



(12) 发明专利申请

(10) 申请公布号 CN 117333627 A

(43) 申请公布日 2024. 01. 02

(21) 申请号 202311631542.4

G06N 3/0475 (2023.01)

(22) 申请日 2023.12.01

G06N 3/094 (2023.01)

(71) 申请人 南方科技大学

地址 518000 广东省深圳市南山区桃源街
道学苑大道1088号

(72) 发明人 张美莹 彭维源 郝祁

(74) 专利代理机构 深圳鼎合诚知识产权代理有
限公司 44281

专利代理师 郭燕 彭家恩

(51) Int. Cl.

G06T 17/00 (2006.01)

G06T 19/20 (2011.01)

G06T 15/00 (2011.01)

G06T 5/50 (2006.01)

G06N 3/045 (2023.01)

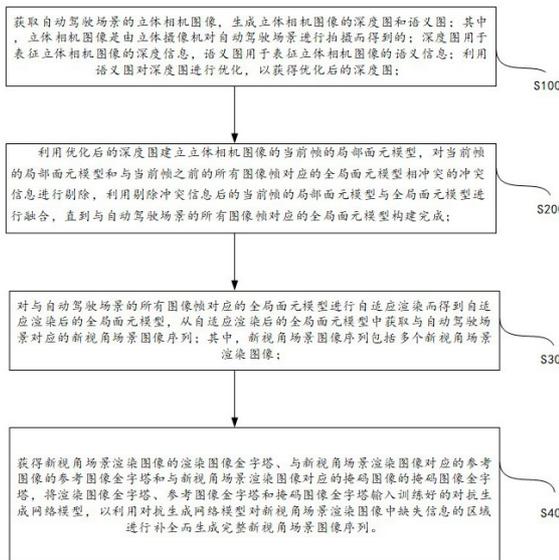
权利要求书4页 说明书20页 附图5页

(54) 发明名称

一种自动驾驶场景的重建与补全方法、系统
及存储介质

(57) 摘要

本发明提供一种自动驾驶场景的重建与补全方法、系统及存储介质。该方法：获取自动驾驶场景的立体相机图像，生成深度图和语义图；利用语义图优化深度图；利用优化后的深度图建立局部面元模型，对局部面元模型中的冲突信息进行剔除，将局部面元模型与全局面元模型进行融合而获得融合后的全局面元模型；待整个场景的全局面元模型构建完成后从中采集新视角场景图像；由于面元模型构建过程中的估计误差与动态目标物的去除，造成新视角场景渲染图像会出现各种空洞/掩码，因此对参考图像进行上下文信息匹配以得到新视角场景渲染图像的上下文关联信息，上下文关联信息包括相似度图和关联图像；利用对抗生成网络模型对新视角场景渲染图像进行补全。



1. 一种自动驾驶场景的重建与补全方法,其特征在于,包括:

获取所述自动驾驶场景的立体相机图像,生成所述立体相机图像的深度图和语义图;其中,所述立体相机图像是由立体摄像机对所述自动驾驶场景进行拍摄而得到的;所述深度图用于表征所述立体相机图像的深度信息,所述语义图用于表征所述立体相机图像的语义信息;利用所述语义图对所述深度图进行优化,以获得优化后的深度图;

利用所述优化后的深度图建立所述立体相机图像的当前帧的局部面元模型,对所述当前帧的局部面元模型和与所述当前帧之前的所有图像帧对应的全局面元模型相冲突的冲突信息进行剔除,利用剔除所述冲突信息后的所述当前帧的局部面元模型与所述全局面元模型进行融合,直到与所述自动驾驶场景的所有图像帧对应的全局面元模型构建完成;

对所述与所述自动驾驶场景的所有图像帧对应的全局面元模型进行自适应渲染而得到自适应渲染后的所述全局面元模型,从所述自适应渲染后的所述全局面元模型中获取与所述自动驾驶场景对应的新视角场景图像序列;其中,所述新视角场景图像序列包括多个新视角场景渲染图像;

获得所述新视角场景渲染图像的渲染图像金字塔、与所述新视角场景渲染图像对应的参考图像的参考图像金字塔和与所述新视角场景渲染图像对应的掩码图像的掩码图像金字塔,将所述渲染图像金字塔、所述参考图像金字塔和所述掩码图像金字塔输入训练好的对抗生成网络模型,以利用所述对抗生成网络模型对所述新视角场景渲染图像中缺失信息的区域进行补全而生成完整新视角场景图像序列;

其中,所述完整新视角场景图像序列包括多个完整新视角场景图像;所述参考图像是来自于所述新视角场景渲染图像的邻近视角的所述立体相机图像,所述掩码图像是基于所述新视角场景渲染图像中缺失信息的区域而得到的。

2. 如权利要求1所述的重建与补全方法,其特征在于,所述对抗生成网络模型包括一个生成器,所述生成器包括多级彼此串联连接的CSD残差模块;

其中,所述将所述渲染图像金字塔、所述参考图像金字塔和所述掩码图像金字塔输入训练好的对抗生成网络模型,以利用所述对抗生成网络模型对所述新视角场景渲染图像中缺失信息的区域进行补全而生成完整新视角场景图像序列,包括:

所述渲染图像金字塔、所述参考图像金字塔和所述掩码图像金字塔中各层的图像结果按照由顶层至底层的顺序分别输入所述生成器内第一级的所述CSD残差模块至最后一级的所述CSD残差模块;

将前一级的所述CSD残差模块的输出作为相邻的后一级的所述CSD残差模块的一种输入,将所述最后一级的CSD残差模块的输出作为所述完整新视角场景图像;

其中,所述生成器内第一级的所述CSD残差模块至最后一级的所述CSD残差模块的网络结构均相同;

其中,所述渲染图像金字塔、所述参考图像金字塔和所述掩码图像金字塔的层数均与所述CSD残差模块的数量相同。

3. 如权利要求2所述的重建与补全方法,其特征在于,所述渲染图像金字塔、所述参考图像金字塔和所述掩码图像金字塔中各层的图像结果按照由顶层至底层的顺序分别输入所述生成器内第一级的所述CSD残差模块至最后一级的所述CSD残差模块;将前一级的所述CSD残差模块的输出作为相邻的后一级的所述CSD残差模块的一种输入,将所述最后一级的

CSD残差模块的输出作为所述完整新视角场景图像,包括:

将所述渲染图像金字塔、所述参考图像金字塔和所述掩码图像金字塔中顶层的图像结果,以及与所述顶层的图像结果的尺度对应的所述新视角场景渲染图像输入第一级的所述CSD残差模块;

将所述渲染图像金字塔、所述参考图像金字塔和所述掩码图像金字塔中除所述顶层外的其余各层的图像结果按照由所述顶层的下一层至所述底层的顺序分别输入所述生成器内第二级的所述CSD残差模块至最后一级的所述CSD残差模块;将所述最后一级的CSD残差模块的输出作为所述完整新视角场景图像。

4. 如权利要求3所述的重建与补全方法,其特征在于,所述CSD残差模块包括第一级的CSD子模块、第二级的CSD子模块和第三级的CSD子模块;

所述CSD残差模块内部的数据处理流程为:

将所述渲染图像金字塔、所述参考图像金字塔和所述掩码图像金字塔中对应层的图像结果,以及输入所述CSD残差模块的特征图分别输入至所述第一级的CSD子模块和所述第三级的CSD子模块,将所述第一级的CSD子模块的输出作为第一ReLU激活函数的输入,对所述第一ReLU激活函数的输出进行卷积操作而得到第一卷积结果,将所述第一卷积结果输入所述第二级的CSD子模块,将所述第二级的CSD子模块的输出作为第二ReLU激活函数的输入,对所述第二ReLU激活函数的输出进行卷积操作而得到第二卷积结果,将所述第三级的CSD子模块的输出作为第三ReLU激活函数的输入,对所述第三ReLU激活函数的输出进行卷积操作而得到第三卷积结果,对所述第二卷积结果和所述第三卷积结果执行单位加操作而得到所述CSD残差模块的输出;

其中,输入第一级的所述CSD残差模块的特征图为与所述顶层的图像结果的尺度对应的所述新视角场景渲染图像,输入相邻的后一级的所述CSD残差模块的特征图为前一级的所述CSD残差模块的输出。

5. 如权利要求4所述的重建与补全方法,其特征在于,所述第一级的CSD子模块、所述第二级的CSD子模块和所述第三级的CSD子模块的数据处理流程均相同;

其中,所述CSD子模块的数据处理流程为:

对输入的所述渲染图像金字塔中对应层的图像结果、所述参考图像金字塔中对应层的图像结果和所述掩码图像金字塔中对应层的图像结果进行相似性匹配,以得到与所述渲染图像金字塔中对应层的图像结果的上下文关联信息;

其中,所述上下文关联信息包括相似度图和所述渲染图像金字塔中对应层的图像结果的关联图像;所述关联图像是基于所述参考图像金字塔中对应层的图像结果与所述渲染图像金字塔中对应层的图像结果最相似的多个区域而生成的,所述相似度图用于表征所述渲染图像金字塔中对应层的图像结果与所述参考图像金字塔中对应层的图像结果之间的相关性;

其中,所述上下文关联信息用于为所述缺失信息的区域提供上下文匹配信息;

对所述关联图像和所述相似度图分别执行第二卷积操作分别得到所述关联图像的空间特征和所述相似度图的空间特征;

对所述渲染图像金字塔中对应层的图像结果执行第二卷积操作而得到所述对应层的所述新视角场景渲染图像的空间特征;

对所述相似度图的空间特征和所述新视角场景渲染图像的空间特征执行逐元素乘法操作而得到逐元素乘法结果；

将所述逐元素乘法结果与所述关联图像的空间特征相加而得到第二仿射参数；其中，所述第二仿射参数的表达式为：

$\beta_{c,h,w} = \alpha_{c,h,w} \odot f_{c,h,w}(I_r) + f_{c,h,w}(I_c)$ ；所述 $\alpha_{c,h,w} \odot f_{c,h,w}(I_r)$ 为所述逐元素乘法结果，所述 $\alpha_{c,h,w}$ 为所述相似度图的空间特征，所述 $f_{c,h,w}(I_c)$ 为所述关联图像的空间特征；其中，所述 c 、 h 和 w 分别为所述对应层的图像结果的特征通道数、长度和宽度，所述 I_r 表示所述新视角场景渲染图像，所述 I_c 为所述对应层的图像结果的关联图像；

对输入所述CSD残差模块的多个特征图进行归一化而得到归一化参数；

将所述归一化参数与所述第二仿射参数相加而得到的归一化结果作为所述CSD子模块的输出。

6. 如权利要求5所述的重建与补全方法，其特征在于，所述对输入的所述渲染图像金字塔中对应层的图像结果、所述参考图像金字塔中对应层的图像结果和所述掩码图像金字塔中对应层的图像结果进行相似性匹配，以得到与所述渲染图像金字塔中对应层的图像结果的上下文关联信息，包括：

利用一个划分框以第一步幅在所述渲染图像金字塔中对应层的图像结果中移动而将所述渲染图像金字塔中对应层的图像结果划分为多个尺寸为 $k \times k$ 的渲染子块，

利用所述划分框以第一步幅在所述掩码图像金字塔中对应层的图像结果中移动而将所述掩码图像金字塔中对应层的图像结果划分为多个尺寸为 $k \times k$ 的掩码子块，

在所述参考图像金字塔中对应层的图像结果中确定与所述渲染子块对应的搜索域，所述搜索域的尺寸为 $k' \times k'$ ，利用所述划分框以第二步幅在所述搜索域中移动而将所述搜索域划分为 $s \times s$ 个尺寸为 $k \times k$ 的参考子块；其中，所述 k 、 k' 和 s 均为预设的常数；

将与所述渲染子块对应的所述掩码子块应用到与所述渲染子块对应的所述参考子块而得到与所述渲染子块对应的参考掩码子块；

计算所述渲染子块和所述与所述渲染子块对应的所述搜索域内各所述参考掩码子块之间的各相似度，将所述各相似度中的最大值作为所述相似度图中与所述渲染子块对应的元素；其中，所述与所述渲染子块对应的所述搜索域的尺寸大于所述渲染子块的尺寸；

