(19)国家知识产权局

(12)发明专利申请

(10)申请公布号 CN 114627176 A
(43)申请公布日 2022.06.14

(21)申请号 202210139037.7

(22)申请日 2022.02.15

(71)申请人 中国科学院深圳先进技术研究院
地址 518055 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

(74)专利代理机构 北京市诚辉律师事务所 11430
专利代理师 成丹 耿慧敏

(51)Int.Cl.
G06T 7/55(2017.01)
G06T 3/00(2006.01)
G06N 3/02(2006.01)
G06K 9/62(2022.01)
G06V 10/774(2022.01)

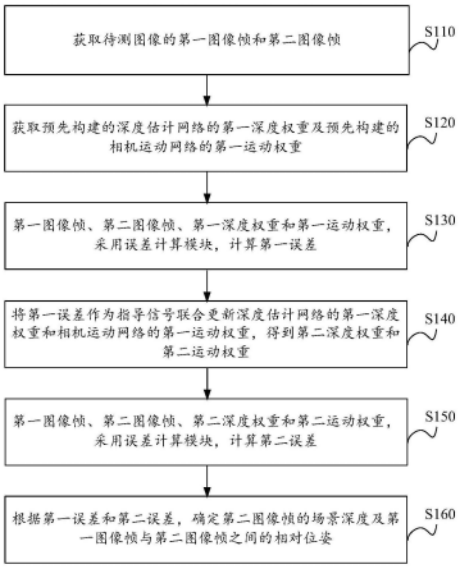
权利要求书4页 说明书16页 附图4页

(54)发明名称

一种基于历史信息的场景深度推理方法、装置及电子设备

(57)摘要

本申请提供一种基于历史信息的场景深度推理方法、装置及电子设备，该方法包括：获取第一图像帧和第二图像帧；获取深度估计网络的第一深度权重及相机运动网络的第一运动权重；第一图像帧、第二图像帧、第一深度权重和第一运动权重，采用误差计算模块，计算第一误差；将第一误差作为指导信号联合更新深度估计网络的第一深度权重和相机运动网络的第一运动权重，得到第二深度权重和第二运动权重；采用误差计算模块，计算第二误差；根据第一误差和第二误差，确定第二图像帧的场景深度及第一图像帧与第二图像帧之间的相对位姿。该方案可以降低因姿态不准确而产生的错误仿射变换的影响。



CN 114627176 A

1.一种基于历史信息的场景深度推理方法,其特征在于,所述方法包括:

获取待测图像的第一图像帧和第二图像帧,所述第一图像帧为所述第二图像帧前一时刻的图像帧;

获取预先构建的深度估计网络的第一深度权重及预先构建的相机运动网络的第一运动权重;

所述第一图像帧、所述第二图像帧、所述第一深度权重和所述第一运动权重,采用误差计算模块,计算第一误差;

将所述第一误差作为指导信号联合更新所述深度估计网络的第一深度权重和所述相机运动网络的第一运动权重,得到第二深度权重和第二运动权重;

所述第一图像帧、所述第二图像帧、所述第二深度权重和所述第二运动权重,采用所述误差计算模块,计算第二误差;

根据所述第一误差和所述第二误差,确定所述第二图像帧的场景深度及所述第一图像帧与所述第二图像帧之间的相对位姿。

2.根据权利要求1所述的方法,其特征在于,所述相机运动网络包括编码器、时间注意力模块和时空相关性模块;

所述编码器用于提取所述堆叠图像帧的特征,得到堆叠特征图;所述堆叠图像帧为所述第一图像帧和所述第二图像帧按通道维度堆叠得到的;

所述时间注意力模块用于将历史记忆单元的信息与当前输入单元的信息建立全局依赖关系,并通过更新单元,将全局相关的所述历史记忆单元中的信息注入到所述当前输入单元,同时将所述当前输入单元中的全局相关信息储存到所述历史记忆单元,作为下一时刻的历史记忆单元;所述当前输入单元包括所述堆叠特征图,所述堆叠特征图通过所述更新单元更新为更新后特征图,所述历史记忆单元包括第一记忆特征图和第一时间特征图,所述下一时刻的历史记忆单元包括第二记忆特征图和第二时间特征图;

所述时空相关性模块用于将所述更新后特征图/所述第二记忆特征图分别建模成为具有空间相关性的第一/第二时空特征图。

3.根据权利要求2所述的方法,其特征在于,所述将历史记忆单元的信息与当前输入单元的信息建立全局依赖关系,并通过更新单元,将全局相关的所述历史记忆单元中的信息注入到所述当前输入单元,同时将所述当前输入单元中的全局相关信息储存到所述历史记忆单元,作为下一时刻的历史记忆单元,包括:

将所述堆叠特征图中的特征信息和所述第一记忆特征图的特征信息注入到所述第一时间特征图,得到第三时间特征图;

根据所述第三时间特征图,确定时间注意力特征向量;

根据所述堆叠特征图,确定第一特征向量;

根据所述第一记忆特征图;确定第二特征向量

根据所述第一特征向量和所述时间注意力特征向量,确定基于时间注意力的输入特征向量;

根据所述第二特征向量和所述时间注意力特征向量,确定基于时间注意力的记忆特征向量;

分别将所述基于时间注意力的输入特征向量和所述基于时间注意力的记忆特征向量,

调整成对应的第一特征图和第二特征图；

根据所述第一特征图和所述第二特征图，确定所述更新后特征图和所述第二记忆特征图；

根据所述更新后特征图和所述第二记忆特征图，将所述第三时间特征图更新为所述第二时间特征图。

4.根据权利要求2所述的方法，其特征在于，将所述更新后特征图建模成具有空间相关性的第一时空特征图，包括：

将所述更新后特征图在通道维度进行切片，得到第一子特征图、第二子特征图和第三子特征图；

分别将所述更新后特征图、所述第一子特征图、所述第二子特征图和所述第三子特征图调整成为第三特征向量、第一子特征向量、第二子特征向量和第三子特征向量；

根据所述第一子特征向量和所述第二子特征向量，计算所述第一子特征图和所述第二子特征图之间的第一空间相关性矩阵；

利用所述第一空间相关性矩阵对所述第三子特征向量进行加权处理，得到第一空间相关特征向量；

根据所述第一空间相关特征向量和所述第三特征向量，确定第一时空特征向量；

将所述第一时空特征向量调整为具有空间相关性的所述第一时空特征图；

将所述第二记忆特征图建模成具有空间相关性的第二时空特征图，包括：

将所述第二记忆特征图在通道维度进行切片，得到第四子特征图、第五子特征图和第六子特征图；

分别将所述第二记忆特征图、所述第四子特征图、所述第五子特征图和所述第六子特征图调整为第四特征向量、第四子特征向量、第五子特征向量和第六子特征向量；

根据所述第四子特征向量和所述第五子特征向量，计算所述第四子特征图和所述第五子特征图之间的第二空间相关性矩阵；

利用所述第二空间相关性矩阵对所述第六子特征向量进行加权处理，得到第二空间相关特征向量；

根据所述第二空间相关特征向量和所述第四特征向量，确定第二时空特征向量；

将所述第二时空特征向量调整为具有空间相关性的所述第二时空特征图。

5.根据权利要求1项所述的方法，其特征在于，

所述第一图像帧、所述第二图像帧、所述第一深度权重和所述第一运动权重，采用误差计算模块，计算第一误差，包括：

将所述第一图像帧和所述第二图像帧输入预先构建的深度估计网络，根据所述第一图像帧和所述第一深度权重，得到所述第一图像帧的第一场景深度和所述第一图像帧的第一编码器特征图，根据所述第二图像帧和所述第一深度权重，得到所述第二图像帧的第二场景深度和所述第二图像帧的第二编码器特征图；

将所述第一图像帧和所述第二图像帧输入预先构建的相机运动网络，根据所述第一图像帧、所述第二图像帧和所述第一运动权重，得到所述第一图像帧和所述第二图像帧之间的第一相对位姿；

所述第一图像帧、所述第二图像帧、所述第一场景深度、所述第二场景深度、所述第一

编码器特征图、所述第二编码器特征图及所述第一相对位姿，采用所述误差计算模块，计算所述第一误差；

所述第一图像帧、所述第二图像帧、所述第二深度权重和所述第二运动权重，采用所述误差计算模块，计算第二误差，包括：

将所述第一图像帧和所述第二图像帧输入预先构建的深度估计网络，根据所述第一图像帧和所述第二深度权重，得到所述第一图像帧的第三场景深度和所述第一图像帧的第三编码器特征图，根据所述第二图像帧和所述第二深度权重，得到所述第二图像帧的第四场景深度和所述第二图像帧的第四编码器特征图；

将所述第一图像帧和所述第二图像帧输入预先构建的相机运动网络，根据所述第一图像帧、所述第二图像帧和所述第二运动权重，得到所述第一图像帧和所述第二图像帧之间的第二相对位姿；

所述第一图像帧、所述第二图像帧、所述第三场景深度、所述第四场景深度、所述第三编码器特征图、所述第四编码器特征图及所述第二相对位姿，采用所述误差计算模块，计算所述第二误差。

6.根据权利要求5所述的方法，其特征在于，所述根据所述第一误差和所述第二误差，确定所述第二图像帧的场景深度及所述第一图像帧与所述第二图像帧之间的相对位姿，包括：

若所述第一误差大于所述第二误差，将所述第二场景深度作为所述第二图像帧的场景深度，将所述第一相对位姿作为所述第一图像帧与所述第二图像帧之间的相对位姿；

若所述第一误差小于或等于所述第二误差，将所述第四场景深度作为所述第二图像帧的场景深度，将所述第二相对位姿作为所述第一图像帧与所述第二图像帧之间的相对位姿。

7.根据权利要求5或6所述的方法，其特征在于，总误差包括所述第一误差和所述第二误差；

所述总误差根据图像合成误差、场景深度结构一致性误差、特征感知损失误差、平滑损失误差确定。

8.根据权利要求7所述的方法，其特征在于，所述总误差根据图像合成误差、场景深度结构一致性误差、特征感知损失误差、平滑损失误差确定，包括：

获取所述第一图像帧的第一图像坐标、所述第二图像帧的第二图像坐标；

根据所述第一图像坐标、相机内参、所述第一场景深度，确定所述第一图像帧的第一世界坐标；

根据所述第二图像坐标、所述相机内参、所述第二场景深度，确定所述第二图像帧的第二世界坐标；

将所述第一图像帧的第一世界坐标仿射变换到所述第二图像帧面板，确定仿射变换后的第三世界坐标；

将所述第二图像帧的第二世界坐标仿射变换到所述第一图像帧面板，确定仿射变换后的第四世界坐标；

将所述第三世界坐标和所述第四世界坐标分别投影到二维平面，得到第一仿射变换后场景深度和第二仿射变换后场景深度及对应的第一仿射变换后图像坐标和第二仿射变换

后图像坐标；

根据所述第一场景深度、所述第二场景深度、所述第一仿射变换后图像坐标、所述第二仿射变换后图像坐标，确定所述场景深度结构一致性误差、第一深度结构不一致性权重和第二深度结构不一致性权重；

