



(12) 发明专利

(10) 授权公告号 CN 113160294 B

(45) 授权公告日 2022. 12. 23

(21) 申请号 202110346713.3

G06N 3/08 (2006.01)

(22) 申请日 2021.03.31

G06N 3/00 (2006.01)

(65) 同一申请的已公布的文献号

审查员 周苏玲

申请公布号 CN 113160294 A

(43) 申请公布日 2021.07.23

(73) 专利权人 中国科学院深圳先进技术研究院

地址 518000 广东省深圳市南山区深圳大

学城学苑大道1068号

(72) 发明人 王飞 程俊 刘鹏磊

(74) 专利代理机构 深圳中一联合知识产权代理

有限公司 44414

专利代理师 张全文

(51) Int.Cl.

G06T 7/50 (2017.01)

G06T 7/80 (2017.01)

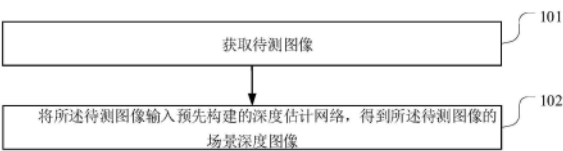
权利要求书5页 说明书23页 附图5页

(54) 发明名称

图像场景深度的估计方法、装置、终端设备和存储介质

(57) 摘要

本申请涉及图像处理技术领域,提出一种图像场景深度的估计方法、装置、终端设备和存储介质。本申请采用的深度估计网络在优化更新参数时,会结合相机姿态估计网络预测输入的样本图像序列的相机姿态向量,该样本图像序列包含目标帧图像和参考帧图像;然后,根据该深度估计网络预测得到的该目标帧图像的场景深度图像、该相机姿态向量、该参考帧图像和对应相机的内参生成与该目标帧图像对应的重建图像;接着,根据该目标帧图像和该重建图像计算得到重建图像时对应的损失函数,最后基于该损失函数构建目标函数并基于该目标函数更新该深度估计网络的参数。通过这样设置,能够充分挖掘目标帧图像和参考帧图像包含的图像信息,降低样本数据采集的成本。



1. 一种图像场景深度的估计方法,其特征在于,包括:

获取待测图像;

将所述待测图像输入预先构建的深度估计网络,得到所述待测图像的场景深度图像;

其中,所述深度估计网络的参数通过以下方式更新:

获取样本图像序列,所述样本图像序列包含目标帧图像和参考帧图像,所述参考帧图像为所述样本图像序列中处于所述目标帧图像之前或之后的一帧以上的图像;

将所述目标帧图像输入所述深度估计网络,得到预测的第一场景深度图像;

将所述样本图像序列输入预先构建的相机姿态估计网络,得到预测的所述目标帧图像和所述参考帧图像之间的相机姿态向量;

根据所述第一场景深度图像、所述相机姿态向量、所述参考帧图像以及拍摄所述样本图像序列采用的相机的内参,生成与所述目标帧图像对应的第一重建图像;

根据所述目标帧图像和所述第一重建图像计算第一图像重建损失,所述第一图像重建损失用于衡量所述目标帧图像和所述第一重建图像之间的差异;

基于所述第一图像重建损失构建目标函数;

根据所述目标函数更新所述深度估计网络的参数;

在获取样本图像序列之后,还包括:

将所述参考帧图像输入所述深度估计网络,得到预测的第二场景深度图像;

根据所述第二场景深度图像、所述相机姿态向量、所述目标帧图像以及拍摄所述样本图像序列采用的相机的内参,生成与所述参考帧图像对应的第二重建图像;

根据所述参考帧图像和所述第二重建图像计算第二图像重建损失,所述第二图像重建损失用于衡量所述参考帧图像和所述第二重建图像之间的差异;

所述基于所述第一图像重建损失构建目标函数,包括:

根据所述第一图像重建损失和所述第二图像重建损失,计算双向图像重建损失;

基于所述双向图像重建损失构建所述目标函数;

其中,所述根据所述第一场景深度图像、所述相机姿态向量、所述参考帧图像以及拍摄所述样本图像序列采用的相机的内参,生成与所述目标帧图像对应的第一重建图像,包括:

根据所述相机姿态向量确定所述目标帧图像转换到所述参考帧图像的第一变换矩阵;

根据所述相机的内参和所述第一场景深度图像,计算所述目标帧图像在世界坐标系下的第一坐标;

基于所述第一变换矩阵对所述第一坐标进行变换,得到所述目标帧图像经转换后在世界坐标系下的第二坐标;

将所述第二坐标转换为在图像坐标系下的第三坐标;

基于所述参考帧图像,以所述第三坐标作为网格点,通过双线性采样机制重建出所述参考帧图像经仿射变换后的图像,并将重建得到的图像确定为所述第一重建图像;

所述根据所述第二场景深度图像、所述相机姿态向量、所述目标帧图像以及拍摄所述样本图像序列采用的相机的内参,生成与所述参考帧图像对应的第二重建图像,包括:

根据所述相机姿态向量确定所述参考帧图像转换到所述目标帧图像的第二变换矩阵;

根据所述相机的内参和所述第二场景深度图像,计算所述参考帧图像在世界坐标系下的第四坐标;

基于所述第二变换矩阵对所述第四坐标进行变换,得到所述参考帧图像经转换后在世界坐标系下的第五坐标;

将所述第五坐标转换为在图像坐标系下的第六坐标;

基于所述目标帧图像,以所述第六坐标作为网格点,通过双线性采样机制重建出所述目标帧图像经仿射变换后的图像,并将重建得到的图像确定为所述第二重建图像;

所述方法还包括:

获取所述目标帧图像在图像坐标系下的第七坐标;

对所述第三坐标和所述第七坐标执行对应元素作差的处理,得到第一前向流坐标;

获取所述参考帧图像在图像坐标系下的第八坐标;

对所述第六坐标和所述第八坐标执行对应元素作差的处理,得到第一后向流坐标;

以所述第三坐标作为网格点,采用双线性采样机制对所述第一后向流坐标进行仿射变换,以合成第二前向流坐标;

以所述第六坐标作为网格点,采用双线性采样机制对所述第一前向流坐标进行仿射变换,以合成第二后向流坐标;

根据所述第一前向流坐标和所述第二前向流坐标计算前向流遮挡掩码,所述前向流遮挡掩码用于衡量所述第一前向流坐标和所述第二前向流坐标之间的匹配程度;

根据所述第一后向流坐标和所述第二后向流坐标计算后向流遮挡掩码,所述后向流遮挡掩码用于衡量所述第一后向流坐标和所述第二后向流坐标之间的匹配程度;

所述根据所述第一图像重建损失和所述第二图像重建损失计算双向图像重建损失,包括:

根据所述第一图像重建损失、所述第二图像重建损失、所述前向流遮挡掩码和所述后向流遮挡掩码计算所述双向图像重建损失。

2. 如权利要求1所述的方法,其特征在于,还包括:

根据所述第二坐标确定所述目标帧图像的第一场景深度值;

根据所述第五坐标确定所述参考帧图像的第二场景深度值;

获取所述第一场景深度图像中与所述第二坐标对应的像素点的第三场景深度值;

获取所述第二场景深度图像中与所述第五坐标对应的像素点的第四场景深度值;

基于所述第三坐标和所述第四场景深度值,通过双线性采样机制重建出所述目标帧图像的第五场景深度值;

基于所述第六坐标和所述第三场景深度值,通过双线性采样机制重建出所述参考帧图像的第六场景深度值;

根据所述第一场景深度值和所述第五场景深度值计算前向场景结构一致性损失,所述前向场景结构一致性损失用于衡量通过多视图几何变换计算得到的所述目标帧图像的场景深度值与重建出的所述目标帧图像的场景深度值之间的差异;

根据所述第二场景深度值和所述第六场景深度值计算后向场景结构一致性损失,所述后向场景结构一致性损失用于衡量通过多视图几何变换计算得到的所述参考帧图像的场景深度值与重建出的所述参考帧图像的场景深度值之间的差异;

根据所述前向场景结构一致性损失和所述后向场景结构一致性损失,计算双向场景结构一致性损失;

所述基于所述双向图像重建损失构建所述目标函数,包括:

基于所述双向图像重建损失和所述双向场景结构一致性损失,构建得到所述目标函数。

3.如权利要求1或2所述的方法,其特征在于,所述深度估计网络包括编码网络,所述方法还包括:

通过所述编码网络获取所述目标帧图像的第一特征图像以及所述参考帧图像的第二特征图像;

基于所述第三坐标和所述第二特征图像,通过双线性采样机制重建出所述目标帧图像的第三特征图像;

基于所述第六坐标和所述第一特征图像,通过双线性采样机制重建出所述参考帧图像的第四特征图像;

根据所述第一特征图像、所述第二特征图像、所述第三特征图像和所述第四特征图像,计算得到双向特征感知损失,所述双向特征感知损失用于衡量通过编码网络获得的所述目标帧图像的特征图像与重建出的所述目标帧图像的特征图像之间的差异,以及通过编码网络获得的所述参考帧图像的特征图像与重建出的所述参考帧图像的特征图像之间的差异;

所述基于所述双向图像重建损失构建所述目标函数,包括:

基于所述双向图像重建损失和所述双向特征感知损失,构建得到所述目标函数。

4.如权利要求3所述的方法,其特征在于,在通过所述编码网络获取所述目标帧图像的第一特征图像以及所述参考帧图像的第二特征图像之后,还包括:

根据所述目标帧图像、所述参考帧图像、所述第一场景深度图像、所述第二场景深度图像、所述第一特征图像和所述第二特征图像,计算得到平滑损失,所述平滑损失用于正则化通过所述深度估计网络获得的场景深度图像和特征图像的梯度;

所述基于所述双向图像重建损失和所述双向特征感知损失,构建得到所述目标函数,包括:

基于所述双向图像重建损失、所述双向特征感知损失和所述平滑损失,构建得到所述目标函数。

5.一种图像场景深度的估计装置,其特征在于,包括:

待测图像获取模块,用于获取待测图像;

场景深度估计模块,用于将所述待测图像输入预先构建的深度估计网络,得到所述待测图像的场景深度图像;

样本获取模块,用于获取样本图像序列,所述样本图像序列包含目标帧图像和参考帧图像,所述参考帧图像为所述样本图像序列中处于所述目标帧图像之前或之后的一帧以上的图像;

第一场景深度预测模块,用于将所述目标帧图像输入所述深度估计网络,得到预测的第一场景深度图像;

相机姿态估计模块,用于将所述样本图像序列输入预先构建的相机姿态估计网络,得到预测的所述目标帧图像和所述参考帧图像之间的相机姿态向量;

第一图像重建模块,用于根据所述第一场景深度图像、所述相机姿态向量、所述参考帧图像以及拍摄所述样本图像序列采用的相机的内参,生成与所述目标帧图像对应的第一重

建图像；

第一图像重建损失计算模块,用于根据所述目标帧图像和所述第一重建图像计算第一图像重建损失,所述第一图像重建损失用于衡量所述目标帧图像和所述第一重建图像之间的差异；

目标函数构建模块,用于基于所述第一图像重建损失构建目标函数；

网络参数更新模块,用于根据所述目标函数更新所述深度估计网络的参数；

第二场景深度预测模块,用于将所述参考帧图像输入所述深度估计网络,得到预测的第二场景深度图像；

第二图像重建模块,用于根据所述第二场景深度图像、所述相机姿态向量、所述目标帧图像以及拍摄所述样本图像序列采用的相机的内参,生成与所述参考帧图像对应的第二重建图像；

第二图像重建损失计算模块,用于根据所述参考帧图像和所述第二重建图像计算第二图像重建损失,所述第二图像重建损失用于衡量所述参考帧图像和所述第二重建图像之间的差异；

所述目标函数构建模块包括：

双向图像重建损失计算单元,用于根据所述第一图像重建损失和所述第二图像重建损失,计算双向图像重建损失；

目标函数构建单元,用于基于所述双向图像重建损失构建所述目标函数；

所述第一图像重建模块包括：

第一变换矩阵确定单元,用于根据所述相机姿态向量确定所述目标帧图像转换到所述参考帧图像的第一变换矩阵；

第一坐标计算单元,用于根据所述相机的内参和所述第一场景深度图像,计算所述目标帧图像在世界坐标系下的第一坐标；

第一坐标变换单元,用于基于所述第一变换矩阵对所述第一坐标进行变换,得到所述目标帧图像经转换后在世界坐标系下的第二坐标；

第一坐标转换单元,用于将所述第二坐标转换为在图像坐标系下的第三坐标；

第一图像重建单元,用于基于所述参考帧图像,以所述第三坐标作为网格点,通过双线性采样机制重建出所述参考帧图像经仿射变换后的图像,并将重建得到的图像确定为所述第一重建图像；

