闭环用户视角采样与误差峰值可视化

Ayaka Yasunaga[1], Hideo Saito[1], and Shohei Mori[2,1]

[1]Keio University [2]University of Stuttgart

ABSTRACT

增强现实（AR）提供了可视化新型视图合成中缺失视角样本的方法。现有方法为新视图样本提供三维标注，并要求用户通过与 AR 显示对齐来拍摄图像。这一数据收集任务已知是精神上耗费的，且由于理想的但限制性的基础采样理论，将捕获区域限制在预定义的小区域内。为了使用户摆脱三维标注和有限场景探索的束缚，我们提出使用局部重建光场并可视化通过插入新视图来移除的误差。我们的结果显示，误差峰值可视化不那么侵入性，减少了最终结果中的失望，并且在我们移动视图合成系统中用更少的视角样本也能令人满意。我们还展示了我们的方法可以贡献于最近的大场景辐射场重建，例如 3D 高斯散射。

Index Terms— 增强现实，多平面图像，用户参与循环，误差峰值，视图合成。

1. 介绍

新型视图合成可以从给定的多视角图像中渲染未见的视图。机器学习的最新进展提供了一种将场景外观编码到训练好的神经网络 [1, 2, 3] 中的方法，或者优化场景代理 [4, 5, 6]。此过程需要几分钟到几小时的时间，而缺失贡献视图样本将会导致进行另一次摄影会话和优化。全光采样理论为无混叠的视图合成提供了最少图像数量的指导 [7]。然而，它仅支持在 2D 平面或光场（LF）[8] 中的视图采样。实际的视图采样会话即使使用多平面图像（MPI）[9] 进行近似也需要数千张图片。

增强现实 (AR) 在三维空间中预计算位置的缺失视图样本可视化方面扮演着重要角色。现有工作提出了一项数据收集任务，该任务测量全息采样理论所需的最小场景深度，并放置 3D 标注 [9]。用户被要求移动跟踪智能手机到这些标注处以触发快门。类似的方法也可应用于球形可视化 [10, 11]。AR 视觉引导的问题包括 (i) 视觉化隐藏了要拍摄的主题，阻碍了摄影的娱乐性，(ii) 需要精确并对精神要求高的对齐任务 [12]，(iii) 全局应用的最小视图间隔可能在当时没有意义，以及 (iv) 固定可视化限制了用户探索超出预定区域的能力。

为了解决这些问题，我们提出了使用局部光场重建或 MPI[9, 13] 的误差峰值可视化方法。类似于数码相机的对焦辅助功能，该功能通过可视化对焦区域让摄影师找到最佳对焦区域，我们的误差峰值可视化则突出显示那些可以通过插入更多视图样本而被削弱的错误像素（参见图 1）。总体而言，当视图合成不再出现伪影时，用户就完成了拍摄会话。这种可视化技术（i）仅干扰伪影，（ii）不需要对齐任务，（iii）使用户能够分析局部重建，并且（iv）允许在 3D 空间中自由探索。演示视频可在 IEEE SigPort 获取。

2. 相关工作

2.1. 光场与记录方法

LF 由 2D 位置（uv 平面）和一个焦平面（st 平面）[8] 的 4D 数据组成，并通过将 4D 数据重采样为 2D 来渲染。这个 4D 数组数据可以通过带有 [8] 转盘、[14] 微透镜或 [15] 摄像头阵列系统来系统地记录。LF 是一个近似的全光照函数 [16]，因此，某个区域的视点样本数量可以通过全光照采样理论 [7] 来解释。较新的方法可以接受假设空间过采样的非结构化视点样本 [11] 或更多几何信息 [17]。

MPI 被视为一种近似的光场，并允许以较稀疏的

Fig. 1. 聚焦峰值、我们的误差峰值以及传统的 AR 可视化。受聚焦峰值启发，我们的方法可视化了底层视图合成中的错误，鼓励插入新视图以减少明显的误差。与传统 AR 可视化相比，我们的方法侵入性较小，并解决了现有问题。

视角采样和 RGB+α 层的软 3D 重建 [9, 18]。具有 D 层的 MPI 与原始光场 [9] 相比，减少了 D^2 个视图样本。

2.2. 现代多平面图像

MPI 是通过多视角分析生成的 [9, 18, 19]，但现代方法可以使用经过大量图像数据训练的神经网络从单个快照中在几秒钟内推断出 MPI[5, 13]。这些方法调整层的位置以最大化每层的像素密度 [13] 并在渲染视角与原始视角不同时对未观察到的像素进行插值 [5]。这些修改超越了全息采样理论中均匀采样的支持范围。