根据所述第一图像帧的第一图像坐标、所述第二仿射变换后图像坐标、所述第二图像帧的第二图像坐标、所述第一仿射变换后图像坐标，确定第一相机流一致性遮挡掩码和第二相机流一致性遮挡掩码；

根据所述第一深度结构不一致性权重、所述第二深度结构不一致性权重、所述第一相机流一致性遮挡掩码和所述第二相机流一致性遮挡掩码，确定所述图像合成误差；

根据所述第一图像帧、所述第二图像帧、所述第一仿射变换后图像坐标和所述第二仿射变换后图像坐标，确定所述特征感知损失误差；

根据所述第一场景深度、所述第二场景深度、所述第一图像帧和所述第二图像帧，确定所述平滑损失误差；

根据所述图像合成误差、场景深度结构一致性误差、特征感知损失误差、平滑损失误差，确定所述总误差。

9.一种基于历史信息的场景深度推理装置，其特征在于，所述装置包括：

第一获取模块，用于获取待测图像的第一图像帧和第二图像帧，所述第一图像帧为所述第二图像帧前一时刻的图像帧；

第二获取模块，用于获取预先构建的深度估计网络的第一深度权重及预先构建的相机运动网络的第一运动权重；

第一处理模块，用于所述第一图像帧、所述第二图像帧、所述第一深度权重和所述第一运动权重，采用误差计算模块，计算第一误差；

更新模块，用于将所述第一误差作为指导信号联合更新所述深度估计网络的第一深度权重和所述相机运动网络的第一运动权重，得到第二深度权重和第二运动权重；

第二处理模块，用于所述第一图像帧、所述第二图像帧、所述第二深度权重和所述第二运动权重，采用所述误差计算模块，计算第二误差；

确定模块，用于根据所述第一误差和所述第二误差，确定所述第二图像帧的场景深度及所述第一图像帧与所述第二图像帧之间的相对位姿。

10.一种电子设备，包括存储器、处理器及存储在存储器上并可在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程序时实现如权利要求1-8中任一所述的基于历史信息的场景深度推理方法。

一种基于历史信息的场景深度推理方法、装置及电子设备

技术领域

[0001] 本申请属于计算机视觉与图像处理技术领域，特别涉及一种基于历史信息的 场景深度推理方法、装置及电子设备。

背景技术

[0002] 从二维图像中准确的恢复出场景深度有助于更好的理解场景的三维结构，从 而更好的完成各种视觉任务。然而，普通的相机在拍摄时获取的都是二维图像， 丢失了场景的深度信息，因此如何从二维图像或者视频序列中恢复出场景深度成 为了计算机视觉领域基础且极具挑战的任务。虽然目前已能从二维图像中恢复出 有竞争的场景深度，但是需要大量的人工标注的数据去训练神经网络，耗时费力， 且一旦完成模型的训练，模型的权重即被冻结，降低了算法对未知场景的泛化能 力。此外，基于全无监督学习的方案从二维图像中恢复场景深度，需要同时从相 邻帧中预测出相机姿态，而不准确的姿态会产生错误的仿射变换结果，直接影响 合成图像的质量，从而影响到恢复的场景深度质量。

发明内容

[0003] 本说明书实施例的目的是提供一种基于历史信息的场景深度推理方法、装置 及电子设备。

[0004] 为解决上述技术问题，本申请实施例通过以下方式实现的：

[0005] 第一方面，本申请提供一种基于历史信息的场景深度推理方法，该方法包括：

[0006] 获取待测图像的第一图像帧和第二图像帧，第一图像帧为第二图像帧前一时 刻的图像帧；

[0007] 获取预先构建的深度估计网络的第一深度权重及预先构建的相机运动网络 的第一运动权重；

[0008] 第一图像帧、第二图像帧、第一深度权重和第一运动权重，采用误差计算模 块，计算第一误差；

[0009] 将第一误差作为指导信号联合更新深度估计网络的第一深度权重和相机运 动网络的第一运动权重，得到第二深度权重和第二运动权重；

[0010] 第一图像帧、第二图像帧、第二深度权重和第二运动权重，采用误差计算模 块，计算第二误差；

[0011] 根据第一误差和第二误差，确定第二图像帧的场景深度及第一图像帧与第二 图像帧之间的相对位姿。

[0012] 第二方面，本申请提供一种基于历史信息的场景深度推理装置，该装置包括：

[0013] 第一获取模块，用于获取待测图像的第一图像帧和第二图像帧，第一图像帧 为第二图像帧前一时刻的图像帧；

[0014] 第二获取模块，用于获取预先构建的深度估计网络的第一深度权重及预先构 建的相机运动网络的第一运动权重；

[0015] 第一处理模块，用于第一图像帧、第二图像帧、第一深度权重和第一运动权 重，采用误差计算模块，计算第一误差；

[0016] 更新模块，用于将第一误差作为指导信号联合更新深度估计网络的第一深度 权重和相机运动网络的第一运动权重，得到第二深度权重和第二运动权重；

[0017] 第二处理模块，用于第一图像帧、第二图像帧、第二深度权重和第二运动权 重，采用误差计算模块，计算第二误差；

[0018] 确定模块，用于根据第一误差和第二误差，确定第二图像帧的场景深度及第 一图像帧与第二图像帧之间的相对位姿。

[0019] 第三方面，本申请提供一种电子设备，包括存储器、处理器及存储在存储器 上并可在处理器上运行的计算机程序，处理器执行程序时实现如第一方面的基于 历史信息的场景深度推理方法。

[0020] 由以上本说明书实施例提供的技术方案可见，该方案：采用全无监督的形式 从二维图像中恢复场景深度，通过时间注意力模块，将记忆单元中的历史帧信息 注入到当前输入单元中，并对时空特征图的空间相关性进行建模来提高相机位姿 的精度，降低因姿态不准确而产生的错误仿射变换的影响；推理期间，利用在线 决策推理来提高算法对未知场景的泛化能力。

附图说明

[0021] 为了更清楚地说明本说明书实施例或现有技术中的技术方案，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 的附图仅仅是本说明书中记载的一些实施例，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。

[0022] 图1为本申请提供的基于历史信息的场景深度推理方法的流程示意图；

[0023] 图2为本申请实施例提供的深度估计网络和相机运动网络联合训练框图；

[0024] 图3为本申请实施例提供的时间注意力模块的原理示意图；

[0025] 图4为本申请实施例提供的时空相关性模块的原理示意图；

[0026] 图5为本申请实施例提供的场景深度推理过程示意图；

[0027] 图6为本申请提供的基于历史信息的场景深度推理装置的结构示意图；

[0028] 图7为本申请提供的电子设备的结构示意图。

具体实施方式

[0029] 为了使本技术领域的人员更好地理解本说明书中的技术方案，下面将结合本 说明书实施例中的附图，对本说明书实施例中的技术方案进行清楚、完整地描述， 显然，所描述的实施例仅仅是本说明书一部分实施例，而不是全部的实施例。基 于本说明书中的实施例，本领域普通技术人员在没有作出创造性劳动前提下所获 得的所有其他实施例，都应当属于本说明书保护的范围。

[0030] 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。

[0031] 在不背离本申请的范围或精神的情况下,可对本申请说明书的具体实施方式 做多种改进和变化,这对本领域技术人员而言是显而易见的。由本申请的说明书 得到的其他实施方式对技术人员而言是显而易见得的。本申请说明书和实施例仅 是示例性的。

[0032] 关于本文中所使用的“包含”、“包括”、“具有”、“含有”等等,均为开放性的 用语,即意指包含但不限于。

[0033] 本申请中的“份”如无特别说明,均按质量份计。

[0034] 下面结合附图和实施例对本发明进一步详细说明。

[0035] 参照图1,其示出了适用于本申请实施例提供的基于历史信息的场景深度推 理方法的流程示意图。

[0036] 如图1所示,基于历史信息的场景深度推理方法,可以包括:

[0037] S110、获取待测图像的第一图像帧和第二图像帧,第一图像帧为第二图像帧 前一时刻的图像帧。

[0038] 其中,待测图像为需要推理预测场景深度的任意一幅图像,待测图像可以为 二维图像。

[0039] 将待测图像按照等时间间隔剪辑为若干相邻帧的图像帧,其中,待测图像的 第一图像帧和第二图像帧为相邻时刻的两个图像帧,且第一图像帧为第二图像帧 前一时刻的图像帧,例如第一图像帧为t-1时刻的图像帧,第二图像帧为t时刻 的图像帧,或第一图像帧为t时刻的图像帧,第二图像帧为t+1时刻的图像帧等。

[0040] S120、获取预先构建的深度估计网络的第一深度权重及预先构建的相机运动 网络的第一运动权重。

[0041] 其中,深度估计网络用于估计二维图像的场景深度,该深度估计网络可以采 用具有编码器解码器结构的神经网络,对于该深度估计网络采用的神经网络的类 型和网络结构不做限定。