所述第二图像重建模块包括：

第二变换矩阵确定单元,用于根据所述相机姿态向量确定所述参考帧图像转换到所述目标帧图像的第二变换矩阵；

第二坐标计算单元,用于根据所述相机的内参和所述第二场景深度图像,计算所述参考帧图像在世界坐标系下的第四坐标；

第二坐标变换单元,用于基于所述第二变换矩阵对所述第四坐标进行变换,得到所述参考帧图像经转换后在世界坐标系下的第五坐标；

第二坐标转换单元,用于将所述第五坐标转换为在图像坐标系下的第六坐标；

第二图像重建单元,用于基于所述目标帧图像,以所述第六坐标作为网格点,通过双线性采样机制重建出所述目标帧图像经仿射变换后的图像,并将重建得到的图像确定为所述

第二重建图像；

所述装置还包括：

第一坐标获取模块，用于获取所述目标帧图像在图像坐标系下的第七坐标；

前向流坐标确定模块，用于对所述第三坐标和所述第七坐标执行对应元素作差的处理，得到第一前向流坐标；

第二坐标获取模块，用于获取所述参考帧图像在图像坐标系下的第八坐标；

后向流坐标确定模块，用于对所述第六坐标和所述第八坐标执行对应元素作差的处理，得到第一后向流坐标；

前向流坐标合成模块，用于以所述第三坐标作为网格点，采用双线性采样机制对所述第一后向流坐标进行仿射变换，以合成第二前向流坐标；

后向流坐标合成模块，用于以所述第六坐标作为网格点，采用双线性采样机制对所述第一前向流坐标进行仿射变换，以合成第二后向流坐标；

前向流遮挡掩码计算模块，用于根据所述第一前向流坐标和所述第二前向流坐标计算前向流遮挡掩码，所述前向流遮挡掩码用于衡量所述第一前向流坐标和所述第二前向流坐标之间的匹配程度；

后向流遮挡掩码计算模块，用于根据所述第一后向流坐标和所述第二后向流坐标计算后向流遮挡掩码，所述后向流遮挡掩码用于衡量所述第一后向流坐标和所述第二后向流坐标之间的匹配程度；

所述双向图像重建损失计算单元具体用于：根据所述第一图像重建损失、所述第二图像重建损失、所述前向流遮挡掩码和所述后向流遮挡掩码计算所述双向图像重建损失。

6. 一种终端设备，包括存储器、处理器以及存储在所述存储器中并可在所述处理器上运行的计算机程序，其特征在于，所述处理器执行所述计算机程序时实现如权利要求1至4中任一项所述的图像场景深度的估计方法。

7. 一种计算机可读存储介质，所述计算机可读存储介质存储有计算机程序，其特征在于，所述计算机程序被处理器执行时实现如权利要求1至4中任一项所述的图像场景深度的估计方法。

图像场景深度的估计方法、装置、终端设备和存储介质

技术领域

[0001] 本申请涉及图像处理技术领域,尤其涉及一种图像场景深度的估计方法、装置、终端设备和存储介质。

背景技术

[0002] 图像的场景深度估计是机器人导航和自动驾驶领域的重要研究方向。随着高性能计算设备的发展,人们通常会使用深度神经网络来对图像的场景深度进行预测。然而,为保证对图像进行场景深度预测的准确性,在训练该深度神经网络时需要大量的样本数据,导致数据采集的成本较高。

发明内容

[0003] 有鉴于此,本申请实施例提供了一种图像场景深度的估计方法、装置、终端设备和存储介质,能够降低样本数据采集的成本。

[0004] 本申请实施例的第一方面提供了一种图像场景深度的估计方法,包括:

[0005] 获取待测图像;

[0006] 将所述待测图像输入预先构建的深度估计网络,得到所述待测图像的场景深度图像;

[0007] 其中,所述深度估计网络的参数通过以下方式更新:

[0008] 获取样本图像序列,所述样本图像序列包含目标帧图像和参考帧图像,所述参考帧图像为所述样本图像序列中处于所述目标帧图像之前或之后的一帧以上的图像;

[0009] 将所述目标帧图像输入所述深度估计网络,得到预测的第一场景深度图像;

[0010] 将所述样本图像序列输入预先构建的相机姿态估计网络,得到预测的所述目标帧图像和所述参考帧图像之间的相机姿态向量;

[0011] 根据所述第一场景深度图像、所述相机姿态向量、所述参考帧图像以及拍摄所述样本图像序列采用的相机的内参,生成与所述目标帧图像对应的第一重建图像;

[0012] 根据所述目标帧图像和所述第一重建图像计算第一图像重建损失,所述第一图像重建损失用于衡量所述目标帧图像和所述第一重建图像之间的差异;

[0013] 基于所述第一图像重建损失构建目标函数;

[0014] 根据所述目标函数更新所述深度估计网络的参数。

[0015] 本申请实施例采用的深度估计网络在优化更新参数时,会结合相机姿态估计网络预测输入的样本图像序列的相机姿态向量,该样本图像序列包含目标帧图像和参考帧图像;然后,根据该深度估计网络预测得到的该目标帧图像的场景深度图像、该相机姿态向量、该参考帧图像和对应相机的内参生成与该目标帧图像对应的重建图像;接着,根据该目标帧图像和该重建图像计算得到重建图像时对应的损失函数,最后基于该损失函数构建目标函数并基于该目标函数更新该深度估计网络的参数。通过这样设置,能够充分挖掘目标帧图像和参考帧图像包含的潜在图像信息,也即采样较少的样本图像即可获得足够的图像

信息以完成该深度估计网络的训练,从而降低样本数据采集的成本。

[0016] 在本申请的一个实施例中,在获取样本图像序列之后,还可以包括:

[0017] 将所述参考帧图像输入所述深度估计网络,得到预测的第二场景深度图像;

[0018] 根据所述第二场景深度图像、所述相机姿态向量、所述目标帧图像以及拍摄所述样本图像序列采用的相机的内参,生成与所述参考帧图像对应的第二重建图像;

[0019] 根据所述参考帧图像和所述第二重建图像计算第二图像重建损失,所述第二图像重建损失用于衡量所述参考帧图像和所述第二重建图像之间的差异;

[0020] 所述基于所述第一图像重建损失构建目标函数,包括:

[0021] 根据所述第一图像重建损失和所述第二图像重建损失,计算双向图像重建损失;

[0022] 基于所述双向图像重建损失构建所述目标函数。

[0023] 通过在深度估计网络的目标函数中添加双向图像重建损失,能够充分挖掘图像数据中的潜在信息,进一步提升深度估计算法的鲁棒性。

[0024] 进一步的,所述根据所述第一场景深度图像、所述相机姿态向量、所述参考帧图像以及拍摄所述样本图像序列采用的相机的内参,生成与所述目标帧图像对应的第一重建图像,可以包括:

[0025] 根据所述相机姿态向量确定所述目标帧图像转换到所述参考帧图像的第一变换矩阵;

[0026] 根据所述相机的内参和所述第一场景深度图像,计算所述目标帧图像在世界坐标系下的第一坐标;

[0027] 基于所述第一变换矩阵对所述第一坐标进行变换,得到所述目标帧图像经转换后在世界坐标系下的第二坐标;

[0028] 将所述第二坐标转换为在图像坐标系下的第三坐标;

[0029] 基于所述参考帧图像,以所述第三坐标作为网格点,通过双线性采样机制重建出所述参考帧图像经仿射变换后的图像,并将重建得到的图像确定为所述第一重建图像;

[0030] 所述根据所述第二场景深度图像、所述相机姿态向量、所述目标帧图像以及拍摄所述样本图像序列采用的相机的内参,生成与所述参考帧图像对应的第二重建图像,包括:

[0031] 根据所述相机姿态向量确定所述参考帧图像转换到所述目标帧图像的第二变换矩阵;

[0032] 根据所述相机的内参和所述第二场景深度图像,计算所述参考帧图像在世界坐标系下的第四坐标;

[0033] 基于所述第二变换矩阵对所述第四坐标进行变换,得到所述参考帧图像经转换后在世界坐标系下的第五坐标;

[0034] 将所述第五坐标转换为在图像坐标系下的第六坐标;

[0035] 基于所述目标帧图像,以所述第六坐标作为网格点,通过双线性采样机制重建出所述目标帧图像经仿射变换后的图像,并将重建得到的图像确定为所述第二重建图像。

[0036] 在本申请的一个实施例中,所述方法还可以包括:

[0037] 获取所述目标帧图像在图像坐标系下的第七坐标;

[0038] 对所述第三坐标和所述第七坐标执行对应元素作差的处理,得到第一前向流坐标;

- [0039] 获取所述参考帧图像在图像坐标系下的第八坐标；
- [0040] 对所述第六坐标和所述第八坐标执行对应元素作差的处理，得到第一后向流坐标；
- [0041] 以所述第三坐标作为网格点，采用双线性采样机制对所述第一后向流坐标进行仿射变换，以合成第二前向流坐标；
- [0042] 以所述第六坐标作为网格点，采用双线性采样机制对所述第一前向流坐标进行仿射变换，以合成第二后向流坐标；
- [0043] 根据所述第一前向流坐标和所述第二前向流坐标计算前向流遮挡掩码，所述前向流遮挡掩码用于衡量所述第一前向流坐标和所述第二前向流坐标之间的匹配程度；
- [0044] 根据所述第一后向流坐标和所述第二后向流坐标计算后向流遮挡掩码，所述后向流遮挡掩码用于衡量所述第一后向流坐标和所述第二后向流坐标之间的匹配程度；
- [0045] 所述根据所述第一图像重建损失和所述第二图像重建损失计算双向图像重建损失，包括：
- [0046] 根据所述第一图像重建损失、所述第二图像重建损失、所述前向流遮挡掩码和所述后向流遮挡掩码计算所述双向图像重建损失。
- [0047] 遮挡掩码用于判断连续的视频帧中是否存在遮挡物体，将遮挡掩码添加到双向图像重建损失的计算中，能够提高该深度估计网络对带有遮挡物体的图像进行深度估计的准确率。
- [0048] 在本申请的一个实施例中，所述方法还可以包括：
- [0049] 根据所述第二坐标确定所述目标帧图像的第一场景深度值；
- [0050] 根据所述第五坐标确定所述参考帧图像的第二场景深度值；
- [0051] 获取所述第一场景深度图像中与所述第二坐标对应的像素点的第三场景深度值；
- [0052] 获取所述第二场景深度图像中与所述第五坐标对应的像素点的第四场景深度值；
- [0053] 基于所述第三坐标和所述第四场景深度值，通过双线性采样机制重建出所述目标帧图像的第五场景深度值；
- [0054] 基于所述第六坐标和所述第三场景深度值，通过双线性采样机制重建出所述参考帧图像的第六场景深度值；
- [0055] 根据所述第一场景深度值和所述第五场景深度值计算前向场景结构一致性损失，所述前向场景结构一致性损失用于衡量通过多视图几何变换计算得到的所述目标帧图像的场景深度值与重建出的所述目标帧图像的场景深度值之间的差异；
- [0056] 根据所述第二场景深度值和所述第六场景深度值计算后向场景结构一致性损失，所述后向场景结构一致性损失用于衡量通过多视图几何变换计算得到的所述参考帧图像的场景深度值与重建出的所述参考帧图像的场景深度值之间的差异；
- [0057] 根据所述前向场景结构一致性损失和所述后向场景结构一致性损失，计算双向场景结构一致性损失；
- [0058] 所述基于所述双向图像重建损失构建所述目标函数，可以包括：
- [0059] 基于所述双向图像重建损失和所述双向场景结构一致性损失，构建得到所述目标函数。
- [0060] 在构建目标函数时，添加双向场景结构一致性损失，能够有效处理待测图像场景

中的遮挡物体与运动对象,从而提高场景深度估计的准确率。

[0061] 在本申请的一个实施例中,所述深度估计网络包括编码网络,所述方法还可以包括:

[0062] 通过所述编码网络获取所述目标帧图像的第一特征图像以及所述参考帧图像的第二特征图像;

[0063] 基于所述第三坐标和所述第二特征图像,通过双线性采样机制重建出所述目标帧图像的第三特征图像;

[0064] 基于所述第六坐标和所述第一特征图像,通过双线性采样机制重建出所述参考帧图像的第四特征图像;

[0065] 根据所述第一特征图像、所述第二特征图像、所述第三特征图像和所述第四特征图像,计算得到双向特征感知损失,所述双向特征感知损失用于衡量通过编码网络获得的所述目标帧图像的特征图像与重建出的所述目标帧图像的特征图像之间的差异,以及通过编码网络获得的所述参考帧图像的特征图像与重建出的所述参考帧图像的特征图像之间的差异;

[0066] 所述基于所述双向图像重建损失构建所述目标函数,包括:

[0067] 基于所述双向图像重建损失和所述双向特征感知损失,构建得到所述目标函数。

[0068] 通过在目标函数中引入双向特征感知损失,能够有效处理待测图像中的弱纹理场景,从而提高场景深度估计的准确率。

[0069] 进一步的,在通过所述编码网络获取所述目标帧图像的第一特征图像以及所述参考帧图像的第二特征图像之后,还可以包括:

[0070] 根据所述目标帧图像、所述参考帧图像、所述第一场景深度图像、所述第二场景深度图像、所述第一特征图像和所述第二特征图像,计算得到平滑损失,所述平滑损失用于正则化通过所述深度估计网络获得的场景深度图像和特征图像的梯度;

[0071] 所述基于所述双向图像重建损失和所述双向特征感知损失,构建得到所述目标函数,包括:

[0072] 基于所述双向图像重建损失、所述双向特征感知损失和所述平滑损失,构建得到所述目标函数。

[0073] 通过在目标函数中引入平滑损失,可以正则化通过所述深度估计网络获得的场景深度图像和特征图像的梯度。

[0074] 本申请实施例的第二方面提供了一种图像场景深度的估计装置,包括:

[0075] 待测图像获取模块,用于获取待测图像;

[0076] 场景深度估计模块,用于将所述待测图像输入预先构建的深度估计网络,得到所述待测图像的场景深度图像;

[0077] 样本获取模块,用于获取样本图像序列,所述样本图像序列包含目标帧图像和参考帧图像,所述参考帧图像为所述样本图像序列中处于所述目标帧图像之前或之后的一帧以上的图像;

[0078] 第一场景深度预测模块,用于将所述目标帧图像输入所述深度估计网络,得到预测的第一场景深度图像;

[0079] 相机姿态估计模块,用于将所述样本图像序列输入预先构建的相机姿态估计网

络,得到预测的所述目标帧图像和所述参考帧图像之间的相机姿态向量;

[0080] 第一图像重建模块,用于根据所述第一场景深度图像、所述相机姿态向量、所述参考帧图像以及拍摄所述样本图像序列采用的相机的内参,生成与所述目标帧图像对应的第一重建图像;

[0081] 第一图像重建损失计算模块,用于根据所述目标帧图像和所述第一重建图像计算第一图像重建损失,所述第一图像重建损失用于衡量所述目标帧图像和所述第一重建图像之间的差异;

[0082] 目标函数构建模块,用于基于所述第一图像重建损失构建目标函数;

[0083] 网络参数更新模块,用于根据所述目标函数更新所述深度估计网络的参数。

[0084] 本申请实施例的第三方面提供了一种终端设备,包括存储器、处理器以及存储在所述存储器中并可在所述处理器上运行的计算机程序,所述处理器执行所述计算机程序时实现如本申请实施例的第一方面提供的图像场景深度的估计方法。

[0085] 本申请实施例的第四方面提供了一种计算机可读存储介质,所述计算机可读存储介质存储有计算机程序,所述计算机程序被处理器执行时实现如本申请实施例的第一方面提供的图像场景深度的估计方法。

[0086] 本申请实施例的第五方面提供了一种计算机程序产品,当计算机程序产品在终端设备上运行时,使得终端设备执行本申请实施例的第一方面所述的图像场景深度的估计方法。

[0087] 可以理解的是,上述第二方面至第五方面的有益效果可以参见上述第一方面中的相关描述,在此不再赘述。

附图说明

[0088] 为了更清楚地说明本申请实施例中的技术方案,下面将对实施例或现有技术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申请的一些实施例,对于本领域普通技术人员来讲,在不付出创造性劳动性的前提下,还可以根据这些附图获得其他的附图。

[0089] 图1是本申请实施例提供的一种图像场景深度的估计方法的一个实施例的流程图;

[0090] 图2是本申请实施例提供的一种优化更新深度估计网络参数的流程示意图;

[0091] 图3是本申请实施例提供的一种深度估计网络的结构示意图;

[0092] 图4是图3所示网络结构中的残差模块的结构示意图;

[0093] 图5是本申请实施例提供的相一种相机姿态估计网络的结构示意图;

[0094] 图6是本申请实施例提供的图像场景深度的估计方法和现有技术的各类算法分别进行单目图像深度预测的结果对比图;

[0095] 图7是本申请实施例提供的一种图像场景深度的估计装置的一个实施例的结构图;

[0096] 图8是本申请实施例提供的一种终端设备的示意图。

具体实施方式

[0097] 以下描述中,为了说明而不是为了限定,提出了诸如特定系统结构、技术之类的具体细节,以便透彻理解本申请实施例。然而,本领域的技术人员应当清楚,在没有这些具体细节的其它实施例中也可以实现本申请。在其它情况中,省略对众所周知的系统、装置、电路以及方法的详细说明,以免不必要的细节妨碍本申请的描述。另外,在本申请说明书和所附权利要求书的描述中,术语“第一”、“第二”、“第三”等仅用于区分描述,而不能理解为指示或暗示相对重要性。

[0098] 本申请提出一种图像场景深度的估计方法、装置、终端设备和存储介质,能够降低样本数据采集的成本。应当理解,本申请各个方法实施例的执行主体为各种类型的终端设备或服务器,比如手机、平板电脑、笔记本电脑、台式电脑和可穿戴设备等。

[0099] 请参阅图1,示出了本申请实施例提供的一种图像场景深度的估计方法,包括:

[0100] 101、获取待测图像;

[0101] 首先,获取待测图像,待测图像是需要预测场景深度的任意一幅图像。

[0102] 102、将所述待测图像输入预先构建的深度估计网络,得到所述待测图像的场景深度图像。

[0103] 在获取待测图像之后,将该待测图像输入一个预先构建的深度估计网络,得到该待测图像的场景深度图像,从而获得该待测图像的场景深度估计结果。具体的,该深度估计网络可以是具有编码器-解码器体系架构的神经网络,而本申请对该深度估计网络采用的神经网络的类型和网络结构不做任何限定。

[0104] 请参阅图2,示出了本申请实施例提供的一种优化更新深度估计网络参数的流程图示意图,包括以下步骤:

[0105] 2.1、获取样本图像序列,所述样本图像序列包含目标帧图像和参考帧图像,所述参考帧图像为所述样本图像序列中处于所述目标帧图像之前或之后的一帧以上的图像;

[0106] 要训练优化深度估计网络,首先需要获取训练集数据,并可以对该训练集数据进行一定的预处理操作。例如,可以获取自动驾驶数据集KITTI作为训练集数据,并对该训练集数据进行随机翻转、随机裁剪、数据归一化等预处理操作,以将该训练集数据转换成指定维度的张量数据,作为深度估计网络的输入。在本申请实施例中,该训练集数据可由大量的样本图像序列组成,其中每个样本图像序列包含目标帧图像和参考帧图像,参考帧图像为该样本图像序列中处于该目标帧图像之前或之后的一帧以上的图像。例如,样本图像序列可以是一个包含5个连续视频帧的视频剪辑,假设为 I_0 、 I_1 、 I_2 、 I_3 、 I_4 ,则 I_2 可作为目标帧图像, I_0 、 I_1 、 I_3 、 I_4 可作为对应的参考帧图像。

[0107] 2.2、将所述目标帧图像输入所述深度估计网络,得到预测的第一场景深度图像;

[0108] 对于该样本图像序列中的目标帧图像,将其输入至该深度估计网络,得到预测的第一场景深度图像,也即该目标帧图像对应的场景深度图像。

[0109] 在本申请的一个实施例中,该深度估计网络的结构示意图如图3所示,其中包含编码器部分和解码器部分。编码器部分用于通过逐层下采样的方式提取输入图像数据的抽象特征,假设对目标帧图像进行预处理后获得维度为 $3*256*832$ 的张量数据,则经过编码器的第一层卷积、归一化和激活函数处理后,获得维度为 $64*128*416$ 的特征图像,完成第一次下采样的处理。然后,该特征图像再经过最大池化和多个残差模块的处理,获得维度为 $256*$

64*208的特征图像,完成第二次下采样的处理。以此类推,经过多次下采样处理后,获得维度为2048*8*26的特征图像。解码器部分用于通过逐层上采样的方式对编码器获得的特征图像进行处理,具体可以用卷积核大小为3*3的卷积层、非线性ELU处理层和最近邻上采样层对编码器获得的特征图像进行处理,获得维度为512*16*52的特征图像。然后,如图3所示,将该512*16*52的特征图像与编码器获得的1024*16*52的特征图像在通道维度上进行拼接,得到维度为1536*16*52的特征图像,至此完成第一次上采样的处理。以此类推,经过多次上采样处理后,最终获得维度为32*256*832的特征图像。接着,将该32*256*832的特征图像依次输入卷积核大小为3*3的卷积层、Sigmoid函数以及 $F(x) = 1/(10*x+0.01)$ 处理,获得最终的场景深度图像,其中,x表示经过Sigmoid函数映射后获得的深度图像,F(x)表示最终获得的场景深度图像。具体的,特征图像经过Sigmoid函数变换后,其中每个像素的取值范围会映射到0至1之间,此处假设实际场景深度范围为0.1m到100m之间,则通过函数 $F(x) = 1/(10*x+0.01)$ 处理可以将估计的深度图像中的像素与实际场景深度建立映射关系,例如若 $x=0$ 对应于现实场景中的100m。因此,经过函数 $F(x) = 1/(10*x+0.01)$ 处理可以将估计的深度图像约束在0.1m到100m之间的合理范围内。

[0110] 图3所示网络结构中的残差模块的结构示意图如图4所示,输入端分为2个分支,其中一个分支依次经过各个卷积层、归一化层BN和ReLU函数的处理后,与另一个分支叠加,从而获得残差模块的输出数据。

[0111] 另外,在图3所示的网络结构中,采用了捷径连接的技术手段,也即将编码器提取的特征图像直接跨越卷积层与解码器获得的相同分辨率的特征图像在通道维度进行拼接。在使用编码器对输入图像进行特征提取的过程中,会采用固定大小的卷积核通过滑动窗口的方式不断地提取图像特征,然而由于卷积核大小的限制及卷积局部图像特征提取的性质,使得浅层网络只能提取图像的局部特征。随着卷积层数的不断增加,提取特征图像的分辨率不断降低,同时特征图像的数量也不断增加,从而能够提取出更加抽象、感受野更大的深度特征。至于解码器部分,直接对编码器最后输出的特征图像进行解码,对该深层次特征进行多次上采样处理,得到不同分辨率的深层次特征图像,如经过第一次上采样处理后得到维度为512*16*52的特征图像,此时将编码器提取的维度为1024*16*52的特征图像直接跨越相应卷积层与解码器获得的512*16*52的特征图像在通道维度进行融合。如图3所示,编码器提取的每种分辨率的特征图像均与对应解码器获得特征图像进行特征融合,实现了图像局部特征与深度特征信息的融合。

[0112] 2.3、将所述样本图像序列输入预先构建的相机姿态估计网络,得到预测的所述目标帧图像和所述参考帧图像之间的相机姿态向量;

[0113] 为了获得该目标帧图像和该参考帧图像之间的相机姿态向量,本申请实施例还预先构建了一个相机姿态估计网络,该网络的结构示意图可以如图5所示,其中包含多个不同参数的卷积层。具体的,假设输入的样本图像序列为 I_0, I_1, I_2, I_3, I_4 总共5帧图像,首先将这5帧图像进行预处理成指定维度的张量数据,作为该相机姿态估计网络的输入;该相机姿态估计网络采用多个指定步长的卷积层来提取图像特征并进行下采样,依次获得相应的特征图像。例如在图5中,输入的张量数据在经过8层卷积处理后,可以得到24维的特征向量,最后将该特征向量调整成 $6*N_{ref}$ 的相机姿态向量,此处的6表示该相机姿态向量是由3个平移向量和3个旋转向量构成的6维向量, $N_{ref}=4$ 表示该样本图像序列输入的参考帧图像的数量

是4。

[0114] 2.4、根据所述第一场景深度图像、所述相机姿态向量、所述参考帧图像以及拍摄所述样本图像序列采用的相机的内参,生成与所述目标帧图像对应的第一重建图像;

[0115] 在获得估计的第一场景深度图像和相机姿态向量之后,需要基于这些数据进行图像重建,获得该目标帧图像对应的第一重建图像,以便后续进行图像重建损失的计算。

[0116] 具体的,所述根据所述第一场景深度图像、所述相机姿态向量、所述参考帧图像以及拍摄所述样本图像序列采用的相机的内参,生成与所述目标帧图像对应的第一重建图像,可以包括:

[0117] (1) 根据所述相机姿态向量确定所述目标帧图像转换到所述参考帧图像的第一变换矩阵;

[0118] (2) 根据所述相机的内参和所述第一场景深度图像,计算所述目标帧图像在世界坐标系下的第一坐标;

[0119] (3) 基于所述第一变换矩阵对所述第一坐标进行变换,得到所述目标帧图像经转换后在世界坐标系下的第二坐标;

[0120] (4) 将所述第二坐标转换为在图像坐标系下的第三坐标;

[0121] (5) 基于所述参考帧图像,以所述第三坐标作为网格点,通过双线性采样机制重建出所述参考帧图像经仿射变换后的图像,并将重建得到的图像确定为所述第一重建图像。