然而，为了更广泛的场景覆盖和再现视点依赖效果，需要多视角采样。我们的误差峰值可视化方法将 MPI 渲染与当前视点进行比较，提供视觉分析，无论 MPI 生成器和采样理论如何。插入新视点的决策完全取决于用户的视觉分析。

2.3. 视图采样的自回归模型

AR 可以提供三维注释来可视化在哪里在空中添加视图样本 [9, 12]。流行的方法包括三维轴线 [9]、二维平面 [12] 和一个半球 [11, 10] 包围目标区域。这些方法（i）可以部分或完全隐藏感兴趣的区域，以及（ii）将拍摄任务重新定位为数据收集任务。整体体验可能需要较高的精神集中度，特别是在较小的区域 [12] 中。

视觉引导的位置是根据全光采样理论 [7] 计算得出的。此过程（iii，iv）涉及整个场景扫描以保证视图

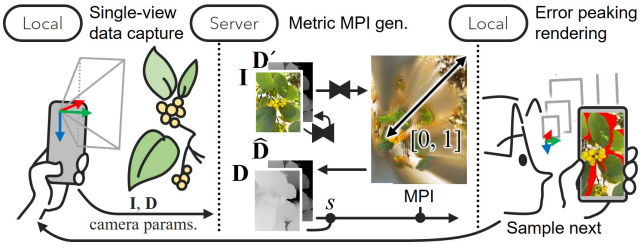

Fig. 2. 用户环视采样系统概述。

采样期间的最小场景距离，因此，一旦会话开始，3D 注释必须固定。

为了找到下一个最佳视图样本，在线评估几何信息增益 [20]，这忽略了视图依赖性。最先进的在线 3D 高斯优化集成到同时定位和映射（SLAM）系统中，抑制了视图依赖性和各向异性高斯分布以提高速度和鲁棒性 [21, 22]。为了获得缺失的视图依赖性，之后必须运行另一个完整的优化程序 [22]。

我们使用 MPI，它似乎在局部区域内实现了像素级精确的重建，可以通过直接比较渲染与当前视图来实现局部视觉评估。要在低分辨率下通过体积或 3D 高斯重建为移动设备提供临时视觉反馈是很难实现的，尽管这些解决方案在消费类应用中很受欢迎 [1]。

3. 用户环路 LF 捕获系统

我们的系统能够通过三个步骤实现误差峰值可视化（图 2）：使用跟踪智能手机进行单视图数据捕获，从单个图像生成 MPI（第 3.1 节），以及对用户评估进行逐像素评估（第 3.2 节）。

3.1. 度量 MPI 生成

单视图数据。我们的系统允许用户通过 AR 支持的手机捕捉单视角数据。单视角数据包括一个 RGB 图像 $\boldsymbol{I}$，一个度量深度图 $\boldsymbol{D}$，以及相机的内参和外参。给定这些数据，系统会根据输入预测 MPI 以提供即时视觉反馈。假设 AR 设备没有足够的计算资源来执行基于 CNN 的 MPI 生成，那么数据会被通过网络路由到能够进行更强大计算的服务器上。在实际操作中，相

[1] 例如，Scanniverse 应用程序 https://scaniverse.com

机参数和度量深度图由 Unity AR Foundation[2] 提供，使用其带有惯性测量单元 (IMU) 的 SLAM 系统。

单视图数据的 MPI。给定大小为 W（宽度）乘以 H（高度）的 $\boldsymbol{I} \in \mathbb{R}^{W\times H\times 3}$ 和 $\boldsymbol{D} \in \mathbb{R}^{W\times H\times 1}$，我们使用最先进的方法 AdaMPI[13] 生成一个 $W \times H \times 4D$ 的 MPI。MPI 具有 D 层（默认为 $D = 32$），每一层由颜色为 $\boldsymbol{c}_i$ 和密度为 σ_i 的像素组成，位于层次深度 ($i \in [1, D]$) 处。$i = 1$ 和 $i = D$ 分别是最接近和最远的层索引。