将所述相似度图中各所述元素所分别对应的所述参考子块作为最佳匹配子块，利用与所述相似度图对应的多个所述最佳匹配子块生成所述关联图像。

7. 如权利要求6所述的重建与补全方法，其特征在于，所述相似度图的表达式为：

$simi(i) = \max_{j \in S_{k \times k}} \langle (p_{k \times k}^r)_i \odot (p_{k \times k}^c)_{i,j}, (p_{k \times k}^l)_i \rangle$ ；其中， $simi(i)$ 表示所述渲染图像金字塔中对应层的图像结果中第 i 个所述渲染子块 $(p_{k \times k}^r)_i$ 和所述参考图像金字塔中对应层的图像结果中与所述第 i 个渲染子块对应的所述搜索域内 $s \times s$ 个所述参考掩码子块 $(p_{k \times k}^l)_i \odot (p_{k \times k}^c)_{i,j}$ 之间的各相似度的最大值， $\langle \cdot, \cdot \rangle$ 为求取余弦相似度的操作， \odot 表示逐元素相乘，此处的所述 $i = 1, 2, \dots, (\frac{H}{n} \times \frac{W}{n})$ ，其中，所述 H 和 W 分别为所述新视角场景渲染图像的高度和宽度，所述 n 为所述第一步幅；所述 I_m 表示所述掩码图像，所述 I_c 表示所述参考图像，所述 j 表示所述搜索域

内所述参考子块 $p_{k \times k}^{I_c}$ 的数量,所述 $p_{k \times k}^{I_m}$ 为所述掩码子块。

8. 如权利要求5所述的重建与补全方法,其特征在于,所述归一化参数的表达式为:

$$\gamma_c \left(\frac{x^i - \mu_c(x^i)}{\sigma_c(x^i)} \right),$$

其中,所述 x^i 为输入第i级的所述CSD残差模块中所述CSD子模块的所述特征图,所述 $x^i \in \mathbb{R}^{N \times C \times H \times W}$,所述C为相应的特征通道的数量;所述 $\mu_c(x^i)$ 和 $\sigma_c(x^i)$ 分别是输入所述第i级的所述CSD残差模块中所述CSD子模块的N个所述特征图的均值和标准差,所述 γ_c 为第一仿射参数,所述 $\gamma_c \in \mathbb{R}^C$;调制仿射参数包括所述第一仿射参数 γ_c 和所述第二仿射参数。

9. 一种自动驾驶场景的重建与补全系统,其特征在于,包括:

立体摄像机,被配置为获取所述自动驾驶场景的立体相机图像;

语义深度生成模块,被配置为生成所述立体相机图像的深度图和语义图;

语义深度增强模块,被配置为利用所述语义图对所述深度图进行优化,以获得优化后的深度图;

面元模型构建模块,被配置为利用所述优化后的深度图建立所述立体相机图像的当前帧的局部面元模型,对所述当前帧的局部面元模型和与所述当前帧之前的所有图像帧对应的全局面元模型相冲突的冲突信息进行剔除,利用剔除所述冲突信息后的所述当前帧的局部面元模型与所述全局面元模型进行融合,直到与所述自动驾驶场景的所有图像帧对应的全局面元模型构建完成;对所述与所述自动驾驶场景的所有图像帧对应的全局面元模型进行自适应渲染而得到自适应渲染后的所述全局面元模型,从所述自适应渲染后的所述全局面元模型中获取与所述自动驾驶场景对应的新视角场景图像序列;其中,所述新视角场景图像序列包括多个新视角场景渲染图像;

对抗网络补全模块,被配置为获得所述新视角场景渲染图像的渲染图像金字塔、与所述新视角场景渲染图像对应的参考图像的参考图像金字塔和与所述新视角场景渲染图像对应的掩码图像的掩码图像金字塔,将所述渲染图像金字塔、所述参考图像金字塔和所述掩码图像金字塔输入训练好的对抗生成网络模型,以利用所述对抗生成网络模型对所述新视角场景渲染图像中缺失信息的区域进行补全而生成完整新视角场景图像序列;其中,所述完整新视角场景图像序列包括多个完整新视角场景图像;所述参考图像是来自于所述新视角场景渲染图像的邻近视角的所述立体相机图像,所述掩码图像是基于所述新视角场景渲染图像中缺失信息的区域而得到的。

10. 一种计算机可读存储介质,其特征在于,包括程序,所述程序能够被处理器执行以实现如权利要求1至8中任一项所述的方法。

一种自动驾驶场景的重建与补全方法、系统及存储介质

技术领域

[0001] 本发明涉及图像场景生成技术领域,具体涉及一种自动驾驶场景的重建与补全方法、系统及存储介质。

背景技术

[0002] 目前,自动驾驶相关应用的开发需要对自动驾驶汽车进行大量的验证和测试,以保证其实施的安全性。其解决方案之一是从不同的轨迹和角度真实地重现大量复杂多样的交通场景,因此所构建的场景流的真实性和一致性对自动驾驶系统测试至关重要。其中,可以通过游戏引擎或高保真计算机图形(CG)模型创建前述场景,如英特尔的CARLA(一个用于自动驾驶研究的开源模拟器)、微软的AirSim(一个跨平台的建立在虚幻引擎上的无人机以及其它自主移动设备的模拟器)、谷歌的CarCraft(一款虚拟重建城市中测试无人驾驶汽车软件)等均可以获得出色的渲染效果,但其合成图像仍然缺乏现实世界图像的丰富性,导致性能下降。另一种选择是数据驱动的方法,依靠感官数据(如摄像头和激光雷达等所获取的数据)重建前述交通场景,所重建的交通场景能够保留关于语义、场景照明和外观的丰富信息。通常,数据驱动的场景生成流程由三个部分组成:首先对前述感官数据进行去噪和细化;然后使用适当的几何代理重构三维模型以生成有序静态场景;最后利用图像增强技术来提高图像质量和一致性。然而,为了实现有效的静态场景,新视角图像合成需要克服以下挑战:

(1) 构建场景几何模型。相机可以捕获比激光雷达更高的测量分辨率和更多的语义信息,但需要更精确的深度估计以用于模型重建。同时,图像的语义信息可以帮助改进估计深度。如何有效地利用立体相机实现高精度、高分辨率的冲浪模型仍然是一个难题;

(2) 生成图像中缺失部分的补全。在构建三维背景模型后,由于估计误差和动态目标的去除,所生成的新视图图像中会出现各种不规则的孔洞。一旦建立了三维背景模型,这些被遮挡的区域从任何角度都无法看到。如何用一定的内容填补这些不可忽视的漏洞,生成合理逼真的图像是一个主要问题;

(3) 所生成的新视图图像之间的时空一致性。对于具有小孔或细孔的高度结构化的新视图图像,可以利用其相邻像素来推断缺失部分,但对于复杂多样的交通场景,这种方法效果不佳。此外,大多数图像补全算法不可避免地会导致时间伪影和抖动。如何从新的视角生成时空一致的新视图图像序列,对于自动驾驶系统中功能模块的验证和测试具有重要意义。

[0003] 移动相机可以通过多视图立体(MVS)技术生成密集的3D点云,但这需要从不同的角度获取同一区域周围的多幅图像,同时移动物体严重降低了它们的性能。与MVS相比,立体匹配方法可用于高分辨率图像的室外环境三维重建,但一个主要问题是缺乏高性能模型的真值数据。为了提高三维重建的质量,研究者们还开发了一些使用语义信息的方法,这些方法通常涉及高计算复杂度和联合学习架构,并且仍然包含许多异常值。

[0004] 在稠密的深度条件下,面元模型(即上述三维背景模型)的构造误差较小,参数可

灵活调整。然而,新视图图像的渲染性能依赖于三维点云的质量,往往不能覆盖整个场景,因此需要先进的图像补全技术。最近,基于对抗生成网络(Generative Adversarial Network, GAN)的方法被开发出来,使用自学习编码器-解码器模型,组成多尺度上下文信息的扩展卷积,以及全局和局部判别器来完成图像。然而,这些方法通常不能处理复杂场景和大孔洞的图像。结合基于patch(补丁)的学习方法,它们可以通过基于patch的注意力机制从周围可见区域借鉴相似特征。然而,这些技术试图从图像内或图像间找到相似的补丁,不适合生成不在可见区域的内容。此外,它们单独完成每个图像,而不考虑上述新视图图像序列之间的空间一致性。然而,条件归一化层(Conditional Normalization Layer, CNL)可以利用附加图像实现全局空间编码和空间变化调整。

[0005] 因此,有必要对上述现有技术进行改进。

发明内容

[0006] 本发明主要解决的技术问题是提供一种自动驾驶场景的重建与补全方法、系统及存储介质,以利用所述优化后的深度图建立初始的局部面元模型,对所述初始的局部面元模型中与初始的全局面元模型相冲突的冲突信息进行剔除,利用剔除所述冲突信息后的所述局部面元模型对所述初始的全局面元模型进行融合而获得融合后的全局面元模型;在获得全局面元模型后,通过自适应方法对新视角场景渲染图像进行渲染,最后利用对抗生成网络模型生成空间结构一致的高质量的完整新视角场景图像序列。

[0007] 根据第一方面,一种实施例中提供一种自动驾驶场景的重建与补全方法。该方法包括:

获取所述自动驾驶场景的立体相机图像,生成所述立体相机图像的深度图和语义图;其中,所述立体相机图像是由立体摄像机对所述自动驾驶场景进行拍摄而得到的;所述深度图用于表征所述立体相机图像的深度信息,所述语义图用于表征所述立体相机图像的语义信息;利用所述语义图对所述深度图进行优化,以获得优化后的深度图;

利用所述优化后的深度图建立所述立体相机图像的当前帧的局部面元模型,对所述当前帧的局部面元模型和与所述当前帧之前的所有图像帧对应的全局面元模型相冲突的冲突信息进行剔除,利用剔除所述冲突信息后的所述当前帧的局部面元模型与所述全局面元模型进行融合,直到与所述自动驾驶场景的所有图像帧对应的全局面元模型构建完成;

对所述与所述自动驾驶场景的所有图像帧对应的全局面元模型进行自适应渲染而得到自适应渲染后的所述全局面元模型,从所述自适应渲染后的所述全局面元模型中获取与所述自动驾驶场景对应的新视角场景图像序列;其中,所述新视角场景图像序列包括多个新视角场景渲染图像;

获得所述新视角场景渲染图像的渲染图像金字塔、与所述新视角场景渲染图像对应的参考图像的参考图像金字塔和与所述新视角场景渲染图像对应的掩码图像的掩码图像金字塔,将所述渲染图像金字塔、所述参考图像金字塔和所述掩码图像金字塔输入训练好的对抗生成网络模型,以利用所述对抗生成网络模型对所述新视角场景渲染图像中缺失信息的区域进行补全而生成完整新视角场景图像序列;

其中,所述完整新视角场景图像序列包括多个完整新视角场景图像;所述参考图

像是来自于所述新视角场景渲染图像的邻近视角的所述立体相机图像,所述掩码图像是基于所述新视角场景渲染图像中缺失信息的区域而得到的。

[0008] 根据第二方面,一种实施例中提供一种自动驾驶场景的重建与补全系统,该重建与补全系统包括:

立体摄像机,被配置为获取所述自动驾驶场景的立体相机图像;

语义深度生成模块,被配置为生成所述立体相机图像的深度图和语义图;

语义深度增强模块,被配置为利用所述语义图对所述深度图进行优化,以获得优化后的深度图;

面元模型构建模块,被配置为利用所述优化后的深度图建立所述立体相机图像的当前帧的局部面元模型,对所述当前帧的局部面元模型和与所述当前帧之前的所有图像帧对应的全局面元模型相冲突的冲突信息进行剔除,利用剔除所述冲突信息后的所述当前帧的局部面元模型与所述全局面元模型进行融合,直到与所述自动驾驶场景的所有图像帧对应的全局面元模型构建完成;对所述与所述自动驾驶场景的所有图像帧对应的全局面元模型进行自适应渲染而得到自适应渲染后的所述全局面元模型,从所述自适应渲染后的所述全局面元模型中获取与所述自动驾驶场景对应的新视角场景图像序列;其中,所述新视角场景图像序列包括多个新视角场景渲染图像;

对抗网络补全模块,被配置为获得所述新视角场景渲染图像的渲染图像金字塔、与所述新视角场景渲染图像对应的参考图像的参考图像金字塔和与所述新视角场景渲染图像对应的掩码图像的掩码图像金字塔,将所述渲染图像金字塔、所述参考图像金字塔和所述掩码图像金字塔输入训练好的对抗生成网络模型,以利用所述对抗生成网络模型对所述新视角场景渲染图像中缺失信息的区域进行补全而生成完整新视角场景图像序列;