[0042] 相机运动网络用于预测相邻图像帧之间的相对位姿。

[0043] 深度估计网络和相机运动网络均是预先构建及训练好的。

[0044] 参照图2,其示出了本申请实施例提供的深度估计网络和相机运动网络联合 训练框图(图2中最左边原始图像、场景深度图片及合成的视图均为彩色图,这 里处理为了灰度图)。可以理解的,要训练深度估计网络和相机运动网络,先获 取训练集数据,本申请中训练集数据为图像中若干相邻三个时刻的图像帧组,例 如,t-1时刻的图像帧、t时刻的图像帧、t+1时刻的图像帧为训练数据集中一个 数据,t-2时刻的图像帧、t-1时刻的图像帧、t时刻的图像帧也为训练数据集中 一个数据,以此类推。按照图2所示的训练框图,以t-1时刻的图像帧、t时刻 的图像帧、t+1时刻的图像帧作为相机运动网络和深度估计网络的输入,利用目 标函数联合训练深度估计网络和相机运动网络,即指导深度估计网络的深度权重和相机运动网络的运动权重的更新。

[0045] 可以理解的,t-1时刻的图像帧、t时刻的图像帧、t+1时刻的图像帧输入深 度估计网络和相机运动网络之前,先进行图像预处理,例如包括对图像数据进行 随机翻转、随机裁剪、数据归一化处理,并将处理后的数据转换成维度为 C×H×W的张量数据,此处省略batch维度,其中,C表示样本的通道维度大小, 其中,训练期间深度估计网络C=3,相机运

动网络中C＝9(测试期间相机运动 网络的输入为相邻时刻的3个图像帧),测试推理期间(即实现本申请基于历史 信息的场景深度推理方法时),相机运动网络中C＝6(推理期间相机运动网络的 输入为相邻时刻的2个图像帧),H表示输入样本图像的高度,示例性的,H＝256, W表示输入样本图像的宽度,示例性的,W＝832。

[0046] 继续参照图2,相机运动网络可以包括编码器、时间注意力模块和时空相关 性模块。

[0047] 其中,编码器用于提取堆叠图像帧的特征,得到堆叠特征图;堆叠图像帧为 第一图像帧和第二图像帧按通道维度堆叠得到的。

[0048] 时间注意力模块用于将历史记忆单元的信息与当前输入单元的信息建立全 局依赖关系,并通过更新单元,将全局相关的历史记忆单元中的信息注入到当前 输入单元,同时将当前输入单元中的全局相关信息储存到历史记忆单元,作为下 一时刻的历史记忆单元;当前输入单元包括堆叠特征图,堆叠特征图通过更新单 元更新为更新后特征图,历史记忆单元包括第一记忆特征图和第一时间特征图, 下一时刻的历史记忆单元包括第二记忆特征图和第二时间特征图。

[0049] 时空相关性模块用于将更新后特征图/第二记忆特征图分别建模成为具有空 间相关性的第一/第二时空特征图。

[0050] 在一个实施例中,时间注意力模块,利用共享的时间注意力权重,将历史记 忆单元的信息与当前输入单元的信息建立全局依赖关系,并通过更新单元,将全 局相关的历史记忆单元中的信息注入到当前输入单元,同时将当前输入单元中的 全局相关信息储存到历史记忆单元,作为下一时刻的历史记忆单元,包括:

[0051] 将堆叠特征图中的特征信息和第一记忆特征图的特征信息注入到第一时间 特征图,得到第三时间特征图;

[0052] 根据第三时间特征图,确定时间注意力特征向量;

[0053] 根据堆叠特征图,确定第一特征向量;

[0054] 根据第一记忆特征图;确定第二特征向量

[0055] 根据第一特征向量和时间注意力特征向量,确定基于时间注意力的输入特征 向量;

[0056] 根据第二特征向量和时间注意力特征向量,确定基于时间注意力的记忆特征 向量;

[0057] 分别将基于时间注意力的输入特征向量和基于时间注意力的记忆特征向量, 调整成对应的第一特征图和第二特征图;

[0058] 根据第一特征图和第二特征图,确定更新后特征图和第二记忆特征图;

[0059] 根据更新后特征图和第二记忆特征图,将第三时间特征图更新为第二时间特 征图。

[0060] 示例性的,参照图3,其示出了本申请实施例提供的时间注意力模块的原理 示意图。为描述方便,假设t时刻的输入特征图(即堆叠特征图)表示为 $X_{(i,t)} \in \Re^{C\times H\times W}$,t-1时刻时间特征图(即第一时间特征图)为$X_{(time,t-1)} \in \Re^{C\times H\times W}$, t-1时刻时间记忆特征图(即第一记忆特征图)为$X_{(m,t-1)} \in \Re^{C\times H\times W}$,其计算过程 如下:

[0061] 1)根据公式(1)将t时刻的输入特征图X_t中的特征信息和t-1时刻记忆特 征图$X_{(m,t-1)}$中的特征信息注入到t-1时刻时间特征图$X_{(time,t-1)}$中,得到第三时间 特征图$\hat{X}_{(time,t-1)} \in \Re^{C\times H\times W}$:

[0062] $$\hat{X}_{(time,t-1)} = \delta_{gelu}(W_{(i,t)}X_{(i,t)} + b_{(i,t)} + W_{(time,t-1)}X_{(time,t-1)} + b_{(time,t-1)} + W_{(m,t-1)}X_{(m,t-1)} + b_{(m,t-1)}) \quad (1)$$

[0063] 其中,$\Re$表示所属特征空间,C表示特征图通道数量,H表示特征图的高 度,W表示特征图的宽度,$\delta_{gelu}(\cdot)$表示激活函数,$W_{(i,t)}$,$W_{(time,t-1)}$,$W_{(m,t-1)}$表示学习 出的对应的权重,$b_{(i,t)}$,$b_{(time,t-1)}$,$b_{(m,t-1)}$表示对应的偏执项。

[0064] 2)根据公式组(2)计算出时间注意力特征向量$x_{(qk_time,t-1)}$:

[0065] $$X_{(qk\_time,t-1)} = \hat{W}_{(time,t-1)}\hat{X}_{(time,t-1)} + \hat{b}_{(time,t-1)}, \quad X_{(qk\_time,t-1)} \in \Re^{2C\times H\times W}$$

[0066] $$X_{(q\_time,t-1)}, X_{(k\_time,t-1)} = F_{split}(X_{(qk\_time,t-1)}), \quad X_{(q\_time,t-1)}, X_{(k\_time,t-1)} \in \Re^{C\times H\times W}$$

[0067] $$x_{(q\_time,t-1)} = F_{reshape}(X_{(q\_time,t-1)}), \quad x_{(q\_time,t-1)} \in \Re^{C\times HW} \quad (2)$$

[0068] $$x_{(k\_time,t-1)} = F_{reshape}(X_{(k\_time,t-1)}), \quad x_{(k\_time,t-1)} \in \Re^{C\times HW}$$

[0069] $$x_{(qk\_time,t-1)} = s * x^T_{(q\_time,t-1)}x_{(k\_time,t-1)}, \quad x_{(qk\_time,t-1)} \in \Re^{HW\times HW}$$

[0070] 其中,s表示标量缩放因子,“*”表示对应元素的乘积,函数$F_{split}(\cdot)$表示按照 通道维度对特征图进行切片,函数$F_{reshape}(\cdot)$用于将特征图或特征向量调整成预设 形状,“T”表示转置,$\hat{W}_{(time,t-1)}$表示对应的权重,$\hat{b}_{(time,t-1)}$表示对应的偏执项。

[0071] 3)根据公式组(3)将t时刻的输入特征图$X_{(i,t)}$调整成特征向量(即第一特 征向量)$x_{(i,t)} \in \Re^{C\times HW}$,将t-1时刻记忆特征图$X_{(m,t-1)}$调整成特征向量(即第二 特征向量)$x_{(m,t-1)} \in \Re^{C\times HW}$:

[0072] $$x_{(i,t)} = F_{reshape}(X_{(i,t)})$$

[0073] $$x_{(m,t-1)} = F_{reshape}(X_{(m,t-1)}) \quad (3)$$

[0074] 4)根据公式组(4)计算出基于时间注意力的输入特征向量$x_{(qk_i,t)} \in \Re^{HW\times C}$,以及基于时间注意力的记忆特征向量$x_{(qk_m,t-1)} \in \Re^{HW\times C}$

[0075] $$x_{(qk\_i,t)} = x_{(qk\_time,t-1)}x^T_{(i,t)}$$

[0076] $$x_{(qk\_m,t-1)} = x_{(qk\_time,t-1)}x^T_{(m,t-1)} \quad (4)$$

[0077] 5)根据公式组(5)分别将特征向量$x_{(qk_i,t)}$和$x_{(qk_m,t-1)}$调整成对应的特征图$X_{(qk_i,t)} \in \Re^{C\times H\times W}$(即第一特征图)和$X_{(qk_m,t-1)} \in \Re^{C\times H\times W}$(即第二特征图):

[0078] $$X_{(qk\_i,t)} = F_{reshape}(x^T_{(qk\_i,t)})$$

[0079] $$X_{(qk\_m,t-1)} = F_{reshape}(x^T_{(qk\_m,t-1)}) \quad (5)$$

[0080] 6)根据公式(6)计算出信息选择门$G_s \in \Re^{C\times H\times W}$,用于有选择的将t-1时刻 记忆特

征图$X_{(qk_m,t-1)}$中的信息注入到t时刻的输入特征图$X_{(qk_i,t)}$中：

[0081] $$G_s=\delta_{sig}(W_{(qk\_ms,t-1)}X_{(qk\_m,t-1)}+b_{(qk\_ms,t-1)}+W_{(qk\_is,t)}X_{(qk\_i,t)}+b_{(qk\_is,t)}) \quad (6)$$

[0082] 其中，函数$\delta_{sig}(\cdot)$表示sigmoid激活函数，$W_{(qk_ms,t-1)}$和$W_{(qk_is,t)}$表示对应的权 重，$b_{(qk_ms,t-1)}$和$b_{(qk_is,t)}$表示对应的偏执项。

[0083] 7)根据公式(7)计算出包含记忆特征图信息的新的特征图$X_{(im,t)}\in\Re^{C\times H\times W}$：

[0084] $$X_{(im,t)}=\delta_{tanh}(W_{(qk\_imi,t)}X_{(qk\_i,t)}+b_{(qk\_imi,t)}+G_s*(W_{(qk\_imm,t-1)}X_{(qk\_m,t-1)}+b_{(qk\_imm,t-1)})) \quad (7)$$

[0085] 其中，函数$\delta_{tanh}(\cdot)$表示tanh激活函数，$W_{(qk_imi,t)}$和$W_{(qk_imm,t-1)}$表示对应的权 重，$b_{(qk_imi,t)}$和$b_{(qk_imm,t-1)}$表示对应的偏执项。