使用 MPI 渲染新视图涉及体积渲染或层析重映和在视图中合成各个层。给定在新颖视角下的 $\boldsymbol{c}_i$ 和 σ_i 以及 i_{th} 的 MPI 层，我们如下计算颜色 $\hat{\boldsymbol{c}}_i$：

$$\hat{\boldsymbol{c}} = \sum_{i=1}^{D}\left(\boldsymbol{c}_i\hat{\alpha}_i\prod_{j=1}^{i-1}(1-\hat{\alpha}_j)\right), \hat{\alpha} = \sum_{i=1}^{D}\left(\alpha_i\prod_{j=1}^{i-1}(1-\alpha_j)\right),$$

其中 $\alpha_i = \exp(-\delta_i\sigma_i)$ 和 δ_i 是平面 i 和平面 $i+1$ 之间的距离图。

混合多个 MPI。为了避免最近的 MPI 与其他 MPI 之间的频繁切换伪影，我们混合了 k 个最近的 MPI 体积（默认为 $K = 3$）。给定一个 k_{th} 渲染颜色 $\hat{\boldsymbol{c}}_k$ 和一个 k_{th} 的渲染 alpha 值 $\hat{\alpha}$，我们根据文献 [9] 在图像空间中混合它们。

$$\boldsymbol{c}_{\text{mpi}} = \frac{\sum_k w_k\hat{\alpha}_k\hat{\boldsymbol{c}}_k}{\sum_k w_k\hat{\alpha}_k},\ \alpha_{\text{mpi}} = \min\left(\sum_k w_k\hat{\alpha}_k, 1\right).$$

我们在此阶段输出 alpha 值，以在第 3.2 节中进行可视化目的。$w_k \propto \exp(-l/\gamma)$。$l$ 是相机位置与注册 MPI 之间的 L^2 距离。γ 是到 K 视图的最大 L^2 距离。

度量尺度对齐。由于单视图成像的尺度歧义 [5, 13]，直接 MPI 输出的层位于 $[0, 1]$ 归一化范围内。为了使尺度与真实场景匹配，我们计算尺度因子 s。对于大小为 w_i（宽度）乘以 h_i（高度）乘以 z_i（距离相机的距离）的 i_{th}MPI 层，我们按照以下方式缩放 3D 矩形。

$$z_i' = sz_i, \quad w_i' = z_i'\frac{w_i}{f}, \quad h_i' = z_i'\frac{h_i}{f}, \tag{1}$$

其中 f 表示相机的焦距。

由于我们有一个度量深度 $\boldsymbol{D}$，并且可以通过将 d_i 替换为 $\boldsymbol{c}_i$ 代入第一个方程，使用 MPI 渲染一个渲染后的深度图像 $\hat{\boldsymbol{D}}$，我们计算这些深度图像之间的比例差异，或 s。受到 Tucker 和 Snavel[5] 的尺度不变合成的启发，我们按以下方式计算 s。

$$s = \exp\left[\frac{1}{|\boldsymbol{D}|}\sum_{(x,y,d)\in\boldsymbol{D}}\left(\ln\hat{\boldsymbol{D}}(x,y) - \ln(d^{-1})\right)\right]. \tag{2}$$

此缩放操作在服务器上执行。MPI 和 s 的结果通过网络发送到 AR 设备。移动设备使用拍摄位置的相机参数将缩放后的 MPI 注册到场景中。

3.2. 误差峰值可视化对于下一视图样本

MPI **渲染**。生成第一个 MPI 后，我们将屏幕切换到统一的黑色背景，并叠加 MPI 渲染。这种方法旨在直接展示当前 MPI 渲染的质量。

所得的颜色值 $\boldsymbol{c}'$ 是渲染像素颜色 $\boldsymbol{c}_{\text{mpi}}$ 覆盖在黑色背景颜色 $\boldsymbol{c}_{\text{blk}} = (0, 0, 0)^\intercal$ 上的结果，权重由渲染的 alpha 值 α_{mpi} 决定。

$$\boldsymbol{c}' = \alpha_{\text{mpi}}\boldsymbol{c}_{\text{mpi}} + (1-\alpha_{\text{mpi}})\boldsymbol{c}_{\text{blk}}. \tag{3}$$