其中,所述完整新视角场景图像序列包括多个完整新视角场景图像;所述参考图像是来自于所述新视角场景渲染图像的邻近视角的所述立体相机图像,所述掩码图像是基于所述新视角场景渲染图像中缺失信息的区域而得到的。

[0009] 根据第三方面,一种实施例中提供一种计算机可读存储介质。该计算机可读存储介质包括程序。所述程序能够被处理器执行以实现如本文中任一实施例所述的方法。

[0010] 本申请的有益效果是:

本申请的重建与补全方法、系统及存储介质通过获取自动驾驶场景的立体相机图像,生成立体相机图像的深度图和语义图;其中,立体相机图像是由立体摄像机对自动驾驶场景进行拍摄而得到的;深度图用于表征立体相机图像的深度信息,语义图用于表征立体相机图像的语义信息;利用语义图对深度图进行优化,以获得优化后的深度图;利用优化后的深度图建立立体相机图像的当前帧的局部面元模型,对当前帧的局部面元模型和与当前帧之前的所有图像帧对应的全局面元模型相冲突的冲突信息进行剔除,利用剔除冲突信息后的当前帧的局部面元模型与全局面元模型进行融合,直到与自动驾驶场景的所有图像帧对应的全局面元模型构建完成;对与自动驾驶场景的所有图像帧对应的全局面元模型进行自适应渲染而得到自适应渲染后的全局面元模型,从自适应渲染后的全局面元模型中获取与自动驾驶场景对应的新视角场景图像序列;其中,新视角场景图像序列包括多个新视角场景渲染图像;获得新视角场景渲染图像的渲染图像金字塔、与新视角场景渲染图像对应的参考图像的参考图像金字塔和与新视角场景渲染图像对应的掩码图像的掩码图像金字

塔,将渲染图像金字塔、参考图像金字塔和掩码图像金字塔输入训练好的对抗生成网络模型,以利用对抗生成网络模型对新视角场景渲染图像中缺失信息的区域进行补全而生成完整新视角场景图像序列;其中,完整新视角场景图像序列包括多个完整新视角场景图像;参考图像是来自于新视角场景渲染图像的邻近视角的立体相机图像,掩码图像是基于新视角场景渲染图像中缺失信息的区域而得到的。

附图说明

[0011] 图1为一种实施例的重建与补全方法的流程示意图;

图2为一种实施例的利用对抗生成网络模型对新视角场景渲染图像中缺失信息的区域进行补全而生成完整新视角场景图像序列的流程示意图;

图3为一种实施例的CSD残差模块内部的数据处理流程的示意图;

图4为一种实施例的CSD子模块的数据处理流程的示意图;

图5为一种实施例的得到与渲染图像金字塔中对应层的图像结果的上下文关联信息的示意图;

图6为一种实施例的渲染子块与渲染子块对应的搜索域内的多个参考掩码子块的相似度值的示意图;

图7为一种实施例的对抗生成网络模型的生成器的模块示意图;

图8为一种实施例的自动驾驶场景的重建与补全系统的模块示意图。

具体实施方式

[0012] 下面通过具体实施方式结合附图对本发明作进一步详细说明。其中不同实施方式中类似元件采用了相关联的类似的元件标号。在以下的实施方式中,很多细节描述是为了使得本申请能被更好的理解。然而,本领域技术人员可以毫不费力的认识到,其中部分特征在不同情况下是可以省略的,或者可以由其他元件、材料、方法所替代。在某些情况下,本申请相关的一些操作并没有在说明书中显示或者描述,这是为了避免本申请的核心部分被过多的描述所淹没,而对于本领域技术人员而言,详细描述这些相关操作并不是必要的,他们根据说明书中的描述以及本领域的一般技术知识即可完整了解相关操作。

[0013] 另外,说明书中所描述的特点、操作或者特征可以以任意适当的方式结合形成各种实施方式。同时,方法描述中的各步骤或者动作也可以按照本领域技术人员所能显而易见的方式进行顺序调换或调整。因此,说明书和附图中的各种顺序只是为了清楚描述某一个实施例,并不意味着是必须的顺序,除非另有说明其中某个顺序是必须遵循的。

[0014] 本文中为部件所编序号本身,例如“第一”、“第二”等,仅用于区分所描述的对象,不具有任何顺序或技术含义。而本申请所说“连接”、“联接”,如无特别说明,均包括直接和间接连接(联接)。

[0015] 本申请所提供的自动驾驶场景的重建与补全方法及其系统的技术构思是:首先,利用立体摄像机所获取的自动驾驶场景的立体相机图像来构建用于表征大规模交通场景的三维几何模型(如下文中的全局面元模型);其中,这是由于可以通过高精度的学习方法(如语义分割、深度预测)对上述立体相机图像进行处理以为前述三维几何模型的构建提供密集的深度信息(如下文中的深度图)和语义信息(如下文中的语义图);然后,利用前述语

义分割的结果(如下文中的语义图)进一步有效地优化前述的深度信息,进而利用优化后的前述深度信息去实现三维几何模型的高质量的重建;之后,针对从前述重建好的三维几何模型获取的新视角场景图像序列中新视角场景渲染图像中缺失信息的区域,采用从所采集的参考图像中搜索与前述新视角场景渲染图像相似的子块或区域,并将多个不同尺度的新视角场景渲染图像与上述参考图像进行上下文内容上相关性的匹配,并把匹配的相关结果转换为后续的特征空间训练归一化的调制仿射参数,以生成上述新视角场景渲染图像中缺失信息的区域的纹理,从而保证最终所生成的完整新视角场景图像序列在内容和结构上的一致性。其中,对于上述三维几何模型的构建,所需的输入包括自动驾驶场景的立体相机图像,利用立体相机图像生成相应的深度图和语义图。利用语义图来优化深度图中边界异常点并去除动态目标,利用优化后的深度图建立初始的局部面元模型,并对初始的局部面元模型与初始的全局面元模型之间的冲突部分进行剔除,全局面元模型随着融合策略(即对初始的全局面元模型进行融合的过程)的增加而递增。在获得全局面元模型后,通过自适应方法对新视角场景渲染图像进行渲染,最后利用对抗生成网络模型生成空间结构一致的高质量完整新视角场景图像序列。

[0016] 下面将结合实施例对本申请的技术方案进行详细说明。

[0017] 本申请提供一种自动驾驶场景的重建与补全方法。请参考图1,该重建与补全方法包括:

步骤S100:获取自动驾驶场景的立体相机图像,生成立体相机图像的深度图和语义图;其中,立体相机图像是由立体摄像机对自动驾驶场景进行拍摄而得到的;深度图用于表征立体相机图像的深度信息,语义图用于表征立体相机图像的语义信息;利用语义图对深度图进行优化,以获得优化后的深度图;

步骤S200:利用优化后的深度图建立立体相机图像的当前帧的局部面元模型,对当前帧的局部面元模型和与当前帧之前的所有图像帧对应的全局面元模型相冲突的冲突信息进行剔除,利用剔除冲突信息后的当前帧的局部面元模型与全局面元模型进行融合,直到与自动驾驶场景的所有图像帧对应的全局面元模型构建完成;

步骤S300:对与自动驾驶场景的所有图像帧对应的全局面元模型进行自适应渲染而得到自适应渲染后的全局面元模型,从自适应渲染后的全局面元模型中获取与自动驾驶场景对应的新视角场景图像序列;其中,新视角场景图像序列包括多个新视角场景渲染图像;

步骤S400:获得新视角场景渲染图像的渲染图像金字塔、与新视角场景渲染图像对应的参考图像的参考图像金字塔和与新视角场景渲染图像对应的掩码图像的掩码图像金字塔,将渲染图像金字塔、参考图像金字塔和掩码图像金字塔输入训练好的对抗生成网络模型,以利用对抗生成网络模型对新视角场景渲染图像中缺失信息的区域进行补全而生成完整新视角场景图像序列。

[0018] 其中,上述步骤S400中的完整新视角场景图像序列包括多个完整新视角场景图像;参考图像是来自于新视角场景渲染图像的邻近视角的立体相机图像,掩码图像是基于新视角场景渲染图像中缺失信息的区域而得到的。

[0019] 上述立体相机图像是由立体摄像机对自动驾驶场景进行拍摄而得到的。上下文关联信息用于为缺失信息的区域提供上下文匹配信息。深度图用于表征立体相机图像的深度

信息,语义图用于表征立体相机图像的语义信息;利用语义图对深度图进行优化,以获得优化后的深度图。

[0020] 上述完整新视角场景图像序列包括与自动驾驶场景对应的多个完整新视角场景图像。参考图像用于为新视角场景渲染图像提供场景上下文信息。

[0021] 上述步骤S100中,立体相机图像是由立体摄像机对前述自动驾驶场景进行拍摄而得到。其中,本领域技术人员可以根据实际需求而对自动驾驶场景进行选择,此处不对自动驾驶场景所包含的具体内容进行限定。

[0022] 一些实施例中,上述步骤S100中,可以利用现有的语义分割模型对前述立体相机图像进行处理,以生成自动驾驶场景的语义图。其中,语义分割模型可以为PSM-Net的预训练模型等。其中,PSM-Net(pyramid stereo matching network,金字塔立体匹配网络)是由空间金字塔池化和三维卷积层组成的网络,将全局的背景信息纳入立体匹配中,以实现遮挡区域、无纹理或重复区域的可靠估计。空间金字塔池化能够通过多尺度积累获取全局语境信息,三维卷积层组成的网络能够获取视差图。

[0023] 需要说明的是,上述步骤S100中利用现有的语义分割模型对前述立体相机图像进行处理以生成自动驾驶场景的语义图的具体过程属于本领域的现有技术,故此处不再对上述具体过程进行赘述。

[0024] 一些实施例中,上述步骤S100中,可以利用现有的深度预测模型对前述立体相机图像进行处理,以生成自动驾驶场景的深度图。其中,深度预测模型可以为PointRend的预训练模型等。PointRend(基于点的渲染)神经网络,是基于迭代细分算法在自适应选择的位置执行基于点的分割预测;其通过在现有最新模型的基础上构建,PointRend可以灵活地应用于实例和语义分割任务。

[0025] 需要说明的是,上述步骤S100中利用深度预测模型对前述立体相机图像进行处理,以生成自动驾驶场景的深度图的具体过程属于本领域的现有技术,故此处不再对上述具体过程进行赘述。

[0026] 需要说明的是,上述步骤S100中利用现有的语义分割模型生成上述语义图的目的有以下两方面:第一,利用上述语义图来优化上述深度图中物体的不太准确的边界深度信息,即,优化上述深度图中所有物体的边界深度信息;第二是利用上述语义图来剔除上述深度图中动态物体。

[0027] 由于预测算法(如前述深度预测模型所使用的深度预测算法)的不确定性,不可避免地会在前述深度图中引入异常值或噪声。例如,一般情况下不同类的物体的深度不同(如树、房子、车等),但深度预测模型输出的深度信息对边界的深度估计误差比较大,因此,上述步骤S100中“利用语义图对深度图进行优化,以获得优化后的深度图”借助更准确的语义信息(即前述语义图)优化前述深度图中所有物体的边界深度信息,即通过优化前述深度图中目标物体的边界来增强前述深度图,进而得到优化后的深度图。具体地,首先确定需要优化的前述深度图中目标物体的边界,即,前述深度图中深度梯度变化超过预设的阈值的区域;其中,一般越靠近目标边界的区域,其深度梯度的变化越大;之后,相应的深度优化到几何上接近语义边界的目标点。根据语义一致区域的深度局部性原则,每个像素深度可以通过对其周围N个像素块的有效部分进行平均来优化,同时确保上述深度图的深度值由语义相同的像素进行融合和深度边界向语义边界方向优化。一些实施例中,还可以过滤掉深度

图中相邻深度相差很大的点。通过比较顺序帧和语义标签可以删除上述深度图中的动态对象,例如行人和汽车。最终实现优化上述深度图中边界异常点并去除动态目标而得到优化后的深度图。

[0028] 三维面元模型(如前述全局面元模型和局部面元模型)的表面被量化为一组定向的小圆盘,即面元。面元的大小由图像的分辨率决定。每个面元被多属性有效地定义和可视化,如位置、法线、半径、颜色、语义和深度等信息。前述面元模型采用面元对三维空间对象的表面进行连续或非连续几何描述和特征描述,而不研究三维空间对象的内部特征。