[0086] 8)根据公式组(8)计算出记忆门$G_r\in\Re^{C\times H\times W}$，用于将t-1时刻记忆特征图$X_{(qk_m,t-1)}$信息更新到t时刻的记忆特征图(第二记忆特征图)$X_{(m,t)}\in\Re^{C\times H\times W}$：

[0087] $$G_r=\delta_{sig}(W_{(qk\_ir,t)}X_{(qk\_i,t)}+b_{(qk\_ir,t)}+W_{(qk\_mr,t-1)}X_{(qk\_m,t-1)}+b_{(qk\_mr,t-1)})$$

[0088] $$X_{(m,t)}=(1-G_r)*X_{(im,t)}+G_r*X_{(qk\_m,t-1)} \quad (8)$$

[0089] 其中，$W_{(qk_ir,t)}$和$W_{(qk_mr,t-1)}$表示对应的权重，$b_{(qk_ir,t)}$和$b_{(qk_mr,t-1)}$表示对应的 偏执项。

[0090] 9)根据公式(9)计算出输出门$G_o\in\Re^{C\times H\times W}$，用于更新输入特征图$X_{(qk_i,t)}$，获得更新后的新特征图(即更新后特征图)$X_{(io,t)}\in\Re^{C\times H\times W}$作为下一时刻的输入 特征图：

[0091] $$G_o=\delta_{sig}(W_{(qk\_io,t)}X_{(qk\_i,t)}+b_{(qk\_io,t)}+W_{(qk\_mo,t-1)}X_{(qk\_m,t-1)}+b_{(qk\_mo,t-1)})$$

[0092] $$X_{(io,t)}=G_o*X_{(im,t)} \quad (9)$$

[0093] 其中，$W_{(qk_io,t)}$和$W_{(qk_mo,t-1)}$表示对应的权重，$b_{(qk_io,t)}$和$b_{(qk_mo,t-1)}$表示对应的 偏执项。

[0094] 10)根据公式(10)将t-1时刻的时间特征图$\hat{X}_{(time,t-1)}$更新到t时刻的时间特 征图(即第二时间特征图)$X_{(time,t)}$：

[0095] $$X_{(time,t)}=\delta_{gelu}(W_{(time,t-1)}\hat{X}_{(time,t-1)}+b_{(time,t-1)}+W_{(io,t)}X_{(io,t)}+b_{(io,t)}+W_{(m,t)}X_{(m,t)}+b_{(m,t)}) \quad (10)$$

[0096] 其中，$W_{(time,t-1)}$、$W_{(io,t)}$和$W_{(m,t)}$表示对应的权重，$b_{(time,t-1)}$、$b_{(io,t)}$和$b_{(m,t)}$表示 对应的偏执项。

[0097] 为了利用特征图的全局空间结构信息以及空间结构之间的依赖关系来推理 相机运动，构造如图4所示的时空相关性模块，利用全局空间相关性权重对空间 上下文信息进行建模，通过堆叠帧的特征图通道之间的依赖关系进行建模，来约 束堆叠帧之间的时序信息。

[0098] 在一个实施例中，时空相关性模块用于将更新后特征图/第二记忆特征图分 别建模成具有空间相关性的第一/第二时空特征图。

[0099] 其中，将更新后特征图建模成具有空间相关性的第一时空特征图，包括：

[0100] 将更新后特征图在通道维度进行切片，得到第一子特征图、第二子特征图和 第三子特征图；

[0101] 分别将更新后特征图、第一子特征图、第二子特征图和第三子特征图，对应 调整

为第三特征向量、第一子特征向量、第二子特征向量和第三子特征向量；

[0102] 根据第一子特征向量和第二子特征向量，计算第一子特征图和第二子特征图 之间的第一空间相关性矩阵；

[0103] 利用第一空间相关性矩阵对第三子特征向量进行加权处理，得到第一空间相 关特征向量；

[0104] 根据第一空间相关特征向量和第三特征向量，确定第一时空特征向量；

[0105] 将第一时空特征向量调整为具有空间相关性的第一时空特征图。

[0106] 其中，将第二记忆特征图建模成具有空间相关性的第二时空特征图，包括：

[0107] 将所述第二记忆特征图在通道维度进行切片，得到第四子特征图、第五子特 征图和第六子特征图；

[0108] 分别将第二记忆特征图、第四子特征图、第五子特征图和第六子特征图调整 为第四特征向量、第四子特征向量、第五子特征向量和第六子特征向量；

[0109] 根据第四子特征向量和第五子特征向量，计算第四子特征图和第五子特征图 之间的第二空间相关性矩阵；

[0110] 利用第二空间相关性矩阵对第六子特征向量进行加权处理，得到第二空间相 关特征向量；

[0111] 根据第二空间相关特征向量和第四特征向量，确定第二时空特征向量；

[0112] 将第二时空特征向量调整为具有空间相关性的第二时空特征图。

[0113] 示例性的，参照图4，计算如下：

[0114] 1)根据公式组(11)对输入特征图(即更新后特征图或第二记忆特征图) $X_{mid} \in \Re^{C\times H\times W}$进行变换，得到特征图$\hat{X}_{mid} \in \Re^{3C\times H\times W}$，并将其均等划分成三个子 特征图(即分别为第一/四子特征图、第二/五子特征图和第三/六子特征图) $\hat{X}_{mid\_1}, \hat{X}_{mid\_2}, \hat{X}_{mid\_3} \in \Re^{C\times H\times W}$，其中函数$F_{split}(\cdot)$表示在通道维度对特征图进行切 片处理。

$$\hat{X}_{mid} = W_{mid} X_{mid} + b_{mid} \tag{11}$$

[0115]

$$\hat{X}_{mid\_1}, \hat{X}_{mid\_2}, \hat{X}_{mid\_3} = F_{split}(\hat{X}_{mid})$$

[0116] 其中，W_{mid}为对应的权重，b_{mid}表示对应的偏执项。

[0117] 2)根据公式组(12)将相应特征图调整成特征向量，得到对应的特征向量 $x_{mid}, \hat{x}_{mid\_1}, \hat{x}_{mid\_2}, \hat{x}_{mid\_3} \in \Re^{C\times HW}$，其中函数$F_{reshape}(\cdot)$用于将特征图或特征向量调整 成预设形状。

[0118] $x_{mid} = F_{reshape}(X_{mid})$

[0119] $\hat{x}_{mid\_1} = F_{reshape}(\hat{X}_{mid\_1})$

[0120] $\hat{x}_{mid\_2} = F_{reshape}(\hat{X}_{mid\_2})$

[0121] $\hat{x}_{mid\_3} = F_{reshape}(\hat{X}_{mid\_3})$ (12)

[0122] 3)根据公式(13)计算出第一/第四子特征图$\hat{X}_{(mid,1)}$与第二/第五子特征图 $\hat{X}_{(mid,2)}$

之间的空间相关性矩阵$W_{corr} \in \Re^{HW \times HW}$，其中s表示标量缩放因子

[0123] $$W_{corr} = s * \hat{x}_{mid\_1}^{T} \hat{x}_{mid\_2} \quad (13)$$

[0124] 4)利用上述计算出的空间相关矩阵对第三/六子特征量$\hat{x}_{mid_3}$进行加权处理，得到空间相关特征向量(包括第一空间相关性特征向量和第二空间相关性特征向 量)$x_{corr} \in \Re^{HW \times C}$，如公式(14)所示

[0125] $$x_{corr} = W_{corr} \hat{x}_{mid\_3}^{T} \quad (14)$$

[0126] 5)根据公式(15)对特征图空间结构之间的依赖关系进行建模，通过建模 堆叠帧的特征图通道之间的依赖关系，来约束堆叠帧之间的时序信息，计算出第 一/二时空特征向量x_{time_corr}：

[0127] $$x_{time\_corr} = x_{mid}^{T} + F_C(x_{corr}), x_{time\_corr} \in \Re^{HW \times C} \quad (15)$$

[0128] 其中，$F_C(\cdot)$由两层核大小为3，步长为1的一维卷积和激活函数组成。

[0129] 6)最后将具有空间相关性的第一/二时空特征向量x_{time_corr}调整成具有空间相 关性的第一/二时空特征图$X_{time_corr} \in \Re^{C \times H \times W}$。

[0130] S130、第一图像帧、第二图像帧、第一深度权重和第一运动权重，采用误差 计算模块，计算第一误差，包括：

[0131] 将第一图像帧和第二图像帧输入预先构建的深度估计网络，根据第一图像帧 和第一深度权重，得到第一图像帧的第一场景深度和第一图像帧的第一编码器特 征图，根据第二图像帧和第一深度权重，得到第二图像帧的第二场景深度和第二 图像帧的第二编码器特征图；

[0132] 将第一图像帧和第二图像帧输入预先构建的相机运动网络，根据第一图像帧、第二图像帧和第一运动权重，得到第一图像帧和第二图像帧之间的第一相对位姿；

[0133] 第一场景深度、第二场景深度及第一相对位姿，采用误差计算模块，计算第 一误差。

[0134] 在一个实施例中，总误差根据图像合成误差、场景深度结构一致性误差、特 征感知损失误差、平滑损失误差确定。其中，总误差包括第一误差和第二误差。

[0135] 具体的，总误差根据图像合成误差、场景深度结构一致性误差、特征感知损 失误差、平滑损失误差确定，包括：

[0136] 获取第一图像帧的第一图像坐标、第二图像帧的第二图像坐标；

[0137] 根据第一图像坐标、相机内参、第一场景深度，确定第一图像帧的第一世界 坐标；

[0138] 根据第二图像坐标、相机内参、第二场景深度，确定第二图像帧的第二世界 坐标；

[0139] 将第一图像帧的第一世界坐标仿射变换到第二图像帧面板，确定仿射变换后 的第三世界坐标；

[0140] 将第二图像帧的第二世界坐标仿射变换到第一图像帧面板，确定仿射变换后 的第四世界坐标；

[0141] 将第三世界坐标和第四世界坐标分别投影到二维平面，得到第一仿射变换后 场景深度和第二仿射变换后场景深度及对应的第一仿射变换后图像坐标和第二 仿射变换后

