

CONSTRUÇÃO DE AMBIENTES IMERSIVOS DO VAREJO - EXPLORANDO CAMPOS DE RADIÂNCIA NEURAL PARA GERAÇÃO AUTOMÁTICA DE OBJETOS 3D

Luis Henrique da Silva Resende¹, Thiago França Naves²

RESUMO: A modelagem 3D é uma tarefa complexa e demorada, exigindo experiência e habilidades técnicas, mesmo entre profissionais. Este artigo propõe uma abordagem inovadora baseada em Campos de Radiância Neural (NeRF) para automatizar a criação de objetos 3D a partir de vídeos, visando simplificar o processo e reduzir o tempo necessário para gerar modelos detalhados. Desde sua introdução em 2020, os NeRF revolucionaram a geração de visualizações 3D ao permitir a criação de novas perspectivas e animações dinâmicas a partir de imagens bidimensionais. Neste estudo, utilizamos a ferramenta Nerfstudio, que fornece uma API robusta para testes, criação e treinamento de diferentes modelos NeRF, e estruturamos o processo em quatro etapas principais: captura de vídeo, onde introduzimos os requisitos mínimos para a câmera e a trajetória ideal de gravação, obtidos através de um estudo piloto; pré-processamento utilizando o COLMAP, algoritmo de Structure-from-Motion (SfM); treinamento do modelo com Nerfacto, que une diversas ideias de diferentes artigos de pesquisa; e exportação das malhas usando o método de Poisson. Avaliamos a qualidade dos objetos gerados por meio de métricas de vértices, faces, arestas e normais, mostrando que cada modelo foi gerado em cerca de três horas, com alta fidelidade aos objetos originais. A pesquisa destaca o potencial do NeRF em aplicações como realidade aumentada, jogos e simulações, e sugere direções futuras para melhorias na limpeza automática das malhas, a fim de otimizar ainda mais o processo de criação 3D.

Palavras-chave: NeRF. Tecnologias Imersivas. Modelagem 3D.

1 INTRODUÇÃO

A modelagem 3D pode ser uma tarefa complexa e desafiadora. Além da curva de aprendizado acentuada frequentemente associada a esses ambientes, a criação de objetos 3D com alto nível de detalhamento pode ser demorada, mesmo para profissionais experientes (PAN et al., 2020). Neste contexto, este artigo propõe uma abordagem baseada em campos de radiância neural (NeRF) para automatizar a geração de objetos 3D a partir de vídeos de entrada. Ao simplificar significativamente o processo de criação, essa técnica promete reduzir tanto o tempo quanto o esforço necessários para produzir modelos tridimensionais detalhados. A proposta visa não apenas melhorar a eficiência, mas também expandir as possibilidades de aplicação em áreas como realidade aumentada e virtual, jogos e simulações.

Os Campos de Radiância Neural (NeRF), introduzidos em agosto de 2020, surgiram como uma tecnologia inovadora para gerar visualizações sintéticas detalhadas de cenas a partir de um conjunto de imagens. Essa abordagem transformou a criação de visualizações 3D, permitindo a geração de novas perspectivas e vídeos dinâmicos a partir de ângulos que não estavam presentes nas imagens originais (MILDENHALL et al., 2020). O NeRF utiliza uma representação contínua 5D, que combina

¹ Bolsista Voluntário de Iniciação Científica e Tecnológica. Universidade Tecnológica Federal do Paraná, Santa Helena, Paraná, Brasil; luisresende@alunos.utfpr.edu.br.

² Docente do Curso de Ciência da Computação, Universidade Tecnológica Federal do Paraná, Santa Helena, Paraná, Brasil; naves@utfpr.edu.br.

as três dimensões espaciais (x , y , z) com dois parâmetros que descrevem a direção da visão (θ , ϕ). Uma rede neural profunda é treinada para mapear essas coordenadas 5D em cores RGB e densidades volumétricas, resultando em imagens fotorealistas por meio de renderização volumétrica. O treinamento é realizado de forma diferenciável, permitindo ajustes baseados em imagens e poses de câmera conhecidas, e a codificação posicional é empregada para capturar detalhes complexos da cena (MILDENHALL et al., 2020). Com a evolução dos modelos NeRF, novas possibilidades foram abertas na geração de malhas 3D, permitindo que a tecnologia se estabeleça e se expanda nesse campo, com aplicações promissoras e avanços contínuos na modelagem tridimensional.

2 METODOLOGIA

A metodologia utilizada baseia-se amplamente no Nerfstudio (TANCIK et al., 2023), uma ferramenta avançada que fornece uma API para a criação, treinamento e teste de diversos modelos de algoritmos generativos 3D, como os *Neural Radiance Fields* (NeRF) e o *Gaussian Splatting* (KERBL et al., 2023). Os experimentos foram conduzidos em um dispositivo equipado com a GPU Nvidia Tesla V100.

