

High-quality 3D Content Creation from Images

Austin

USTC

目录

1	Why 3D content creation from images(为什么我们要从图片生成三维内容)	3
1.1	What are 3D contents(什么是三维内容)	4
1.2	Rendering and inverse rendering(渲染与反渲染)	4
1.3	Three key factors in inverse rendering(反渲染中的三个重要因素)	5
1.3.1	Shape representations	6
1.3.2	Appearance representations	6

1.3.3	Rendering operators	7
2	Why NeRF is a big thing(为什么 NeRF 很重要)	7
2.1	Break NeRF into 3 components(NeRF 的三个组成部分)	8
2.2	Take-home message from NeRF(从 NeRF 中能了解什么)	8
2.2.1	Soft shape	8
2.2.2	Fourier features fixes the spectral bias of MLPs(傅里叶特征法修复了 MLP 的隐含偏差)	9
2.2.3	Neural fields for inverse rendering(用于反渲染的神经场)	9
2.3	Unbounded scene and anti-aliasing(无界场景与抗锯齿)	9
2.3.1	Five capture scenarios(五种拍摄场景)	10
2.3.2	NeRF++	10
2.3.3	Conclusion	10
2.3.4	NeRF's anti-aliasing issue	11
3	Editable 3D contents(可编辑三维内容)	11

3.1	Relighting, material re-touching(重打光与材料编辑)	12
3.1.1	What NeRF can and can't do	12
3.1.2	Stage of NeRF	13
3.2	Stylized appearance(风格化的外观)	13
3.2.1	Artistic 3D content creation through stylization(通过风格化进行艺术 3D 内容创作)	13
3.2.2	Nearest Neighbor Feature Matching(NNFM) style loss	14
3.3	Other things(其他内容)	14
3.3.1	Paper and resourses	14

1 Why 3D content creation from images(为什么我们要从图片生成三维内容)

1.1 What are 3D contents(什么是三维内容)

3D contents are combinations of shape and appearance models that canbe rendered into 2D images from different viewpoints.

3D 内容是形状和外观组合模型，可以从不同视点渲染为 2D 图像。

Shape and appearance, where appearance consists of material and lighting, can create RGB images through rendering.

形状和外观（其中外观由材质和照明组成）可以通过渲染创建 RGB 图像。

1.2 Rendering and inverse rendering(渲染与反渲染)

Creating a 3D content requires strong expertise and is time-consuming.

创建 3D 内容需要强大的专业知识，而且耗时。

Images are cheap to acquire, and this leads to inverse rendering.

图片的获取成本很低，使得我们关注反渲染技术。

计算机图形学：3D contents \Rightarrow RGB images

Rendering — efficient, high-quality.

渲染高效，并且生成的图像质量很高。

计算机视觉：Images \Rightarrow 3D contents

Inverse Rendering — change lighting, material, combination of contents etc. \Rightarrow new images.

反渲染能够改变原有的打光、材质、内容的组合等，生成新的图像。这些图像可能是三维内容在不同的打光环境下生成的，也可能是改变了三维内容的材质，或者是对内容进行重新组合后生成的。

Question 问答环节

三维重建算反渲染吗？

三维重建是反渲染的一种，重点在重建三维内容的 Shape 方面，关注于打造高质量的三维几何模型，与 NeRF 存在不同。目前三维重建也在向 Appearance 方面进军，对内容的材质和照明也有关注。

1.3 Three key factors in inverse rendering(反渲染中的三个重要因素)

- What shape representations to use?
- What appearance representations to use?
- What rendering operator to use?

1.3.1 Shape representations

- Mesh 网格
- Point cloud 点云
- Occupancy field 占领场

- Signed distance field(SDF) 有向距离场

1.3.2 Appearance representations

- Material texture map & environment lighting
- Radiance field(surface light field[1])

第一种方法使用 unwrap 展开，在每个点存储材料。光线由 360 image 全景图生成。这是理想的外观表征方式，但是实际情况中，会变得难解和复杂。

第二种使用辐射场 (或者表明光线场)，打包材料和光线，使得外观表征变得简单。但是它难编辑，并且难以表示内容在新环境下的情况。

$$(x, y, z, \theta, \phi) \rightarrow rgb$$

1.3.3 Rendering operators

- Depend on the specific shape and appearance representations(图像的渲染取决于三维内容特定的形状和外观表征)
- Need to be fully differentiable(ideal) (渲染在理想情况下是完全可区分的)

Question 问答环节

三维重建和 NeRF 哪个效果比较好？

三维重建的效果比较好。三维重建的目标就是建立高质量的三维几何模型，而 NeRF 所建立的模型表面较为 noisy。

2 Why NeRf is a big thing(为什么 NeRF 很重要)

NeRF can generate highly photo-realistic novel views.