图像坐标；

[0142] 根据第一场景深度、第二场景深度、第一仿射变换后图像坐标、第二仿射变 换后图像坐标，确定场景深度结构一致性误差、第一深度结构不一致性权重和第 二深度结构不一致性权重；

[0143] 根据第一图像帧的第一图像坐标、第二仿射变换后图像坐标、第二图像帧的 第二图像坐标、第一仿射变换后图像坐标，确定第一相机流一致性遮挡掩码和第 二相机流一致性遮挡掩码；

[0144] 根据第一深度结构不一致性权重、第二深度结构不一致性权重、第一相机流 一致性遮挡掩码和第二相机流一致性遮挡掩码，确定图像合成误差；

[0145] 根据第一图像帧、第二图像帧、第一仿射变换后图像坐标和第二仿射变换后 图像坐标，确定特征感知损失误差；

[0146] 根据第一场景深度、第二场景深度、第一图像帧和第二图像帧，确定平滑损 失误差；

[0147] 根据图像合成误差、场景深度结构一致性误差、特征感知损失误差、平滑损 失误差，确定总误差。

[0148] 示例性的，为方便描述，假设已经训练好的场景深度拟合函数是 $D_t=F_D(I_t|W_D)$，其中I_t表示需要恢复t时刻的场景深度的二维图像，W_D表示场 景深度拟合函数$F_D(\cdot)$中学习好的权重参数，D_t表示恢复出的t时刻的二维图像 I_t的场景深度，$X_{(enc,t)}$表示编码器输出的t时刻图像帧I_t的特征图， $T_{t-1}^{t}=F_T([I_{t-1},I_t]|W_T)$表示从t-1时刻的图像帧$I_{t-1}$变换到t时刻的图像帧$I_t$的位 姿，$W_T$表示位姿变换函数$F_T(\cdot)$中学习好的权重参数，相机内参表示为K，用 $P_{(xy,t-1)}$表示图像帧I_{t-1}的图像坐标，$P_{(xyz,t-1)}$图像帧I_{t-1}的世界坐标，$P_{(xy,t)}$表示图像帧I_t的图像坐标，$P_{(xyz,t)}$表示图像帧I_t的世界坐标。总误差计算过程如下：

[0149] 1)根据公式(16)计算出图像帧I_{t-1}的世界坐标(即第一世界坐标)$P_{(xyz,t-1)}$， 图像帧I_t的世界坐标(即第二世界坐标)$P_{(xyz,t)}$：

[0150] $P_{(xyz,t-1)}=D_{t-1}*K^{-}P_{(xy,t-1)}$

[0151] $P_{(xyz,t)}=D_t*K^{-}P_{(xy,t)}$ (16)

[0152] 其中，“*”号表示矩阵对应元素的乘积。

[0153] 2)根据公式(17)计算出将图像帧I_{t-1}的世界坐标$P_{(xyz,t-1)}$仿射变换到图像帧 I_t面板，获得仿射变换后的世界坐标(即第三世界坐标)$P_{(proj\_xyz,t)}$，以及将图像 帧I_t的世界坐标$P_{(xyz,t)}$仿射变换到图像帧I_{t-1}面板，获得仿射变换后的世界坐标 (即第四世界坐标)$P_{(proj\_xyz,t-1)}$：

[0154] $$P_{(proj\_xyz,t)}=KT_{t-1}^{t}P_{(xyz,t-1)}$$

[0155] $$P_{(proj\_xyz,t-1)}=KT_{t}^{t-1}P_{(xyz,t)} \quad (17)$$

[0156] 3)将仿射变换计算出的世界坐标$P_{(proj_xyz,t)}$、$P_{(proj_xyz,t-1)}$分别投影到二维平面，获得仿射变换后的场景深度$D_{(proj,t)}$(即第一仿射变换后场景深度)和$D_{(proj,t-1)}$(即 第二仿射变换后场景深度)，以及对应的仿射变换后的图像坐标$P_{(proj_xy,t)}$(即第 一仿射变换后图像坐标)、$P_{(proj\_xy,t-1)}$(即第二仿射变换后图像坐标)。

[0157] 4)根据图像帧I_{t-1}和$P_{(proj_xy,t-1)}$合成图像$I_{(syn,t)}$；根据编码器特征图$X_{(enc,t-1)}$和

$P_{(proj_xy,t-1)}$合成特征图$X_{(syn_enc,t)}$；根据估计的深度图D_{t-1}和$P_{(projxy,t-1)}$合成深度图 $D_{(syn,t)}$；根据图像帧I_t和$P_{(proj_xy,t)}$合成图像$I_{(syn,t-1)}$；根据编码器特征图$X_{(enc,t)}$和 $P_{(proj_xy,t)}$合成特征图$X_{(syn_enc,t-1)}$；根据估计的深度图D_t和$P_{(projxy,t)}$合成深度图 $D_{(syn,t-1)}$；根据I_{t-1}的图像坐标和$P_{(proj_xy,t-1)}$计算出前向相机流$U_{forward}$；根据I_t的图 像坐标和$P_{(proj_xy,t)}$计算出后向相机流$U_{backward}$；根据$U_{forward}$和$P_{(proj_xy,t)}$合成前向相 机流$U_{syn_forward}$；根据$U_{backward}$和$P_{(projxy,t-1)}$合成后向相机流$U_{syn_backward}$。

[0158] 5)根据公式组(18)分别计算第一相机流一致性遮挡掩码$M_{(occ,t-1)}$和第二相 机流一致性遮挡掩码$M_{(occ,t)}$：

[0159] $M_{(occ,t-1)}=\Gamma\left(\|U_{syn\_forward}+U_{backward}\|^2,\alpha_1\left(\|U_{syn\_forward}\|^2+\|U_{backward}\|^2\right)+\alpha_2\right)$

[0160] $M_{(occ,t)}=\Gamma\left(\|U_{syn\_backward}+U_{forward}\|^2,\alpha_1\left(\|U_{syn\_backward}\|^2+\|U_{forward}\|^2\right)+\alpha_2\right)$

[0161] $$\Gamma(a,b)=\begin{cases}1, a<b\\ 0, otherwise\end{cases} \quad (18)$$

[0162] 6)根据公式组(19)计算场景深度结构一致性误差E_D以及第一深度结构不 一致性权重$M_{(D,t-1)}$、第二深度结构不一致性权重$M_{(D,t)}$。

[0163] $$E_D=\frac{\|D_{(proj,t-1)}-D_{(syn,t-1)}\|}{D_{(proj,t-1)}+D_{(syn,t-1)}}+\frac{\|D_{(proj,t)}+D_{(syn,t)}\|}{D_{(proj,t)}+D_{(syn,t)}} \quad (19)$$

[0164] $$M_{(D,t-1)}=1-\frac{\|D_{(proj,t-1)}-D_{(syn,t-1)}\|}{D_{(proj,t-1)}+D_{(syn,t-1)}}$$

[0165] $$M_{(D,t)}=1-\frac{\|D_{(proj,t)}+D_{(syn,t)}\|}{D_{(proj,t)}+D_{(syn,t)}}$$

[0166] 7)根据公式(20)计算出图像合成误差E_I：

[0167] $$E_I=M_{(D,t)}*M_{(occ,t)}*\left(0.85*\frac{1-SSIM(I_t,I_{(syn,t)})}{2}+0.15*ERF(I_t,I_{(syn,t)})\right)$$
$$+M_{(D,t-1)}*M_{(occ,t-1)}*\left(0.85*\frac{1-SSIM(I_{t-1},I_{(syn,t-1)})}{2}+0.15*ERF(I_{t-1},I_{(syn,t-1)})\right) \quad (20)$$

[0168] 其中，$ERF(a,b)=\sqrt{\|a-b\|^2+\epsilon}$，$\epsilon=0.01$。

[0169] 8)根据公式(21)计算出特征感知损失误差E_X：

[0170] $E_X=ERF\left(X_{(enc,t)},X_{(syn\_enc,t)}\right)+ERF\left(X_{(enc,t-1)},X_{(syn\_enc,t-1)}\right)$ (21)

[0171] 9)根据公式(22)计算出平滑损失误差E_S：

[0172] $$E_S=\sum|\partial D_{t-1}|*e^{-|\partial I_{t-1}|}+\sum|\partial D_t|*e^{-|\partial I_t|}+\sum|\partial X_{(enc,t)}|*e^{-|\partial I_t|}+\sum|\partial X_{(enc,t-1)}|*e^{-|\partial I_{t-1}|} \quad (22)$$

[0173] 10)根据公式(23)计算出总误差E：

[0174] $E=\lambda_I E_I+\lambda_D E_D+\lambda_X E_X+\lambda_S E_S$ (23)

[0175] S140、将第一误差作为指导信号联合更新深度估计网络的第一深度权重和相 机运动网络的第一运动权重，得到第二深度权重和第二运动权重。

[0176] S150、第一图像帧、第二图像帧、第二深度权重和第二运动权重，采用误差 计算模

块,计算第二误差,可以包括:

[0177] 将第一图像帧和第二图像帧输入预先构建的深度估计网络,根据第一图像帧 和第二深度权重,得到第一图像帧的第三场景深度和第一图像帧的第三编码器特 征图,根据第二图像帧和第二深度权重,得到第二图像帧的第四场景深度和第二 图像帧的第四编码器特征图;

[0178] 将第一图像帧和第二图像帧输入预先构建的相机运动网络,根据第一图像帧、第二图像帧和第二运动权重,得到第一图像帧和第二图像帧之间的第二相对位姿;

[0179] 第一图像帧、第二图像帧、第三场景深度、第四场景深度、第三编码器特征 图、第四编码器特征图及第二相对位姿,采用误差计算模块,计算第二误差。

[0180] 该步骤参考S130步骤的具体计算过程,只是将S130中的第一深度权重和 第一运动权重替换为第二深度权重和第二运动权重即可,这里不再赘述。

[0181] S160、根据第一误差和第二误差,确定第二图像帧的场景深度及第一图像帧 与第二图像帧之间的相对位姿。

[0182] 具体的,若第一误差大于第二误差,将第二场景深度作为第二图像帧的场景 深度,将第一相对位姿作为第一图像帧与第二图像帧之间的相对位姿;