[0029] 虽然上述步骤S100对深度图进行了优化,但是在将重建的新局部模型(即上述初始的局部面元模型)合并到现有的全局模型(即上述初始的全局面元模型)之前,还需要对上述初始的局部面元模型进行进一步的去噪。因此,上述步骤S200中“利用优化后的深度图建立初始的局部面元模型,对初始的局部面元模型中与初始的全局面元模型相冲突的冲突信息进行剔除,利用剔除冲突信息后的局部面元模型对初始的全局面元模型进行融合而获得融合后的全局面元模型”的具体流程的论述如下。由于当目标距离较远时,深度预测的性能会下降,因此距离相机(如前述立体摄像机)的光学中心较近的深度测量精度更高。基于此假设,现有初始的局部面元模型中的深度小于初始的全局面元模型的映射深度,即局部面元比全局面元更靠近相机光学中心,这是由于相机向目标移动且视场小于180度。如果在一定阈值范围内,局部面元的深度距离相机光学中心的距离远于全局面元模型中全局面元的深度距离相机光学中心的距离,则设置为较低的置信度,否则设置为较高的置信度。

[0030] 一些实施例中,可以通过映射的方式找到“深度冲突比较大的面元”(即上述冲突信息),对上述初始的局部面元模型中的冲突信息进行剔除之后需要利用剔除冲突信息后的局部面元模型对初始的全局面元模型进行融合。一些实施例中,局部面元元素的预测深度没有置信度,或者局部面元点比全局面元点离相机光学中心远得多(如超出预设的阈值),这些局部面元点将作为冲突点被删除,从而进一步融合全局面元模型而获得融合后的全局面元模型。

[0031] 需要说明的是,上述深度的置信度可以直接采用现有技术进行初始化。例如,上述初始化可以采用现有的文献“M.F.A.Eldesokey and F.S.Khan,“Confidence propagation through cnns forguided sparse depthregression,”IEEEtransactions on pattern analysis and machine intelligence,vol.42,no.10,pp.2423-2436,2019”所记载的方法。由于上述初始化可以直接采用本领域的现有技术,故此处不再对其进行赘述。

[0032] 需要说明的是,上述局部面元模型的重建是依靠深度信息(即前述深度图)重建的,因此在将初始的局部面元模型中冲突深度(即上述冲突信息)剔除后,留下来的局部面元模型(本申请将该留下来的局部面元模型视为是比较准确的)再融合到上述全局面元模型。

[0033] 一些实施例中,上述步骤S200中“利用剔除冲突信息后的当前帧的局部面元模型与全局面元模型进行融合”时,为了避免在模型融合过程中出现异常值,只有那些局部面元模型中能够匹配到颜色和语义属性相似且置信度高的全局面元的局部面元才会被融合到全局面元模型中。当局部面元比全局面元更接近相机光学中心时,只保留局部面元元素。若局部面元和全局面元都在预设的距离阈值内,则使用前述置信度加权平均值的方式将全局面元与局部面元进行融合。若没有局部面元和全局面元配对,新的局部面元将作为不稳

定面元元素添加到前述全局面元模型中,其置信度也设置为相对较低的数值。

[0034] 需要说明的是,一个自动驾驶场景包括多个立体相机图像的图像帧,例如,多个连续的图像帧。上述步骤S200中,“与当前帧之前的所有图像帧对应的全局面元模型”是一帧一帧融合而成的,即,当前帧的局部面元模型构建好后,对当前帧的局部面元模型和与当前帧之前的所有图像帧对应的全局面元模型相冲突的冲突信息进行剔除,然后将当前帧对应的局部面元模型融合进与当前帧之前的所有图像帧对应的全局面元模型;以此类推,由于融合的帧数越来越多,全局面元模型将会越来越大,直到与自动驾驶场景的所有图像帧对应的全局面元模型构建完成。需要注意的是,在构建第一帧(即第一个当前帧)的局部面元模型时,是没有全局面元模型的;全局面元模型是从第二帧(即第二个当前帧)开始构建的,即,将与第一个当前帧对应的局部面元模型作为上述与当前帧之前的所有图像帧对应的全局面元模型,以此类推。直到将自动驾驶场景的最后一帧(即最后一个当前帧)融合进前述与当前帧之前的所有图像帧对应的全局面元模型,至此,与自动驾驶场景的所有图像帧对应的全局面元模型构建完成。其中,与当前帧之前的所有图像帧均对应的是上述立体相机图像。

[0035] 一些实施例中,上述步骤S300中,可以直接从与自动驾驶场景的所有图像帧对应的全局面元模型获取与自动驾驶场景对应的新视角场景图像序列。

[0036] 一些实施例中,上述步骤S300中,可以对与自动驾驶场景的所有图像帧对应的全局面元模型进行自适应渲染而得到自适应渲染后的全局面元模型,从自适应渲染后的全局面元模型获取与自动驾驶场景对应的新视角场景图像序列。

[0037] 需要说明的是,从上述全局面元模型获取与自动驾驶场景对应的新视角场景图像序列的具体过程属于本领域的现有技术和公知常识,故此处不再赘述。

[0038] 需要说明的是,前述相机采集的视角一般称为“原视角”,在重建3D面元模型(如全局面元模型)过程中也是根据该“原视角”建立的,3D面元模型建立好以后,可以从3D面元模型中采集“新视角”(即“非原视角”)图片,像视频一样连续采集进而得到新视角场景图像序列。

[0039] 在构建好前述全局面元模型之后,可以从不完整的前述全局面元模型中获取新视角图像序列。相机的新视点接近原始视角,其中,一些实施例中,光轴被限制在 ± 15 度,基线相对较宽(约 $\pm 1\text{m}$)。一些实施例中,为了更好地利用前述全局面元模型的高保真可视化,本申请提供了一种对与自动驾驶场景的所有图像帧对应的全局面元模型进行自适应渲染的处理方式。该处理方式包含两种主要方案,分别用于调整前述全局面元模型中面元的法线和半径。第一,前述全局面元模型中远视点的面元将根据沿观测方向的法向量调整。如此的操作可以确保那些远距离的面元点也能有效地参与渲染。第二,前述全局面元模型中一个特定的半径往往会在较短的距离内从一个新的视角产生模糊的纹理。为了克服这个问题,一些实施例中,可以将使用面元的法线与新视角的观测方向之间的夹角来调整冲浪半径。一个面元点覆盖多个像素,并且法线在观测方向的一定范围内(如 $\pi/2$)时,为了避免与其它

面元的冗余重叠,其半径可以由 $r_i(u)$ 大幅缩小为 $e^{\left(\frac{\alpha}{\pi/2}-1\right)} r_i(u)$ 。其中, u 为面元覆盖的

像素的位置, α 为面元的法线与新视角的观测方向的夹角,此处的 i 表示面元的编号。因此,

一些实施例中,可以根据沿上述观测方向的法向量调整全局面元模型中面元(如远视点的面元)的法线和/或半径。

[0040] 上述步骤S400主要用于完成交通场景(即上述新视角场景渲染图像)的图像补全。上述全局面元模型是基于真实世界的传感器数据(如上述立体相机图像)而构建的。由于上述全局面元模型重建的不完整以及上述动态物体的移除,场景背景渲染图像(即上述新视角场景渲染图像)中出现各种掩膜或空洞。本申请主要考虑到空间和内容上的一致性,采用一定的内容去填充这些掩膜/空洞(即前述新视角场景渲染图像中缺失信息的区域),使用下文的对抗生成网络模型合成高质量、逼真的交通场景序列(即前述完整新视角场景图像序列)。对抗生成网络模型的输入包括不同尺度(即不同分辨率的)的新视角场景渲染图像、不同尺度的掩码图像和不同尺度的参考图像。对抗生成网络模型的整体处理过程包括两个阶段:

第一阶段,首先将对应尺度的新视角场景渲染图像以子块(patch)的方式划分为多个渲染子块,将渲染子块在参考图像上通过下文的掩膜图像表征关联机制搜索近似上下文而得到上下文关联信息;其中,为了增加上述近似上下文相关信息的匹配精度,将对应的掩膜图像的掩膜子块应用到参考图像对应的参考子块后得到参考掩膜子块,再进行渲染子块与参考掩膜子块匹配;

第二阶段,将多个尺度的渲染图像作为输入,通过对抗生成网络模型来对新视角场景渲染图像中缺失信息的区域(如空洞等)进行补全而生成空间结构一致的、高质量的完整新视角场景图像序列;其中,将第一阶段预测的近似上下文相关信息(即,下文的关联图像和相似度图)转移到多个CSD子模块,以保持所生成的完整新视角场景图像序列之间的一致性。通过上述两阶段的图像补全处理,进而使得所生成的交通场景图像(即完整新视角场景图像序列)更加逼真、在内容上更加连贯。

[0041] 需要说明的是,请参考图4,掩码图像 I_m 是基于新视角场景渲染图像 I_r 中缺失信息的区域而得到的。这是由于从自适应渲染后的全局面元模型中获取的新视角场景渲染图像 I_r 本身就会出现各种不规则空洞/掩码,若直接使用新视角场景渲染图像与来自邻近视角的另一个新视角场景渲染图像(即参考图像)进行相似性匹配,很可能造成相似性匹配的准确度不高(例如,需补全的新视角场景渲染图像的一个区域存在空洞,而来自邻近视角的参考图像的同一区域并不存在该空洞,若直接进行相似性匹配,可以理解的是,该区域很难匹配到相似的上下文信息),因此,可以首先将新视角场景渲染图像 I_r 的各种不规则空洞/掩码单独提取出来而制作成与新视角场景渲染图像 I_r 对应的掩码图像 I_m (例如,新视角场景渲染图像 I_r 中汽车的后视镜区域存在不规则空洞,本申请就将该汽车的后视镜区域存在不规则空洞的区域单独提取出来而生成掩码图像 I_m),然后再将掩码图像 I_m 应用于上述参考图像而得到对应的参考掩码子块,如此便可以使得参考图像中与新视角场景渲染图像 I_r 的各种不规则空洞/掩码对应的区域也自动地出现相同的不规则空洞/掩码等缺失信息的区域,进而可以显著地提高上述近似上下文相关信息的匹配精度(如下述渲染子块和参考子块进行匹配的精度);反之,若不将掩码图像 I_m 应用于参考图像而得到对应的参考掩码子块并直接采用参考掩码子块与渲染子块进行相似性匹配,则容易造成后续生成的关联图像 I_c 的准确度不高,进而无法完成高质量地对新视角场景渲染图像中缺失信息的区域进行补全。

[0042] 请参考图2,上述步骤S400中,将渲染图像金字塔、参考图像金字塔和掩码图像金

字塔输入训练好的对抗生成网络模型,以利用对抗生成网络模型对新视角场景渲染图像中缺失信息的区域进行补全而生成完整新视角场景图像序列,包括:

步骤S410:渲染图像金字塔、参考图像金字塔和掩码图像金字塔中各层的图像结果按照由顶层至底层的顺序分别输入生成器内第一级的CSD残差模块至最后一级的CSD残差模块;将前一级的CSD残差模块的输出作为相邻的后一级的CSD残差模块的一种输入,将最后一级的CSD残差模块的输出作为完整新视角场景图像;其中,生成器内第一级的CSD残差模块至最后一级的CSD残差模块的网络结构均相同;其中,渲染图像金字塔、参考图像金字塔和掩码图像金字塔的层数均与CSD残差模块的数量相同。

[0043] 一些实施例中,上述步骤S400中,可以对新视角场景渲染图像、与新视角场景渲染图像对应的参考图像和与新视角场景渲染图像对应的掩码图像分别进行下采样等现有技术而分别获得新视角场景渲染图像的渲染图像金字塔、参考图像的参考图像金字塔和掩码图像的掩码图像金字塔。

[0044] 图像金字塔是由一系列不同分辨率的图像构成的集合,主要采用上采样和下采样两种常用的运算。下采样是指由高分辨率图像向低分辨率图像采样并进行高斯滤波平滑处理。上采样是指由低分辨率图像向高分辨率图像插值并进行高斯滤波平滑处理。对原始图像进行上采样而得到的图像的分辨率变为原始图像的分辨率的两倍。对原始图像进行下采样得到的图像的分辨率变为原始图像的分辨率的一半。

[0045] 一些实施例中,向下采样而构建图像金字塔的简要过程可以为:对于给定的图像先做一次高斯滤波平滑处理,也就是对上述图像进行一个卷积操作;对上述图像进行下采样,其中,可以在上述图像的行方向取奇数列,在上述图像的列方向取偶数列;或者,可以在上述图像的行方向取偶数列,在上述图像的列方向取奇数列;对采样后的图像,重复前两步操作即可得到图像金字塔。需要说明的是,一些实施例中,对新视角场景渲染图像、参考图像和掩码图像分别进行下采样而分别得到新视角场景渲染图像的渲染图像金字塔、参考图像的参考图像金字塔和掩码图像的掩码图像金字塔的具体过程属于本领域的现有技术,故此处不再赘述。

