

Gaussian Splatting SLAM

Revisor

Daniel Morais

danielmduarte@gmail.com

Arqueóloga

Fernanda Viviani

fernandaviviani.professional@gmail.com

Hacker

Thiago Baldivieso

thiagojmb@gmail.com

Doutorando

Thiago Baldivieso

thiagojmb@gmail.com

1. Revisão

O artigo apresenta uma contribuição significativa para os campos de visão computacional e computação gráfica através do desenvolvimento de um dos primeiros sistemas SLAM baseados em Gaussian Splatting 3D.

1.1. Problema Abordado

O artigo aborda as limitações fundamentais dos sistemas SLAM existentes, especialmente em cenários monoculares. Os sistemas atuais enfrentam desafios significativos na representação de cenas 3D, frequentemente comprometendo a qualidade visual pela velocidade de processamento, ou vice-versa. Além disso, as representações tradicionais têm dificuldades específicas com objetos transparentes e estruturas finas.

1.2. Motivação

A motivação principal dos autores foi de desenvolver um sistema SLAM que possa oferecer reconstrução 3D de alta fidelidade em tempo real usando apenas uma câmera monocular. O objetivo foi de superar as limitações das representações existentes (como grades de voxels, malhas e campos neurais) que apresentam problemas com uso de memória, eficiência computacional ou qualidade de renderização.

1.3. Resumo do Método

O método proposto utiliza Gaussian Splatting 3D como única representação para todo o pipeline SLAM. O sistema:

- Implementa um Jacobiano analítico no grupo de Lie para

otimização direta da pose da câmera, permitindo rastreamento robusto.

- Utiliza regularização isotrópica da forma Gaussiana para manter consistência geométrica durante a reconstrução incremental.
- Emprega um sistema de alocação de recursos e poda Gaussiana para manter a qualidade da geometria e permitir rastreamento preciso.
- Integra estas técnicas em um pipeline que opera a 3 fps, permitindo reconstrução 3D em tempo quase real.

1.4. Contribuições

As principais contribuições do artigo são:

- O primeiro sistema SLAM em tempo quase real usando exclusivamente representação Gaussian Splatting 3D, funcionando com entrada monocular.
- Desenvolvimento de novas técnicas dentro da estrutura SLAM, incluindo otimização de pose por Jacobiano analítico e regularização isotrópica.
- Demonstração de desempenho superior em diversos conjuntos de dados, especialmente em cenários do mundo real.

Este trabalho representa uma contribuição substancial para o campo, pois:

- Introduz uma nova abordagem para SLAM que supera limitações fundamentais das técnicas existentes.
- Demonstra resultados estado-da-arte em reconstrução 3D e rastreamento de câmera.
- Oferece uma solução unificada que pode lidar com casos desafiadores como objetos transparentes e estruturas finas.
- Estabelece uma nova direção para pesquisa em SLAM, combinando eficiência computacional com alta qualidade

de reconstrução.

A contribuição do artigo para a área de visão computacional é particularmente significativa considerando que o método funciona com entrada monocular, o cenário mais desafiador em SLAM, e ainda assim alcança resultados competitivos com sistemas que utilizam sensores mais sofisticados.

1.5. Método

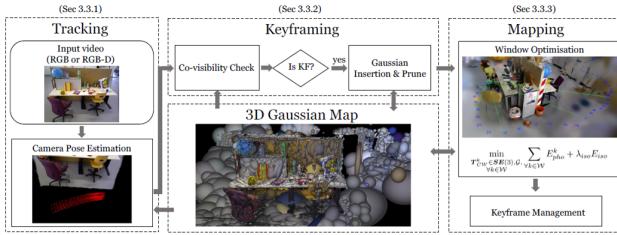


Figure 1. SLAM System Overview. Fonte: Autor.

O método (Figura 1) apresentado no artigo pode ser analisado através de suas principais componentes e fórmulas matemáticas. Vou examinar cada aspecto fundamental e suas justificativas.

A fórmula fundamental para síntese de cor de pixel (Equação 1) é:

$$c_p = \sum_{i \in \mathcal{N}} c_i \alpha_i \prod_{j=1}^{i-1} (1 - \alpha_j)$$

Esta equação representa como os pixels são sintetizados através da combinação de Gaussianas 3D. A clareza desta fórmula poderia ser melhorada com:

- Exemplos visuais mostrando como diferentes valores afetam o resultado
- Uma derivação passo a passo de como a fórmula foi obtida

As equações 2 mostram a transformada projetiva do modelo, ou seja, a relação entre Gaussianas 3D e 2D:

$$\mu_I = \pi(T_{CW} \cdot \mu_W), \Sigma_I = \mathbf{J} \mathbf{W} \Sigma_W \mathbf{W}^T \mathbf{J}^T$$

