

视频生成

技术领域

[01]本说明书实施例属于计算机技术领域，特别的涉及视频生成方法及装置。

背景技术

[02]虚拟现实技术(Virtual Reality, VR)全景沉浸式展示方案利用先进的虚拟现实技术，使用户可以身临其境地感受虚拟场景。无论是参观名胜古迹、探索自然风光，还是参加会展活动、体验产品，用户都可以在虚拟世界中获得真实感受，仿佛亲临现场。这种身临其境的体验，为展示活动增添了更多趣味和吸引力。

[03]VR 全景沉浸式展示方案的全景视频通常需要专业的摄影团队进行拍摄，一次性采集高质量的场景图像数据，采集过程复杂。并且，需要额外进行训练优化和渲染轨迹选取。大型 AI 绘画工具能够基于单图和本文描述生成多帧的变化视角的全景视频，由于训练数据不包含多视角图像信息以及对应的语义标注，生成的全景视频中包含的视角信息不够丰富，且存在多视角一致性的情况。

发明内容

[04]本说明书实施例提供了一种视频生成方法、装置、存储介质及电子设备，其技术方案如下。

[05]第一方面，本说明书实施例提供了一种视频生成方法，包括：

获取目标场景的样本图像集合，所述样本图像集合包括多张关于所述目标场景的原始样本图像；

基于所述样本图像集合，生成所述目标场景的浏览轨迹和所述目标场景的神经辐射场模型；

基于所述浏览轨迹和所述神经辐射场模型，生成所述目标场景对应于所述浏览轨迹的渲染图像；

对所述渲染图像进行质量评估，得到各所述渲染图像的质量评估结果；若所述渲染图像的质量评估结果低于预期，获取对应于所述渲染图像的补充样本图像；将所述补充样本图像添加至所述样本图像集合，得到更新后的样本图像集合；基于更新后的样本图

像集合重新生成所述目标场景的神经辐射场模型，得到更新后的神经辐射场模型；基于所述浏览轨迹和更新后的神经辐射场模型，重新生成所述目标场景对应于所述浏览轨迹的渲染图像；重复本步骤，直至所述目标场景对应于所述浏览轨迹的渲染图像的质量评估结果均不低于预期；

基于所述目标场景对应于所述浏览轨迹的渲染图像，生成所述目标场景的全景视频。

[06]第二方面，本说明书实施例提供了一种视频生成装置，包括：

图像获取模块，获取目标场景的样本图像集合，所述样本图像集合包括多张关于所述目标场景的原始样本图像；

图像处理模块，基于所述样本图像集合，生成所述目标场景的浏览轨迹和所述目标场景的神经辐射场模型；

渲染模块，基于所述浏览轨迹和所述神经辐射场模型，生成所述目标场景对应于所述浏览轨迹的渲染图像；

质量评估及增量采集模块，对所述渲染图像进行质量评估，得到各所述渲染图像的质量评估结果；若所述渲染图像的质量评估结果低于预期，获取对应于所述渲染图像的补充样本图像；将所述补充样本图像添加至所述样本图像集合，得到更新后的样本图像集合；基于更新后的样本图像集合重新生成所述目标场景的神经辐射场模型，得到更新后的神经辐射场模型；基于所述浏览轨迹和更新后的神经辐射场模型，重新生成所述目标场景对应于所述浏览轨迹的渲染图像；重复本步骤，直至所述目标场景对应于所述浏览轨迹的渲染图像的质量评估结果均不低于预期；

视频生成模块，基于所述目标场景对应于所述浏览轨迹的渲染图像，生成所述目标场景的全景视频。

[07]第三方面，本说明书实施例提供了一种计算机可读存储介质，其上存储有计算机程序，所述计算机程序被处理器执行时实现上述第一方面所述的方法的步骤。

[08]第四方面，本说明书实施例提供了一种电子设备，包括：

一个或多个处理器，以及

与所述一个或多个处理器关联的存储器，所述存储器用于存储程序指令，所述程序指令在被所述一个或多个处理器读取执行时，执行上述第一方面所述的方法的步骤。

[09]本说明书的一个或者多个实施例提供的技术方案带来的有益效果如下。

[10]通过采集补充样本图像对目标场景的神经辐射场模型进行增量学习训练，不断补充视角信息来学习优化对目标场景的表征能力，基于较低的图像数据采集和训练成本，生成较高质量的全景视频。

附图说明

[11]为了更清楚地说明本说明书实施例中的技术方案，下面将对实施例中所需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的一些实施例，对于本领域普通技术人员来讲，在不付出创造性劳动的前提下，还可以根据这些附图获得其他的附图。

[12]图 1 为本说明书实施例提供一种视频生成方法的流程示意图。

[13]图 2 为图 1 中步骤 102 的流程示意图。

[14]图 3 为本说明书实施例提供一种视频生成方法的示意性应用场景示意图。

[15]图 4 为本说明书实施例提供一种视频生成装置的结构示意图。

[16]图 5 为示例性的电子设备的架构示意图。

具体实施方式

[17]为使本申请的目的、技术方案和优点更加清楚，下面将结合本说明书实施例及相应的附图，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，所描述的实施例仅是本申请一部分实施例，而不是全部的实施例。基于本申请中的实施例，本领域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，都属于本申请保护的范围。