NeRF 可以生成高度真实的新视图。

例 2.1 28 captured photos \rightarrow NeRF(3D content) \rightarrow render at unseen viewpoints.

2.1 Break NeRF into 3 components(NeRF 的三个组成部分)

- Shape

Soft opacity field(fog) 软不透明度场

$$(x, y, z) \rightarrow \sigma$$

- Appearance

Radiance field(surface light field)

$$(x, y, z, \theta, \phi) \rightarrow rgb$$

- Rendering

$$(x, y, z, \theta, \phi) \rightarrow F_{\Theta} \rightarrow (RGB\sigma)$$

2.2 Take-home message from NeRF(从 NeRF 中能了解什么)

2.2.1 Soft shape

- initialized as nothingness(初始化为空)
- grow as needed(依据需要而变)
- compare with hard geometry, e.g., DVR[2], IDR[3].
 - ✓ Not require object segmentation masks: genus issue.
 - ✓ No boundary discontinuity: easy for differentiable rendering.
 - ✗ Rendering is more expensive; editing is hard.
- without neural networks, but soft shape: fast.[4][5]

2.2.2 Fourier features fixes the spectral bias of MLPs(傅里叶特征法修复了 MLP 的隐含偏差)

$$\gamma(\mathbf{v}) = [\cdots, \cos(2\pi\sigma^{j/m}\mathbf{v}), \sin(2\pi\sigma^{j/m}\mathbf{v}), \cdots]^T$$

2.2.3 Neural fields for inverse rendering(用于反渲染的神经场)

- Compact: each NeRF(3D content) is only about 10 MB.
 - small
 - free to transport
- Continuous: scenes are not discretized.
 - high-quality
- Flexible and easy for optimization: implicit regularization from neural networks.

2.3 Unbounded scene and anti-aliasing(无界场景与抗锯齿)

2.3.1 Five capture scenarios(五种拍摄场景)

- 360 Inward-facing (no background)
- 360 Outward-facing
- Foreward-facing
- Unconstrained 360 Outward-facing (irregular)
- 360 Inward + Outward-facing

2.3.2 NeRF++

NeRF's resolution for 360 capture of unbounded scene: local or overall?

NeRF++: separate foreground and background; semmetry between 360 inward-facing and 360 outward-facing captures.[6]

- Map unbounded space outside a sphere into a unit cube.
- Farther-away points get more "squeezed" after mapping.

2.3.3 Conclusion

- NeRF faces resolution for 360 capture of unbounded scene.
- NeRF++ separates foreground and background modelling, as NeRF is compositional.
- NeRF++ uses inverted sphere parametrization to establish semmetry between 360 inward-facing and 360 outward-facing captures.

2.3.4 NeRF's anti-aliasing issue

Nyquist-Shannon sampling theorem($f_s > 2B$)

key: low pass filter

MipNeRF: low-pass filter the Fourier features(Fourier features are vector functions of the spatial location).

$$\mathbf{v} = (x, y, z)$$

.

- NeRF suffers from anti-aliasing issue.
- Images are discrete signals. MipNeRF introduces low-pass filtering over the Fourier features, where the filter size is controlled by the cone size.[7]

3 Editable 3D contents(可编辑三维内容)

3.1 Relighting, material re-touching(重打光与材料编辑)

3.1.1 What NeRF can and can't do

- NeRF is great 3D content in term of synthesizing highly photo realistic novel views.
- Future application:
 - Put mutiple NeRF in the same environment
 - Retouch appearance
 - Create artistic 3D contents
 - Interactive editings

- Physical simulation
- Object insertion
- Animations

Mesh and material textures remain dominate 3D content format today.