Estas fórmulas são fundamentais para o processo de projeção, mas sua apresentação poderia ser aprimorada com:

- Diagramas ilustrando a transformação geométrica

- Uma explicação mais clara do significado físico de cada termo
- Demonstração de como estas equações se relacionam com o processo de estimada projetiva

O artigo apresenta as equações 7 e 8 para resíduos fotométricos e geométricos no processo de otimização de pose da câmera:

$$E_{pho} = \|I(\mathcal{G}, T_{CW}) - \bar{I}\|_1$$

$$E_{geo} = \|D(\mathcal{G}, T_{CW}) - \bar{D}\|_1$$

A exposição destas fórmulas é adequada, mas poderia ser melhorada com:

- Uma explicação mais detalhada do processo de otimização
- Visualizações do processo de convergência

A equação 10 para regularização isotrópica é:

$$E_{iso} = \sum_{i=1}^{|G|} \|\mathbf{s}_i - \tilde{\mathbf{s}}_i \cdot \mathbf{1}\|_1$$

Esta fórmula poderia ser mais clara se incluisse:

- Comparação com outras possíveis abordagens de regularização
- Exemplos visuais mostrando o efeito da regularização

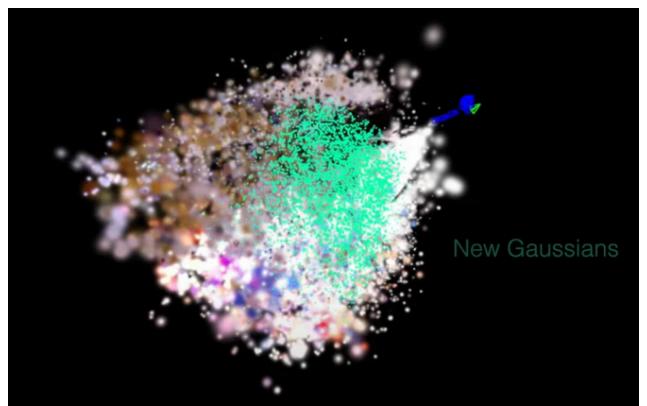


Figure 2. Gaussians pruning. Fonte: Autor.

Em linhas gerais, o artigo poderia incluir mais exemplos práticos para exposição do encadeamento do uso estatístico das fórmulas, dado que o modelo se baseia na análise das

médias dos keypoints para a execução do método de pruning do modelo (Figura 2).

1.6. Pontos positivos

- Permite trabalhar com representações mínimas dos parâmetros de movimento (6 graus de liberdade para movimentos 3D)
- Evita problemas de singularidades que podem ocorrer com outras representações
- Permitem representar rotações e transformações de forma computacionalmente eficiente
- Funciona bem com entrada monocular
- Unifica todo o pipeline de SLAM em uma única representação

1.7. Pontos negativos

- A qualidade final ainda não é perfeita, mas já é um grande avanço no campo
- A introdução de novas gaussianas a cada keyframe pode tornar o sistema lento
- Ainda não é capaz de lidar com cenas em larga escala.

1.8. Avaliação

A partir das considerações aqui expostas, consideramos que o artigo deve ser classificado com nota 5 dada a importância dos resultados e a relevância do método construído para a área de SLAM em visão computacional e robótica.

O tipo de abordagem de se construir um modelo 3D à semelhança do trabalho de um escultor, lapidando a forma 3D a cada keyframe, é de fato muito interessante e empolgante de se observar.

2. Arqueóloga

2.1. Onde o artigo se encaixa no contexto de trabalhos anteriores e posteriores

O método apresentado pelo paper adota uma abordagem centrada em mapas utilizando Gaussianas 3D como representação SLAM. Ao contrário de métodos como Elastic Fusion e Point Fusion, ao usar rasterização diferenciável, seu sistema Slam pode capturar detalhes de cena com alta fidelidade e representar propriedades desafiadoras de objetos.

Comparado a trabalhos anteriores, a principal fraqueza do Nerf é sua velocidade de treinamento. Desenvolvimen-

tos recentes introduziram estruturas de volumes explícitos, como grade de voxel multirresolução ou funções hash para melhorar o desempenho. Neste contexto, o artigo mostra as vantagens do 3DGS porque, em contraste com o Nerf e outras técnicas anteriores, ele realiza rasterização diferenciável e arquiva uma representação que é expressiva para capturar cenas 3D de alta fidelidade, ao mesmo tempo em que oferece renderização significativamente mais rápida.