[0122] 假设目标帧图像为 I_{tgt} ,参考帧图像为 I_{ref} ,对应相机的内参矩阵为 K ,则可以通过上文所述的深度估计网络估计出 I_{tgt} 对应的第一场景深度图像为 D_{tgt} ,由上文所述的相机姿态估计网络估计出两帧图像之间的相机姿态,得到从目标帧图像 I_{tgt} 转换到参考帧图像 I_{ref} 的第一变换矩阵 T (由旋转向量和平移向量组成)。然后,根据该相机内参矩阵 K 、第一场景深度图像 D_{tgt} 和该目标帧图像 I_{tgt} ,可以计算出目标帧图像 I_{tgt} 在世界坐标系下的坐标(第一坐标)。例如,假设目标帧图像 I_{tgt} 中某一像素点的图像坐标为 $img_{xy}^{tgt} = (x, y)$,根据该第一场景深度图像 D_{tgt} 确定该像素点的深度为 d_{tgt} ,则可以由以下公式组计算得到该像素在世界坐标系下的坐标:

$$[0123] \quad img_{xyz}^{tgt} = (X, Y, Z)$$

$$[0124] \quad X = \frac{d_{tgt}}{f}(x - c_x)$$

$$[0125] \quad Y = \frac{d_{tgt}}{f}(y - c_y)$$

$$[0126] \quad Z = d_{tgt}$$

$$[0127] \quad img_{xyz}^{tgt} = D_{tgt} K^{-1} img_{xy}^{tgt}$$

[0128] 其中, $img_{xyz}^{tgt} = (X, Y, Z)$ 表示该像素点在世界坐标系下的坐标, (c_x, c_y, f) 是相机内参矩阵中的参数, c_x 和 c_y 表示主点偏移量, f 表示焦距。

[0129] 然后,基于该第一变换矩阵 T 对第一坐标 $img_{xyz}^{tgt} = (X, Y, Z)$ 进行变换,获得该目标帧图像 I_{tgt} 经转换后在世界坐标系下的第二坐标 $img_{xyz}^{ref} = (\hat{X}_{ref}, \hat{Y}_{ref}, \hat{Z}_{ref})$,具体可以采用以下公

式计算：

$$[0130] \quad \text{img}_{xyz}^{\text{ref}} = R_x R_y R_z \text{img}_{xyz}^{\text{tgt}} + t$$

[0131] 其中, $(R_x, R_y, R_z, t) \in \text{SE3}$ 是3D旋转角度和平移向量, 可以通过该第一变换矩阵T获得。 R_x, R_y, R_z 分别表示相对于世界坐标系中的x轴、y轴和z轴的旋转量, t 表示x轴、y轴和z轴的平移量, SE3表示特殊欧式群。

[0132] 接着, 将该第二坐标转换为在图像坐标系下的第三坐标 $\text{img}_{xy}^{\text{ref}} = (\hat{x}, \hat{y})$, 具体可以通过以下公式组进行转换:

$$[0133] \quad \hat{x} = \frac{f}{Z} + c_x$$

$$[0134] \quad \hat{y} = \frac{f}{Z} + c_y$$

$$[0135] \quad \text{img}_{xy}^{\text{ref}} = K T_{\text{tgt} \rightarrow \text{ref}} \text{img}_{xyz}^{\text{ref}}$$

[0136] 其中, $T_{\text{tgt} \rightarrow \text{ref}}$ 表示由旋转矩阵和平移矩阵构成的相机外参矩阵。

[0137] 在获得第三坐标 $\text{img}_{xy}^{\text{ref}}$ 之后, 可以基于参考帧图像 I_{ref} , 以该第三坐标作为网格点, 通过双线性采样机制重建出参考帧图像 I_{ref} 经仿射变换后的图像 \hat{I}_{ref} , 并将重建得到的图像 \hat{I}_{ref} 确定为该第一重建图像。其中, 双线性采样机制的原理可以参照现有技术, 在此不再赘述。

[0138] 2.5、根据所述目标帧图像和所述第一重建图像计算第一图像重建损失, 所述第一图像重建损失用于衡量所述目标帧图像和所述第一重建图像之间的差异;

[0139] 具体的, 可以采用以下公式计算得到第一图像重建损失:

$$[0140] \quad L_{\text{photo}}^{\text{ref} \rightarrow \text{tgt}}(I_{\text{tgt}}, \hat{I}_{\text{ref}}) = \alpha \frac{1 - \text{SSIM}(I_{\text{tgt}}, \hat{I}_{\text{ref}})}{2} + (1 - \alpha) \text{ERF}(I_{\text{tgt}}, \hat{I}_{\text{ref}})$$

[0141] 其中, $L_{\text{photo}}^{\text{ref} \rightarrow \text{tgt}}(I_{\text{tgt}}, \hat{I}_{\text{ref}})$ 表示该第一图像重建损失, α 为预设的权重参数, 例如可以为0.85。SSIM(*) 为以下所示的结构相似性度量函数:

$$[0142] \quad \text{SSIM}(a, b) = \frac{(2\mu_a \mu_b + c_1)(2\delta_{ab} + c_2)}{(\mu_a^2 + \mu_b^2 + c_1)(\delta_a^2 + \delta_b^2 + c_2)}$$

[0143] 上式中, μ, δ 分别是像素均值和方差, $c_1 = 0.01^2$, $c_2 = 0.03^2$ 。

[0144] ERF(*) 为以下所示的鲁棒性误差度量参数:

$$[0145] \quad \text{ERF}(m, n) = \sqrt{(m - n)^2 + \epsilon^2}$$

[0146] 上式中, $\epsilon = 0.01$ 。

[0147] 2.6、基于所述第一图像重建损失构建目标函数;

[0148] 在获得该第一图像重建损失之后, 可以基于该第一图像重建损失构建一个目标函数, 以便完成该深度估计网络的参数更新。

[0149] 在本申请的一个实施例中, 在获取样本图像序列之后, 还可以包括:

[0150] (1) 将所述参考帧图像输入所述深度估计网络,得到预测的第二场景深度图像;

[0151] (2) 根据所述第二场景深度图像、所述相机姿态向量、所述目标帧图像以及拍摄所述样本图像序列采用的相机的内参,生成与所述参考帧图像对应的第二重建图像;

[0152] (3) 根据所述参考帧图像和所述第二重建图像计算第二图像重建损失,所述第二图像重建损失用于衡量所述参考帧图像和所述第二重建图像之间的差异。

[0153] 具体的,所述根据所述第二场景深度图像、所述相机姿态向量、所述目标帧图像以及拍摄所述样本图像序列采用的相机的内参,生成与所述参考帧图像对应的第二重建图像,可以包括:

[0154] (2.1) 根据所述相机姿态向量确定所述参考帧图像转换到所述目标帧图像的第二变换矩阵;

[0155] (2.2) 根据所述相机的内参和所述第二场景深度图像,计算所述参考帧图像在世界坐标系下的第四坐标;

[0156] (2.3) 基于所述第二变换矩阵对所述第四坐标进行变换,得到所述参考帧图像经转换后在世界坐标系下的第五坐标;

[0157] (2.4) 将所述第五坐标转换为在图像坐标系下的第六坐标;

[0158] (2.5) 基于所述目标帧图像,以所述第六坐标作为网格点,通过双线性采样机制重建出所述目标帧图像经仿射变换后的图像,并将重建得到的图像确定为所述第二重建图像。

[0159] 与上述计算第一图像重建损失的方法类似,在计算第二图像重建损失时,假设目标帧图像为 I_{tgt} ,参考帧图像为 I_{ref} ,对应相机的内参矩阵为 K ,则可以通过上文所述的深度估计网络估计出 I_{ref} 对应的第二场景深度图像为 D_{ref} ,由上文所述的相机姿态估计网络估计出两帧图像之间的相机姿态,得到从参考帧图像 I_{ref} 转换到目标帧图像 I_{tgt} 的第二变换矩阵 T_{inv} ,该第二变换矩阵是从目标帧图像 I_{tgt} 转换到参考帧图像 I_{ref} 的第一变换矩阵 T 的逆变换矩阵。然后,根据该相机内参矩阵 K 、第二场景深度图像 D_{ref} 和该参考帧图像 I_{ref} ,可以计算出参考帧图像 I_{ref} 在世界坐标系下的坐标(第四坐标)。然后,基于该第二变换矩阵 T_{inv} 对该第四坐标进行转换,获得该参考帧图像 I_{ref} 经转换后在世界坐标系下的第五坐标,接着计算出此第五坐标在图像坐标系的第六坐标,具体的坐标变换步骤可以参照前文所述的计算第一图像重建损失的相关内容。最后,可以基于目标帧图像 I_{tgt} ,以该第六坐标为网格点,通过双线性采样机制重建出目标帧图像 I_{tgt} 经仿射变换后的图像 \hat{I}_{tgt} ,并将重建得到的图像 \hat{I}_{tgt} 确定为该第二重建图像。计算第二图像重建损失可以采用以下公式:

$$[0160] \quad L_{photo}^{tgt \rightarrow ref}(\hat{I}_{tgt}, I_{ref}) = \alpha \frac{1 - SSIM(I_{ref}, \hat{I}_{tgt})}{2} + (1 - \alpha) ERF(I_{ref}, \hat{I}_{tgt})$$

[0161] 关于上述公式中各个参数的定义,可以参照前文所述的计算第一图像重建损失的公式中的说明。

[0162] 第一图像重建损失可以定义为前向图像重建损失,第二图像重建损失可以定义为后向图像重建损失,那么可以基于这两个图像重建损失构建得到双向图像重建损失,具体的计算公式可以如下:

$$[0163] \quad L_{photo} = L_{photo}^{ref \rightarrow tgt}(I_{tgt}, \hat{I}_{ref}) + L_{photo}^{tgt \rightarrow ref}(\hat{I}_{tgt}, I_{ref})$$

[0164] 之后,可以根据基于该双向图像重建损失构建目标函数。通过在深度估计网络的目标函数中添加双向图像重建损失,能够充分挖掘图像数据中的潜在信息,进一步提升深度估计算法的鲁棒性。

[0165] 在本申请的一个实施例中,所述方法还可以包括:

[0166] (1) 获取所述目标帧图像在图像坐标系下的第七坐标;

[0167] (2) 对所述第三坐标和所述第七坐标执行对应元素作差的处理,得到第一前向流坐标;

[0168] (3) 获取所述参考帧图像在图像坐标系下的第八坐标;

[0169] (4) 对所述第六坐标和所述第八坐标执行对应元素作差的处理,得到第一后向流坐标;

[0170] (5) 以所述第三坐标作为网格点,采用双线性采样机制对所述第一后向流坐标进行仿射变换,以合成第二前向流坐标;

[0171] (6) 以所述第六坐标作为网格点,采用双线性采样机制对所述第一前向流坐标进行仿射变换,以合成第二后向流坐标;

[0172] (7) 根据所述第一前向流坐标和所述第二前向流坐标计算前向流遮挡掩码,所述前向流遮挡掩码用于衡量所述第一前向流坐标和所述第二前向流坐标之间的匹配程度;

[0173] (8) 根据所述第一后向流坐标和所述第二后向流坐标计算后向流遮挡掩码,所述后向流遮挡掩码用于衡量所述第一后向流坐标和所述第二后向流坐标之间的匹配程度。

[0174] 这个过程可以概括为双向流一致性的检验,其中包括前向流一致性检验和后向流一致性检验。首先,获取目标帧图像在图像坐标系下的第七坐标 $img_{xy}^{tgt} = (x, y)$ 以及前文所述的第三坐标(即 $img_{xy}^{ref} = (\hat{x}, \hat{y})$), 然后对该第三坐标和该第七坐标执行对应元素作差的处理,得到第一前向流坐标 $u_{cam}^{tgt \rightarrow ref}$, 如以下公式所示:

$$[0175] \quad u_{cam}^{tgt \rightarrow ref} = (\hat{x} - x, \hat{y} - y)$$

[0176] 类似的,获取该参考帧图像在图像坐标系下的第八坐标 $img_{xy}^{ref} = (x_{ref}, y_{ref})$ 以及前文所述的第六坐标(可以表示为 $img_{xy}^{tgt} = (\tilde{x}, \tilde{y})$), 然后对该第六坐标和该第八坐标执行对应元素作差的处理,得到第一后向流坐标 $u_{cam}^{ref \rightarrow tgt}$, 如以下公式所示:

$$[0177] \quad u_{cam}^{ref \rightarrow tgt} = (\tilde{x} - x_{ref}, \tilde{y} - y_{ref})$$

[0178] 然后,以该第三坐标作为网格坐标,采用双线性采样机制对该第一后向流坐标 $u_{cam}^{ref \rightarrow tgt}$ 进行仿射变换,以合成第二前向流坐标 $\hat{u}_{cam}^{tgt \rightarrow ref}$, 在理想的情况下,合成的前向流坐标 $\hat{u}_{cam}^{tgt \rightarrow ref}$ 和计算得到的前向流坐标 $u_{cam}^{tgt \rightarrow ref}$ 的大小相同,而方向相反,此为前向流一致性。