Input: photomatric images → Output: 3D assets

But mesh and material textures are hard to optimize directly.

- Fixed topology, fixed resolution
- Hard to avoid self-intersections

Leverage the insights from NeRF to ease optimization?

- During optimization time, use neural field for shape and appearance representations.
 - Use soft shape for quickly instantiating geometry from nothingness.
 - Use hard shape for physics-based rendering.
- During test-time, use mesh and material textures.

3.1.2 Stage of NeRF

- Optimize a thick surface via volumetric rendering.
- Refine a surface and solve for material & lighting.

- Edge-aware surface rendering
 - Shading at edge pixels(visiblity discontinuities) contain shape cues.
 - Sub-optimal shape reconstruction without accounting for such edges.
 - Split an edge pixel into havles, and blend color of both halves.

3.2 Stylized appearance(风格化的外观)

3.2.1 Artistic 3D content creation through stylization(通过风格化进行艺术 3D 内容创作)

Style loss $\xleftarrow{\text{render \& compare 3D content like NeRF}}$ $\xrightarrow{\text{render \& compare L1/L2 loss}}$

Style loss measures mismatch in feature sets.

3.2.2 Nearest Neighbor Feature Matching(NNFM) style loss

$$\ell_{nnfm}(\mathbf{F}_{render}, \mathbf{F}_{style}) = \frac{1}{N} \sum_{i,j} \min_{i',j'} D(\mathbf{F}_{render}(i,j), \mathbf{F}_{style}(i',j'))$$

$$D(\mathbf{v}_1, \mathbf{v}_2) = 1 - \mathbf{v}_1^T \mathbf{v}_2 \sqrt{\mathbf{v}_1^T \mathbf{v}_1 \mathbf{v}_2^T \mathbf{v}_2}$$

3.3 Other things(其他内容)

ARF project page: <https://www.cs.cornell.edu/projects/arf>

3.3.1 Paper and resourses

- Mildenhall et al. ECCV 2020. NeRF: Representing Scenes as Neural Radiance Fields for View Synthesis.
- Zhang et al. ArXiv 2020. Nerf++: Analyzing and Improving Neural Radiance Fields.
- Barron et al. CVPR 2021. Mip-NeRF: A Multiscale Reprmentation for Anti-aliasing Neural Radiance Fields.
- Zhang et al. CVPR 2022. IRON: Inverse Rendering by Optimizing Neural SDFs and Materials from Photometric Images.
- Zhang et al. ECCV 2022. ARF: Artistic Radiance Fields.
- Awesome NeRF: <http://github.com/yenchelin/awesome-NeRF>.

参考文献

- [1] WOOD D N, AZUMA D I, ALDINGER K, et al. Surface light fields for 3d photography[C]//Proceedings of the 27th annual conference on Computer graphics and interactive techniques. [S.l.: s.n.], 2000: 287-296.
- [2] NIEMEYER M, MESCHEDER L, OECHSLE M, et al. Differentiable volumetric rendering: Learning implicit 3d representations without 3d supervision[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. [S.l.: s.n.], 2020: 3504-3515.

- [3] YARIV L, KASTEN Y, MORAN D, et al. Multiview neural surface reconstruction by disentangling geometry and appearance[J]. Advances in Neural Information Processing Systems, 2020, 33: 2492-2502.
- [4] FRIDOVICH-KEIL S, YU A, TANCIK M, et al. Plenoxels: Radiance fields without neural networks[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. [S.l.: s.n.], 2022: 5501-5510.
- [5] SUN C, SUN M, CHEN H T. Direct voxel grid optimization: Super-fast convergence for radiance fields reconstruction[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. [S.l.: s.n.], 2022: 5459-5469.
- [6] ZHANG K, RIEGLER G, SNAVELY N, et al. Nerf++: Analyzing and improving neural radiance fields[J]. arXiv preprint arXiv:2010.07492, 2020.
- [7] BARRON J T, MILDENHALL B, TANCIK M, et al. Mip-nerf: A multiscale representation for anti-aliasing neural radiance fields[C]//Proceedings of the IEEE/CVF International Conference on Computer Vision. [S.l.: s.n.], 2021: 5855-5864.