[0046] 一些实施例中,上述步骤S400中,上述渲染图像金字塔/参考图像金字塔和掩码图像金字塔中各层图像结果的尺度(或空间分辨率)按照由顶层至底层的顺序可以为 $1/32$ 、 $1/16$ 、 $1/8$ 、 $1/4$ 、 $1/2$,其中, $1/2$ 表示当前层的尺度(或空间分辨率)为原始尺度(或空间分辨率)的 $1/2$,以此类推。本领域技术人员可以根据实际需求而对上述渲染图像金字塔/参考图像金字塔和掩码图像金字塔中各层图像结果的尺度(或空间分辨率)进行调整。

[0047] 一些实施例中,对抗生成网络模型包括一个生成器。请参考图7,生成器包括多级彼此串联连接的CSD残差模块(图7中由左至右的CSD残差模块分别为第一级的CSD残差模块、第二级的CSD残差模块至最后一级的CSD残差模块)。其中,CSD(Contextual and Spatial Denormalization)表示上下文和空间非规范化。

[0048] 请参考图7,上述步骤S410中,渲染图像金字塔、参考图像金字塔和掩码图像金字塔中各层的图像结果按照由顶层至底层的顺序分别输入生成器内第一级的CSD残差模块至最后一级的CSD残差模块;将前一级的CSD残差模块的输出作为相邻的后一级的CSD残差模块的一种输入,将最后一级的CSD残差模块的输出作为完整新视角场景图像,包括:

将渲染图像金字塔、参考图像金字塔和掩码图像金字塔中顶层的图像结果,以及

与顶层的图像结果的尺度对应的新视角场景渲染图像输入第一级的CSD残差模块；

将渲染图像金字塔、参考图像金字塔和掩码图像金字塔中除顶层外的其余各层的图像结果按照由顶层的下一层至底层的顺序分别输入生成器内第二级的CSD残差模块至最后一级的CSD残差模块；将最后一级的CSD残差模块的输出作为完整新视角场景图像。

[0049] 一些实施例中,可以将渲染图像金字塔、参考图像金字塔和掩码图像金字塔中顶层的图像结果,以及批量的与顶层的图像结果的尺度对应的新视角场景渲染图像输入第一级的CSD残差模块。本领域技术人员可以自行确定上述每个批量(batchsize=N)的数量,如,每个批量的数量为五个。

[0050] 上述CSD残差模块包括第一级的CSD子模块、第二级的CSD子模块和第三级的CSD子模块。其中,CSD(Contextual and Spatial Denormalization)子模块,即,上下文和空间非规范化子模块。

[0051] 请参考图3,上述CSD残差模块内部的数据处理流程为:

将渲染图像金字塔、参考图像金字塔和掩码图像金字塔中对应层的图像结果(请参考图4,如对应尺度的掩码图像 I_m 、参考图像 I_c 和新视角场景渲染图像 I_r),以及输入CSD残差模块的特征图分别输入至第一级的CSD子模块和第三级的CSD子模块,将第一级的CSD子模块的输出作为第一ReLU激活函数的输入,对第一ReLU激活函数的输出进行卷积操作而得到第一卷积结果,将第一卷积结果输入第二级的CSD子模块,将第二级的CSD子模块的输出作为第二ReLU激活函数的输入,对第二ReLU激活函数的输出进行卷积操作而得到第二卷积结果,将第三级的CSD子模块的输出作为第三ReLU激活函数的输入,对第三ReLU激活函数的输出进行卷积操作而得到第三卷积结果,对第二卷积结果和第三卷积结果执行单位加操作而得到CSD残差模块的输出;

其中,输入第一级的CSD残差模块的特征图为与顶层的图像结果的尺度对应的新视角场景渲染图像(即需要补全的新视角场景渲染图像),输入相邻的后一级的CSD残差模块的特征图(请参考图3中的 x^i)为前一级的CSD残差模块的输出。

[0052] 例如,输入第二级的CSD残差模块的特征图为第一级的CSD残差模块的输出,输入第三级的CSD残差模块的特征图为第二级的CSD残差模块的输出,输入其他级的CSD残差模块的特征图以此类推。

[0053] 一些实施例中,渲染图像金字塔的层数、参考图像金字塔的层数、掩码图像金字塔的层数均与生成器中CSD残差模块的数量相同。上述第一ReLU激活函数、第二ReLU激活函数和第三ReLU激活函数,即本领域现有常用的ReLU激活函数。

[0054] 上述第一级的CSD子模块、第二级的CSD子模块和第三级的CSD子模块的数据处理流程均相同。请参考图4,图4中的BatchNorm表示归一化层或归一化操作,conv表示下述的卷积操作(如第一卷积操作),MPC表示下文的掩膜图像表征关联机制,element-wise表示逐元素乘法操作。

[0055] 请参考图4,CSD子模块的数据处理流程为:

对输入的渲染图像金字塔中对应层的图像结果(如对应尺度的新视角场景渲染图像 I_r)、参考图像金字塔中对应层的图像结果(如对应尺度的参考图像 I_c)和掩码图像金字塔中对应层的图像结果(如对应尺度的掩码图像 I_m)进行相似性匹配,以得到与渲染图像金字塔中对应层的图像结果的上下文关联信息;

其中,上下文关联信息包括相似度图和渲染图像金字塔中对应层的图像结果(即对应尺度的 I_r)的关联图像 I_c ;关联图像是基于参考图像金字塔中对应层的图像结果(即对应尺度的 I_c)与渲染图像金字塔中对应层的图像结果(即对应尺度的 I_r)最相似的多个区域而生成的,相似度图用于表征渲染图像金字塔中对应层的图像结果(即对应尺度的 I_r)与参考图像金字塔中对应层的图像结果(即对应尺度的 I_c)之间的相关性;

其中,上下文关联信息用于为缺失信息的区域提供上下文匹配信息;

对关联图像 I_c 和相似度图分别执行第二卷积操作而分别得到关联图像 I_c 的空间特征 $f(I_c)$ 和相似度图的空间特征;

对渲染图像金字塔中对应层的图像结果 I_r 执行第二卷积操作得到对应层的新视角场景渲染图像的空间特征 $f(I_r)$;

对相似度图的空间特征 $\alpha_{c,h,w}$ 和新视角场景渲染图像的空间特征 $f(I_r)$ 执行逐元素乘法操作而得到逐元素乘法结果;

将逐元素乘法结果与关联图像 I_c 的空间特征 $f(I_c)$ 相加而得到第二仿射参数 $\beta_{c,h,w}$;其中,第二仿射参数的表达式为:

$$\beta_{c,h,w} = \alpha_{c,h,w} \odot f_{c,h,w}(I_r) + f_{c,h,w}(I_c);$$

$\alpha_{c,h,w} \odot f_{c,h,w}(I_r)$ 为逐元素乘法结果, $\alpha_{c,h,w}$ 为相似度图的空间特征,

$f_{c,h,w}(I_c)$ 为关联图像 I_c 的空间特征;其中,所述 c 、 h 和 w 分别为所述对应层的图像结果的特征通道数、长度和宽度,所述 I_r 表示所述新视角场景渲染图像,所述 I_c 为所述对应层的图像结果的关联图像;

对输入CSD残差模块的多个特征图进行归一化而得到归一化参数;

将归一化参数与第二仿射参数 $\beta_{c,h,w}$ 相加而得到的归一化结果(即下文的BN'(xⁱ))作为CSD子模块的输出。

[0056] 需要说明的是,对关联图像 I_c 和相似度图分别执行第二卷积操作分别是由不同的简单两层卷积网络来实现的。对渲染图像金字塔中对应层的图像结果(即 I_r)执行第二卷积操作也是由简单两层卷积网络来实现的。但是,需要注意的是,与渲染图像金字塔中对应层的图像结果对应的简单两层卷积网络和上述与关联图像对应的简单两层卷积网络的操作相同且共享参数。

[0057] 一些实施例中,请参考图5,对输入的渲染图像金字塔中对应层的图像结果、参考图像金字塔中对应层的图像结果和掩码图像金字塔中对应层的图像结果进行相似性匹配,以得到与渲染图像金字塔中对应层的图像结果的上下文关联信息,包括:

利用一个划分框以第一步幅在渲染图像金字塔中对应层的图像结果(即对应尺度的 I_r)中移动而将渲染图像金字塔中对应层的图像结果划分为多个尺寸为 $k \times k$ 的渲染子块,

利用划分框以第一步幅在掩码图像金字塔中对应层的图像结果(即对应尺度的 I_m)中移动而将掩码图像金字塔中对应层的图像结果划分为多个尺寸为 $k \times k$ 的掩码子块,

在参考图像金字塔中对应层的图像结果(即对应尺度的 I_c)中确定与渲染子块对

应的搜索域,搜索域的尺寸为 $k' \times k'$,利用划分框以第二步幅在搜索域中移动而将搜索域划分为 $s \times s$ 个尺寸为 $k \times k$ 的参考子块; k, k' 和 s 均为预设的常数;

将与渲染子块对应的掩码子块应用到与渲染子块对应的参考子块而得到与渲染子块对应的参考掩码子块;

计算渲染子块和与渲染子块对应的搜索域内各参考掩码子块之间的各相似度,将各相似度中的最大值作为相似度图中与渲染子块对应的元素;

其中,与渲染子块对应的搜索域的尺寸大于渲染子块的尺寸;

将相似度图中各元素所分别对应的参考子块作为最佳匹配子块,利用与相似度图对应的多个最佳匹配子块生成关联图像(即 I_c)。

[0058] 一些实施例中,当 $k=3$,第二步幅 $m=1$, $s=3$ 时,而搜索域的尺寸 $k' \times k' = [k+m \times (s-1)] \times [k+m \times (s-1)]$,因此上述 $k'=5$ 。本领域技术人员可以根据实际需求而对上述参数 k, m, s 和 k' 进行调整。

[0059] 一些实施例中,上述得到与渲染图像金字塔中对应层的图像结果的上下文关联信息的流程中,划分框是一个虚拟的方框。可以理解的是,可以将上述划分框看作一个卷积核,上述划分框可以按照第一步幅在渲染图像金字塔中对应层的图像结果(即对应尺度的 I_r)中移动(例如从一行中的最左侧至最右侧,再从上一行至下一行,等移动方式),而将上述划分框所包围的渲染图像金字塔中对应层的图像结果的区域作为一个渲染子块,进而将渲染图像金字塔中对应层的图像结果划分为多个尺寸为 $k \times k$ 的渲染子块。

[0060] 需要说明的是,本领域技术人员可以根据实际需求而对上述第一步幅 n 的大小进行调整,此处不对第一步幅的大小进行限制。例如,第一步幅 $n=1$ 。

[0061] 需要说明的是,上述得到与渲染图像金字塔中对应层的图像结果的上下文关联信息的流程中“将与渲染子块对应的掩码子块应用到与渲染子块对应的参考子块而得到与渲染子块对应的参考掩码子块”指的是利用对应的掩码子块去遮挡对应的参考子块进而生成带有掩码的参考子块(即上述参考掩码子块),然后再计算渲染子块与该渲染子块的搜索域内的多个参考掩码子块之间的相似度。

[0062] 如前文所述,从上述全局面元模型获取的新视角场景渲染图像并不完整,但在同一采集视角存在一张或多张与新视角场景渲染图像对应的原始真实图像。一些实施例中,定义上述 $I_m \in L^{H \times W}$ 为二进制掩码图像(即上述掩码图像),其中 H 和 W 分别为新视角场景渲染图像对应的掩码图像的高度和宽度。请参考图5,参考图像 I_c ,即条件样本,是由相机(如前述立体摄像机)采集的原始图像(即前述立体相机图像)。参考图像 I_c 来自新视角场景渲染图像 I_r 的邻近视点。参考图像 I_c 为新视角场景渲染图像 I_r 的补全提供场景上下文信息。新视角场景渲染图像 I_r 和对应的参考图像 I_c 之间的相关性计算方法与卷积类似,将输入的新视角场景渲染图像 I_r 划分成多个尺寸为 $k \times k$ 个渲染子块,其掩码图像相应地划分为相同的一组掩码子块。新视角场景渲染图像 I_r 的每个渲染子块在条件样本中对应的搜索域中作为一个卷积核来操作。搜索域的尺寸大于渲染子块的尺寸。在进行匹配之前,所有的参考子块都要被赋予相应的二进制掩码(即掩码子块),以提高上述相关性计算的匹配精度。