[18]本说明书和权利要求书及上述附图中的术语“包括”和“具有”以及它们任何变形，意图在于覆盖不排他的包含。例如包含了一系列步骤或单元的过程、方法、系统、产品或设备没有限定于已列出的步骤或单元，而是可选地还包括没有列出的步骤或单元，或可选地还包括对于这些过程、方法、产品或设备固有的其他步骤或单元。取决于语境，如在此所适用的词语“如果”可以被解释成为“在……时”或“当……时”或“响应于确定”或“响应于检测”。

[19]针对传统全景视频的合成通常需要一次性采集高质量的场景图像数据和额外的训练优化及渲染轨迹选取，大型 AI 绘画工具基于单图和本文描述生成的全景视频中包含的

视角信息不足、多视角之间的一致性不足等技术问题，本说明书实施例，提出一种采集补充样本图像对目标场景的神经辐射场模型进行增量学习训练的视频生成方法，通过采集补充样本图像补充视角信息来学习优化对目标场景的表征能力，实现较低的图像数据采集和训练成本下，生成较高质量的全景视频。

[20]以下结合附图，详细说明本说明书各实施例提供的技术方案。

[21]图 1 为本说明书实施例提供的一种视频生成方法的流程示意图。参见图 1，该方法可以包括以下步骤 102 至 110。

[22]102、获取目标场景的样本图像集合，所述样本图像集合包括多张关于所述目标场景的原始样本图像。

[23]目标场景可以是任意需要展示的对象，例如为街道、商场、公路、商店、广场等场景，也可以是鞋子、桌子、汽车等物体。在沉浸式全景显示应用场景下中，需要目标场景的三维模型，可以基于目标场景的三维模型确定不同视角下看到的模型内容，进而基于该模型内容渲染出不同视角下的图像，从而生成目标场景的全景视频。在对目标场景进行三维重建之前，获取目标场景的样本图像集合，样本图像集合包括多张关于目标场景的原始样本图像，以便于基于原始样本图像进行目标场景的三维重建。在一些实施例中，可以对处于真实世界中的目标场景从不同拍摄角度进行拍摄，得到关于该目标场景的多张原始样本图像或者得到该目标场景的视频，从视频中提取关于该目标场景的多张原始样本图像。另一些实施例中，可以在目标场景内部或者外部环绕一周的方式进行拍摄，以提高重建出目标场景的三维模型的准确度，进而提高基于三维模型的图像渲染质量。不同的原始样本图像可以对应不同的相机位姿，相机位姿包括拍摄设备在拍摄图像时的位置和姿态。其中，拍摄设备可以是能够获取目标场景图像数据的任何设备，例如可以是但不限于相机、具有拍摄功能的手机、平板电脑、可穿戴设备等。例如，可以采用普及率较高的具有拍照功能的手机设备，实现 1-2 分钟内对目标场景的视频拍摄，从视频中提取关于目标场景的多个原始样本图像，原始样本图像可以是 RGB 格式的图像数据。

[24]104、基于所述样本图像集合，生成所述目标场景的浏览轨迹和所述目标场景的神经辐射场模型。

[25]基于样本图像集合可以生成目标场景的浏览轨迹和神经辐射场模型。目标场景的表征模型可以是一个全连接神经网络，全连接神经网络又称多层感知器（Multi-Layer

Perceptron, MLP)。本说明书的一个实施例中,目标场景可以表征为一个神经辐射场(Neural Radiance Field, NERF)。神经辐射场模型能够基于输入的空间点的空间坐标和视角信息,预测输出该空间点的体积密度和颜色信息。其中,神经辐射场模型可以表达为:

$$\sigma, c = F(d, X)$$

[26]其中, $X = (x, y, z)$, X 记为空间点的空间坐标 (x, y, z) ; $d = (\theta, \phi)$ 记为空间点的视角信息 (θ, ϕ) , θ 为方位角, ϕ 为仰角; $c = (r, g, b)$, c 记为空间点的颜色信息 (r, g, b) , r 是指红色(Red, R), g 是指绿色(Green, G), b 是指蓝色(Blue, B); σ 记为空间点的体积密度。在获取到目标场景的样本图像集合后,分别计算各原始样本图像对应的相机位姿,根据每张原始样本图像对应的相机位姿和相机内参等数据确定相机在拍摄每张原始样本图像时发射出来的多条视线以及每条视线的视角信息。在每条视线上进行空间点采样,得到多个空间点。应理解,从同一条视线上采样得到的空间点的视角信息均是该视线的视角信息。在得到多个空间点之后,利用多个空间点的空间坐标及其视角信息进行基于神经网络的三维重建。该过程可以是分批多次执行的过程,最终可得到目标场景的神经辐射场模型。

[27]目标场景的浏览轨迹通常是指相机或者观察者在目标场景中移动的路径,该路径记录相机或观察者在观察目标场景时的位置和姿态变化,浏览轨迹可以是直线、曲线或者复杂不规则的路径。浏览路径可以表示为多个离散的浏览点组成的时间序列,也可以表示为与时间相关的浏览点的函数。如图2所示,在一些实施例中,步骤102可以包括2022至2026。