当用户看到如堆栈卡片伪影等渲染伪影时，这些伪影会在显示设备改变拍摄位置时揭示分层结构，预计用户会添加另一个 MPI。然而，对于不了解根据 MPI 生成方法而变化的伪影的新手用户来说，这将是一个挑战。这就是我们为何增加错误峰值可视化的原因。

误差峰值叠加。我们计算视频帧与 MPI 渲染在屏幕空间中的差异，并用红色突出显示存在显著错误的像素。这向用户传达了错误的位置和程度。然后，用户添加另一个 MPI 并检查红色像素是否减少。

$$\boldsymbol{c}' = \begin{cases}\boldsymbol{c}_{\text{err}}, & \text{if } L(\boldsymbol{c}_{\text{mpi}}, \boldsymbol{c}_{\text{vid}}) > t\\ \boldsymbol{c}_{\text{vid}}, & \text{otherwise.}\end{cases} \tag{4}$$

这里，$\boldsymbol{c}_{\text{err}}$ 表示预定义的错误高亮像素颜色，$\boldsymbol{c}_{\text{vid}}$ 是视频帧的一个像素颜色，$L(\cdot)$ 计算两个像素值之间的距离，t 是一个阈值。

我们在真实图像上观察到了由于 SLAM 跟踪不完美而导致的错位错误高亮。因此，我们将可视化的误差叠加到 MPI 渲染上。用户从未看到原始场景，而是通过数字双胞来看它们。

$$\boldsymbol{c}' = \begin{cases}\boldsymbol{c}_{\text{err}}, & \text{if } L(\boldsymbol{c}_{\text{mpi}}, \boldsymbol{c}_{\text{vid}}) > t\\ \alpha_{\text{mpi}}\boldsymbol{c}_{\text{mpi}} + (1-\alpha_{\text{mpi}})\boldsymbol{c}_{\text{blk}}, & \text{otherwise.}\end{cases} \tag{5}$$

[2]Unity AR 基础 https://unity.com/solutions/xr/ar

4. 评估

4.1. 用户研究

目标。我们在用户研究中验证了用户在我们的错误峰值可视化下自发采样视图的能力。我们将我们的系统与基于 plenoptic 采样理论的 LLFF[9] 进行比较，以观察不同的可视化如何影响主观和定量因素。我们设计了一个重复测量的受试者内研究来识别两个系统的特征，我们的和 LLFF。我们也评估了我们的系统在三维高斯溅射（3DGS）[6] 中视图采样的适用性。

场景。我们设计了一个具有不同深度变化的办公场景进行场景捕获。场景深度范围大约为 0.5 米到 3.0 米（图 3）。最近的深度出现在 AR 设备位于左侧时，并随着设备向右移动而逐渐消失。我们在两侧各放置了一根柱子，每根相距 0.15 米，以调节捕获区域并告知参与者范围。

实现细节。我们使用 Unity 为 Apple iPhone 15 Pro 开发了我们的移动应用。我们使用了 AdaMPI，但用更近的 Depth Anything v2[23] 替换了深度估计器 [24] 以提高准确性。每个 MPI 是 $W \times H \times D = 294 \times 639 \times 32$ 像素，在纵向模式下导致了一个 45.93° 的水平视场。为了避免超出机载 RAM 的限制，我们只为最近的 k 个视图分配内存空间，并在组合的 k 个视图发生变化时交换内容。$L(\cdot)$ 计算了 RGB 值的 L1 范数，以及 $t = 0.4$。采样一个 MPI 需要 3.6 秒（标准差 = 0.5）。

基线实现。我们将我们的方法与基于 plenoptic 采样理论在 2D 网格上显示 3D 轴的 LLFF 视觉引导 [9] 进行了比较。LLFF 要求用户提前扫描整个场景，以计算最小深度 z_{min} 和最少视图数量 N 来确定相机间隔 Δ_u。

$$N = \left(\frac{SW}{2Dz_{min} \tan(\theta/2)} \right)^2, \ \Delta_u = S/\sqrt{N}, \qquad (6)$$

其中，S 是捕获区域的边长（即我们设置中的 0.15 米），θ 是相机视野。请注意，我们的方法在设计上不需要这样的准备。

指标。我们使用系统可用性量表（SUS）[25]、NASA-TLX[26]、完成时间（TCT）、排除通信时间的净完成时间（NCT），以及捕获图像的数量（捕获计数）评估了这些系统。我们还从我们的任务导向问卷中收集了三个额外分数：自信心（S-Conf.）：“捕获的结果将符合我的预期。”；满意度（Satis.）：“结果符合我的预期。”；以及场景聚焦：“我在捕获时专注于感兴趣的场景。”，评分范围为 1 到 7 分，从 (1) 强烈不同意 到 (7) 强烈同意。