[0179] 以该第六坐标作为网格坐标,采用双线性采样机制对该第一前向流坐标 $u_{cam}^{tgt \rightarrow ref}$ 进行仿射变换,以合成第二后向流坐标 $\hat{u}_{cam}^{ref \rightarrow tgt}$, 在理想的情况下,合成的后向流坐标 $\hat{u}_{cam}^{ref \rightarrow tgt}$ 和计算得到的后向流坐标 $u_{cam}^{ref \rightarrow tgt}$ 的大小相同,而方向相反,此为后向流一致性。

[0180] 接下来,可以根据该第一前向流坐标和该第二前向流坐标计算得到前向流遮挡掩码 $mask_{occ}^{tgt \rightarrow ref}$,该掩码用于衡量第一前向流坐标和第二前向流坐标之间的匹配程度,具体可以采用以下公式计算:

$$[0181] \quad mask_{occ}^{tgt \rightarrow ref} = \Gamma(\|u_{cam}^{tgt \rightarrow ref} + \hat{u}_{cam}^{tgt \rightarrow ref}\|^2, \alpha_1(\|u_{cm}^{tgt \rightarrow ref}\|^2 + \|\hat{u}_{cam}^{tgt \rightarrow ref}\|^2) + \alpha_2)$$

$$[0182] \quad \text{其中, } \Gamma(a, b) = \begin{cases} = 1, & a < b \\ = 0, & a \geq b \end{cases}, \text{参数 } \alpha_1 = 0.01 \alpha_2 = 0.5。$$

[0183] 可以根据该第一后向流坐标和该第二后向流坐标计算得到后向流遮挡掩码 $mask_{occ}^{ref \rightarrow tgt}$ 该掩码用于衡量第一后向流坐标和第二后向流坐标之间的匹配程度,具体可以采用以下公式计算:

$$[0184] \quad mask_{occ}^{ref \rightarrow tgt} = \Gamma(\|u_{cam}^{ref \rightarrow tgt} + \hat{u}_{cam}^{ref \rightarrow tgt}\|^2, \alpha_1(\|u_{cm}^{ref \rightarrow tgt}\|^2 + \|\hat{u}_{cam}^{ref \rightarrow tgt}\|^2) + \alpha_2)$$

[0185] 其中,各个参数的定义可以参照前文所述。

[0186] 在计算得到两个流遮挡掩码之后,所述根据所述第一图像重建损失和所述第二图像重建损失计算双向图像重建损失,可以包括:

[0187] 根据所述第一图像重建损失、所述第二图像重建损失、所述前向流遮挡掩码和所述后向流遮挡掩码计算所述双向图像重建损失。

[0188] 遮挡掩码用于判断连续的视频帧中是否存在遮挡物体,将遮挡掩码添加到双向图像重建损失的计算中,能够提高该深度估计网络对带有遮挡物体的图像进行深度估计的准确率。

[0189] 进一步的,所述方法还可以包括:

[0190] (1) 根据所述第二坐标确定所述目标帧图像的第一场景深度值;

[0191] (2) 根据所述第五坐标确定所述参考帧图像的第二场景深度值;

[0192] (3) 获取所述第一场景深度图像中与所述第二坐标对应的像素点的第三场景深度值;

[0193] (4) 获取所述第二场景深度图像中与所述第五坐标对应的像素点的第四场景深度值;

[0194] (5) 基于所述第三坐标和所述第四场景深度值,通过双线性采样机制重建出所述目标帧图像的第五场景深度值;

[0195] (6) 基于所述第六坐标和所述第三场景深度值,通过双线性采样机制重建出所述参考帧图像的第六场景深度值;

[0196] (7) 根据所述第一场景深度值和所述第五场景深度值计算前向场景结构一致性损失,所述前向场景结构一致性损失用于衡量通过多视图几何变换计算得到的所述目标帧图像的场景深度值与重建出的所述目标帧图像的场景深度值之间的差异;

[0197] (8) 根据所述第二场景深度值和所述第六场景深度值计算后向场景结构一致性损失,所述后向场景结构一致性损失用于衡量通过多视图几何变换计算得到的所述参考帧图像的场景深度值与重建出的所述参考帧图像的场景深度值之间的差异;

[0198] (9) 根据所述前向场景结构一致性损失和所述后向场景结构一致性损失,计算双向场景结构一致性损失。

[0199] 上述步骤用于计算双向场景结构一致性损失,首先,根据前文所述的第二坐标 $img_{xyz}^{ref} = (\hat{X}_{ref}, \hat{Y}_{ref}, \hat{Z}_{ref})$,可以得到对应场景的深度值为 \hat{Z}_{ref} (第一场景深度值);根据前文所述的第五坐标 $img_{xyz}^{tgt} = (\hat{X}_{tgt}, \hat{Y}_{tgt}, \hat{Z}_{tgt})$ 可以得到对应场景的深度值为 \hat{Z}_{tgt} (第二场景深度值)。然后,根据该第一场景深度图像,可以估计出目标帧图像 I_{tgt} 中图像坐标为 $img_{xy}^{tgt} = (x, y)$ 处的像素点的深度值 d_{tgt} (第三场景深度值);根据该第二场景深度图像,可以估计出参考帧图像 I_{ref} 中图像坐标为 $img_{xy}^{ref} = (x_{ref}, y_{ref})$ 处的像素点的深度值 d_{ref} (第四场景深度值)。接着,基于第三坐标和深度值 d_{ref} ,可以通过双线性采样机制重建出目标帧图像的第五场景深度值 \hat{d}_{ref} ;基于第六坐标和深度值 d_{tgt} ,可以通过双线性采样机制重建出参考帧图像的第六场景深度值 \hat{d}_{tgt} 。

[0200] 在理论上,第一场景深度值 \hat{Z}_{ref} 和第五场景深度值 \hat{d}_{ref} 应该相等,第二场景深度值 \hat{Z}_{tgt} 和第六场景深度值 \hat{d}_{tgt} 应该相等。然而通过实验测试发现,它们之间并不总是相等,因此可以通过以下2个公式分别计算前向场景结构误差 $depth_{diff}^{ref \rightarrow tgt}$ 以及后向场景结构误差 $depth_{diff}^{tgt \rightarrow ref}$,进而对场景结构施加一致性约束:

$$[0201] \quad depth_{diff}^{ref \rightarrow tgt} = \frac{ERF(\hat{Z}_{ref}, \hat{d}_{ref})}{\hat{Z}_{ref} + \hat{d}_{ref}}$$

$$[0202] \quad depth_{diff}^{tgt \rightarrow ref} = \frac{ERF(\hat{Z}_{tgt}, \hat{d}_{tgt})}{\hat{Z}_{tgt} + \hat{d}_{tgt}}$$

[0203] 通过对场景结构施加一致性约束,可以定位出图像场景中的运动对象和遮挡物体的位置。例如, $depth_{diff}^{ref \rightarrow tgt}$ 和 $depth_{diff}^{tgt \rightarrow ref}$ 的值越大的位置,表示该位置越可能存在运动对象和遮挡物体。

[0204] 然后,计算前向场景结构一致性损失,其用于衡量通过多视图几何变换计算得到的目标帧图像的场景深度值与重建出的目标帧图像的场景深度值之间的差异,具体可以采用以下公式计算:

$$[0205] \quad L_{dsc}^{ref \rightarrow tgt} = \frac{\sum depth_{diff}^{ref \rightarrow tgt}}{N_{ref}}$$

[0206] 其中, N_{ref} 表示参考帧图像中有效网格坐标的数量。

[0207] 计算后向场景结构一致性损失,其用于衡量通过多视图几何变换计算得到的参考帧图像的场景深度值与重建出的参考帧图像的场景深度值之间的差异,具体可以采用以下公式计算:

$$[0208] \quad L_{dsc}^{tgt \rightarrow ref} = \frac{\sum depth_{diff}^{tgt \rightarrow ref}}{N_{tgt}}$$

[0209] 其中, N_{tgt} 表示目标帧图像中有效网格坐标的数量。

[0210] 最后, 根据前向场景结构一致性损失和后向场景结构一致性损失, 可以计算双向场景结构一致性损失如下:

$$[0211] \quad L_{dsc} = L_{dsc}^{ref \rightarrow tgt} + L_{dsc}^{tgt \rightarrow ref}$$

[0212] 所述基于所述双向图像重建损失构建所述目标函数, 可以包括:

[0213] 基于所述双向图像重建损失和所述双向场景结构一致性损失, 构建得到所述目标函数。

[0214] 在构建目标函数时, 添加双向场景结构一致性损失, 能够有效处理待测图像场景中的遮挡物体与运动对象, 从而提高场景深度估计的准确率。

[0215] 另一方面, 在计算双向图像重建损失时, 可以同时引入前文所述的两个遮挡掩码以及两个场景结构误差, 例如可以采用以下公式计算:

$$[0216] \quad L_{photo} = (1 - depth_{diff}^{ref \rightarrow tgt}) * (1 - mask_{occ}^{ref \rightarrow tgt}) * L_{photo}^{ref \rightarrow tgt} + (1 - depth_{diff}^{tgt \rightarrow ref}) * (1 - mask_{occ}^{tgt \rightarrow ref}) * L_{photo}^{tgt \rightarrow ref}$$

[0217] 通过使用 $1 - depth_{diff}^{ref \rightarrow tgt}$ 和 $1 - depth_{diff}^{tgt \rightarrow ref}$ 对图像重建损失函数进行加权处理, 能够达到处理遮挡和运动对象的目的。具体的, 使用所述前向流遮挡掩码和所述前向场景结构不一致权重对第一图像重建损失进行加权处理; 使用所述后向流遮挡掩码和后向场景结构不一致权重对第二图像重建损失进行加权处理; 基于加权处理后的第一图像重建损失和加权处理后的第二图像重建损失构建所述双向图像重建损失。

[0218] 在本申请的一个实施例中, 所述深度估计网络包括编码网络, 所述方法还可以包括:

[0219] (1) 通过所述编码网络获取所述目标帧图像的第一特征图像以及所述参考帧图像的第二特征图像;

[0220] (2) 基于所述第三坐标和所述第二特征图像, 通过双线性采样机制重建出所述目标帧图像的第三特征图像;

[0221] (3) 基于所述第六坐标和所述第一特征图像, 通过双线性采样机制重建出所述参考帧图像的第四特征图像;

[0222] (4) 根据所述第一特征图像、所述第二特征图像、所述第三特征图像和所述第四特征图像, 计算得到双向特征感知损失, 所述双向特征感知损失用于衡量通过编码网络获得的所述目标帧图像的特征图像与重建出的所述目标帧图像的特征图像之间的差异, 以及通过编码网络获得的所述参考帧图像的特征图像与重建出的所述参考帧图像的特征图像之间的差异。

[0223] 上述步骤用于计算双向特征感知损失, 相比于原始RGB图像, 经过编码器提取的特征在弱纹理区域具有更好的区分性。本申请利用编码网络提取的最高分辨率特征图来处理弱纹理区域, 通过该深度估计网络中的编码网络, 可以提取目标帧图像的特征图像 f_{tgt} (第一特征图像) 和参考帧图像的特征图像 f_{ref} (第二特征图像)。然后, 基于前文所述的第三坐标以及该参考帧图像的特征图像 f_{ref} , 可以通过双线性采样机制将该特征图像 f_{ref} 进行仿射变换, 重建出该目标帧图像的第三特征图像 \hat{f}_{ref} ; 基于前文所述的第六坐标以及该目标帧图像的特征图像 f_{tgt} , 可以通过双线性采样机制将该特征图像 f_{tgt} 进行仿射变换, 重建出该参

考帧图像的第四特征图像 \hat{f}_{tgt} 。接着,可以采用以下公式计算得到双向特征感知损失:

$$[0224] \quad L_{feat} = \|f_{tgt} - \hat{f}_{tgt}\| + \|f_{ref} - \hat{f}_{tgt}\|$$

[0225] 双向特征感知损失 L_{feat} 用于衡量通过编码网络获得的目标帧图像的特征图像与重建出的目标帧图像的特征图像之间的差异,以及通过编码网络获得的参考帧图像的特征图像与重建出的参考帧图像的特征图像之间的差异。

[0226] 所述基于所述双向图像重建损失构建所述目标函数,可以包括:

[0227] 基于所述双向图像重建损失和所述双向特征感知损失,构建得到所述目标函数。

[0228] 通过在目标函数中引入双向特征感知损失,能够有效处理待测图像中的弱纹理场景,从而提高场景深度估计的准确率。

[0229] 进一步的,在通过所述编码网络获取所述目标帧图像的第一特征图像以及所述参考帧图像的第二特征图像之后,还可以包括:

[0230] 根据所述目标帧图像、所述参考帧图像、所述第一场景深度图像、所述第二场景深度图像、所述第一特征图像和所述第二特征图像,计算得到平滑损失,所述平滑损失用于正则化通过所述深度估计网络获得的场景深度图像和特征图像的梯度。

[0231] 所述基于所述双向图像重建损失和所述双向特征感知损失,构建得到所述目标函数,可以包括:

[0232] 基于所述双向图像重建损失、所述双向特征感知损失和所述平滑损失,构建得到所述目标函数。

[0233] 为了正则化通过所述深度估计网络获得的场景深度图像和特征图像的梯度,可以在目标函数中引入平滑损失 L_s ,具体可以采用以下公式计算:

$$[0234] \quad L_s = \sum |\partial d_{ref}| * e^{-|\partial I_{ref}|} + \sum |\partial d_{tgt}| * e^{-|\partial I_{tgt}|} + \sum |\partial f_{tgt}| * e^{-|\partial I_{tgt}|} + \sum |\partial f_{ref}| * e^{-|\partial I_{ref}|}$$

[0235] 其中, $|\partial d_{ref}|$ 表示对深度估计网络估计的参考帧深度图 d_{ref} 计算偏导数,然后计算每个元素位置处的绝对值, $|\partial I_{ref}|$ 表示对参考帧图像 I_{ref} 计算偏导数,然后计算每个元素位置的绝对值, $e^{-|\partial I_{ref}|}$ 表示以 $-|\partial I_{ref}|$ 为幂的自然指数,以此类推。

[0236] 前文提出了四种类型的损失函数,分别为双向图像重建损失、平滑损失、双向场景结构一致性损失以及双向特征感知损失,可以基于这几个损失函数构建得到最终的目标函数。例如,某个目标函数 L 的表达式如下:

$$[0237] \quad L = \lambda_{photo} L_{photo} + \lambda_s L_s + \lambda_{dsc} L_{dsc} + \lambda_{feat} L_{feat}$$

[0238] 其中,各个 λ 为设定的系数,例如可以为 $\lambda_{photo} = 1.0$, $\lambda_s = 0.001$, $\lambda_{dsc} = 0.5$, $\lambda_{feat} = 0.05$ 。

[0239] 另外,在前文所述计算各个损失函数的过程中,举例说明的是单个参考帧图像的计算结果,而若参考帧图像有多个,则每个参考帧图像都可以采用前文所述相同的方式计算得到对应的损失值,最后可以用这些参考帧图像对应损失值的平均值作为最后构建目标函数时采用的损失值。

[0240] 2.7、根据所述目标函数更新所述深度估计网络的参数。

[0241] 在构建出目标函数之后,可以根据该目标函数更新该深度估计网络的参数,以达到优化和训练网络的目的。具体的,可以利用AdamW优化器求解出该目标函数相对于深度估计网络权重的梯度,并以此梯度来更新深度估计网络的权重,如此不断迭代,直至达到设定的最大迭代次数,完成该深度估计网络的训练。

[0242] 进一步的,该目标函数可以同时用来对前文所述的相机姿态估计网络进行训练。同样的,可以利用AdamW优化器求解出该目标函数相对于相机姿态估计网络权重的梯度,并以此梯度来更新相机姿态估计网络的权重,如此不断迭代,直至达到设定的最大迭代次数,完成该相机姿态估计网络的训练。总的来说,在构建好目标函数后,可以用该目标函数作为监督信号来联合指导深度估计网络和相机姿态估计网络的训练。具体的,可以利用AdamW优化器求解出该目标函数相对于深度估计网络权重的梯度以及该目标函数相对于相机姿态估计网络权重的梯度,并以此梯度来同时更新深度估计网络和相机姿态估计网络的权重,如此不断迭代,直至达到设定的最大迭代次数,完成深度估计网络和相机姿态估计网络的联合训练。

[0243] 在完成两个网络的训练后,就可以使用单目图像(例如待测图像)作为该深度估计网络的输入,直接计算出对应的场景深度图像。也可以使用连续的图像序列(例如任意连续的5帧单目图像)作为该相机姿态估计网络的输入,计算得到对应的相机姿态向量。需要说明的是,深度估计网络和相机姿态估计网络仅在训练期间需要联合优化,训练完成后网络的权重即固定了,测试期间不需要进行反向传播,只需要进行前向传播即可,因而测试期间两个网络可以单独使用。

[0244] 本申请实施例采用的深度估计网络在优化更新参数时,会结合相机姿态估计网络预测输入的样本图像序列的相机姿态向量,该样本图像序列包含目标帧图像和参考帧图像;然后,根据该深度估计网络预测得到的该目标帧图像的场景深度图像、该相机姿态向量、该参考帧图像和对应相机的内参生成与该目标帧图像对应的重建图像;接着,根据该目标帧图像和该重建图像计算得到重建图像时对应的损失函数,最后基于该损失函数构建目标函数并基于该目标函数更新该深度估计网络的参数。通过这样设置,能够充分挖掘目标帧图像和参考帧图像包含的潜在图像信息,也即采样较少的样本图像即可获得足够的图像信息以完成该深度估计网络的训练,从而降低样本数据采集的成本。

[0245] 另外,通过在目标函数中添加双向的图像重建损失、双向场景结构一致性损失、双向特征感知损失和平滑损失,能够进一步挖掘图像中包含的潜在信息,降低样本数据的采集成本,并能够有效的处理视频帧中存在的运动对象、遮挡等问题以及提高对弱纹理环境的鲁棒性。

[0246] 以下内容为通过仿真结果来说明本申请提出的图像场景深度估计以及相机姿态估计的技术效果。其中,采用Eigen划分的测试集来作为深度估计网络的评估数据,使用KITTI Odometry数据集中的09-10序列作为相机姿态估计网络的评估数据。

[0247] 深度估计网络采用的评估标准包括:绝对误差(AbsRel)、均方根误差(Rmse)、均方误差(SqRel)、对数均方根误差(Rmse_{log})和阈值(δ_t);相机姿态估计网络采用的评估指标为绝对轨迹误差(ATE)。经过仿真测试,本申请提出的方法与现有技术的算法进行比较的测试结果如以下的表1-表3所示。

[0248] 表1

[0249]

Method	Cap(m)	AbsRel	SqRel	RMSE	RMSE log	$\delta < 1.25$	$\delta < 1.25^2$	$\delta < 1.25^3$
Tinghui Zhou	80	0.208	1.768	6.856	0.283	0.678	0.885	0.957
Zhichao Yin	80	0.155	1.296	5.857	0.233	0.793	0.931	0.973
Reza Mahjourian	80	0.163	1.240	6.220	0.250	0.762	0.916	0.968
Yuliang Zou	80	0.150	1.124	5.507	0.223	0.806	0.933	0.973
Clement Godard	80	0.132	1.044	5.142	0.210	0.845	0.948	0.977
Anurag Ranjan	80	0.140	1.070	5.326	0.217	0.826	0.941	0.975
JiaWang Bian	80	0.137	1.089	5.439	0.217	0.830	0.942	0.975
Chenxu Luo	-	0.141	1.029	5.350	0.216	0.816	0.941	0.976
Wang Zhao	-	0.130	0.893	5.026	0.205	0.832	0.949	0.981
本发明	80	0.1199	0.9474	4.9405	0.1965	0.8630	0.9569	0.9814

[0250] 表1为在80m的深度范围内,对单目图像进行场景深度预测的结果对比。其中,绝对误差(AbsRel)、均方根误差(Rmse)、均方误差(SqRel)、对数均方根误差(Rmse log)绝对值表示算法误差值,用于衡量算法的精度,误差值越小表示精度越高,阈值(δ_t)表示预测的场景深度与真实值的接近程度,阈值越高表示算法稳定性越好。通过表1中的测试结果可以发现,本申请提出的方法与现有技术的算法相比,能够获得更高的场景深度预测精度,以及更好的算法稳定性。

[0251] 表2

[0252]

Method	Cap(m)	AbsRel	SqRel	RMSE	RMSE log	$\delta < 1.25$	$\delta < 1.25^2$	$\delta < 1.25^3$
Tinghui Zhou	50	0.201	1.391	5.181	0.264	0.696	0.900	0.966
Zhichao Yin	50	0.147	0.936	4.348	0.218	0.810	0.941	0.977
Reza Mahjourian	50	0.155	0.927	4.549	0.231	0.781	0.931	0.975
Chenxu Luo	-	0.141	1.029	5.350	0.216	0.816	0.941	0.976
Wang Zhao	-	0.130	0.893	5.026	0.205	0.832	0.949	0.981
本发明	50	0.1155	0.8169	4.0249	0.1876	0.8758	0.9619	0.9830

[0253] 表2为在50m的深度范围内,对单目图像进行场景深度预测的结果对比。通过表2中的测试结果同样可以发现,本申请提出的方法与现有技术的算法相比,能够获得更高的场景深度预测精度,以及更好的算法稳定性,因此能够更加鲁棒的预测出单目图像的场景深度和更多细节。

[0254] 表3

[0255]

Method	Seq.09	Seq.10
ORB-SLAM(short)	0.064±0.141	0.064±0.130
ORB-SLAM(full)	0.014±0.008	0.012±0.011
Tinghui Zhou	0.021±0.017	0.020±0.015
Yuliang Zou	0.017±0.007	0.015±0.009
Chenxu Luo	0.013±0.007	0.012±0.008
Reza Mahjourian	0.013±0.010	0.012±0.011
Anurag Ranjan	0.012±0.007	0.012±0.008
本发明	0.0120±0.0068	0.0118±0.0081

[0256] 表3中的绝对轨迹误差(ATE)表示相机位姿的真实值与预测的相机位姿之间的差值,误差值越小,预测的相机位姿越准确。仿真结果表明,同现有的各类算法相比,本申请提出的相机姿态估计的方法预测的相机位姿更准确。

[0257] 另外,图6是本申请提出的图像场景深度的估计方法和现有技术的各类算法分别

进行单目图像深度预测的结果对比图,其中Ground Truth深度图是通过可视化激光雷达数据获得的深度图。

[0258] 应理解,上述各个实施例中各步骤的序号的大小并不意味着执行顺序的先后,各过程的执行顺序应以其功能和内在逻辑确定,而不应对本申请实施例的实施过程构成任何限定。

[0259] 上面主要描述了一种图像场景深度的估计方法,下面将对一种图像场景深度的估计装置进行描述。

[0260] 请参阅图7,本申请实施例中一种图像场景深度的估计装置的一个实施例包括:

[0261] 待测图像获取模块701,用于获取待测图像;

[0262] 场景深度估计模块702,用于将所述待测图像输入预先构建的深度估计网络,得到所述待测图像的场景深度图像;

[0263] 样本获取模块703,用于获取样本图像序列,所述样本图像序列包含目标帧图像和参考帧图像,所述参考帧图像为所述样本图像序列中处于所述目标帧图像之前或之后的一帧以上的图像;

[0264] 第一场景深度预测模块704,用于将所述目标帧图像输入所述深度估计网络,得到预测的第一场景深度图像;

[0265] 相机姿态估计模块705,用于将所述样本图像序列输入预先构建的相机姿态估计网络,得到预测的所述目标帧图像和所述参考帧图像之间的相机姿态向量;

[0266] 第一图像重建模块706,用于根据所述第一场景深度图像、所述相机姿态向量、所述参考帧图像以及拍摄所述样本图像序列采用的相机的内参,生成与所述目标帧图像对应的第一重建图像;

[0267] 第一图像重建损失计算模块707,用于根据所述目标帧图像和所述第一重建图像计算第一图像重建损失,所述第一图像重建损失用于衡量所述目标帧图像和所述第一重建图像之间的差异;

[0268] 目标函数构建模块708,用于基于所述第一图像重建损失构建目标函数;

[0269] 网络参数更新模块709,用于根据所述目标函数更新所述深度估计网络的参数。

[0270] 在本申请的一个实施例中,所述装置还可以包括:

[0271] 第二场景深度预测模块,用于将所述参考帧图像输入所述深度估计网络,得到预测的第二场景深度图像;

[0272] 第二图像重建模块,用于根据所述第二场景深度图像、所述相机姿态向量、所述目标帧图像以及拍摄所述样本图像序列采用的相机的内参,生成与所述参考帧图像对应的第二重建图像;

[0273] 第二图像重建损失计算模块,用于根据所述参考帧图像和所述第二重建图像计算第二图像重建损失,所述第二图像重建损失用于衡量所述参考帧图像和所述第二重建图像之间的差异;

[0274] 所述目标函数构建模块可以包括:

[0275] 双向图像重建损失计算单元,用于根据所述第一图像重建损失和所述第二图像重建损失,计算双向图像重建损失;

[0276] 目标函数构建单元,用于基于所述双向图像重建损失构建所述目标函数。

[0277] 进一步的,所述第一图像重建模块可以包括:

[0278] 第一变换矩阵确定单元,用于根据所述相机姿态向量确定所述目标帧图像转换到所述参考帧图像的第一变换矩阵;

[0279] 第一坐标计算单元,用于根据所述相机的内参和所述第一场景深度图像,计算所述目标帧图像在世界坐标系下的第一坐标;

[0280] 第一坐标变换单元,用于基于所述第一变换矩阵对所述第一坐标进行变换,得到所述目标帧图像经转换后在世界坐标系下的第二坐标;

[0281] 第一坐标转换单元,用于将所述第二坐标转换为在图像坐标系下的第三坐标;

[0282] 第一图像重建单元,用于基于所述参考帧图像,以所述第三坐标作为网格点,通过双线性采样机制重建出所述参考帧图像经仿射变换后的图像,并将重建得到的图像确定为所述第一重建图像;