[28]2022、对所述样本图像集合进行特征点匹配和位姿估计。通过特征点匹配,能在两个或多个图像之间寻找具有相似特征的点,并建立它们之间的对应关系,从而在多个原始样本图像之间找到相同的物体或场景。位姿估计能够通过已知的特征点匹配,计算出各原始样本图像的相机位姿。在一些实施例中,可以采用 SLAM(simultaneous localization and mapping, 即时定位与地图构建)算法来更加准确计算每张样本图像对应的相机位姿。具体的,SLAM 算法在计算相机位姿时,首先提取每张样本图像中的特征点,接着建立相邻两张样本图像的特征点之间的匹配关系,根据相邻两张样本图像的特征点之间的匹配关系计算相邻两张样本图像之间的相对相机位姿,根据两两样本图像质检的相对相机位姿计算每张样本图像对应的相机位姿。在一些实施例中,可以使用神经网络模型对样本图像集合中的各样本图像进行处理,提取各样本图像中的特征点。特征点可以是角点

(如 Harris 角点、Shi-Tomasi 角点)、边缘、半点、SIFT 特征点等具有显著性质的点。对样本图像集中的两个或多个样本图像,分别使用神经网络提取的特征点进行匹配,以确定特征点之间的对应关系。匹配算法可以为基于距离的匹配(如最近邻匹配、最近邻距离比匹配)、基于集合关系的匹配(如基于 RANSAC 的鲁棒匹配)等。通过特征点匹配的对应关系,使用开源的计算机视觉软件包 Colmap 进行位姿估计,从而得到原始样本图像与其位姿的数据对。在一些实施例中,还可以使用全局位姿优化算法对位姿估计的结果进行优化,以提高相机位姿估计和目标场景三维重建的精度和稳定性。

[29]2024、基于所述特征点匹配和所述位姿估计的结果,生成所述目标场景的粗重建表示,基于所述目标场景的粗重建表示,生成所述目标场景的浏览轨迹。在一些实施例中,目标场景可以被粗重建表示为基于稀疏三维特征点的几何模型。例如,可以基于所述特征点匹配和所述位姿估计的结果,提取所述目标场景的稀疏三维点云数据,基于所述稀疏三维点云数据,生成所述目标场景的粗重建模型;基于所述目标场景的粗重建模型,生成所述目标场景的浏览轨迹。稀疏三维点云数据是从目标场景中提取出的一部分离散的、具有代表性的三维点,这些点可以提供足够的信息来描述场景的几何形状和结构。通过提取稀疏的关键点进行三角化计算,可以获得较小的点云数据集,从而减少计算量和存储需求。

[30]目标场景的浏览轨迹可以基于样本图像集合对应的采集轨迹进行处理得到。基于对样本图像集合的位姿估计,得到样本图像集合中各样本图像的相机位姿,每个相机位姿即对应一个样本图像的采集点。样本图像集合中各样本图像的相机位姿的时间序列,则构成了样本图像集合的采集轨迹。对于采用视频拍摄方式采集的原始样本图像,视频拍摄时的拍摄轨迹通常是符合用户观察习惯的,样本图像集合的采集轨迹即用户拍摄视频时的拍摄轨迹。在一些实施例中,对采集轨迹进行插值处理,可以得到符合用户观察习惯的浏览轨迹。其中,插值处理的方法可以是线性插值、样本插值、贝塞尔曲线插值等。另一些实施例中,可以计算出采集轨迹的中心和朝向,生成圆周轨迹作为浏览轨迹。其中,采集轨迹的中心点用于确定所生成浏览轨迹的圆心,采集轨迹的朝向用于确定生成的浏览轨迹的起始角度。

[31]2026、基于所述特征点匹配和所述位姿估计的结果,生成所述目标场景的神经辐射场模型。本说明书实施例中,基于特征点匹配和位姿估计,能够得到样本图像与相机位姿的数据对,将样本图像集合中所有的样本图像与其相机位姿的数据对用于目标场景的神经辐射场的训练,能够得到目标场景的神经辐射场模型。

[32]106、基于所述浏览轨迹和所述神经辐射场模型，生成所述目标场景对应于所述浏览轨迹的渲染图像。

[33]在一些实施例中，浏览点包括浏览位置和浏览视角，浏览位置用于记录相机或者观察者在观察目标场景时的位置，其可以表示为目标场景中的一个空间点坐标，浏览视角用于记录相机或者观察者在观察目标场景时的姿态变化，其可以表示为空间点的视角信息。将浏览位置所表示的空间点坐标和浏览视角所表示的视角信息输入到目标场景的隐式三维表征模型进行预测，可以得到与空间坐标的体积密度和颜色信息，从而渲染出目标场景对应浏览点的渲染图像。

[34]108、对所述渲染图像进行质量评估，得到各所述渲染图像的质量评估结果；若所述渲染图像的质量评估结果低于预期，获取对应于所述渲染图像的补充样本图像；将所述补充样本图像添加至所述样本图像集合，得到更新后的样本图像集合；基于更新后的样本图像集合重新生成所述目标场景的隐式三维表征模型，得到更新后的隐式三维表征模型；基于所述浏览轨迹和更新后的隐式三维表征模型，重新生成所述目标场景对应于所述浏览轨迹的渲染图像；重复本步骤，直至所述目标场景对应于所述浏览轨迹的渲染图像的质量评估结果均不低于预期。