参与者。我们收集了 20 名参与者（5 名女性和 15 名男性，年龄 $\bar{X} = 22.4$（标准差 = 1.4），所有参与者均为右手利手且视力矫正）。所有参与者都是主修计算机科学的大学生，并在 AR 经验中自评为 $\bar{X} = 3.7$，标准差为 = 1.9 在 $[1, 5]$。

结果。我们对收集的数据进行了球形度、正态性和方差同质性测试。只有在所有初步测试都满足的情况下，我们才使用 T 检验，否则我们将使用 Wilcoxon 符号秩检验（图 4）。统计分析显示 TCT 存在显著差异 (LLFF:$\bar{X}$ = 37.6, 标准差 SD= 15.6; 我们的:$\bar{X}$ = 99.9, 标准差 SD= 31.7;$p < 0.001$; Cohen's d= 2.4), NCT (LLFF:$\bar{X}$ = 37.6, 标准差 SD= 15.6; 我们的:$\bar{X}$ = 62.2, 标准差 SD= 27.70;$p = 0.02$; Cohen's d= 1.1), 收集的图像数量 (LLFF:M = 14.0; 我们的:M = 11.0;$p < 0.001$; RBC= 0.9), S-Conf. (LLFF:M = 4.5; 我们的:M = 5.0;$p = 0.04$; RBC= −0.5), 满足。(LLFF:M = 4.5; 我们的:M = 6.0;$p = 0.005$; RBC= −0.78), 以及场景聚焦 (LLFF:$M = 2.0$; 我们的:$M = 6.5$;$p < 0.001$; RBC= −1.0)。

分析揭示了 SUS-Q4（LLFF：$M = 2.0$；我们的方法：$M = 3.0$；$p = 0.004$；RBC= −0.9）、SUS-Q5（LLFF：$M = 3.0$；我们的方法：$M = 4.0$；$p = 0.01$；RBC= −0.9）、SUS-Q10（LLFF：$M = 2.0$；我们的方法：$M = 2.0$；$p = 0.03$；RBC= −0.7）、TLX-时间需求：$M = 32.5$；我们的方法：$M = 15.0$；$p = 0.02$；RBC= 0.7）和 TLX-挫败感：$\bar{X} = 42.0$，标准差 = 28.0；我们的方法：$\bar{X}$ = 28.5，标准差 = 16.6；p = 0.04；Cohen's d= 0.6）的显著差异。

总体而言，我们的在视图采样期间提供了更高的自信心，更好地聚焦场景，并且在最终渲染质量方面减少了失望感，同时使用较少的视图。然而，它需要更长的时间段和学习过程。结果验证了第 1 节中的声明 (i–iii)。

