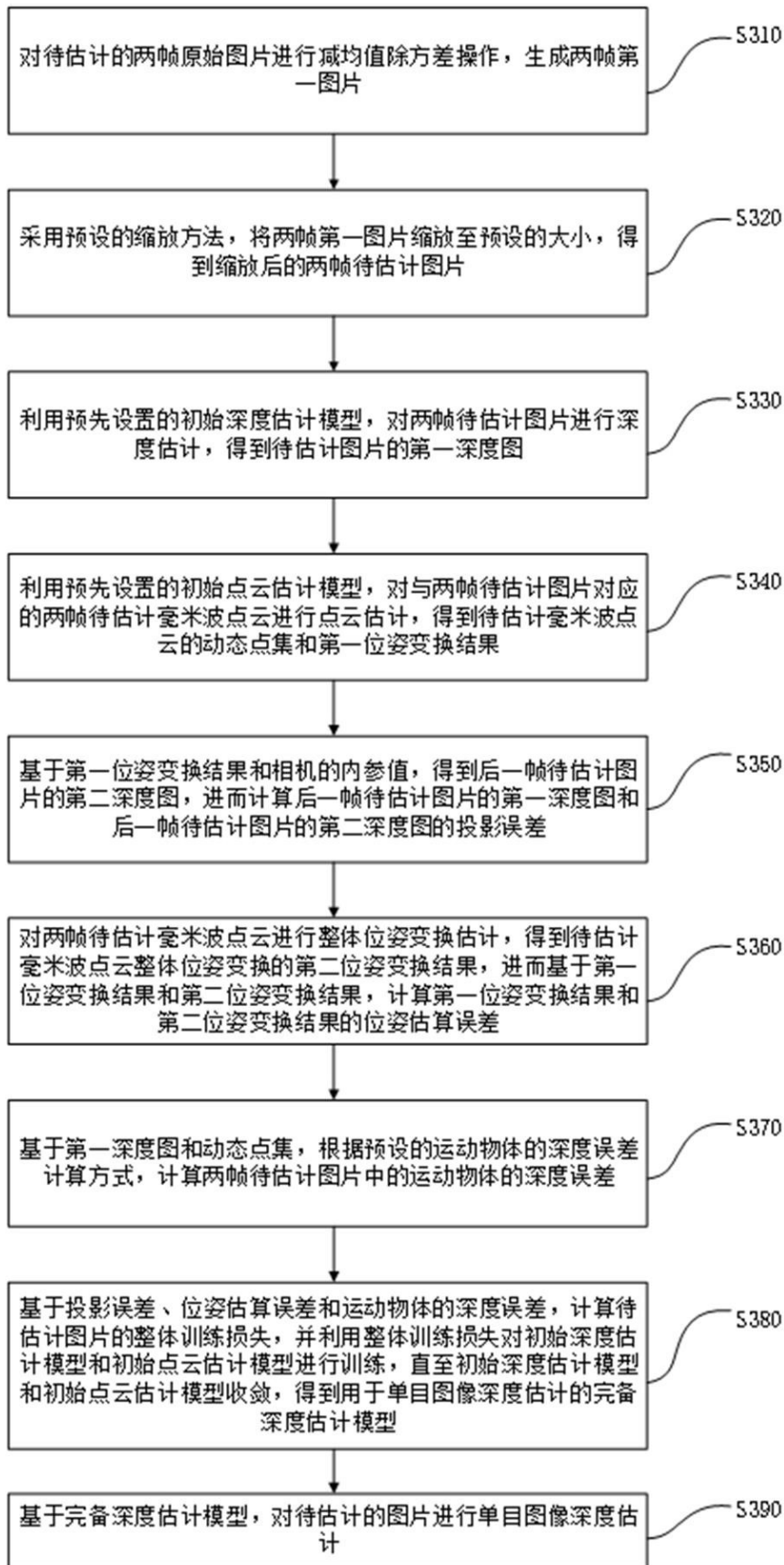


图 3

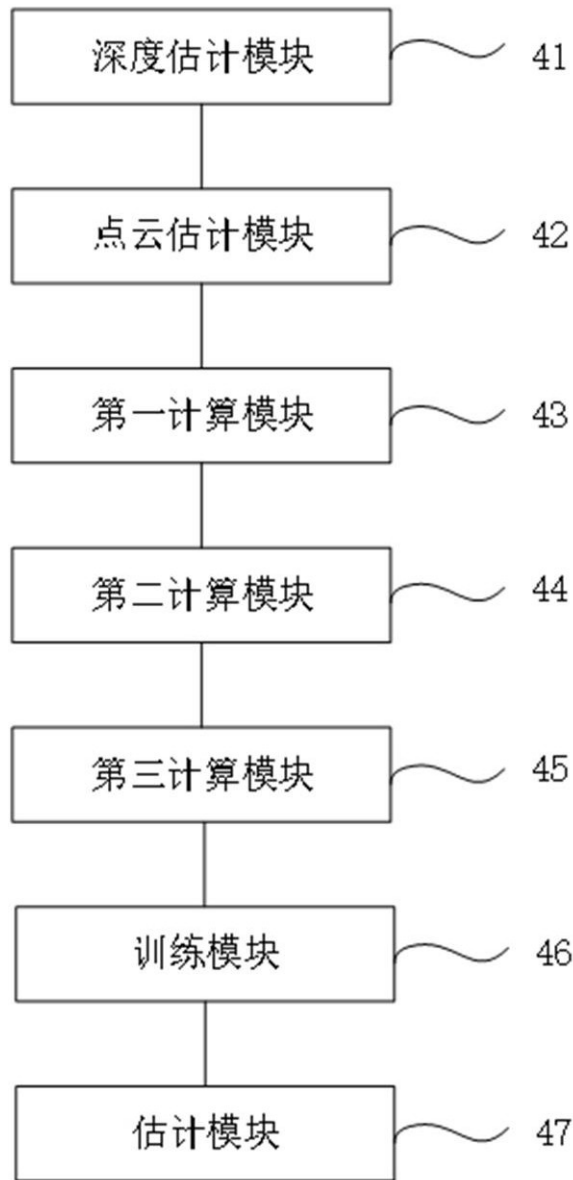


图 4