[35]本说明书实施例中，对步骤 104 获取的目标场景的原始样本图像的质量和数量要求不做限制，以提高获取目标场景的样本图像集合的效率，降低样本采集成本。因此，样本图像集合中原始样本图像采集的数量和视角有限，基于原始样本图像训练得到的神经辐射场模型通常是视角不全的低质量神经辐射场模型，该神经辐射场模型的不同视角下的渲染图像质量不均衡。因此，可能会出现浏览轨迹中的部分浏览点的渲染图像较低的情况。

[36]在一些实施例中，用于对渲染图像进行质量评估的指标可以包括峰值信噪比（PSNR）、结构相似度（SSIM）、图像锐度（Sharpness）等。例如可以将用于质量评估的指标进行加权求和，并归一化处理后得到渲染图像的质量评估结果。将质量评估结果预设的质量评估指标值进行比较：如果质量评估结果大于或等于预设的质量评估指标值，表示渲染图像的质量评估结果不低于预期；如果质量评估结果小于预设的质量评估指标值，则表示渲染图像的质量评估结果低于预期。渲染图像的质量评估结果低于预期，则意味着需要对步骤 102 生成的目标场景的神经辐射场模型进行增量学习。

[37]在一些实施例中，若所述渲染图像的质量评估结果低于预期，获取对应于所述渲染图像的补充样本图像，包括：获取质量评估结果低于预期的渲染图像对应的浏览点；基

于获取的浏览点，发起所述目标场景对应于所述浏览点的样本图像补充采集，以获取所述渲染图像的补充样本图像。例如，可以在目标场景的实时渲染平台中标出浏览点的浏览位置，或者输出质量评估结果较差的渲染图像，从而指示图像补充采集时需采用的视角，使得补充采集的补充样本图像的相机位置和相机姿态，与渲染图像的浏览位置和浏览视角相同或相近。

[38]需要说明的是，步骤 106 和步骤 108 可以交替进行，也可以先后进行。即，可以每生成一张渲染图像即对生成的渲染图像进行质量评估，根据质量评估的结果判断是否需要增量学习。也可以在步骤 106 中生成浏览轨迹的所有渲染视图后，再由步骤 108 对生成的各渲染视图进行质量评估，根据质量评估的结果判断是否需要增量学习。本说明书实施例中采用后者，步骤 108 中将所有质量评估结果不符合预期的渲染图像对应的各浏览点均进行样本图像补充采集。本说明书实施例对各浏览点的补充样本图像数量不做限制，可以采集一张也可以采集多张。具体来说，还可以根据多个浏览点生成辅助补充采集的指导轨迹，用于指示视频拍摄的采集轨迹，根据该采集轨迹拍摄的视频可以包括质量评估结果不符合预期的各渲染图像对应的浏览点的补充样本图像。

[39]本实施例中，将所述补充样本图像添加至所述样本图像集合，得到更新后的样本图像集合。更新后的图像样本集合中包括了原本渲染质量较差的视角对应的样本图像。对更新后的样本图像集合采用步骤 2022 类似的手段进行特征点匹配和位姿估计，得到各样本图像的特征点和对应的相机位姿。基于特征匹配和位姿估计的结果，利用更新后的样本图像集再次训练目标场景的神经辐射场模型。其中，对神经辐射场模型的再次训练可以在之前训练得到的神经辐射场模型基础上增加训练数据的继续训练，也可以是头从开始训练新的神经辐射场模型。因为补充样本图像是针对渲染质量低于预期的渲染图像的浏览点采集的，基于更新后的样本图像集训练对神经辐射场模型的增量训练能够更加有针对性和有效率得弥补表征缺陷，更新后的神经辐射场模型在对应浏览点下的表征效果得到提升。如此经过对本步骤的重复，可以不断修补目标场景的神经辐射场模型在不同视角下的表征缺陷。最终，目标场景对应于所述浏览轨迹的所有渲染图像的质量评估结果均能够达到预期。最终，能够得到一个目标场景的高质量的神经辐射场模型。

[40]110、基于所述目标场景对应于所述浏览轨迹的渲染图像，生成所述目标场景的全景视频。

[41]在一些实施例中，使用图像拼接算法将对应于浏览轨迹的所有渲染图像拼接成目标场景的全景图像。图像拼接算法可以根据图像中的重叠区域和相机参数来自动对图像堆

砌和融合。例如，特征匹配、图像配准和图像融合等算法。在一些实施例中，使用视频处理库将目标场景的全景图像转换成全景视频。在另一些实施例中，还可以对生成的全景视频进行二次编辑，比如风格化处理、分辨率调优处理等，得到目标场景的高质量全景视频。

[42]本说明书实施例提供的视频服务方法，能够针对渲染图像的质量评估结果针对性的补充采集样本图像添加到样本图像集中对目标场景的神经辐射场进行增量学习，提高全景视频的生成效率和质量。