[0183] 若第一误差小于或等于第二误差,将第四场景深度作为第二图像帧的场景深 度,将第二相对位姿作为第一图像帧与第二图像帧之间的相对位姿。

[0184] 参照图5,其示出了场景深度推理过程示意图。推理场景深度的过程如下:

[0185] 1)以训练所得的深度估计网络的权重W_D,相机运动网络权重W_T作为推理 期间深度估计网络和相机运动网络的模型权重,根据S130,计算出总误差E、历史帧到当前帧的位姿变换矩阵T_{t-1}^{t}、当前帧的场景深度D_t。

[0186] 2)、以上述1)计算出的总误差作为指导信号来更新深度估计网络和相机运 动网络的权重,得到新的模型权重$\hat{W}_D$和$\hat{W}_T$。

[0187] 3)、根据上述2)所得模型权重$\hat{W}_D$和$\hat{W}_T$以及S150计算出此时的总误差$\hat{E}$、历史帧到当前帧的位姿变换矩阵$\hat{T}_{t-1}^{t}$、当前帧的场景深度$\hat{D}_t$。

[0188] 4)通过比较总误差E和$\hat{E}$的大小决策最终输出的当前帧的场景深度。

[0189] 本申请实施例,采用全无监督的形式从二维图像中恢复场景深度,通过时间 注意力模块,将记忆单元中的历史帧信息注入到当前输入单元中,并对时空特 征图的空间相关性进行建模来提高相机位姿的精度,降低因姿态不准确而产生 的错误仿射变换的影响;推理期间,利用在线决策推理提高算法对未知场景的 泛化能力。

[0190] 参照图6,其示出了根据本申请一个实施例描述的基于历史信息的场景深度 推理装置的结构示意图。

[0191] 如图6所示,基于历史信息的场景深度推理装置600,可以包括:

[0192] 第一获取模块610,用于获取待测图像的第一图像帧和第二图像帧,第一图 像帧为第二图像帧前一时刻的图像帧;

[0193] 第二获取模块620,用于获取预先构建的深度估计网络的第一深度权重及预 先构建的相机运动网络的第一运动权重;

[0194] 第一处理模块630，用于第一图像帧、第二图像帧、第一深度权重和第一运 动权重，采用误差计算模块，计算第一误差；

[0195] 更新模块640，用于将第一误差作为指导信号联合更新深度估计网络的第一 深度权重和相机运动网络的第一运动权重，得到第二深度权重和第二运动权重；

[0196] 第二处理模块650，用于第一图像帧、第二图像帧、第二深度权重和第二运 动权重，采用误差计算模块，计算第二误差；

[0197] 确定模块660，用于根据第一误差和第二误差，确定第二图像帧的场景深度 及第一图像帧与第二图像帧之间的相对位姿。

[0198] 可选的，相机运动网络包括编码器、时间注意力模块和时空相关性模块；

[0199] 编码器用于提取堆叠图像帧的特征，得到堆叠特征图；堆叠图像帧为第一图 像帧和第二图像帧按通道维度堆叠得到的；

[0200] 时间注意力模块用于将历史记忆单元的信息与当前输入单元的信息建立全 局依赖关系，并通过更新单元，将全局相关的历史记忆单元中的信息注入到当前 输入单元，同时将当前输入单元中的全局相关信息储存到历史记忆单元，作为下 一时刻的历史记忆单元；当前输入单元包括堆叠特征图，堆叠特征图通过更新单 元更新为更新后特征图，历史记忆单元包括第一记忆特征图和第一时间特征图， 下一时刻的历史记忆单元包括第二记忆特征图和第二时间特征图；

[0201] 时空相关性模块用于将更新后特征图/第二记忆特征图分别建模成为具有空 间相关性的第一/二时空特征图。

[0202] 可选的，基于历史信息的场景深度推理装置600，还用于：

[0203] 将堆叠特征图中的特征信息和第一记忆特征图的特征信息注入到第一时间 特征图，得到第三时间特征图；

[0204] 根据第三时间特征图，确定时间注意力特征向量；

[0205] 根据堆叠特征图，确定第一特征向量；

[0206] 根据第一记忆特征图；确定第二特征向量；

[0207] 根据第一特征向量和时间注意力特征向量，确定基于时间注意力的输入特征 向量；

[0208] 根据第二特征向量和时间注意力特征向量，确定基于时间注意力的记忆特征 向量；

[0209] 分别将基于时间注意力的输入特征向量和基于时间注意力的记忆特征向量， 调整成对应的第一特征图和第二特征图；

[0210] 根据第一特征图和第二特征图，确定更新后特征图和第二记忆特征图；

[0211] 根据更新后特征图和第二记忆特征图，将第三时间特征图更新为第二时间特 征图。

[0212] 可选的，基于历史信息的场景深度推理装置600，还用于：

[0213] 将更新后特征图在通道维度进行切片，得到第一子特征图、第二子特征图和 第三子特征图；

[0214] 分别将更新后特征图、第一子特征图、第二子特征图和第三子特征图，对应 调整为第三特征向量、第一子特征向量、第二子特征向量和第三子特征向量；

[0215] 根据第一子特征向量和第二子特征向量，计算第一子特征图和第二子特征图 之间的第一空间相关性矩阵；

[0216] 利用第一空间相关性矩阵对第三子特征向量进行加权处理，得到第一空间相 关特征向量；

[0217] 根据第一空间相关特征向量和第三特征向量，确定第一时空特征向量；

[0218] 将第一时空特征向量调整为具有空间相关性的第一时空特征图；

[0219] 将第二记忆特征图在通道维度进行切片，得到第四子特征图、第五子特征图 和第六子特征图；

[0220] 分别将第二记忆特征图、第四子特征图、第五子特征图和第六子特征图调整 为第四特征向量、第四子特征向量、第五子特征向量和第六子特征向量；

[0221] 根据第四子特征向量和第五子特征向量，计算第四子特征图和第五子特征图 之间的第二空间相关性矩阵；

[0222] 利用第二空间相关性矩阵对第六子特征向量进行加权处理，得到第二空间相 关特征向量；

[0223] 根据第二空间相关特征向量和第四特征向量，确定第二时空特征向量；

[0224] 将第二时空特征向量调整为具有空间相关性的所述第二时空特征图。

[0225] 可选的，第一处理模块630，还用于：

[0226] 将第一图像帧和第二图像帧输入预先构建的深度估计网络，根据第一图像帧 和第一深度权重，得到第一图像帧的第一场景深度和第一图像帧的第一编码器特 征图，根据第二图像帧和第一深度权重，得到第二图像帧的第二场景深度和第二 图像帧的第二编码器特征图；

[0227] 将第一图像帧和第二图像帧输入预先构建的相机运动网络，根据第一图像帧、第二图像帧和第一运动权重，得到第一图像帧和第二图像帧之间的第一相对位姿；

[0228] 第一图像帧、第二图像帧、第一场景深度、第二场景深度、第一编码器特征 图、第二编码器特征图及第一相对位姿，采用误差计算模块，计算第一误差；

[0229] 可选的，第二处理模块650，还用于：

[0230] 将第一图像帧和第二图像帧输入预先构建的深度估计网络，根据第一图像帧 和第二深度权重，得到第一图像帧的第三场景深度和第一图像帧的第三编码器特 征图，根据第二图像帧和第二深度权重，得到第二图像帧的第四场景深度和第二 图像帧的第四编码器特征图；

[0231] 将第一图像帧和第二图像帧输入预先构建的相机运动网络，根据第一图像帧、第二图像帧和第二运动权重，得到第一图像帧和第二图像帧之间的第二相对位姿；

[0232] 第一图像帧、第二图像帧、第三场景深度、第四场景深度、第三编码器特征 图、第四编码器特征图及第二相对位姿，采用误差计算模块，计算第二误差。

[0233] 可选的，确定模块660还用于：

[0234] 若第一误差大于第二误差，将第二场景深度作为第二图像帧的场景深度，将 第一相对位姿作为第一图像帧与第二图像帧之间的相对位姿；

[0235] 若第一误差小于或等于第二误差，将第四场景深度作为第二图像帧的场景深 度，将第二相对位姿作为第一图像帧与第二图像帧之间的相对位姿。

[0236] 可选的，总误差包括第一误差和第二误差；总误差根据图像合成误差、场景 深度结构一致性误差、特征感知损失误差、平滑损失误差确定。

[0237] 可选的，第一处理模块630或第二处理模块650还用于：

[0238] 获取第一图像帧的第一图像坐标、第二图像帧的第二图像坐标；

[0239] 根据第一图像坐标、相机内参、第一场景深度，确定第一图像帧的第一世界 坐标；

[0240] 根据第二图像坐标、相机内参、第二场景深度，确定第二图像帧的第二世界 坐标；

[0241] 将第一图像帧的第一世界坐标仿射变换到第二图像帧面板，确定仿射变换后 的第三世界坐标；

[0242] 将第二图像帧的第二世界坐标仿射变换到第一图像帧面板，确定仿射变换后 的第四世界坐标；

[0243] 将第三世界坐标和第四世界坐标分别投影到二维平面，得到第一仿射变换后 场景深度和第二仿射变换后场景深度及对应的第一仿射变换后图像坐标和第二 仿射变换后图像坐标；

[0244] 根据第一场景深度、第二场景深度、第一仿射变换后图像坐标、第二仿射变 换后图像坐标，确定场景深度结构一致性误差、第一深度结构不一致性权重和第 二深度结构不一致性权重；

[0245] 根据第一图像帧的第一图像坐标、第二仿射变换后图像坐标、第二图像帧的 第二图像坐标、第一仿射变换后图像坐标，确定第一相机流一致性遮挡掩码和第 二相机流一致性遮挡掩码；