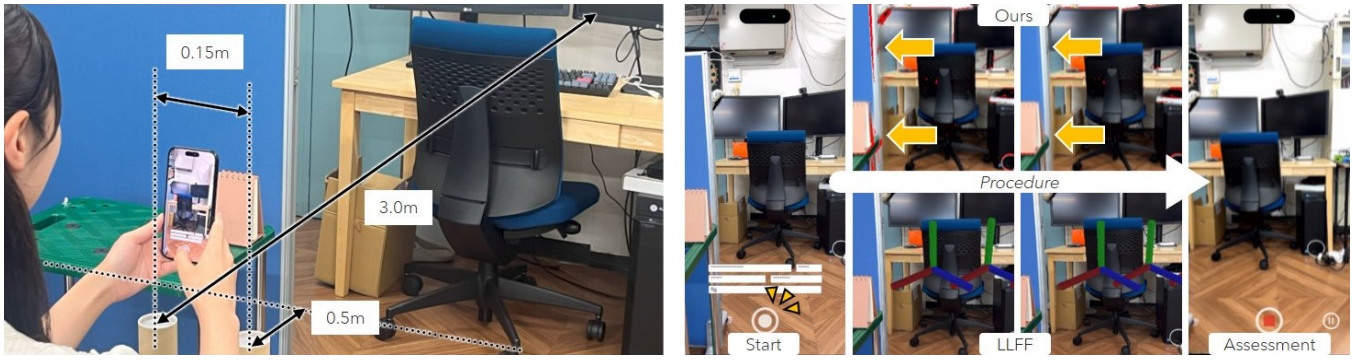

Fig. 3. 用户研究。（左）实验设置深度变化为 [0.5, 3.0] 米。（右）实验程序。

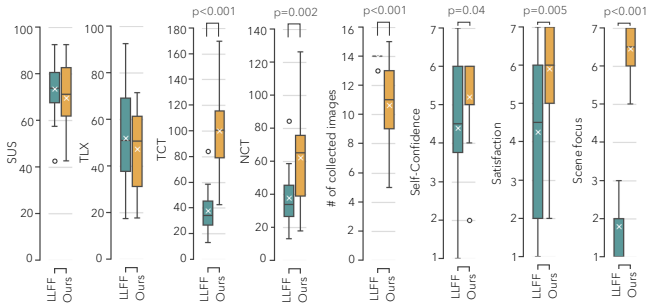

Fig. 4. 用户研究结果。

Table 1. 不同视图采样策略的视图合成质量。以比率形式评估，以便考虑场景、捕获方法和重建技术的变化。

Method (MPI)	PSNR (↑)	SSIM (↑)	LPIPS (↓)
Random	0.89 (0.73)	0.92 (0.56)	1.09 (0.67)
Uniform	0.97 (0.87)	0.96 (0.78)	1.03 (0.86)
Ours	1.00 (1.00)	1.00 (1.00)	1.00 (1.00)
Method (3DGS)	**PSNR (↑)**	**SSIM (↑)**	**LPIPS (↓)**
Random	0.92 (1.74)	0.986 (1.98)	1.30 (1.96)
Uniform	0.96 (1.27)	0.997 (1.12)	1.09 (1.28)
Ours	1.00 (1.00)	1.00 (1.00)	1.00 (1.00)

4.2. 定量评估

数据集。我们记录了三个具有不同深度和相机运动的场景的真实图像数据集，以验证第 1 节中的陈述 (iv)。我们评估了 MPI 和 3DGS[6] 的渲染质量，以验证我们的方法的可扩展性。我们密集地记录了场景，并选择子集来形成以下虚拟方法：(我们的) 我们计算了用户研究中所有视点和参与者上的平均误差值 (= 4.28%)。每次误差值超过此阈值时，我们都录制图像。(均匀) 我们记录了所有图像，并使用与我们的相同数量的图像，尽可能保持视图之间的等距。(随机) 我们从均匀和我们的中随机选择了相同数量的图像。我们使用 COLMAP[27] 来获取相机参数和 3DGS 的 3D 点。真实图像位于均匀的每两个视图之间。

结果。由于强烈依赖于场景、捕获策略和重建方法，我们计算了竞争对手图像指标相对于我们的比率，而不是使用绝对的图像质量指标 (表 1)。我们的在 PSNR、SSIM 和 LPIPS 比率上优于其他方法。3DGS 的结果表明，我们的方法可以作为即时局部反馈用于大型场景的其他视图合成方法中。图 5 展示了 3DGS 结果的定性比较。在这里，我们将“完整”视图合成与所有可用视图样本进行对比，不包括评估数据集。我们的方法产生的伪影比均匀和随机少，这是由于 MPI 的误差分析识别出了潜在的错误像素。

5. 限制和未来方向

我们的用户研究和定量评估结果使用局部 MPI 和光场重建来描述我们错误峰值可视化的特点。然而，我们讨论了当前实现面临的一些已知限制。

表面的深度。我们使用了 Depth Anything v2 作为 AdaMPI 的输入深度图。我们发现基于神经网络的深度估计器总是恢复表面的深度像素，而不考虑物体材料，在合成金属和透明物体视图时表现不佳。在这些区域中，误差可视化可能会过于显著，用户可以通过