[0283] 所述第二图像重建模块可以包括:

[0284] 第二变换矩阵确定单元,用于根据所述相机姿态向量确定所述参考帧图像转换到所述目标帧图像的第二变换矩阵;

[0285] 第二坐标计算单元,用于根据所述相机的内参和所述第二场景深度图像,计算所述参考帧图像在世界坐标系下的第四坐标;

[0286] 第二坐标变换单元,用于基于所述第二变换矩阵对所述第四坐标进行变换,得到所述参考帧图像经转换后在世界坐标系下的第五坐标;

[0287] 第二坐标转换单元,用于将所述第五坐标转换为在图像坐标系下的第六坐标;

[0288] 第二图像重建单元,用于基于所述目标帧图像,以所述第六坐标作为网格点,通过双线性采样机制重建出所述目标帧图像经仿射变换后的图像,并将重建得到的图像确定为所述第二重建图像。

[0289] 在本申请的一个实施例中,所述装置还可以包括:

[0290] 第一坐标获取模块,用于获取所述目标帧图像在图像坐标系下的第七坐标;

[0291] 前向流坐标确定模块,用于对所述第三坐标和所述第七坐标执行对应元素作差的处理,得到第一前向流坐标;

[0292] 第二坐标获取模块,用于获取所述参考帧图像在图像坐标系下的第八坐标;

[0293] 后向流坐标确定模块,用于对所述第六坐标和所述第八坐标执行对应元素作差的处理,得到第一后向流坐标;

[0294] 前向流坐标合成模块,用于以所述第三坐标作为网格点,采用双线性采样机制对所述第一后向流坐标进行仿射变换,以合成第二前向流坐标;

[0295] 后向流坐标合成模块,用于以所述第六坐标作为网格点,采用双线性采样机制对所述第一前向流坐标进行仿射变换,以合成第二后向流坐标;

[0296] 前向流遮挡掩码计算模块,用于根据所述第一前向流坐标和所述第二前向流坐标计算前向流遮挡掩码,所述前向流遮挡掩码用于衡量所述第一前向流坐标和所述第二前向流坐标之间的匹配程度;

[0297] 后向流遮挡掩码计算模块,用于根据所述第一后向流坐标和所述第二后向流坐标计算后向流遮挡掩码,所述后向流遮挡掩码用于衡量所述第一后向流坐标和所述第二后向流坐标之间的匹配程度;

[0298] 双向图像重建损失计算单元具体可以用于：根据所述第一图像重建损失、所述第二图像重建损失、所述前向流遮挡掩码和所述后向流遮挡掩码计算所述双向图像重建损失。

[0299] 在本申请的一个实施例中，所述装置还可以包括：

[0300] 第一场景深度值确定模块，用于根据所述第二坐标确定所述目标帧图像的第一场景深度值；

[0301] 第二场景深度值确定模块，用于根据所述第五坐标确定所述参考帧图像的第二场景深度值；

[0302] 第三场景深度值确定模块，用于获取所述第一场景深度图像中与所述第二坐标对应的像素点的第三场景深度值；

[0303] 第四场景深度值确定模块，用于获取所述第二场景深度图像中与所述第五坐标对应的像素点的第四场景深度值；

[0304] 第一场景深度值重建模块，用于基于所述第三坐标和所述第四场景深度值，通过双线性采样机制重建出所述目标帧图像的第五场景深度值；

[0305] 第二场景深度值重建模块，用于基于所述第六坐标和所述第三场景深度值，通过双线性采样机制重建出所述参考帧图像的第六场景深度值；

[0306] 前向场景结构一致性损失计算模块，用于根据所述第一场景深度值和所述第五场景深度值计算前向场景结构一致性损失，所述前向场景结构一致性损失用于衡量通过多视图几何变换计算得到的所述目标帧图像的场景深度值与重建出的所述目标帧图像的场景深度值之间的差异；

[0307] 后向场景结构一致性损失计算模块，用于根据所述第二场景深度值和所述第六场景深度值计算后向场景结构一致性损失，所述后向场景结构一致性损失用于衡量通过多视图几何变换计算得到的所述参考帧图像的场景深度值与重建出的所述参考帧图像的场景深度值之间的差异；

[0308] 双向场景结构一致性损失计算模块，用于根据所述前向场景结构一致性损失和所述后向场景结构一致性损失，计算双向场景结构一致性损失；

[0309] 所述目标函数构建单元具体可以用于：基于所述双向图像重建损失和所述双向场景结构一致性损失，构建得到所述目标函数。

[0310] 在本申请的一个实施例中，所述深度估计网络包括编码网络，所述装置还可以包括：

[0311] 特征图像获取模块，用于通过所述编码网络获取所述目标帧图像的第一特征图像以及所述参考帧图像的第二特征图像；

[0312] 第一特征图像重建模块，用于基于所述第三坐标和所述第二特征图像，通过双线性采样机制重建出所述目标帧图像的第三特征图像；

[0313] 第二特征图像重建模块，用于基于所述第六坐标和所述第一特征图像，通过双线性采样机制重建出所述参考帧图像的第四特征图像；

[0314] 双向特征感知损失计算模块，用于根据所述第一特征图像、所述第二特征图像、所述第三特征图像和所述第四特征图像，计算得到双向特征感知损失，所述双向特征感知损失用于衡量通过编码网络获得的所述目标帧图像的特征图像与重建出的所述目标帧图像

的特征图像之间的差异,以及通过编码网络获得的所述参考帧图像的特征图像与重建出的所述参考帧图像的特征图像之间的差异;

[0315] 所述目标函数构建单元具体可以用于:基于所述双向图像重建损失和所述双向特征感知损失,构建得到所述目标函数。

[0316] 进一步的,所述装置还可以包括:

[0317] 平滑损失计算模块,用于根据所述目标帧图像、所述参考帧图像、所述第一场景深度图像、所述第二场景深度图像、所述第一特征图像和所述第二特征图像,计算得到平滑损失,所述平滑损失用于正则化通过所述深度估计网络获得的场景深度图像和特征图像的梯度;

[0318] 所述目标函数构建单元具体可以用于:基于所述双向图像重建损失、所述双向特征感知损失和所述平滑损失,构建得到所述目标函数。

[0319] 本申请实施例还提供一种计算机可读存储介质,所述计算机可读存储介质存储有计算机程序,所述计算机程序被处理器执行时实现如图1表示的任意一种图像场景深度的估计方法。

[0320] 本申请实施例还提供一种计算机程序产品,当该计算机程序产品在终端设备上运行时,使得终端设备执行实现如图1表示的任意一种图像场景深度的估计方法。

[0321] 图8是本申请一实施例提供的终端设备的示意图。如图8所示,该实施例的终端设备8包括:处理器80、存储器81以及存储在所述存储器81中并可在所述处理器80上运行的计算机程序82。所述处理器80执行所述计算机程序82时实现上述各个图像场景深度的估计方法的实施例中的步骤,例如图1所示的步骤101至102。或者,所述处理器80执行所述计算机程序82时实现上述各装置实施例中各模块/单元的功能,例如图7所示模块701至709的功能。

[0322] 所述计算机程序82可以被分割成一个或多个模块/单元,所述一个或者多个模块/单元被存储在所述存储器81中,并由所述处理器80执行,以完成本申请。所述一个或多个模块/单元可以是能够完成特定功能的一系列计算机程序指令段,该指令段用于描述所述计算机程序82在所述终端设备8中的执行过程。

[0323] 所称处理器80可以是中央处理单元(Central Processing Unit,CPU),还可以是其他通用处理器、数字信号处理器(Digital Signal Processor,DSP)、专用集成电路(Application Specific Integrated Circuit,ASIC)、现成可编程门阵列(Field-Programmable Gate Array,FPGA)或者其他可编程逻辑器件、分立门或者晶体管逻辑器件、分立硬件组件等。通用处理器可以是微处理器或者该处理器也可以是任何常规的处理器等。

[0324] 所述存储器81可以是所述终端设备8的内部存储单元,例如终端设备8的硬盘或内存。所述存储器81也可以是所述终端设备8的外部存储设备,例如所述终端设备8上配备的插接式硬盘,智能存储卡(Smart Media Card,SMC),安全数字(Secure Digital,SD)卡,闪存卡(Flash Card)等。进一步地,所述存储器81还可以既包括所述终端设备8的内部存储单元也包括外部存储设备。所述存储器81用于存储所述计算机程序以及所述终端设备8所需的其他程序和数据。所述存储器81还可以用于暂时地存储已经输出或者将要输出的数据。

[0325] 所属领域的技术人员可以清楚地了解到,为了描述的方便和简洁,仅以上述各功

能单元、模块的划分进行举例说明,实际应用中,可以根据需要而将上述功能分配由不同的功能单元、模块完成,即将所述装置的内部结构划分成不同的功能单元或模块,以完成以上描述的全部或者部分功能。实施例中的各功能单元、模块可以集成在一个处理单元中,也可以是各个单元单独物理存在,也可以两个或两个以上单元集成在一个单元中,上述集成的单元既可以采用硬件的形式实现,也可以采用软件功能单元的形式实现。另外,各功能单元、模块的具体名称也只是为了便于相互区分,并不用于限制本申请的保护范围。上述系统中单元、模块的具体工作过程,可以参考前述方法实施例中的对应过程,在此不再赘述。

[0326] 所属领域的技术人员可以清楚地了解到,为描述的方便和简洁,上述描述的系统,装置和单元的具体工作过程,可以参考前述方法实施例中的对应过程,在此不再赘述。

[0327] 在上述实施例中,对各个实施例的描述都各有侧重,某个实施例中未详述或记载的部分,可以参见其它实施例的相关描述。

[0328] 本领域普通技术人员可以意识到,结合本文中所公开的实施例描述的各示例的单元及算法步骤,能够以电子硬件、或者计算机软件和电子硬件的结合来实现。这些功能究竟以硬件还是软件方式来执行,取决于技术方案的特定应用和设计约束条件。专业技术人员可以对每个特定的应用来使用不同方法来实现所描述的功能,但是这种实现不应认为超出本申请的范围。

[0329] 在本申请所提供的实施例中,应该理解到,所揭露的装置和方法,可以通过其它的方式实现。例如,以上所描述的系统实施例仅仅是示意性的,例如,所述模块或单元的划分,仅仅为一种逻辑功能划分,实际实现时可以有另外的划分方式,例如多个单元或组件可以结合或者可以集成到另一个系统,或一些特征可以忽略,或不执行。另一点,所显示或讨论的相互之间的耦合或直接耦合或通讯连接可以是通过一些接口,装置或单元的间接耦合或通讯连接,可以是电性,机械或其它的形式。

[0330] 所述作为分离部件说明的单元可以是或者也可以不是物理上分开的,作为单元显示的部件可以是或者也可以不是物理单元,即可以位于一个地方,或者也可以分布到多个网络单元上。可以根据实际的需要选择其中的部分或者全部单元来实现本实施例方案的目的。

[0331] 另外,在本申请各个实施例中的各功能单元可以集成在一个处理单元中,也可以是各个单元单独物理存在,也可以两个或两个以上单元集成在一个单元中。上述集成的单元既可以采用硬件的形式实现,也可以采用软件功能单元的形式实现。

[0332] 所述集成的单元如果以软件功能单元的形式实现并作为独立的产品销售或使用时,可以存储在一个计算机可读取存储介质中。基于这样的理解,本申请实现上述实施例方法中的全部或部分流程,也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计算机可读存储介质中,该计算机程序在被处理器执行时,可实现上述各个方法实施例的步骤。其中,所述计算机程序包括计算机程序代码,所述计算机程序代码可以为源代码形式、对象代码形式、可执行文件或某些中间形式等。所述计算机可读介质可以包括:能够携带所述计算机程序代码的任何实体或装置、记录介质、U盘、移动硬盘、磁碟、光盘、计算机存储器、只读存储器(ROM, Read-Only Memory)、随机存取存储器(RAM, Random Access Memory)、电载波信号、电信信号以及软件分发介质等。需要说明的是,所述计算机可读介质包含的内容可以根据司法管辖区内立法和专利实践的要求进行适当的增减,例如

在某些司法管辖区,根据立法和专利实践,计算机可读介质不包括是电载波信号和电信信号。

[0333] 以上所述实施例仅用以说明本申请的技术方案,而非对其限制;尽管参照前述实施例对本申请进行了详细的说明,本领域的普通技术人员应当理解:其依然可以对前述各实施例所记载的技术方案进行修改,或者对其中部分技术特征进行等同替换;而这些修改或者替换,并不使相应技术方案的本质脱离本申请各实施例技术方案的精神和范围,均应包含在本申请的保护范围之内。