O processo de geração de objetos 3D com NeRFs foi estruturado em quatro etapas principais.

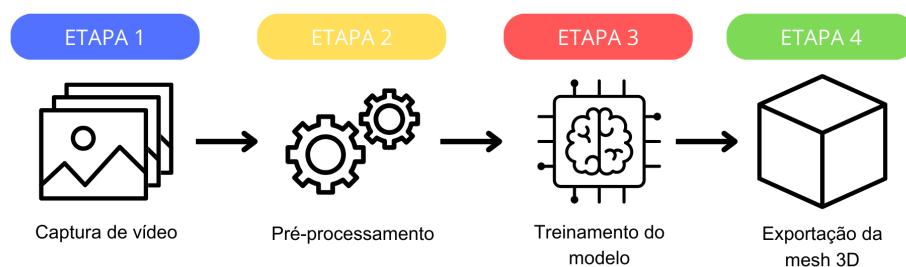


Figura 1 – Representação visual da metodologia. Fonte: Autoria própria (2024).

2.1 Captura do vídeo

Para realizar o treinamento de um modelo NeRF, é fundamental construir uma base de dados sólida, composta por imagens capturadas de diferentes ângulos do objeto que se deseja modelar. No Nerfstudio, é possível utilizar um vídeo como fonte de dados, a partir do qual a ferramenta extrai frames para compor essa base de imagens. No entanto, para garantir a qualidade do resultado, algumas boas práticas de captura devem ser seguidas, já que a qualidade dos frames extraídos pode ser impactada pelo estilo de filmagem e pelas características do equipamento utilizado. Após um estudo piloto onde diversos vídeos foram gravados e processados a fim de encontrar o melhor padrão de gravação, foram estabelecidos alguns requisitos mínimos para a captura de vídeo, garantindo assim, uma base de dados robusta.

Além dos requisitos descritos na Tabela 1, foi definida uma trajetória ideal para a câmera durante a captura do vídeo, visando maximizar a diversidade de ângulos do objeto e garantir uma reconstrução tridimensional de alta qualidade. Durante a gravação, o objeto deve permanecer centralizado no vídeo,

Tabela 1 – Requisitos mínimos para a captura de vídeo

Requisito	Descrição
Resolução	1920x1080 pixels.
Taxa de quadros (FPS)	30 frames por segundo.
Abertura da lente	0.5x
Estabilização	Técnicas de estabilização para evitar movimentos bruscos.

Fonte: Autoria própria (2024).

e a recomendação é realizar três voltas completas ao seu redor, com foco em diferentes partes a cada volta: a primeira volta deve focar na base do objeto, a segunda volta na parte intermediária, e a terceira no topo, como ilustrado na Figura 2. O conjunto de dados utilizado neste estudo é composto por 30 vídeos de diferentes objetos e em diversos ambientes. As gravações foram realizadas com um smartphone Samsung Galaxy A54.

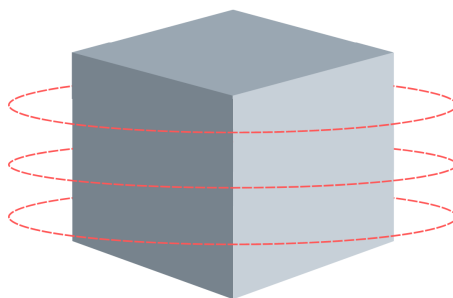


Figura 2 – Trajetória da câmera durante a captura. Os tracejados vermelhos representam a trajetória da câmera, enquanto o cubo representa o objeto que está sendo gravado. Fonte: Autoria própria (2024).

2.2 Pré-processamento

No processo de pré-processamento, o Nerfstudio utiliza o COLMAP, uma ferramenta que implementa um algoritmo de *Structure-from-Motion* (SfM). Essa técnica é usada para reconstruir uma cena tridimensional a partir de múltiplas imagens bidimensionais. O princípio fundamental do algoritmo é identificar correspondências entre imagens que capturam a mesma cena. Com base nessas correspondências, o COLMAP estima as poses das câmeras e a estrutura tridimensional dos pontos presentes na cena (SCHÖNBERGER; FRAHM, 2016).

Como resultado desse processo, o COLMAP fornece uma métrica que indica a porcentagem de imagens para as quais as poses foram corretamente estimadas. O cenário ideal é alcançar 100% de acerto, algo que pode ser conseguido na maioria dos casos ao seguir as boas práticas descritas na Seção 2.1, conforme demonstrado na Tabela 3.