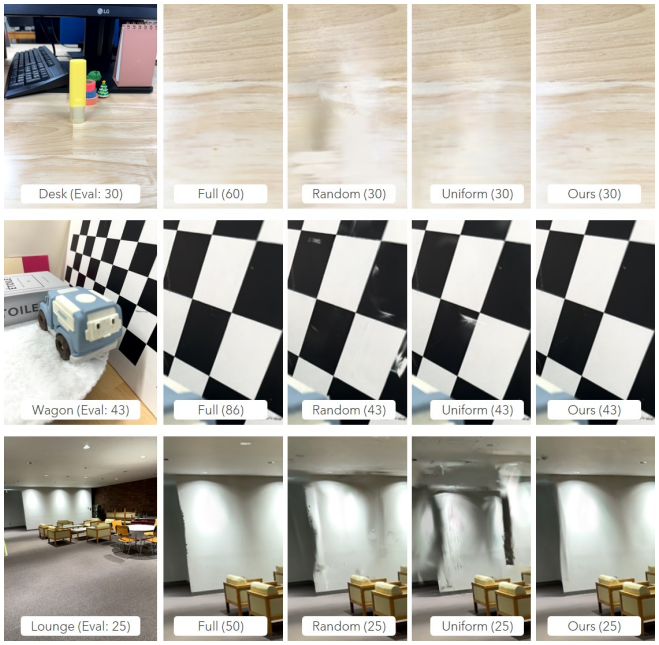

Fig. 5. 3DGS 渲染结果。

拍摄更多图像来减轻这些伪影。

通信和网络推理的延迟。我们方法的优势在于通过视图合成实现实时错误可视化。现有的单视角 MPI 生成方法是为台式机 GPU 设计的，无法在移动捕捉系统上实际实施。因此，我们试图在一个服务器-客户端系统上实现它，以在网络推理时从 AR 设备中卸载任务。然而，我们的系统仍然由于网络推理而存在延迟问题，此外还有服务器和客户端之间的通信所需的时间。进一步的性能改进将带来更好的结果。

进一步的用户分析。文献中描述的采样理论 [9] 忽略了我们的 AR 显示配置。因此，该理论自身无法很好地解释我们的发现。缺失的因素包括屏幕像素分辨率、AR 显示器与用户眼睛之间的距离以及用户在观察渲染结果时的动作。分析眼动追踪数据将通过量化距离和用户对屏幕的注意力提供更多的解释性措施。

6. 结论

在这篇论文中，我们讨论了现有 AR 注释在新视角合成中的视图采样问题。为了解决这些问题，我们提出了一种用户参与的方法，该方法要求用户通过插入新的视图来减少底层 MPI 重建的错误高亮显示，而不是依赖于严格的采样理论。我们的结果显示，误差峰值可视化不那么侵入性，减少了对最终结果的失望，并且在我们的移动视图合成系统中使用较少的视图也令人满意。此外，我们的结果表明，使用我们系统的照片可以为现代 3D 高斯投射带来益处。

致谢 本工作得到了德国研究基金会 (DFG，德国研究协会) 的支持，资助项目为德国卓越战略 – EXC 2120/1 – 390831618，并部分由下一代发起的先驱研究支持 (JST Support for Pioneering Research Initiated by the Next Generation (# JPMJSP2123))。

7. REFERENCES

[1] Ben Mildenhall, Pratul P. Srinivasan, Matthew Tancik, Jonathan T. Barron, Ravi Ramamoorthi, and Ren Ng, "NeRF: Representing scenes as neural radiance fields for view synthesis," in ECCV, 2020.

[2] Qianqian Wang, Zhicheng Wang, Kyle Genova, Pratul Srinivasan, Howard Zhou, Jonathan T. Barron, Ricardo Martin-Brualla, Noah Snavely, and Thomas Funkhouser, "IBRNet: Learning multi-view image-based rendering," in CVPR, 2021.

[3] Vincent Sitzmann, Semon Rezchikov, William T. Freeman, Joshua B. Tenenbaum, and Fredo Durand, "Light field networks: Neural scene representations with single-evaluation rendering," in NeurIPS, 2021.

[4] Sara Fridovich-Keil, Alex Yu, Matthew Tancik, Qinhong Chen, Benjamin Recht, and Angjoo Kanazawa, "Plenoxels: Radiance fields without neural networks," in CVPR, 2022, pp. 5501–5510.

[5] Richard Tucker and Noah Snavely, "Single-view view synthesis with multiplane images," in CVPR, 2020.

[6] Bernhard Kerbl, Georgios Kopanas, Thomas Leimkühler, and George Drettakis, "3D Gaussian Splatting for real-time radiance field rendering.," ACM TOG, vol. 42, no. 4, pp. 139–1, 2023.