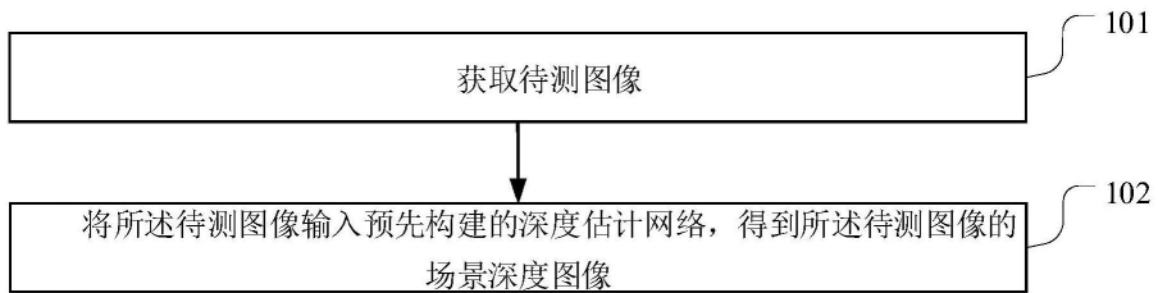


图1

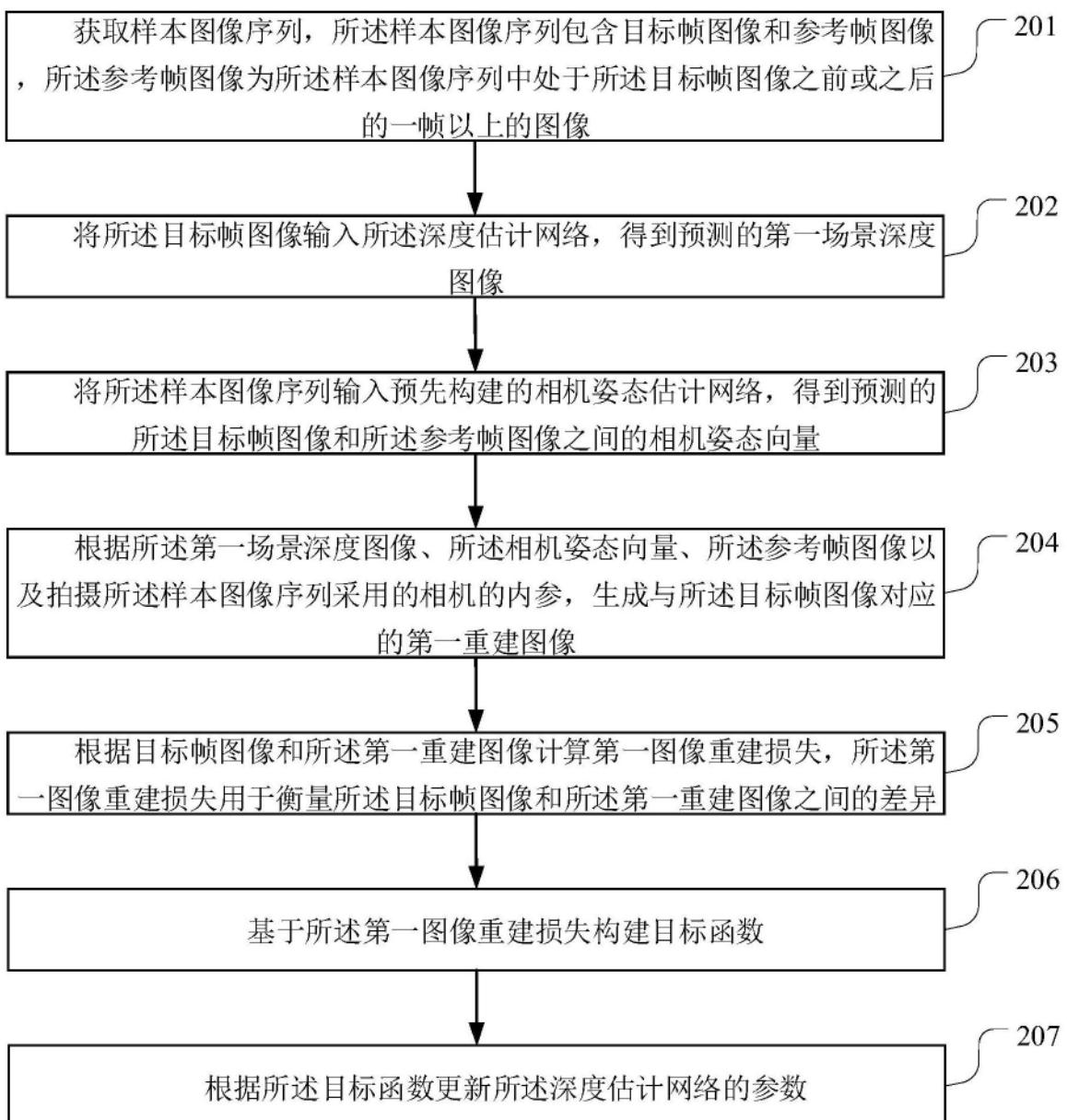


图2

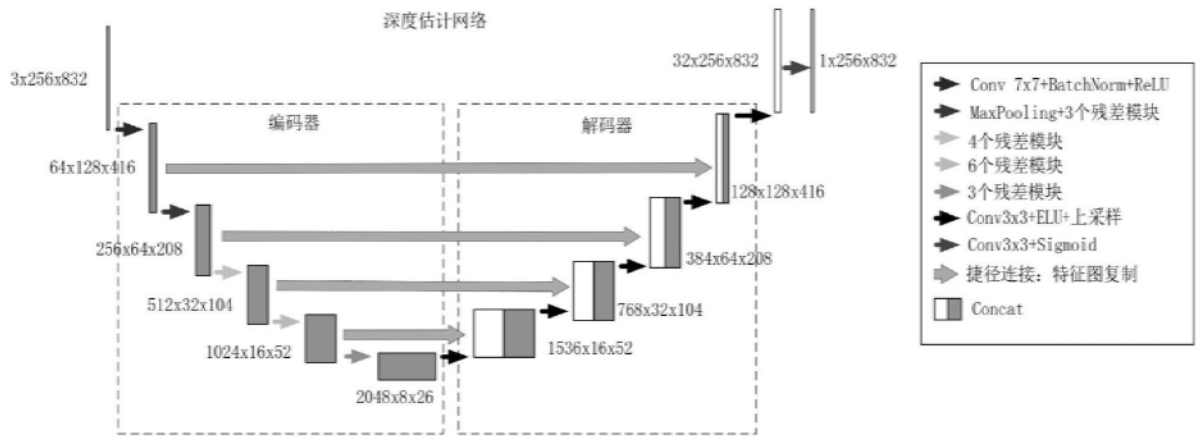


图3

残差模块

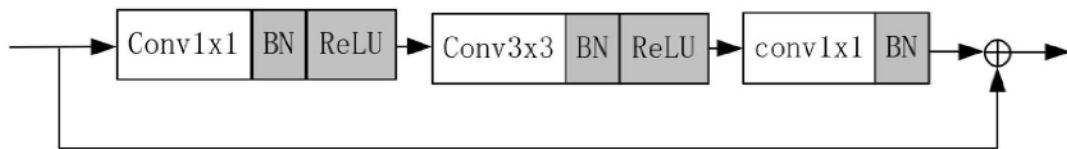


图4

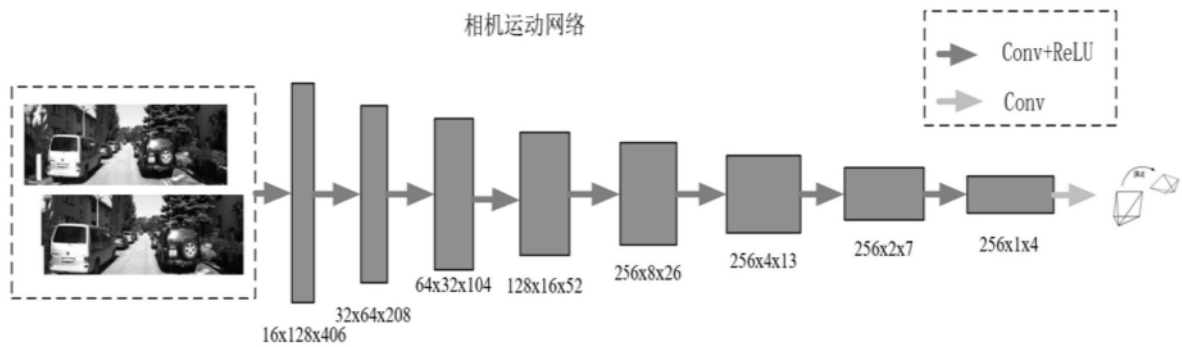


图5



图6

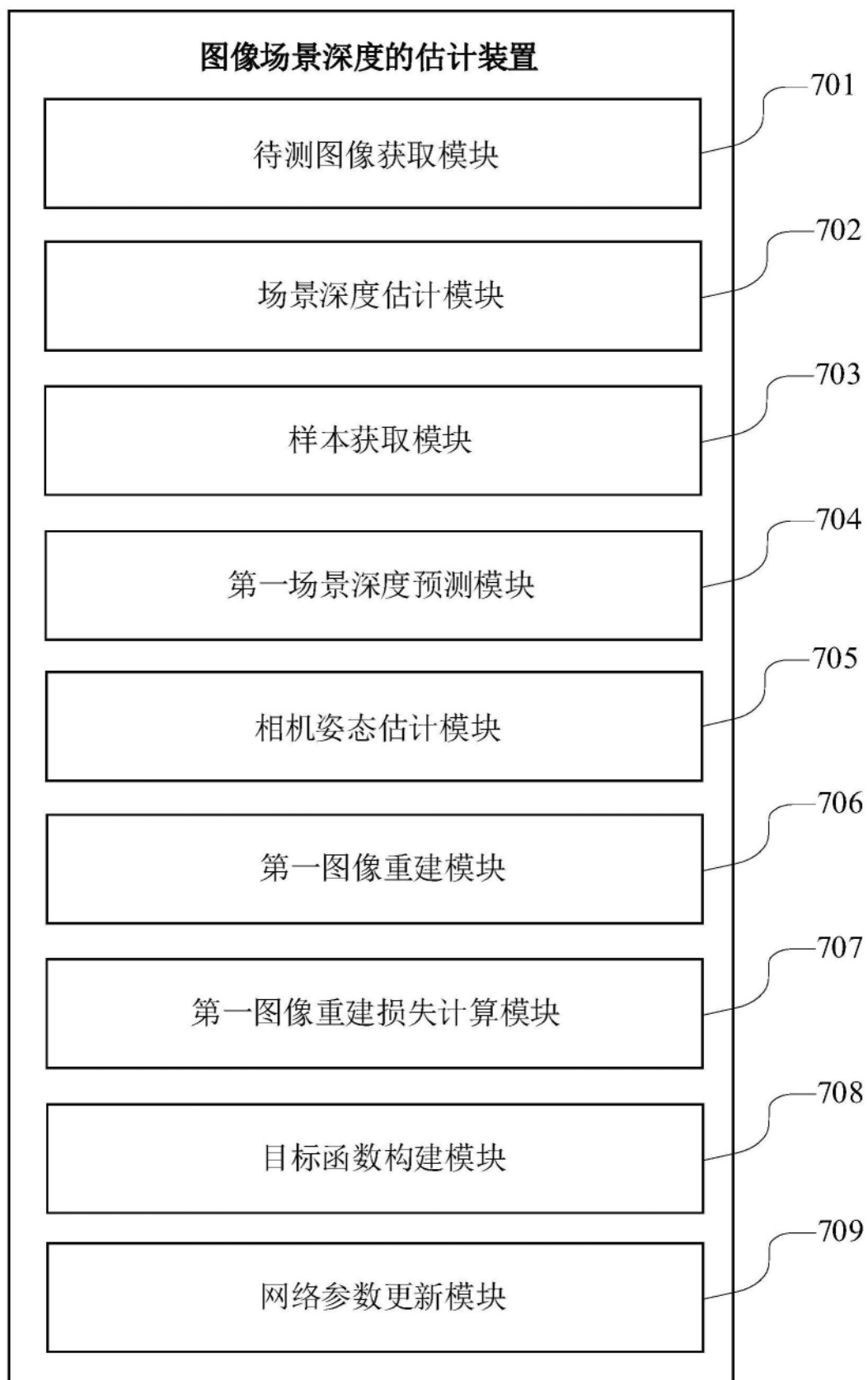


图7

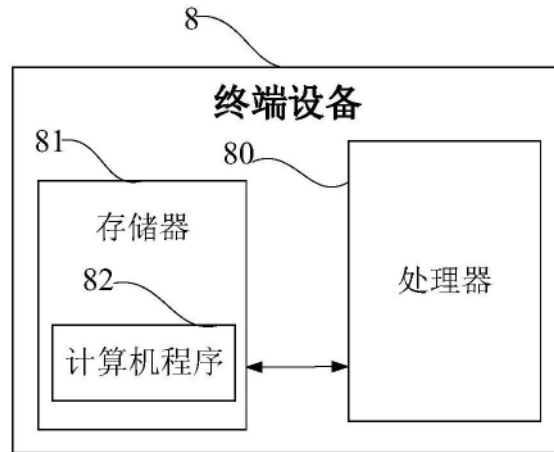


图8