[43]为了便于理解，下面介绍集中场景实施例对本申请实施例提供的视频生成方法进行详细说明。

[44]场景实施例 1。

[45]在电商场景中，可以使用本申请实施例提供的视频生成方法制作商品或者店铺的沉浸式全景视频。参见图 3，实际应用中，商家可以使用诸如手机、平板电脑、可穿戴式智能设备、智能家居设备、摄像机等拍摄装置 302 环绕目标场景 301（可以是商品或者商铺对象等）的方式拍摄一段视频，商家可在客户端 303 提供的人机交互界面（例如为 Web 界面）上发起图片或视频上传操作，以将包括多张目标场景的图片或者视频上传至执行视频生成方法的服务端 304。其中，客户端 303 可以是独立于拍摄装置 302 的，例如，安装有视频服务客户端 app 或者小程序的手机、平板电脑、计算机等设备。客户端 303 也可以和拍摄装置 302 集成在同一设备中，例如具有拍摄功能并且安装有客户端 app 或者小程序的手机、平板电脑等。服务端 304 为单个服务器或多个服务器组成的分布式服务器集群，进一步可选的，服务端 304 可以为云端服务器。服务端 304 基于多张目标场景的图片或者视频，运行本实施例的视频生成方法，生成目标场景的全景视频并发送给客户端，以供客户端展示。

[46]场景实施例 2。

[47]在 AI（Artificial intelligent，人工智能）商场领域，为了给消费者的线上购物提供 3D 场景化的沉浸式体验，可以为店铺创建三维模型。为此，设计人员可以用手机环绕真实场景中的店铺（目标场景）拍摄一段视频，并将该视频上传至执行视频生成方法的装置，视频生成装置基于视频中的多张店铺相关的图片生成店铺的全景视频，以完成店铺的 AI 场景创建任务。

[48]此外，与上述图 1 所示的视频生成方法相对应地，本说明书实施例还提供一种视频

生成装置。图 4 是本说明书实施例提供的一种视频生成装置 400 的结构示意图。包括：

[49]图像获取模块 401，获取目标场景的样本图像集合，所述样本图像集合包括多张关于所述目标场景的原始样本图像。在对目标场景进行三维重建之前，获取目标场景的样本图像集合，样本图像集合包括多张关于目标场景的原始样本图像，以便于基于原始样本图像进行目标场景的三维重建。

[50]图像处理模块 402，基于所述样本图像集合，生成所述目标场景的浏览轨迹和所述目标场景的神经辐射场模型。神经辐射场模型能够基于输入的空间点的空间坐标和视角信息，预测输出该空间点的体积密度和颜色信息。目标场景的浏览轨迹通常是指相机或者观察者在目标场景中移动的路径，该路径记录相机或观察者在观察目标场景时的位置和姿态变化，浏览轨迹可以是直线、曲线或者复杂不规则的路径。浏览路径可以表示为多个离散的浏览点组成的时间序列，也可以表示为与时间相关的浏览点的函数。

[51]在一些实施例中，图像处理模块 402 包括特征匹配及位姿估计单元 4021、轨迹生成单元 4022、模型训练单元 4023。

[52]特征匹配及位姿估计单元 4021，对所述样本图像集合进行特征点匹配和位姿估计。通过特征点匹配，能在两个或多个图像之间寻找具有相似特征的点，并建立它们之间的对应关系，从而在多个原始样本图像之间找到相同的物体或场景。位姿估计能够通过已知的特征点匹配，计算出各原始样本图像的相机位姿，从而得到原始样本图像与其位姿的数据对。

[53]轨迹生成单元 4022，基于所述特征点匹配和所述位姿估计的结果，生成所述目标场景的粗重建表示，基于所述目标场景的粗重建表示，生成所述目标场景的浏览轨迹。在本说明书的一个实施例中，轨迹生成单元，包括三维点云处理子单元和轨迹生成子单元。

[54]三维点云处理子单元，所述基于所述特征点匹配和所述位姿估计的结果，提取所述目标场景的稀疏三维点云数据，基于所述稀疏三维点云数据，生成所述目标场景的粗重建模型。稀疏三维点云数据是从目标场景中提取出的一部分离散的、具有代表性的三维点，这些点可以提供足够的信息来描述场景的几何形状和结构。通过提取稀疏的关键点进行三角化计算，可以获得较小的点云数据集，从而减少计算量和存储需求。

[55]轨迹生成子单元，基于所述目标场景的粗重建模型，生成所述目标场景的浏览轨迹。

[56]目标场景的浏览轨迹可以基于样本图像集合对应的采集轨迹进行处理得到。基于对样本图像集合的位姿估计，得到样本图像集合中各样本图像的相机位姿，每个相机位姿

即对应一个样本图像的采集点。样本图像集合中各样本图相的相机位姿的时间序列，则构成了样本图像集合的采集轨迹。对于采用视频拍摄方式采集的原始样本图像，视频拍摄时的拍摄轨迹通常是符合用户观察习惯的，样本图像集合的采集轨迹即用户拍摄视频时的拍摄轨迹。在一些实施例中，轨迹生成子单元对采集轨迹进行插值处理，可以得到符合用户观察习惯的浏览轨迹。其中，插值处理的方法可以是线性插值、样本插值、贝塞尔曲线插值等。另一些实施例中，轨迹生成子单元可以计算出采集轨迹的中心和朝向，生成圆周轨迹作为浏览轨迹。其中，采集轨迹的中心点用于确定所生成浏览轨迹的圆心，采集轨迹的朝向用于确定生成的浏览轨迹的起始角度。

[57]模型训练单元 4023，基于所述特征点匹配和所述位姿估计的结果，生成所述目标场景的神经辐射场模型。基于特征点匹配和位姿估计，能够得到样本图像与相机位姿的数据对，将样本图像集合中所有的样本图像与其相机位姿的数据对用于目标场景的神经辐射场的训练，能够得到目标场景的神经辐射场模型。