[7] Jin-Xiang Chai, Xin Tong, Shing-Chow Chan, and Heung-Yeung Shum, "Plenoptic sampling," in SIGGRAPH, 2000, pp. 307–318.

[8] Marc Levoy and Pat Hanrahan, "Light field rendering," in SIGGRAPH, 1996, p. 31 – 42.

[9] Ben Mildenhall, Pratul P Srinivasan, Rodrigo Ortiz-Cayon, Nima K. Kalantari, Ravi Ramamoorthi, Ren Ng, and Abhishek Kar, "Local light field fusion: Practical view synthesis with prescriptive sampling guidelines," ACM TOG, vol. 38, no. 4, pp. 1–14, 2019.

[10] Peter Mohr, Shohei Mori, Tobias Langlotz, Bruce H Thomas, Dieter Schmalstieg, and Denis Kalkofen, "Mixed reality light fields for interactive remote assistance," in CHI, 2020, pp. 1–12.

[11] Abe Davis, Marc Levoy, and Fredo Durand, "Unstructured light fields," Computer Graphics Forum, vol. 31, no. 2pt1, pp. 305–314, 2012.

[12] Reina Ishikawa, Hideo Saito, Denis Kalkofen, and Shohei Mori, "Multi-layer scene representation from composed focal stacks," IEEE TVCG, 2023.

[13] Yuxuan Han, Ruicheng Wang, and Jiaolong Yang, "Single-view view synthesis in the wild with learned adaptive multiplane images," in SIGGRAPH, 2022.

[14] Ren Ng, Marc Levoy, Mathieu Brédif, Gene Duval, Mark Horowitz, and Pat Hanrahan, "Light field photography with a hand-held plenoptic camera," Tech. Rep., Stanford Tech Report CTSR 2005-02 Light, 2005.

[15] Vaibhav Vaish, Bennett Wilburn, Neel Joshi, and Marc Levoy, "Using plane + parallax for calibrating dense camera arrays," in CVPR, 2004, pp. 2–9.

[16] Edward H. Adelson and James R. Bergen, "The plenoptic function and the elements of early vision," in Computational Models of Vis. Proc. 1991, pp. 3–20, MIT Press.

[17] Chris Buehler, Michael Bosse, Leonard McMillan, Steven Gortler, and Michael Cohen, "Unstructured lumigraph rendering," in SIGGRAPH, 2001, pp. 425–432.

[18] Eric Penner and Li Zhang, "Soft 3D reconstruction for view synthesis," ACM TOG, vol. 36, no. 6, 2017.

[19] Richard Szeliski and Polina Golland, "Stereo matching with transparency and matting," IJCV, vol. 32, no. 1, pp. 45–61, 1999.

[20] Yuetao Li, Zijia Kuang, Ting Li, Guyue Zhou, Shaohui Zhang, and Zike Yan, "ActiveSplat: High-fidelity scene reconstruction through active Gaussian Splatting," in ICRA, 2025.

[21] Hidenobu Matsuki, Riku Murai, Paul H. J. Kelly, and Andrew J. Davison, "Gaussian Splatting SLAM," in CVPR, 2024.

[22] Nikhil Keetha, Jay Karhade, Krishna Murthy Jatavallabhula, Gengshan Yang, Sebastian Scherer, Deva Ramanan, and Jonathon Luiten, "SplaTAM: Splat, track & map 3D Gaussians for dense RGB-D SLAM," in CVPR, 2024.

[23] Lihe Yang, Bingyi Kang, Zilong Huang, Zhen Zhao, Xiaogang Xu, Jiashi Feng, and Hengshuang Zhao, "Depth Anything V2," in NeurIPS, 2024.

[24] René Ranftl, Alexey Bochkovskiy, and Vladlen Koltun, "Vision transformers for dense prediction," in ICCV, 2021, pp. 12159–12168.

[25] John Brooke, "SUS: A 'quick and dirty' usability scale," Usability Eval. in Ind., vol. 189, no. 194, pp. 4–7, 1996.

[26] Sandra G Hart and Lowell E Staveland, "Development of NASA-TLX (task load index): Results of empirical and theoretical research," in Advances in Psychology, vol. 52, pp. 139–183. 1988.

[27] Johannes Lutz Schönberger and Jan-Michael Frahm, "Structure-from-motion revisited," in CVPR, 2016.