[0063] 需要说明的是,上述搜索域的尺寸大于渲染子块的尺寸的原因是:由于“新视角场景渲染图像 I_r ”与“原视角”在视角上存在偏差,因此需要在做上下文匹配的时候搜索更大的区域,才能匹配到对应的上下文信息。

[0064] 上述相似度图的表达式为:

$$simi(i) = \max_{j \in s \times s} \langle (p_{k \times k}^{I_m})_i \odot (p_{k \times k}^{I_c})_{i,j}, (p_{k \times k}^{I_r})_i \rangle;$$

其中, $simi(i)$ 表示所述渲染图像金字塔中对应层的图像结果中第 i 个渲染子块 $(p_{k \times k}^{I_r})_i$ 和参考图像金字塔中对应层的图像结果 (即对应尺度的 I_c) 中与第 i 个渲染子块对应的搜索域内 $s \times s$ 个参考掩码子块 $(p_{k \times k}^{I_m})_i \odot (p_{k \times k}^{I_c})_{i,j}$ 之间的各相似度的最大值, $\langle \cdot, \cdot \rangle$ 为求取余弦相似度的操作, \odot 表示逐元素相乘, 此处的 $i = 1, 2, \dots, (\frac{H}{n} \times \frac{W}{n})$, 其中, H 和 W 分别为新视角场景渲染图像的高度和宽度, n 为上述第一步幅; I_m 表示掩码图像, I_c 表示参考图像, I_r 表示新视角场景渲染图像, j 表示搜索域内参考子块 $p_{k \times k}^{I_c}$ 的数量, $p_{k \times k}^{I_m}$ 为掩码子块。

[0065] 上述归一化参数的表达式为: $\gamma_c \left(\frac{x^i - \mu_c(x^i)}{\sigma_c(x^i)} \right)$, 其中, x^i 为输入第 i 级的CSD残

差模块中CSD子模块的一批特征图 (如 N 个特征图), $x^i \in \mathbb{R}^{N \times C \times H \times W}$, C 为相应的特征通道的数量, 例如, x^i 可以指输入第 i 级CSD残差模块中第一级的CSD子模块和第三级的CSD子模块的特征图; $\mu_c(x^i)$ 和 $\sigma_c(x^i)$ 分别是输入第 i 级的CSD残差模块中CSD子模块的 N 个特征图的均值和标准差, γ_c 为第一仿射参数, $\gamma_c \in \mathbb{R}^C$; 调制仿射参数包括第一仿射参数 γ_c 和第二仿射参数 $\beta_{c,h,w}$ 。其中, 均值和标准差的表达式分别为:

$$\mu_c(x^i) = \frac{1}{NHW} \sum_{n,h,w} x_{n,c,h,w}^i,$$

$$\sigma_c(x^i) = \sqrt{\frac{1}{NHW} \sum_{n,h,w} (x_{n,c,h,w}^i - \mu_c(x^i))^2 + \varepsilon};$$
 其中, 此处的 $n \in N, c \in C, h \in H, w$

$\in W, \varepsilon$ 为预设的系数, 该预设的系数的设置属于本领域的公知常识。

[0066] 一些实施例, 本领域技术人员可以根据实际需求而灵活地设定上述搜索域的尺寸 (即大小), 例如, 搜索域的大小可以为 5×5 , 此时搜索域为以参考图像 I_c 中与该渲染子块对应的参考掩码子块为中心, 边长为 5 的区域。通常情况下, 上述 H 和 W 的数值相同。

[0067] 需要说明的是, 在通过上述相似度图的表达式, 为了提高相似度计算的精度, 需要将上述掩码图像对应的掩码子块应用到参考图像中搜索域内的各参考子块而得到各参考

子块对应的参考掩码子块 (即, 上述 $(p_{k \times k}^{I_m})_i \odot (p_{k \times k}^{I_c})_{i,j}$ 表示将第 i 个掩码子块

分别赋予参考图像 I_c 中与第 i 个渲染子块对应的搜索域内各参考子块而得到的对应的搜索域内各参考掩码子块),之后再计算上述各参考掩码子块与对应的渲染子块的相似度。

[0068] 一些实施例中,可以记录最佳匹配子块的索引。最佳匹配子块是指相似度图中各元素所分别对应的参考子块,即,渲染子块对应的搜索域中渲染子块与多个参考掩码子块之间的各相似度的最大值所对应的参考子块。上述索引用于标记最佳匹配子块在上述渲染子块对应的搜索域中对应的位置。对于新视角场景渲染图像 I_r 中缺失信息的较大区域,由于该区域对应的参考子块完全被掩码子块填充,因而上述渲染子块与多个参考掩码子块之间的各相似度中的一个或者多个可能为零,进而导致后续不正确的关联。为了缓解上述这一技术问题,一些实施例中,在逐个确定相似图中的各元素(请参考图5中的元素S1至元素S9)的所对应的最佳匹配子块的索引时,假设一个渲染子块对应的搜索域中含有9个参考掩码子块,请参考图6,a1至a9可以分别看作该渲染子块与渲染子块对应的搜索域内的9个参考掩码子块的相似度值,本申请采用非零索引所对应的的相似度值的平均值代替该零值索引的相似度值。例如,若计算出的a1至a9全部是0(其中,a1至a9所分别对应的索引为0至8),此时零值索引分别为0至8,那么该渲染子块和与渲染子块对应的搜索域内各参考掩码子块之间的各相似度的最大值就是0,此种情况下默认该渲染子块的最佳匹配子块是该搜索域内的第0个参考掩码子块所对应的参考子块,即a1,但是a1所对应的参考子块事实上并非最佳匹配子块。因此,本申请采用零值索引周围的某几个非零索引(如非零索引可以为零值索引的上方、下方、左侧和右侧的4个临近的索引)所对应的相似度值来计算出平均值,并使用该平均值来代替零值索引所对应的相似度值,之后,再根据与该渲染子块对应的搜索域内各参考掩码子块之间的各相似度的最大值的索引找出最佳匹配子块。例如,计算得到的最佳匹配子块的索引是5,而该索引值5指的是a6,即,该渲染子块的最佳匹配子块是该渲染子块对应的搜索域内a6所对应的参考掩码子块对应的参考子块。本领域技术人员可以根据实际需求而对最佳匹配子块的具体确定过程进行适应性调整,如灵活设置零值索引周围的非零索引的具体数量等。

[0069] 一些实施例中,上述得到与渲染图像金字塔中对应层的图像结果的上下文关联信息的流程中“将相似度图中各元素所分别对应的参考子块作为最佳匹配子块,利用与相似度图对应的多个最佳匹配子块生成关联图像”的流程可以为:第一步,通过上述索引获取上述相似度图中各元素所分别对应的最佳匹配子块,其中,这些最佳匹配子块都是相互独立的;第二步,需要把这些最佳匹配子块再拼接成一副图;其中,具体的拼接方式与反卷积的方式相同,故此处可以将最佳匹配子块称为“反卷积滤波器”;第三步,在拼回一幅图的过程中会出现部分重叠的区域,对这些部分重叠的区域中分别对应的多个像素值求平均,并将计算得到的平均值作为上述部分重叠的区域内对应像素点的数值,进而最终得到上述参考图像 I_c 对应的关联图像 I_c' 。

[0070] 需要说明的是,采用反卷积的方式将上述最佳匹配子块拼接成一副图的具体过程属于本领域的现有技术,故此处不再对该具体过程进行赘述。

[0071] 一些实施例中,请参考图3,上述CSD残差模块包括第一级的CSD子模块、第二级的CSD子模块和第三级的CSD子模块。

[0072] 在给定场景上下文信息(即,上述条件样本)的情况下,本申请通过在上述第一级的CSD子模块、第二级的CSD子模块和第三级的CSD子模块中归一化层中加入上下文关联信

息(即,上述对应的关联图像和对应的相似图),以用于补全上述新视角场景渲染图像中缺失信息的区域。

[0073] 需要说明的是,上述给定场景上下文信息和上下文关联信息能够提供上下文匹配信息,以使得对抗生成网络模型计算得出新视角场景渲染图像中缺失信息的具体区域。

[0074] 一些实施例中,请参考图4,在上述归一化层中加入上下文关联信息的灵感来自于条件批归一化(BatchNormalization, BN),其激活层被归一化为零均值和单位偏差,然后通过调整来自外部数据的仿射参数进行反规范化。与尺度对应的调制仿射参数可以用来控制对抗生成网络模型的全局输出,同时保留图像内容的空间结构,自适应实例归一化(AdaIN)也证明了这一点。需要说明的是,全局输出可以包括对抗生成网络模型所最终输出图片(如完整新视角场景图像序列)的内容和风格。本申请提出的各CSD子模块(即,上下文和空间非规范化子模块)中各归一化层利用空间变化的仿射变换使其适合于图像合成。与批量归一化(BN)相似,本申请的各CSD子模块对每个单独的特征通道($n \in N, c \in C, h \in H, w \in W$)的均值和标准差进行归一化。其中,上述归一化结果的表达式为:

$$BN'(x^i) = \gamma_c \left(\frac{x^i - \mu_c(x^i)}{\sigma_c(x^i)} \right) + \beta_{c,h,w}。上述归一化结果作为各CSD子模$$

块的最终输出。

[0075] 需要说明的是,每个CSD残差模块中所有归一化层的调制仿射参数(即上述与不同尺度分别对应的调制仿射参数)均是从各CSD子模块学习到的。本申请将新视角场景渲染图像 I_r 和参考图像 I_c 输入到各CSD子模块的归一化层,而不是整个对抗生成网络模型的第一层网络,这是由于各CSD子模块比普通归一化层能更好地保存输入信息。事实上,各CSD子模块学习到的调制仿射参数已经编码了足够的输入图像信息,因此本申请抛弃了通常在编-解码生成对抗网络中使用的生成器的编码部分。

[0076] 该对抗生成网络模型的生成器由几个带有上采样层的CSD残差模块(即CSD-ResBlk)组成,每个CSD残差模块的输出都是相对应维度的特征。

[0077] 需要说明的是,上述归一化层具体对应图4中的符号“BatchNorm”。

[0078] 本申请首先对输入的新视角场景渲染图像、参考图像和对应的掩码图像进行下采样,建立各自的图像金字塔,然后通过每个CSD残差模块进行上下文信息相关,在不同尺度下匹配空间分辨率,通过设置不同的子块(如渲染子块和参考掩码子块)和不同的搜索域来灵活调整匹配范围。

[0079] 需要说明的是,上述“通过对每个CSD残差模块进行上下文信息相关,在不同尺度下匹配空间分辨率,通过设置不同的子块(如渲染子块和参考掩码子块)和不同的搜索域来灵活调整匹配范围”操作的原因是可以灵活设置上述子块和搜索域的大小。比如,当图片输入的尺度缩小一倍之后,而上述子块和搜索域的大小不变,因此,对应的CSD残差模块所匹配的感受野就增大一倍。而本申请希望可以尽可能扩大搜索域,这样可以处理视角偏差较大的场景(如转弯场景),但此种场景的计算量很大,因此本申请通过不同尺度的输入来灵活调整。上述操作的大致步骤为:第一步,将输入图片(即新视角场景渲染图像、参考图像和掩码图像)缩放至不同尺度而得到不同尺度的图片(即上述渲染图像金字塔中对应层的图

像结果、参考图像金字塔中对应层的图像结果和掩码图像金字塔中对应层的图像结果)；第二步，将上述不同尺度的图片输入到对应的CSD子模块中并通过上述掩膜图像表征关联机制(MPC)进行上下文关联匹配；第三步，将获得的匹配信息通过各CSD子模块输送到对抗生成网络模型的整个网络。

[0080] 一些实施例中，本申请的对抗生成网络模型与pix2pixHD类似，使用多尺度鉴别器/判别器训练生成器。即，本申请的对抗生成网络模型还可以包括多个判别器和VGG卷积网络子模型。其中，pix2pixHD是pix2pix的重要升级，可以实现高分辨率图像生成和图像的语义编辑。对于一个生成对抗网络(GAN)，学习的关键就是理解生成器、判别器和损失函数这三部分。pix2pixHD的生成器和判别器都是多尺度的，单一尺度的生成器和判别器的结构和现有的pix2pix是一样的。pix2pix是一种基于条件式生成对抗网络(CGAN)的图像转译模型，而条件式生成抵抗网络是生成对抗网络的一种扩展，它通过在生成器和判别器中引入条件信息来实现有条件的图像生成。pix2pix的生成器采用U-Net网络结构，融合底层细粒度特征和高层抽象；判别器采用patchGAN网络结构，在图块尺度提取纹理等高频信息。