[0246] 根据第一深度结构不一致性权重、第二深度结构不一致性权重、第一相机流 一致性遮挡掩码和第二相机流一致性遮挡掩码，确定图像合成误差；

[0247] 根据第一图像帧、第二图像帧、第一仿射变换后图像坐标和第二仿射变换后 图像坐标，确定特征感知损失误差；

[0248] 根据第一场景深度、第二场景深度、第一图像帧和第二图像帧，确定平滑损 失误差；

[0249] 根据图像合成误差、场景深度结构一致性误差、特征感知损失误差、平滑损 失误差，确定总误差。

[0250] 本实施例提供的一种基于历史信息的场景深度推理装置，可以执行上述方 法的实施例，其实现原理和技术效果类似，在此不再赘述。

[0251] 图7为本发明实施例提供的一种电子设备的结构示意图。如图7所示，示出 了适于用来实现本申请实施例的电子设备300的结构示意图。

[0252] 如图7所示，电子设备300包括中央处理单元(CPU)301，其可以根据存 储在只读存储器(ROM)302中的程序或者从存储部分308加载到随机访问存储 器(RAM)303中的程序而执行各种适当的动作和处理。在RAM 303中，还存 储有设备300操作所需的各种程序和数据。CPU 301、ROM 302以及RAM 303 通过总线304彼此相连。输入/输出(I/O)接口305也连接至总线304。

[0253] 以下部件连接至I/O接口305：包括键盘、鼠标等的输入部分306；包括诸 如阴极射线管(CRT)、液晶显示器(LCD)等以及扬声器等的输出部分307； 包括硬盘等的存储部分308；以及包括诸如LAN卡、调制解调器等的网络接口 卡的通信部分309。通信部分309经由

诸如因特网的网络执行通信处理。驱动器 310也根据需要连接至I/O接口305。可拆卸介质311，诸如磁盘、光盘、磁光盘、 半导体存储器等等，根据需要安装在驱动器310上，以便于从其上读出的计算机 程序根据需要被安装入存储部分308。

[0254] 特别地，根据本公开的实施例，上文参考图1描述的过程可以被实现为计算 机软件程序。例如，本公开的实施例包括一种计算机程序产品，其包括有形地包 含在机器可读介质上的计算机程序，计算机程序包含用于执行上述基于历史信息 的场景深度推理方法的程序代码。在这样的实施例中，该计算机程序可以通过通 信部分309从网络上被下载和安装，和/或从可拆卸介质311被安装。

[0255] 附图中的流程图和框图，图示了按照本发明各种实施例的系统、方法和计算 机程序产品的可能实现的体系架构、功能和操作。在这点上，流程图或框图中的 每个方框可以代表一个模块、程序段、或代码的一部分，前述模块、程序段、或 代码的一部分包含一个或多个用于实现规定的逻辑功能的可执行指令。也应当注 意，在有些作为替换的实现中，方框中所标注的功能也可以以不同于附图中所标 注的顺序发生。例如，两个接连地表示的方框实际上可以基本并行地执行，它们 有时也可以按相反的顺序执行，这依所涉及的功能而定。也要注意的是，框图和 /或流程图中的每个方框、以及框图和/或流程图中的方框的组合，可以用执行规 定的功能或操作的专用的基于硬件的系统来实现，或者可以用专用硬件与计算机 指令的组合来实现。

[0256] 描述于本申请实施例中所涉及到的单元或模块可以通过软件的方式实现，也 可以通过硬件的方式来实现。所描述的单元或模块也可以设置在处理器中。这些 单元或模块的名称在某种情况下并不构成对该单元或模块本身的限定。

[0257] 上述实施例阐明的系统、装置、模块或单元，具体可以由计算机芯片或实体 实现，或者由具有某种功能的产品来实现。一种典型的实现设备为计算机。具体 的，计算机例如可以为个人计算机、笔记本电脑、行动电话、智能电话、个人数 字助理、媒体播放器、导航设备、电子邮件设备、游戏控制台、平板计算机、可 穿戴设备或者这些设备中的任何设备的组合。

[0258] 作为另一方面，本申请还提供了一种存储介质，该存储介质可以是上述实施 例中前述装置中所包含的存储介质；也可以是单独存在，未装配入设备中的存储 介质。存储介质存储有一个或者一个以上程序，前述程序被一个或者一个以上的 处理器用来执行描述于本申请的基于历史信息的场景深度推理方法。

[0259] 存储介质包括永久性和非永久性、可移动和非可移动媒体可以由任何方法 或技术来实现信息存储。信息可以是计算机可读指令、数据结构、程序的模块 或其他数据。计算机的存储介质的例子包括，但不限于相变内存(PRAM)、静态 随机存取存储器(SRAM)、动态随机存取存储器(DRAM)、其他类型的随机存取 存储器(RAM)、只读存储器(ROM)、电可擦除可编程只读存储器(EEPROM)、 快闪记忆体或其他内存技术、只读光盘只读存储器(CD-ROM)、数字多功能光 盘(DVD)或其他光学存储、磁盒式磁带，磁带磁磁盘存储或其他磁性存储设备 或任何其他非传输介质，可用于存储可以被计算设备访问的信息。按照本文中 的界定，计算机可读介质不包括暂存电脑可读媒体(transitory media)，如调制的 数据信号和载波。

[0260] 需要说明的是，术语“包括”、“包含”或者其任何其他变体意在涵盖非排他性 的包

含,从而使得包括一系列要素的过程、方法、商品或者设备不仅包括那些要 素,而且还包括没有明确列出的其他要素,或者是还包括为这种过程、方法、商 品或者设备所固有的要素。在没有更多限制的情况下,由语句“包括一个……” 限定的要素,并不排除在包括要素的过程、方法、商品或者设备中还存在另外的 相同要素。

[0261] 本说明书中的各个实施例均采用递进的方式描述,各个实施例之间相同相 似的部分互相参见即可,每个实施例重点说明的都是与其他实施例的不同之处。尤其,对于系统实施例而言,由于其基本相似于方法实施例,所以描述的比较简 单,相关之处参见方法实施例的部分说明即可。

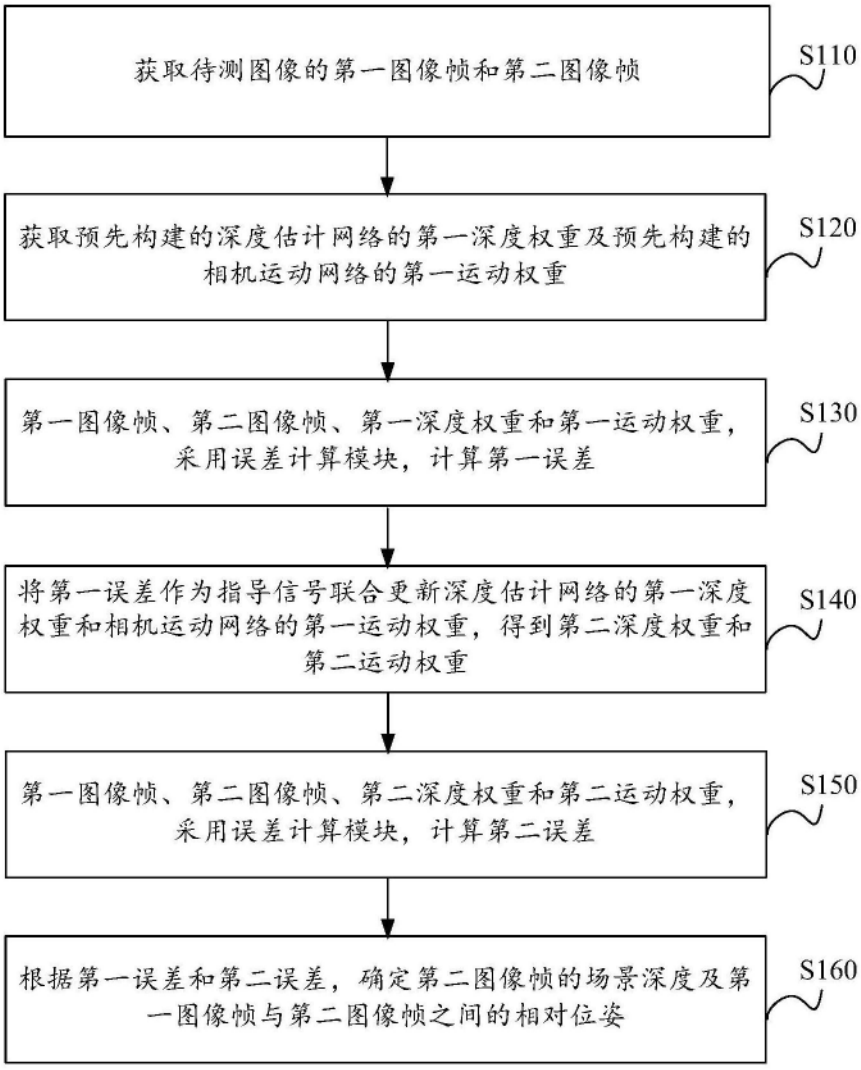

获取待测图像的第一图像帧和第二图像帧
S110
获取预先构建的深度估计网络的第一深度权重及预先构建的相机运动网络的第一运动权重
S120
第一图像帧、第二图像帧、第一深度权重和第一运动权重，采用误差计算模块，计算第一误差
S130
将第一误差作为指导信号联合更新深度估计网络的第一深度权重和相机运动网络的第一运动权重，得到第二深度权重和第二运动权重
S140
第一图像帧、第二图像帧、第二深度权重和第二运动权重，采用误差计算模块，计算第二误差
S150
根据第一误差和第二误差，确定第二图像帧的场景深度及第一图像帧与第二图像帧之间的相对位姿
S160