2.3 Treinamento do modelo

O Nerfstudio oferece uma ampla gama de modelos NeRF, mas neste estudo utilizamos exclusivamente o modelo padrão da ferramenta, o Nerfacto. Este modelo destaca-se pela sua integração de múltiplas

ideias de diversos artigos de pesquisa (TANCIK et al., 2023). A capacidade de unir diferentes abordagens em um único modelo contribui significativamente para a sua eficácia e performance.

O Nerfacto também se destaca na geração de malhas 3D ao prever as normais de uma nuvem de pontos durante a reconstrução, oferecendo uma precisão superior em comparação com previsões baseadas na triangulação de um conjunto de pontos, como acontece na maioria dos modelos. Após os primeiros testes alguns parâmetros padrões do algoritmo foram alterados, visando obter resultados de alta qualidade.

Tabela 2 – Parâmetros ajustados para o algoritmo Nerfacto

Parâmetro	Valor Padrão	Valor Alterado	Descrição
–max-num-iterations	30000	45000	Define o número de iterações durante o treinamento.
–model.pipeline.predict-normals	False	True	Habilita a predição de normais.

Fonte: Autoria própria (2024).

2.4 Exportação da mesh 3D

Dentre os métodos de exportação disponíveis no Nerfstudio, o Poisson (KAZHDAN; BOLITHO; HOPPE, 2006) destacou-se nos testes realizados, oferecendo as reconstruções de melhor qualidade.

A aplicação da textura nos objetos é realizada por meio da amostragem densa da imagem de textura, utilizando interpolação baricêntrica para localizar os pontos 3D correspondentes na malha. Em seguida, os raios NeRF são renderizados próximos à superfície, ao longo das normais preditas, para obter os valores RGB (TANCIK et al., 2023).

Por fim, extraímos algumas métricas utilizando a linguagem de programação Python para avaliar a qualidade dos resultados.

3 RESULTADOS E DISCUSSÃO

A Tabela 3 apresenta uma seleção dos resultados obtidos após a etapa de pré-processamento. Os valores correspondem à porcentagem de poses encontradas após o processamento de cada vídeo. Ao observar o desempenho do vídeo intitulado "Calculadora", nota-se que nem todos os vídeos, mesmo seguindo o padrão ideal de captura, atingem 100% de poses encontradas. Já na Figura 3, são exibidos alguns dos melhores resultados finais obtidos. Cada imagem ilustra a renderização do objeto em questão, utilizando o Blender, uma ferramenta de modelagem 3D.

Ao analisar os três primeiros resultados (Flores, Troféu e Xícaras) apresentados na Figura 3, observa-se que, além de gerar a malha do objeto-alvo, o algoritmo também reconstruiu a base sobre a qual o objeto estava apoiado durante a captura, que, nos casos mencionados, corresponde a uma mesa. No quarto resultado da imagem, para evitar essa questão, o objeto foi suspenso por um fio de nylon, resultando em uma malha final sem a inclusão de qualquer base. Embora essa abordagem seja provisória e possa não ser aplicável a todos os cenários, ela se mostra um método válido para contornar o problema.

Tabela 3 – Percentual de poses identificadas durante o pré-processamento em uma amostra do dataset

Título do vídeo	Percentual de poses encontradas
Banjo	100%
Calculadora	0.58%
Xícaras	100%
Pandeiro	100%
Troféu	100%
Flores	100%

Fonte: Autoria própria (2024).

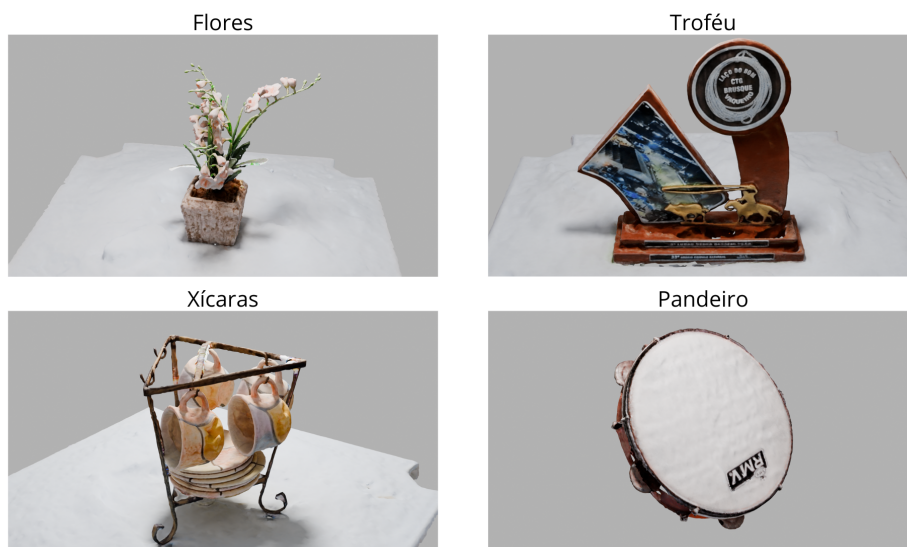


Figura 3 – Renderização dos resultados: Flores, Troféu, Xícaras e Pandeiro. Fonte: Autoria própria (2024).