[0081] 此外，本申请在判别器中引入了相同的特征匹配损失，以改善对抗损失并稳定训练；与VGG的损失函数类似，感知损失 L_{VGG} 被联合使用以进一步提高性能。因此，本申请的对抗生成网络模型的总目标损失函数为：

$$L_{total} = \min_G \left(\max_{D_1, D_2} \lambda_1 \sum_{i=1,2} L_{GAN}(G, D_i) + \lambda_2 \sum_{i=1,2} L_{FM}(G, D_i) + \lambda_3 L_{VGG}(G) \right);$$

其中， $\lambda_1=1, \lambda_2=10, \lambda_3=10$ ； L_{GAN} 表示对抗损失，主要控制图片输出的真假； L_{FM} 表示特征损失，主要控制输出图片在鉴别器/判别器中的特征一致； L_{VGG} 表示感知损失，主要控制VGG卷积网络子模型中特征的一致性。

[0082] 需要说明的是，对抗损失 L_{GAN} 、特征损失 L_{FM} 和感知损失 L_{VGG} 的具体获取过程属于本领域的现有技术。本领域技术人员可以参考现有的pix2pixHD模型等的具体结构和训练方式。此处不再对本申请的对抗生成网络模型的具体训练过程进行赘述。

[0083] 需要说明的是，特征损失 L_{FM} 和感知损失 L_{VGG} 这两个损失函数分别来自判别器和预训练的VGG-19模型的层特征映射：

$$L_{FM}(G, D_i) = \sum_{j=1}^T \frac{1}{C_j H_j W_j} \|f_{D_i}^{(j)}(I_r, I_g) - f_{D_i}^{(j)}(I_r, G(I_r))\|_1;$$

$$L_{VGG}(G) = \sum_{j=1}^{T'} \frac{1}{C_j H_j W_j} \|f_{VGG}^{(j)}(I_r) - f_{VGG}^{(j)}(I_g)\|_1;$$

其中， T 表示判别器 D_i 的网络层数， T' 表示VGG卷积网络子模型的网络层数， $f_{D_i}^{(j)}$ 表示判别器 D_i 的第 j 层特征， $f_{VGG}^{(j)}$ 表示VGG卷积网络子模型的第 j 层特征。 G 是指生成器(Generator)， D_i 是指第 i 个判别器(Discriminator)。生成器是一个神经网络模型； D 是采用与现有pix2pixHD中判别器类似的神经网络模型。 I_g 表示采集的原图像(即新视角场景渲染图像 I_r 所对应的立体相机图像)，该原图像作为真值； C 表示判别器 D_i 或VGG卷积网络子模型中相应的特征通道的数量， C_j 为判别器 D_i 或VGG卷积网络子模型中第 j 层网络的特征通道数。

[0084] 可以看出，本申请所提出的重建与补全方法采用的是一种基于数据驱动的自动驾驶场景合成框架，主要包含三维面元模型(如上述全局面元模型)的重建和图像补全网络

(即对抗生成网络模型)。本重建与补全方法利用由立体摄像机对自动驾驶场景进行拍摄而得到的立体相机图像,并获得立体相机图像

的深度图和语义图,基于该语义图对该深度图进行优化和滤波而得到优化后的深度图,可以有效地上述重建三维面元模型。

[0085] 一些实施例中,该三维面元模型采用自适应渲染机制,能够以合理的质量呈现新视角图像。

[0086] 本申请所提出的图像补全网络可以利用所提出的掩膜-图像匹配技术(即上述掩膜图像表征关联机制MPC)从场景上下文中找到最近似的信息。

[0087] 本申请所提出的CSD子模块可以增强合成图像序列(即上述完整新视角场景图像序列)的内容一致性。实验结果表明,本申请所提出的重建与补全方法在图像质量方面高于现有最先进的方法约5%~73%的精度,在视频质量方面高出45%~162%,在2D检测方面在指标mAP@0.5下高出约6%~24%,在定位精度方面高出38%~41%。其中,mAP@0.5表示是在IOU阈值0.5下计算的mAP。mAP(mean Average Precision)在机器学习中的目标检测领域是十分重要的衡量指标,用于衡量目标检测算法的性能。

[0088] 可以看出,本申请所提出的重建与补全方法可以用于重建自动驾驶的静态场景序列;其中,利用上述语义图中语义信息对深度图进行优化而获得优化后的深度图,利用优化后的深度图构建具有自适应渲染表示的全局面元模型,从该全局面元模型获取新视角场景渲染图像。

[0089] 可以看出,本申请所提出的重建与补全方法,为了保持生成图像(即完整新视角场景图像序列)的空间内容的连续性,本申请提供了一个掩膜图像匹配机制,该机制可以从采集的场景数据中匹配到与渲染图像(即上述新视角场景渲染图像)最相似的上下文近似信息。

[0090] 可以看出,通过CSD子模块将匹配到的上下文近似信息输入特征域(即前述的各CSD子模块)的归一化层中。其中,上述“将匹配到的上下文近似信息输入特征域的归一化层中”指的是经过卷积层和归一化操作,其只是表示一个信息流动过程。CSD子模块可以学习到上述渲染图像(即新视角场景渲染图像)与参考图像之间的先验相似信息(如上述调制仿射参数),并将其传播到对抗生成网络模型的整个网络。

[0091] 以上就是关于一种自动驾驶场景的重建与补全方法的一些说明。请参考图8,本申请一些实施例中还公开了一种自动驾驶场景的重建与补全系统,包括:

立体摄像机100,被配置为获取所述自动驾驶场景的立体相机图像;

语义深度生成模块200,被配置为生成所述立体相机图像的深度图和语义图;

语义深度增强模块300,被配置为利用所述语义图对所述深度图进行优化,以获得优化后的深度图;

面元模型构建模块400,被配置为利用所述优化后的深度图建立所述立体相机图像的当前帧的局部面元模型,对所述当前帧的局部面元模型和与所述当前帧之前的所有图像帧对应的全局面元模型相冲突的冲突信息进行剔除,利用剔除所述冲突信息后的所述当前帧的局部面元模型与所述全局面元模型进行融合,直到与所述自动驾驶场景的所有图像帧对应的全局面元模型构建完成;对所述与所述自动驾驶场景的所有图像帧对应的全局面元模型进行自适应渲染而得到自适应渲染后的所述全局面元模型,从所述自适应渲染后的

所述全局面元模型中获取与所述自动驾驶场景对应的新视角场景图像序列;其中,所述新视角场景图像序列包括多个新视角场景渲染图像;

对抗网络补全模块500,被配置为获得所述新视角场景渲染图像的渲染图像金字塔、与所述新视角场景渲染图像对应的参考图像的参考图像金字塔和与所述新视角场景渲染图像对应的掩码图像的掩码图像金字塔,将所述渲染图像金字塔、所述参考图像金字塔和所述掩码图像金字塔输入训练好的对抗生成网络模型,以利用所述对抗生成网络模型对所述新视角场景渲染图像中缺失信息的区域进行补全而生成完整新视角场景图像序列;

其中,所述完整新视角场景图像序列包括多个完整新视角场景图像;所述参考图像是来自于所述新视角场景渲染图像的邻近视角的所述立体相机图像,所述掩码图像是基于所述新视角场景渲染图像中缺失信息的区域而得到的。

[0092] 需要说明的是,语义深度增强模块300的作用是优化深度图中物体边界的深度和剔除动态物体(例如场景中行动的车和人),即,通过语义深度增强模块300改进了原深度图中不同物体边界的深度。其中,语义深度增强模块300根据语义一致区域的深度局部性原则来优化边界深度的具体内容已在前文进行过论述,该部分是本申请根据现有深度图的深度误差大的问题而提出来的解决方案。

[0093] 需要说明的是,本自动驾驶场景的重建与补全系统的具体执行流程和技术效果可以参考本自动驾驶场景的重建与补全方法的具体执行流程和技术效果。此处不再对其具体执行流程和技术效果进行赘述。

[0094] 以上就是关于一种自动驾驶场景的重建与补全系统的一些说明。本申请一些实施例中还公开了一种计算机可读存储介质,包括程序,该程序能够被处理器执行以实现如本文中任一项实施例的方法。

[0095] 本文参照了各种示范实施例进行说明。然而,本领域的技术人员将认识到,在不脱离本文范围的情况下,可以对示范性实施例做出改变和修正。例如,各种操作步骤以及用于执行操作步骤的组件,可以根据特定的应用或考虑与系统的操作相关联的任何数量的成本函数以不同的方式实现(例如一个或多个步骤可以被删除、修改或结合到其他步骤中)。

[0096] 在上述实施例中,可以全部或部分地通过软件、硬件、固件或者其任意组合来实现。另外,如本领域技术人员所理解的,本文的原理可以反映在计算机可读存储介质上的计算机程序产品中,该可读存储介质预装有计算机可读程序代码。任何有形的、非暂时性的计算机可读存储介质皆可被使用,包括磁存储设备(硬盘、软盘等)、光学存储设备(CD至ROM、DVD、Blu Ray盘等)、闪存和/或诸如此类。这些计算机程序指令可被加载到通用计算机、专用计算机或其他可编程数据处理设备上以形成机器,使得这些在计算机上或其他可编程数据处理装置上执行的指令可以生成实现指定的功能的装置。这些计算机程序指令也可以存储在计算机可读存储器中,该计算机可读存储器可以指示计算机或其他可编程数据处理设备以特定的方式运行,这样存储在计算机可读存储器中的指令就可以形成一件制造品,包括实现指定功能的实现装置。计算机程序指令也可以加载到计算机或其他可编程数据处理设备上,从而在计算机或其他可编程设备上执行一系列操作步骤以产生一个计算机实现的进程,使得在计算机或其他可编程设备上执行的指令可以提供用于实现指定功能的步骤。

[0097] 虽然在各种实施例中已经示出了本文的原理,但是许多特别适用于特定环境和操作要求的结构、布置、比例、元件、材料和部件的修改可以在不脱离本披露的原则和范围内

使用。以上修改和其他改变或修正将被包含在本文的范围之内。

[0098] 前述具体说明已参照各种实施例进行了描述。然而,本领域技术人员将认识到,可以在不脱离本披露的范围的情况下进行各种修正和改变。因此,对于本披露的考虑将是说明性的而非限制性的意义上的,并且所有这些修改都将被包含在其范围内。同样,有关于各种实施例的优点、其他优点和问题的解决方案已如上所述。然而,益处、优点、问题的解决方案以及任何能产生这些的要素,或使其变得更明确的解决方案都不应被解释为关键的、必需的或必要的。本文中所用的术语“包括”和其任何其他变体,皆属于非排他性包含,这样包括要素列表的过程、方法、文章或设备不仅包括这些要素,还包括未明确列出的或不属于该过程、方法、系统、文章或设备的其他要素。此外,本文中所使用的术语“耦合”和其任何其他变体都是指物理连接、电连接、磁连接、光连接、通信连接、功能连接和/或任何其他连接。