[58]渲染模块 403，基于所述浏览轨迹和所述神经辐射场模型，生成所述目标场景对应于所述浏览轨迹的渲染图像。在一些实施例中，浏览点包括浏览位置和浏览视角，浏览位置用于记录相机或者观察者在观察目标场景时的位置，其可以表示为目标场景中的一个空间点坐标，浏览视角用于记录相机或者观察者在观察目标场景时的姿态变化，其可以表示为空间点的视角信息。将浏览位置所表示的空间点坐标和浏览视角所表示的视角信息输入到目标场景的隐式三维表征模型进行预测，可以得到与空间坐标的体积密度和颜色信息，从而渲染出目标场景对应应该浏览点的渲染图像。

[59]本说明书实施例中，图像获取模块 401 获取的目标场景的原始样本图像的质量和数量要求不做限制，以提高获取目标场景的样本图像集合的效率，降低样本采集成本。因此，样本图像集合中原始样本图像采集的数量和视角有限，基于原始样本图像训练得到的神经辐射场模型通常是视角不全的低质量神经辐射场模型，该神经辐射场模型的不同视角下的渲染图像质量不均衡。因此，可能会出现浏览轨迹中的部分浏览点的渲染图像较低的情况。

[60]质量评估及补充采集模块 404，对所述渲染图像进行质量评估，得到各所述渲染图像的质量评估结果；若所述渲染图像的质量评估结果低于预期，获取对应于所述渲染图像的补充样本图像；将所述补充样本图像添加至所述样本图像集合，得到更新后的样本图像集合；基于更新后的样本图像集合重新生成所述目标场景的神经辐射场模型，得到更新后的神经辐射场模型；基于所述浏览轨迹和更新后的神经辐射场模型，重新生成所

述目标场景对应于所述浏览轨迹的渲染图像；重复本步骤，直至所述目标场景对应于所述浏览轨迹的渲染图像的质量评估结果均不低于预期。

[61]对于质量评估不符合预期的渲染图像，质量评估及补充采集模块 404 将所述补充样本图像添加至所述样本图像集合，得到更新后的样本图像集合。更新后的图像样本集合中包括了原本渲染质量较差的视角对应的样本图像。特征匹配及位姿估计单元 4021 对更新后的样本图像集合进行特征点匹配和位姿估计，得到各样本图像的特征点和对应的相机位姿。基于特征匹配和位姿估计的结果，模型训练单元 4023 利用更新后的样本图像集再次训练目标场景的神经辐射场模型。因为补充样本图像是针对渲染质量低于预期的渲染图像的浏览点采集的，基于更新后的样本图像集训练对神经辐射场模型的增量训练能够更加有针对性和有效率得弥补表征缺陷，更新后的神经辐射场模型在对应浏览点下的表征效果得到提升。如此经过对循环，可以不断修补目标场景的神经辐射场模型在不同视角下的表征缺陷。最终，渲染模块 403 生成的目标场景对应于所述浏览轨迹的所有渲染图像的质量评估结果均能够达到预期。最终，能够得到一个目标场景的高质量的神经辐射场模型。

[62]视频生成模块 405，基于所述目标场景对应于所述浏览轨迹的渲染图像，生成所述目标场景的全景视频。

[63]本说明书实施例的视频生成装置，能够针对渲染图像的质量评估结果针对性的补充采集样本图像添加到样本图像集合中对目标场景的神经辐射场进行增量学习，提高全景视频的生成效率和质量。

[64]在一些实施例中，视频生成装置还包括：对所述全景视频进行风格化处理的风格化处理模块。

[65]在一些实施例中，视频生成装置还包括：对所述全景视频进行超分辨调优处理的调优模块。

[66]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相互参见即可，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，对于装置实施例而言，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。

[67]另外，本申请实施例还提供另一种计算机可读存储介质，其上存储有计算机程序，该计算机程序被处理器执行时实现前述实施例方法中任一项所述的方法的步骤。上述装

置的各组成模块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，可以存储在上述计算机可读存储介质中。

[68]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。当使用软件实现时，可以全部或部分地以计算机程序产品的形式实现。所述计算机程序产品包括一个或多个计算机指令。在计算机上加载和执行所述计算机程序指令时，全部或部分地产生按照本说明书实施例所述的流程或功能。所述计算机可以是通用计算机、专用计算机、计算机网络、或者其他可编程装置。所述计算机指令可以存储在计算机可读存储介质中，或者通过所述计算机可读存储介质进行传输。所述计算机指令可以从一个网站站点、计算机、服务器或数据中心通过有线（例如同轴电缆、光纤、数字用户线（Digital Subscriber Line，DSL））或无线（例如红外、无线、微波等）方式向另一个网站站点、计算机、服务器或数据中心进行传输。所述计算机可读存储介质可以是计算机能够存取的任何可用介质或者是包含一个或多个可用介质集成的服务器、数据中心等数据存储设备。所述可用介质可以是磁性介质，（例如，软盘、硬盘、磁带）、光介质（例如，数字多功能光盘（Digital Versatile Disc，DVD））、或者半导体介质（例如，固态硬盘（Solid State Disk，SSD））等。