图1

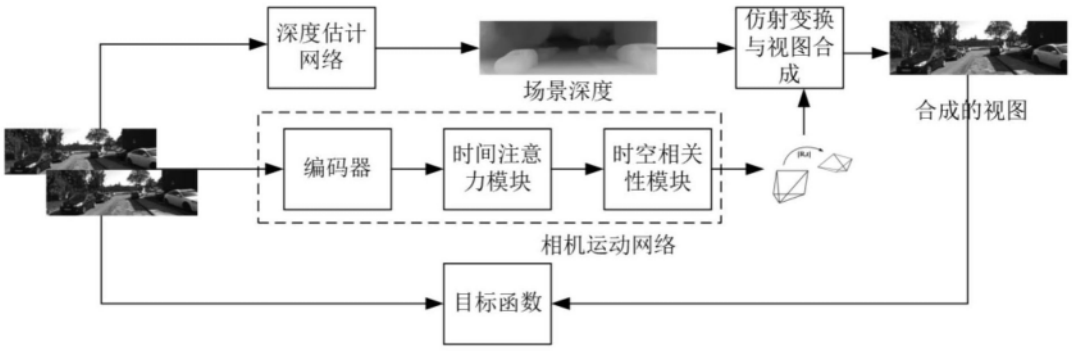

深度估计网络
场景深度
仿射变换与视图合成
合成的视图
编码器
时间注意力模块
时空相关性模块
相机运动网络
目标函数

图2

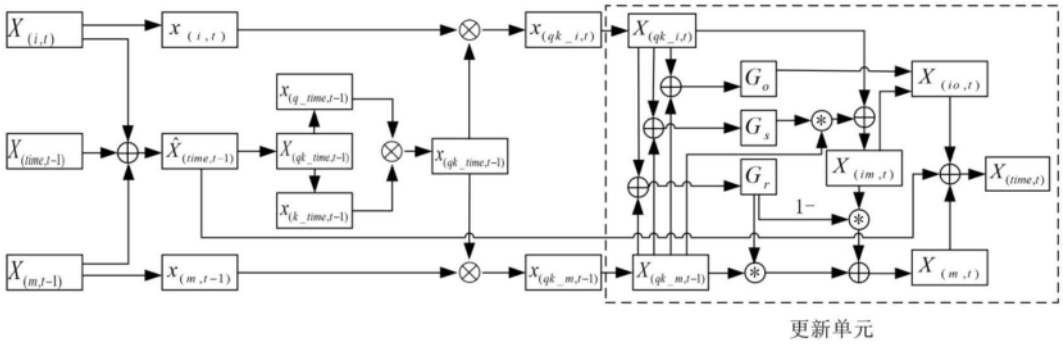

更新单元

图3

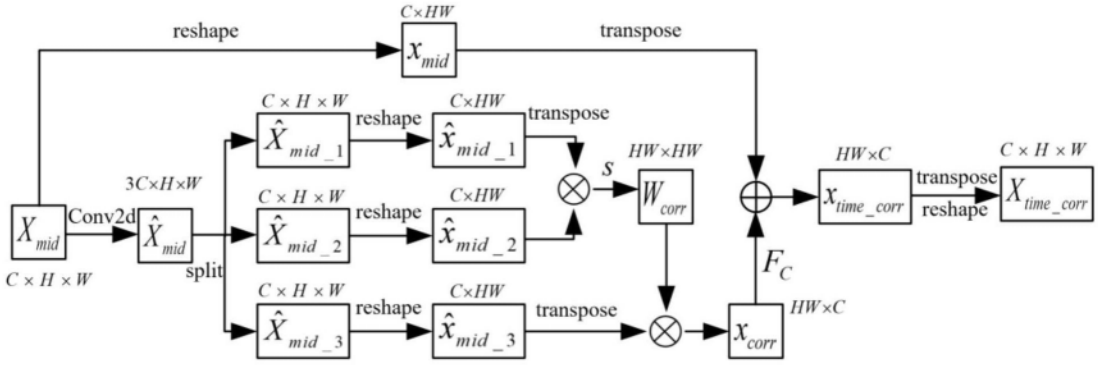

reshape
transpose
Conv2d
split
reshape
transpose
reshape
transpose
reshape
transpose
C × H × W
3C×H×W
C×HW
HW×HW
HW×C
s

图4

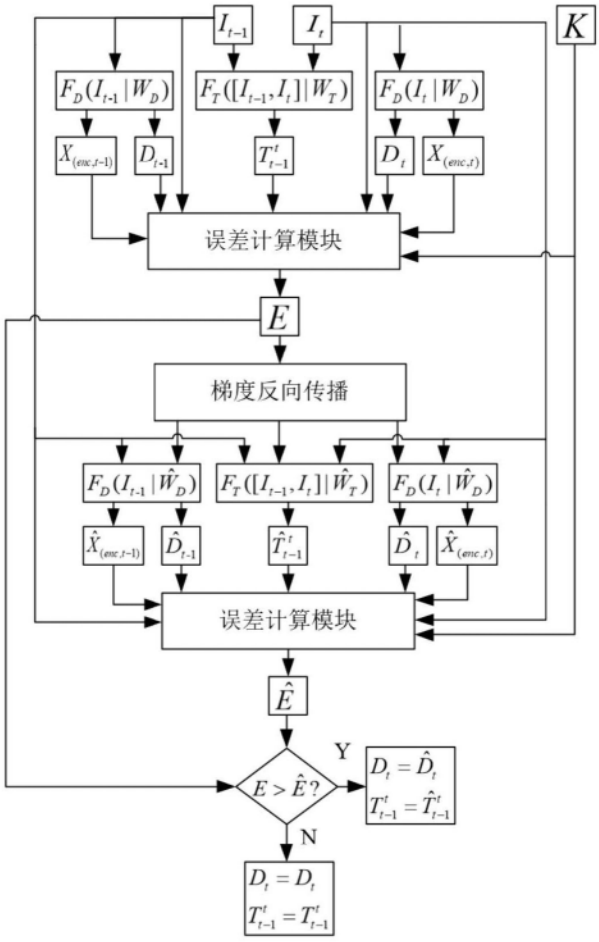

I_{t-1}
I_t
K
$F_D(I_{t-1}|W_D)$
$F_T([I_{t-1},I_t]|W_T)$
$F_D(I_t|W_D)$
$X_{(enc,t-1)}$
D_{t-1}
T^t_{t-1}
D_t
$X_{(enc,t)}$
误差计算模块
E
梯度反向传播
$F_D(I_{t-1}|\hat{W}_D)$
$F_T([I_{t-1},I_t]|\hat{W}_T)$
$F_D(I_t|\hat{W}_D)$
$\hat{X}_{(enc,t-1)}$
$\hat{D}_{t-1}$
$\hat{T}^t_{t-1}$
$\hat{D}_t$
$\hat{X}_{(enc,t)}$
误差计算模块
$\hat{E}$
$E > \hat{E}$?
Y
$D_t = \hat{D}_t$
$T^t_{t-1} = \hat{T}^t_{t-1}$
N
$D_t = D_t$
$T^t_{t-1} = T^t_{t-1}$

图5

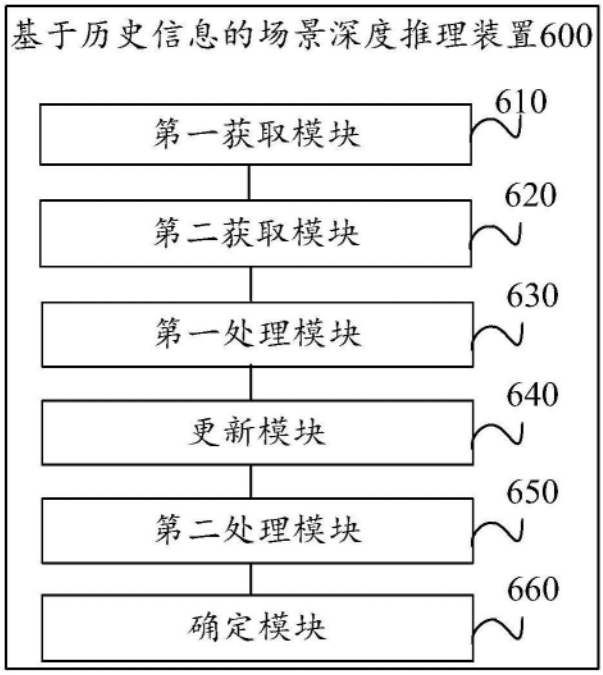

基于历史信息的场景深度推理装置600
第一获取模块
610
第二获取模块
620
第一处理模块
630
更新模块
640
第二处理模块
650
确定模块
660

图6

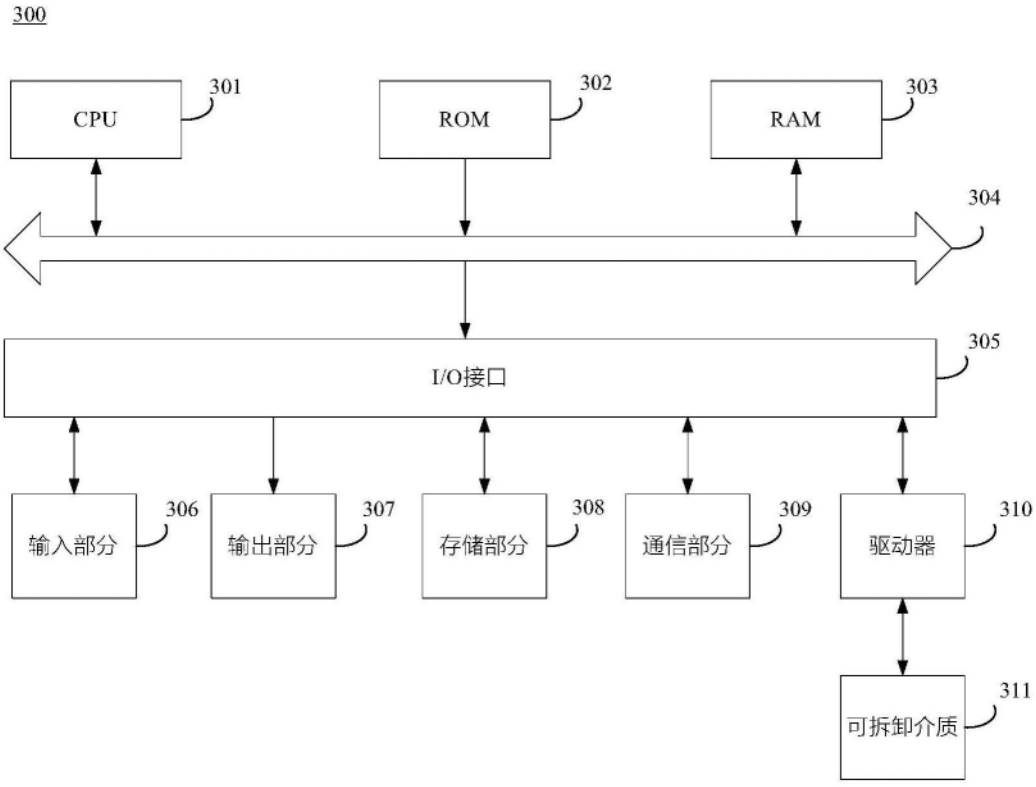

300
CPU
301
ROM
302
RAM
303
304
I/O接口
305
输入部分
306
输出部分
307
存储部分
308
通信部分
309
驱动器
310
可拆卸介质
311

图7