[0099] 具有本领域技术的人将认识到,在不脱离本发明的基本原理的情况下,可以对上述实施例的细节进行许多改变。因此,本发明的范围应仅由权利要求确定。



图 1

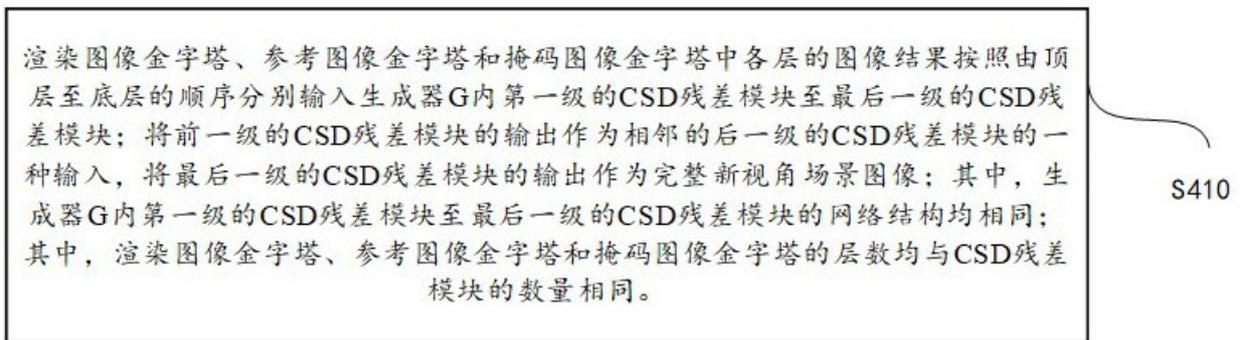


图 2

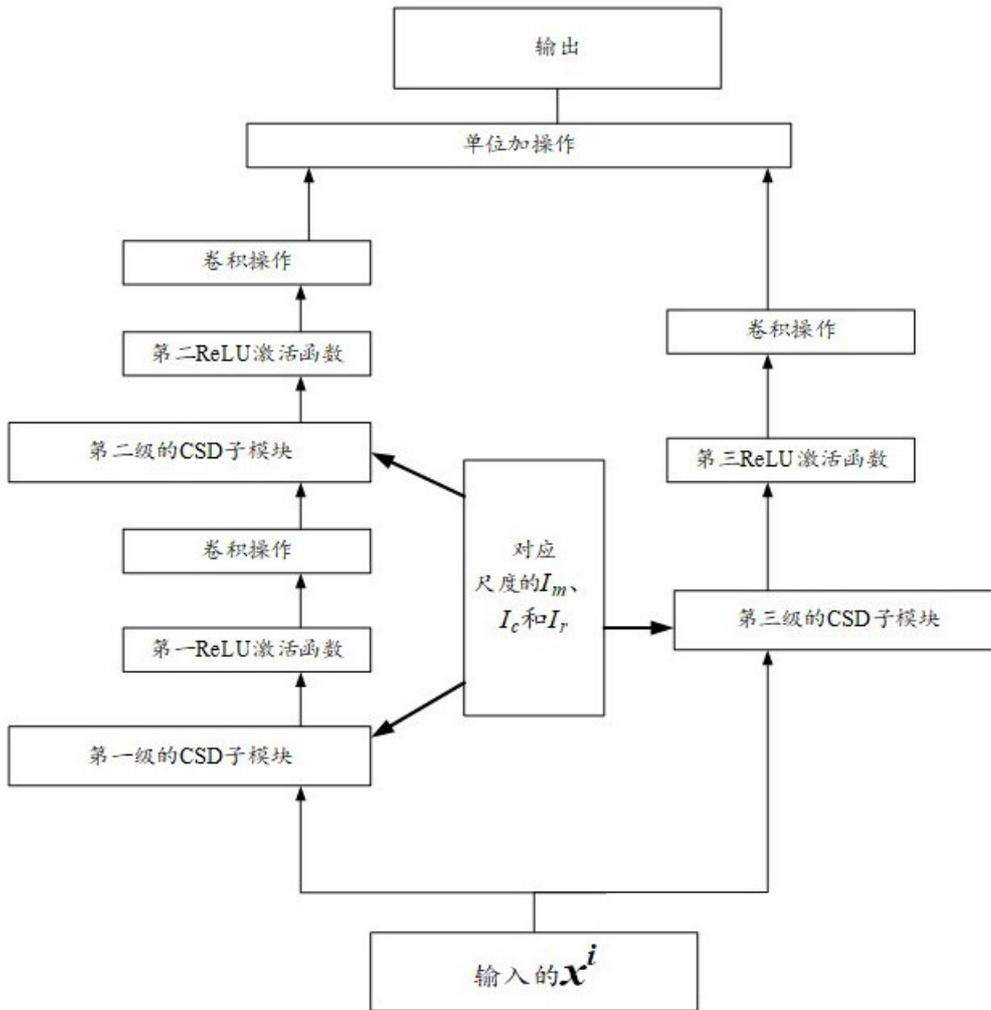


图 3

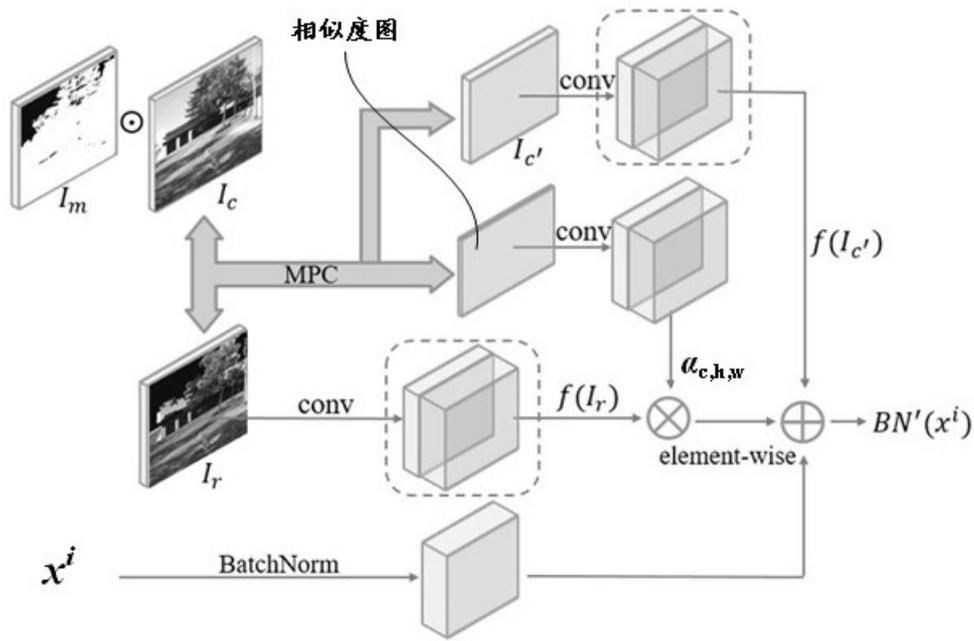


图 4

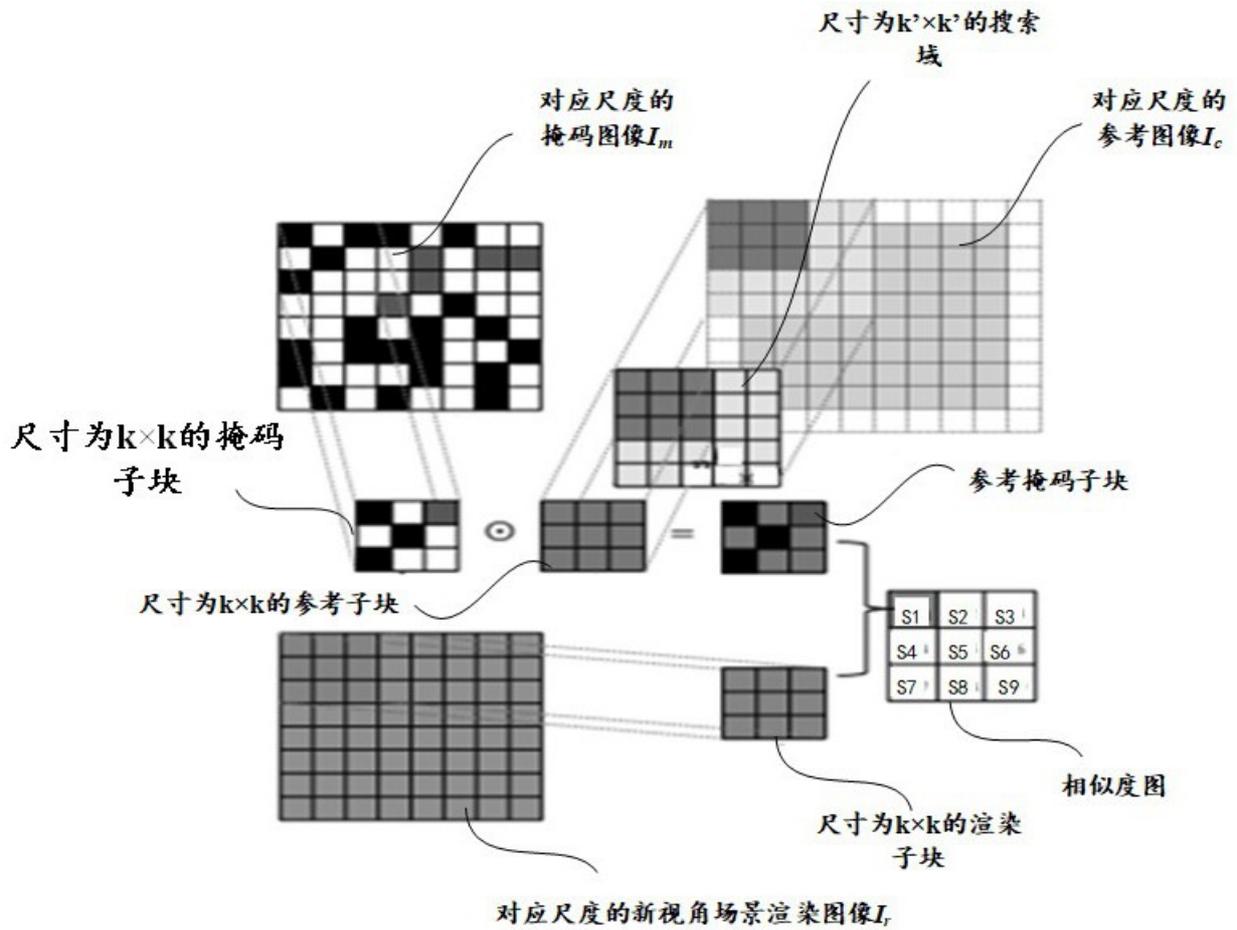


图 5

a1	a2	a3
a4	a5	a6
a7	a8	a9

图 6

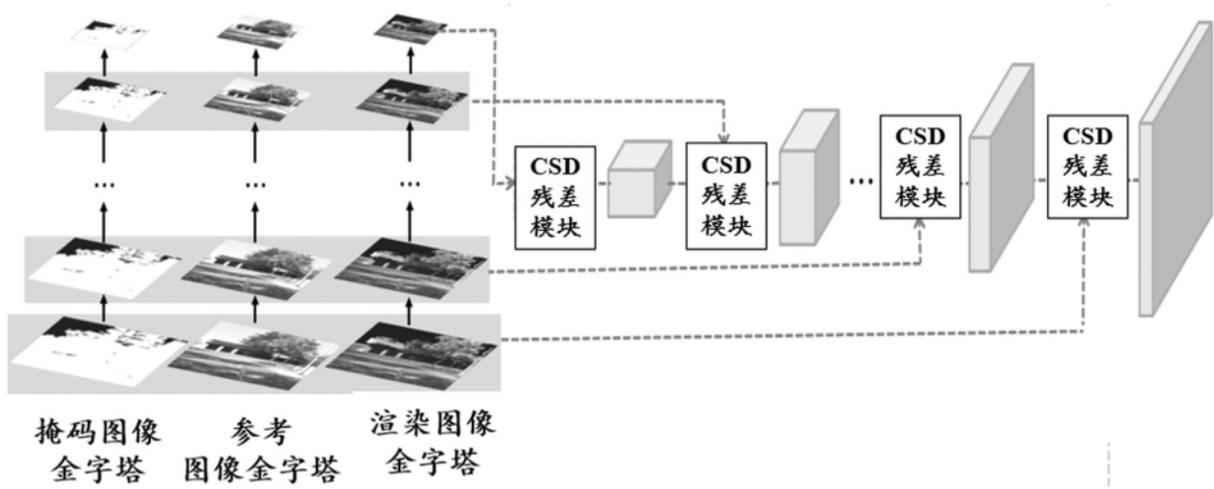


图 7

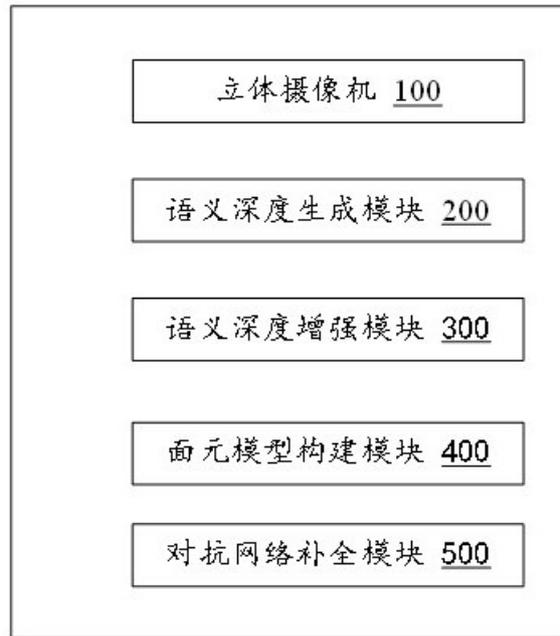


图 8