[69]本申请实施例还提供一种电子设备，包括：

一个或多个处理器，以及

与上述一个或多个处理器关联的存储器，所述存储器用于存储程序指令，所述程序指令在被上述一个或多个处理器读取执行时，执行前述方法实施例中任一项所述的方法的步骤。

[70]本申请还提供了一种计算机程序产品，包括计算机程序，该计算机程序在被处理器执行时实现前述方法实施例中任一项所述的方法的步骤。

[71]其中，图5示例性的示出了电子设备500的架构，具体可以包括：处理器510，磁盘驱动器520，输入/输出接口530，网络接口540，以及存储器550。上述处理器510、磁盘驱动器520、输入/输出接口530、网络接口540，与存储器550之间可以通过通信总线560进行通信连接。

[72]其中，处理器510可以采用通用的CPU、微处理器、应用专用集成电路（Application Specific Integrated Circuit，ASIC）、或者一个或多个集成电路等方式实现，用于执行相关程序，以实现本申请所提供的技术方案。

[73]存储器 550 可以采用 ROM (Read Only Memory, 只读存储器)、RAM (Read Access Memory, 随机存取存储器)、静态存储器、动态存储设备等形式实现。存储器 550 可以存储用于控制电子设备 500 运行的操作系统 551, 用于控制电子设备 500 的低级别操作的基本输入输出系统 (BIOS) 552。另外, 还可以存储网页浏览器 553, 数据存储管理系统 554 等。总之, 在通过软件或固件来实现本申请所提供的技术方案时, 相关的程序代码保存在存储器 550 中, 并由处理器 510 来调用执行。

[74]输入/输出接口 530 用于连接输入/输出模块, 以实现信息输入及输出。输入/输出模块可以作为组件配置在设备中 (图中未示出), 也可以外接于设备以提供相应功能。其中输入设备可以包括键盘、鼠标、触摸屏、麦克风、各类传感器等, 输出设备可以包括显示器、扬声器、振动器、提示灯等。

[75]网络接口 540 用于连接通信模块 (图中未示出), 以实现设备与其他设备的通信交互。其中, 通信模块可以通过有线方式 (例如 USB、网线等) 实现通信, 也可以通过无线方式 (例如移动网络、WIFI、蓝牙等) 实现通信。

[76]总线 560 包括一通路, 在设备的各个组件 (例如处理器 510、磁盘驱动器 520、输入/输出接口 530、网络接口 540, 与存储器 550) 之间传输信息。

[77]需要说明的是, 尽管上述设备仅示出了处理器 510、磁盘驱动器 520、输入/输出接口 530、网络接口 540, 与存储器 550、总线 560 等, 但是在具体实施过程中, 该设备还可以包括实现正常运行所必须的其他组件。此外, 本领域的技术人员可以理解的是, 上述设备中可以仅包含实现本申请方法所必需的组件, 而不必包含图中所示的全部组件。

[78]本领域普通技术人员可以理解实现上述实施例方法中的全部或部分流程, 可以通过计算机程序来指令相关的硬件来完成, 该程序可存储于计算机可读取存储介质中, 该程序在执行时, 可包括如上述各方法的实施例的流程。而前述的存储介质包括: ROM、RAM、磁碟或者光盘等各种可存储程序代码的介质。在不冲突的情况下, 本实施例和实施方案中的技术特征可以任意组合。

[79]以上所述的实施例仅仅是本说明书的优选实施例方式进行描述, 并非对本说明书的范围进行限定, 在不脱离本说明书的设计精神的前提下, 本领域普通技术人员对本说明书的技术方案作出的各种变形及改进, 均应落入本说明书的权利要求书确定的保护范围内。

权利要求书

1. 视频生成服务方法，包括：

获取目标场景的样本图像集合，所述样本图像集合包括多张关于所述目标场景的原始样本图像；

基于所述样本图像集合，生成所述目标场景的浏览轨迹和所述目标场景的神经辐射场模型；

基于所述浏览轨迹和所述神经辐射场模型，生成所述目标场景对应于所述浏览轨迹的渲染图像；

对所述渲染图像进行质量评估，得到各所述渲染图像的质量评估结果；若所述渲染图像的质量评估结果低于预期，获取对应于所述渲染图像的补充样本图像；将所述补充样本图像添加至所述样本图像集合，得到更新后的样本图像集合；基于更新后的样本图像集合重新生成所述目标场景的神经辐射场模型，得到更新后的神经辐射场模型；基于所述浏览轨迹和更新后的神经辐射场模型，重新生成所述目标场景对应于所述浏览轨迹的渲染图像；重复本步骤，直至所述目标场景对应于所述浏览轨迹的渲染图像的质量评估结果均不低于预期；

基于所述目标场景对应于所述浏览轨迹的渲染图像，生成所述目标场景的全景视频。

2. 根据权利要求1所述的方法，所述基于所述样本图像集合生成所述目标场景的浏览轨迹和所述目标场景的神经辐射场模型，包括：

对所述样本图像集合进行特征点匹配和位姿估计；

基于所述特征点匹配和所述位姿估计的结果，生成所述目标场景的粗重建表示，基于所述目标场景的粗重建表示，生成所述目标场景的浏览轨迹；

基于所述特征点匹配和所述位姿估计的结果，生成所述目标场景的神经辐射场模型。

3. 根据权利要求2所述的方法，所述基于所述特征点匹配和所述位姿估计的结果，生成所述目标场景的粗重建表示，基于所述目标场景的粗重建表示，生成所述目标场景的浏览轨迹，包括：