Uma vez que os resultados obtidos se referem a reconstruções 3D de objetos, a metrificação da qualidade visual torna-se uma tarefa complexa. Nesse contexto, a Tabela 4 apresenta diversas métricas que possibilitam a comparação entre os objetos reconstruídos, oferecendo uma abordagem mais objetiva para a avaliação de suas características visuais.

Tabela 4 – Métricas das malhas 3D geradas

Métrica	Pandeiro	Xícaras	Banjo	Flores	Troféu
Número de Vértices	323487	50051	39093	99125	49141
Número de Faces	637219	98085	76453	194024	95766
Área Total	1601094.71	181044.87	182315.46	382062.85	100137.98
Área Média por Face	2.51	1.85	2.38	1.97	1.05
Densidade de Vértices	0.202	0.276	0.214	0.259	0.491
Comprimento Médio das Arestas	2.57	2.20	2.42	2.29	1.62
Normal Média das Faces	0.034	0.071	0.031	0.036	0.063
Variância das Normais	0.333	0.332	0.333	0.333	0.332

Fonte: Autoria própria (2024).

A análise das métricas obtidas revela diferenças significativas na qualidade e na complexidade das malhas 3D reconstruídas. O número de vértices e faces varia amplamente entre os objetos, com

o "pandeiro" apresentando a maior quantidade, o que pode indicar um nível de detalhe superior em comparação aos outros modelos. A densidade de vértices, que reflete a compactação da malha, varia de 0,202 a 0,491, sugerindo que alguns objetos, como o "Troféu", têm uma estrutura mais densa, potencialmente resultando em uma representação mais fiel. A área média por face também apresenta variações, o que pode impactar a suavidade das superfícies e a complexidade visual. Além disso, as métricas das normais das faces mostram uma baixa média, o que pode indicar a necessidade de ajustes para garantir a uniformidade na iluminação e nos reflexos, essencial para uma representação fotorealista. Em suma, os dados obtidos oferecem insights valiosos sobre a qualidade das malhas e suas possíveis aplicações em contextos que exigem alta fidelidade visual.

4 CONSIDERAÇÕES FINAIS

Conclui-se que a aplicação de Campos de Radiância Neural para a criação de objetos 3D demonstrou ser viável, permitindo a geração de representações que se assemelham aos objetos de referência. Para trabalhos futuros, recomenda-se explorar abordagens de limpeza automática dos objetos resultantes, com o objetivo de aprimorar e automatizar ainda mais o processo de geração de malhas tridimensionais.

AGRADECIMENTOS

Ao Laboratório de Aprendizado de Máquina Aplicado à Indústria - LAMIA¹, pelo fomento à pesquisa e suporte. A Universidade Tecnológica Federal do Paraná, pelo apoio fundamental e recursos disponibilizados para a realização deste trabalho de pesquisa.

REFERÊNCIAS

- KAZHDAN, Michael; BOLITHO, Matthew; HOPPE, Hugues. Poisson surface reconstruction. In: PROCEEDINGS of the Fourth Eurographics Symposium on Geometry Processing. Cagliari, Sardinia, Italy: Eurographics Association, 2006. (SGP '06), p. 61–70. ISBN 3905673363.
- KERBL, Bernhard et al. 3D Gaussian Splatting for Real-Time Radiance Field Rendering. **ACM Transactions on Graphics**, v. 42, n. 4, jul. 2023. Disponível em: [🔗](#).
- MILDENHALL, Ben et al. **NeRF: Representing Scenes as Neural Radiance Fields for View Synthesis**. [S.l.: s.n.], 2020. arXiv: [2003.08934 \[cs.CV\]](#). Disponível em: [🔗](#).
- PAN, Dongdong et al. 3D scene and geological modeling using integrated multi-source spatial data: Methodology, challenges, and suggestions. **Tunnelling and Underground Space Technology**, Elsevier, v. 100, p. 103393, 2020.
- SCHÖNBERGER, Johannes Lutz; FRAHM, Jan-Michael. Structure-from-Motion Revisited. In: CONFERENCE on Computer Vision and Pattern Recognition (CVPR). [S.l.: s.n.], 2016.
- TANCIK, Matthew et al. Nerfstudio: A Modular Framework for Neural Radiance Field Development. In: ACM SIGGRAPH 2023 Conference Proceedings. [S.l.: s.n.], 2023. (SIGGRAPH '23).

¹ <https://www.lamia-edu.com>