Dentro do contexto de artigos posteriores podemos mencionar o fato de que o paper de acordo com o Google Scholar foi citado por outros 146 artigos desde a sua data de publicação. Listamos cronologicamente os principais assuntos relacionados com o paper e contidos em outros trabalhos:

Eve Splatting (2002)

Slam (2012)

Real-time 3D Reconstruction (2017)

RGB-D Slam(2017)

Mapping (2017)

Nerf (2020)

Imap (2021)

3d gaussian splatting (2023)

Generative gaussian splatting for efficient 3d content creation (2024)

4d gaussian splatting (2024)

Além disso, existem artigos citados e publicados em 2024 que os tópicos apresentam uma evolução ao tema central proposto pelo paper porque abordam 4D Gaussian Splatting: 4d gaussian splatting for real-time dynamic scene rendering (CVPR) [1] E Real-time photorealistic dynamic scene representation and rendering with 4d gaussian splatting (ICLR) [2].

2.2. Encontrar e relatar sobre um artigo mais antigo citado pelo atual e explicar como ele se relaciona:

(E. Sucar, S. Liu, J. Ortiz, and A. J. Davison. iMAP: Implicit mapping and positioning in real-time. In Proceedings of the International Conference on Computer Vision (ICCV), 2021.)

Recentemente, além das primitivas gráficas clássicas, as representações de mapas baseadas em redes neurais são uma alternativa promissora. O iMAP demonstrou as pro-

priedades interessantes da representação neural, como o preenchimento sensível de buracos de geometria não observada. O paper mais antigo que analisamos foi publicado em 2021. O algoritmo iMAP usa uma estrutura de keyframe e um fluxo de computação multiprocessado, com amostragem de pixels guiada por informações dinâmicas para maior velocidade.

2.3. Artigo mais recente que cita o atual

Wu, T., Yuan, YJ., Zhang, LX. et al. Recent advances in 3D Gaussian splatting. Comp. Visual Media 10, 613–642 (2024).

Para estabilizar a localização e o mapeamento, o Gaussian Splatting SLAM usa uma perda de regularização de escala Gaussiana extra para encorajar Gaussianos isotrópicos. A comparação quantitativa dos novos resultados de síntese de visualização, obtidos por diferentes métodos SLAM no conjunto de dados Replica, utilizando as métricas PSNR, SSIM e LPIPS, demonstra que o método apresentado neste artigo, GS-SLAM, alcançou resultados superiores em relação aos demais abordados no artigo posterior (Recent Advances in 3D Gaussian Splatting).

2.4. Referências adequadas

O paper não cita referências importantes como EWA Volume Splatting [3]. Além disso, outro paper publicado no mesmo ano no CVPR [4] foi além das suas referências, incluindo outros papers com temas relevantes sobre o assunto como flexible techniques for differentiable rendering with 3d [5].

2.5. Referências Bibliográficas utilizadas pela Arqueóloga:

[1] Guanjun Wu, Taoran Yi, Jiemin Fang, Lingxi Xie, Xiaopeng Zhang, Wei Wei, Wenyu Liu, Qi Tian, and Xinggang Wang. 4d gaussian splatting for real-time dynamic scene rendering. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2024

[2] Zeyu Yang, Hongye Yang, Zijie Pan, Xiatian Zhu, and Li Zhang. Real-time photorealistic dynamic scene representation and rendering with 4d gaussian splatting. Proceedings of the International Conference on Learning Representations (ICLR), 2024

[3] Matthias Zwicker, Hanspeter Pfister, Jeroen van Baar, and Markus H. Gross. Caused volume splatting. Proceedings Visualization, 2001. VIS '01., pages 29–538, 2001

[4] Yan, Chi, et al. "Gs-slam: Dense visual slam with 3d gaussian splatting." Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2024.

[5] Leonid Keselman and Martial Hebert. Flexible techniques for differentiable rendering with 3d gaussians. arXiv preprint arXiv:2308.14737, 2023.

3. Código e experimentos

O Hacker precisa fornecer um DEMO do paper o mais rápido possível. Logo, deve avaliar a reproduibilidade do método e implementar uma pequena parte do artigo ou uma versão bem simplificada (eg. 2D em vez de 3D).

- O trabalho poderia ser reproduzido por um ou mais estudantes de pós-graduação? Execute o código do github associado ao paper escolhido e teste em outros datasets;
- Compare as formulas implementadas no código com as equações do paper. Todos os detalhes importantes de algoritmos ou sistemas são discutidos adequadamente?
- "Rode" os experimentos apresentados no paper;
- (Adicional) Pensar em outros experimentos. Discutir a possibilidade no discord primeiro (Participação do Doutorand@).

4. Projeto de doutorado

4.1. Título: Aplicação de GS-SLAM para Preservação Ambiental e Documentação de Patrimônio Histórico

A preservação ambiental e a documentação de patrimônio histórico são questões de crescente relevância na sociedade contemporânea, especialmente em face das mudanças climáticas e da urbanização acelerada [6]. Este projeto de doutorado propõe a aplicação do método GS-SLAM (Gaussian Splatting SLAM) com uma abordagem inovadora para a reconstrução tridimensional de ambientes, visando aprimorar a captura e análise de cenas em contextos outdoor.