基于所述特征点匹配和所述位姿估计的结果，提取所述目标场景的稀疏三维点云数据，基于所述稀疏三维点云数据，生成所述目标场景的粗重建模型；

基于所述目标场景的粗重建模型，生成所述目标场景的浏览轨迹。

4. 根据权利要求2所述的方法，所述基于所述目标场景的粗重建表示，生成所述目标场景的浏览轨迹，包括：

基于所述目标场景的粗重建表示，生成所述样本图像集合的采集轨迹；所述采集轨

迹包括多个采集点;

对所述采集轨迹进行插值处理,得到所述浏览轨迹。

5.根据权利要求2所述的方法,所述基于所述目标场景的粗重建表示,生成所述目标场景的浏览轨迹,包括:

基于所述目标场景的粗重建表示,生成所述样本图像集合的采集轨迹;

基于所述采集轨迹的中心和朝向,生成圆周轨迹作为所述浏览轨迹。

6.根据权利要求1所述的方法,所述浏览轨迹包括多个浏览点,所述浏览点包括浏览位置、所述浏览位置对应的浏览视角;

所述基于所述浏览轨迹和所述神经辐射场模型,生成所述目标场景对应于所述浏览轨迹的渲染图像,包括:

将浏览点的浏览位置和浏览视角输入所述目标场景的神经辐射场模型,得到所述目标场景对应于所述浏览点的渲染颜色和渲染密度数据,基于所述渲染颜色和渲染密度数据生成所述目标场景对应于所述浏览点的渲染图像。

7.根据权利要求6所述的方法,若所述渲染图像的质量评估结果低于预期,获取对应于所述渲染图像的补充样本图像,包括:

获取质量评估结果低于预期的渲染图像对应的浏览点;

基于获取的浏览点,发起所述目标场景对应于所述浏览点的样本图像补充采集,以获取所述渲染图像的补充样本图像。

8.根据权利要求1所述的方法,所述基于所述目标场景对应于所述浏览轨迹的渲染图像,生成所述目标场景的全景视频之后,还包括:

对所述全景视频进行风格化处理。

9.根据权利要求1所述的方法,所述基于所述目标场景对应于所述浏览轨迹的渲染图像,生成所述目标场景的全景视频之后,还包括:

对所述全景视频进行超分辨率调优处理。

10.视频生成装置,包括:

图像获取模块,获取目标场景的样本图像集合,所述样本图像集合包括多张关于所述目标场景的原始样本图像;

图像处理模块,基于所述样本图像集合,生成所述目标场景的浏览轨迹和所述目标场景的神经辐射场模型;

渲染模块,基于所述浏览轨迹和所述神经辐射场模型,生成所述目标场景对应于所述浏览轨迹的渲染图像;

质量评估及补充采集模块, 对所述渲染图像进行质量评估, 得到各所述渲染图像的质量评估结果; 若所述渲染图像的质量评估结果低于预期, 获取对应于所述渲染图像的补充样本图像; 将所述补充样本图像添加至所述样本图像集合, 得到更新后的样本图像集合; 基于更新后的样本图像集合重新生成所述目标场景的神经辐射场模型, 得到更新后的神经辐射场模型; 基于所述浏览轨迹和更新后的神经辐射场模型, 重新生成所述目标场景对应于所述浏览轨迹的渲染图像; 重复本步骤, 直至所述目标场景对应于所述浏览轨迹的渲染图像的质量评估结果均不低于预期;

视频生成模块, 基于所述目标场景对应于所述浏览轨迹的渲染图像, 生成所述目标场景的全景视频。

11. 根据权利要求 10 所述的装置, 所述图像处理模块包括:

特征匹配及位姿估计单元, 对所述样本图像集合进行特征点匹配和位姿估计;

轨迹生成单元, 基于所述特征点匹配和所述位姿估计的结果, 生成所述目标场景的粗重建表示, 基于所述目标场景的粗重建表示, 生成所述目标场景的浏览轨迹;

模型训练单元, 基于所述特征点匹配和所述位姿估计的结果, 生成所述目标场景的神经辐射场模型。

12. 根据权利要求 11 所述的装置, 所述轨迹生成单元, 包括:

三维点云处理子单元, 所述基于所述特征点匹配和所述位姿估计的结果, 提取所述目标场景的稀疏三维点云数据, 基于所述稀疏三维点云数据, 生成所述目标场景的粗重建模型;

轨迹生成子单元, 基于所述目标场景的粗重建模型, 生成所述目标场景的浏览轨迹。

13. 根据权利要求 10 所述的装置, 还包括:

风格化处理模块, 对所述全景视频进行风格化处理。

14. 根据权利要求 10 所述的装置, 还包括:

调优模块, 对所述全景视频进行超分辨调优处理。

15. 计算机可存储介质, 其上存储有计算机程序, 所述计算机程序被处理器执行时实现权利要求 1-9 中任一项所述的方法的步骤。

16. 电子设备, 包括:

一个或多个处理器, 以及

与所述一个或多个处理器关联的存储器, 所述存储器用于存储程序指令, 所述程序指令在被所述一个或多个处理器读取执行时, 执行所述权利要求 1-9 中任一项所述的方法的步骤。