O GS-SLAM representa um avanço significativo em relação a métodos tradicionais de SLAM, utilizando uma representação gaussiana 3D que permite a fusão de dados em tempo real e a adaptação do modelo às condições variáveis do ambiente. Essa capacidade de otimização em tempo real possui relevância para a captura de dados em ambientes pequenos como apresentado no artigo, para ambientes outdoor existem outras problemáticas, onde fatores como iluminação e obstruções podem impactar a qualidade da reconstrução tridimensional.

Objetivos do Projeto

- Desenvolver um Protocolo de Captura: Criar um protocolo que utilize drones equipados com câmeras de alta resolução para capturar imagens de áreas ambientais sensíveis e construções históricas. O GS-SLAM será

utilizado para processar essas imagens em tempo real, gerando modelos 3D precisos.

- Análise de Impacto Ambiental: Utilizar os modelos 3D gerados para analisar mudanças em ecossistemas ao longo do tempo, permitindo a identificação de áreas que necessitam de intervenção para preservação.
- Documentação de Patrimônio Histórico: Aplicar a tecnologia para documentar e preservar digitalmente construções históricas, criando um registro que pode ser utilizado para restaurações futuras e para fins educacionais.

Viabilidade: Parceria com o LIARC (Laboratório de Inteligência Artificial e Cibernética) do IME (Instituto Militar de Engenharia) que contém apoio tecnológico de dispositivos robóticos, drones da Marca DJI como Mavic 3 enterprise thermal, Matrice 350 (com cameras intercambiáveis que possuem tecnologias de LiDAR e cameras próprias para mapeamento, drones de uso semi profissional como mini 2 SE e Mini 3 Pro, computadores com placas gráficas de alto desempenho e pesquisadores da área para apoio.

Metodologia

A metodologia será dividida em três etapas principais:

- Captura de Dados: Utilizar drones para realizar voos sobre áreas selecionadas, capturando imagens em alta resolução. O GS-SLAM será implementado para processar essas imagens, permitindo a reconstrução em tempo real.
- Processamento e Análise: Os dados capturados são processados utilizando o pipeline de renderização de splatting diferenciável proposto no artigo. Isso permitirá a criação de modelos 3D que serão analisados para identificar mudanças ambientais e documentar características arquitetônicas.
- Validação e Disseminação: Os resultados são validados através de comparações com dados históricos e análises de especialistas em preservação ambiental e arquitetura. Os modelos 3D resultantes serão disponibilizados em plataformas digitais para acesso público e uso educacional.

Justificativa:

A realização deste projeto é viável devido as inovações apresentadas no artigo "Gaussian Splatting SLAM". A capacidade de otimizar a captura e reconstrução de cenas complexas em tempo real é fundamental para a eficácia do projeto, permitindo uma abordagem mais dinâmica e responsiva às necessidades de preservação. Além disso, a representação gaussiana 3D oferece uma solução robusta

para lidar com a complexidade das cenas outdoor, que muitas vezes apresentam desafios significativos para métodos tradicionais de SLAM.

Contribuições: Este projeto de pesquisa não apenas contribuirá para a preservação ambiental e a documentação de patrimônio histórico, mas também demonstrará a aplicabilidade prática das inovações trazidas pelo GS-SLAM. A combinação de tecnologia avançada e objetivos de conservação pode resultar em um impacto significativo na forma como interagimos e protegemos nosso ambiente e herança cultural.

4.2. Referências Bibliográficas

- [6] Perles, A., Fuster-López, L., Bosco, E. (2024). Preventive conservation, predictive analysis and environmental monitoring. *Heritage Science*, 12, Article number: 11.

5. Conclusões

O paper apresenta uma aplicação de 3D Gaussian Splatting no SLAM monocular. O método deles, que opera em tempo real a 3 fps, utiliza Gaussianas como única representação 3D, unificando as etapas necessárias para rastreamento preciso e eficiente, mapeamento e renderização de alta qualidade. A abordagem, projetada para cenários desafiadores com câmera monocular, também pode ser estendida para SLAM RGB-D quando um sensor de profundidade externo está disponível. Eles destacam que várias inovações são necessárias para reconstruir cenas 3D continuamente com alta fidelidade em tempo real.

Além disso, aproveitaram a natureza explícita das Gaussianas para introduzir verificação geométrica e regularização, capazes de lidar com ambiguidades presentes na reconstrução densa incremental em 3D. Por fim, desenvolveram um sistema completo de SLAM que atinge resultados de última geração na reconstrução de objetos pequenos, incluindo os que são transparentes.