

Comparison of Low-Cost 3D Reconstruction Techniques for Change Detection in Indoor Environments

Comparação de Técnicas de Reconstrução 3D de Baixo Custo para Detecção de Mudanças em Ambientes Internos

Walter Bueno de Brito Neto¹, Michel M. Silva¹, Thiago L. Gomes^{1*}

Resumo: This work evaluates low-cost 3D capture techniques for change detection in indoor environments. We compared a traditional Structure-from-Motion method (COLMAP) with a deep learning approach (CUT3R), using RGB images captured in two sessions of the same laboratory, one with visual changes and one without. The pipeline included 3D reconstruction, point cloud post-processing, alignment, and point-to-point (Cloud-to-Cloud) change detection. Results show that CUT3R produces visually richer reconstructions with significantly lower processing time but suffers from scale inconsistencies and object duplication, compromising precise change detection. COLMAP, although slower, maintains spatial coherence and provides more reliable identification of alterations. These findings indicate that traditional methods are still preferable for rigorous change detection, while deep learning approaches are advantageous for scenarios requiring fast or visually detailed reconstructions.

Keywords: 3D reconstruction — change detection — point clouds — deep learning

Resumo: Este trabalho avalia técnicas de baixo custo para captura 3D aplicadas à detecção de mudanças em ambientes internos. Foram comparados um método tradicional de reconstrução baseado em Structure-from-Motion (COLMAP) e uma abordagem de deep learning (CUT3R), utilizando imagens RGB obtidas em duas coletas do mesmo laboratório, realizadas com e sem alterações visuais. O pipeline incluiu reconstrução 3D, pós-processamento das nuvens de pontos, alinhamento e detecção de mudanças ponto a ponto (Cloud-to-Cloud). Os resultados mostram que o CUT3R gera reconstruções visualmente mais completas e com tempo de processamento significativamente menor, mas apresenta inconsistências de escala e duplicação de elementos, comprometendo a detecção precisa de mudanças. O COLMAP, embora mais lento, mantém coerência espacial e oferece maior confiabilidade na identificação de alterações. Esses achados indicam que métodos tradicionais ainda são mais adequados para detecção de mudanças rigorosa, enquanto abordagens de deep learning podem ser vantajosas em cenários que exigem reconstruções rápidas ou visualmente detalhadas.

Palavras-Chave: reconstrução 3D — detecção de mudanças — nuvens de pontos — deep learning

¹ Departamento de Informática, Universidade Federal de Viçosa (UFV), Brasil

*Corresponding author: thiago.luange@ufv.br

DOI: <http://dx.doi.org/10.22456/2175-2745.150907> • Received: 14/10/2025 • Accepted: 27/10/2025

CC BY-NC-ND 4.0 - This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License.

1. Introdução

A análise de mudanças em ambientes por meio de técnicas de captura 3D tem ganhado destaque como uma área de pesquisa relevante [1, 2, 3], impulsionada pelo desenvolvimento de sensores, algoritmos e plataformas capazes de adquirir dados tridimensionais com alta precisão. Essas tecnologias permitem identificar alterações estruturais em ambientes pela comparação das reconstruções tridimensionais (3D) em momentos diferentes, facilitando aplicações em áreas como monitoramento urbano, robótica autônoma e gestão ambiental.

No contexto das cidades, detectar mudanças geométricas

é crucial para monitorar obras em andamento e identificar construções irregulares [4, 1]. Em sistemas de vigilância autônoma, algoritmos que analisam a densidade espacial podem identificar alterações em cenários, mesmo sob a influência de ruídos [5]. Na robótica, sensores como câmeras estéreo e escâneres a laser (LiDAR, do inglês *Light Detection And Ranging*) são amplamente usados para reconstrução de cenas em tempo real e detecção de mudanças [6]. Métodos modernos, muitas vezes apoiados por algoritmos de inteligência artificial, possibilitam a criação de mapas métrico-semânticos, otimizando a percepção e a navegação em cenários dinâmicos [7].

Para captura 3D, diversas abordagens vêm sendo empregadas, como o uso de LiDAR [8], câmeras de alta resolução e sensores de profundidade (e.g., Intel RealSense) [9]. Essas ferramentas oferecem variações em termos de custo, precisão e aplicabilidade, exigindo uma análise criteriosa para selecionar a melhor opção conforme os objetivos. Entre os métodos baseados em imagens, destacam-se a inferência de estrutura através do movimento (SfM, do inglês *Structure-from-Motion*) [10] e o múltiplas visões capturadas com par estéreo (MVS, do inglês *Multi-View Stereo*) [11], que transformam imagens bidimensionais em modelos tridimensionais.

Apesar desses avanços, ainda existem desafios importantes na área. Enquanto métodos baseados em LiDAR se destacam pela alta precisão, seu custo elevado e complexidade de operação limitam sua aplicação em projetos menores. Por outro lado, técnicas baseadas em imagens, como SfM e MVS, enfrentam dificuldades em ambientes de baixa textura visual e demandam um grande poder de processamento e tempo para realizar as reconstruções.

Nas últimas décadas, o campo da reconstrução 3D tem sido significativamente impactado pelos avanços em técnicas baseadas em aprendizado profundo [12, 13], com destaque inicial para as redes neurais convolucionais (CNNs, do inglês *Convolutional Neural Networks*), amplamente empregadas em tarefas como classificação, detecção de objetos e reconstrução tridimensional. Mais recentemente, os *Transformers* visuais [14] surgiram como uma alternativa promissora às arquiteturas tradicionais, sendo inspirados nos *Transformers* utilizados no processamento de linguagem natural. Nesse contexto, novos modelos de aprendizado profundo como CUT3R [15] se propõem a realizar reconstruções 3D contínuas a partir de múltiplas imagens, unificando percepção e reconstrução em um único *pipeline*.

No entanto, ainda há uma escassez de estudos comparativos aprofundados que avaliem não apenas a qualidade visual das reconstruções geradas por novos modelos, mas também sua aplicabilidade em tarefas de detecção de mudanças. Diante dessa lacuna, este trabalho investiga a viabilidade de técnicas de reconstrução 3D de baixo custo, baseadas exclusivamente em imagens RGB, voltadas à detecção de mudanças em cenários com captura não controlada. Para isso, adotamos uma abordagem fundamentada em nuvens de pontos, que envolve etapas de remoção de ruídos, alinhamento e comparação ponto a ponto, permitindo uma avaliação comparativa entre métodos de aprendizado profundo e abordagens clássicas, com o intuito de fornecer subsídios para aplicações práticas de monitoramento e análise de alterações tridimensionais.

2. Trabalhos Relacionados

Nesta seção, listamos trabalhos nas áreas relacionadas ao trabalho proposto, Detecção de Mudança e Reconstrução 3D.

2.1 Detecção de Mudança

A detecção de mudança é referida por Singh *et al.* [16] como o processo de identificar diferenças no estado de um objeto

ou fenômeno ao observá-lo em momentos distintos. A detecção de mudança 3D, por sua vez, é um subconjunto dessa abordagem e constitui um tópico relativamente novo. Esse campo tem ganhado destaque devido à crescente facilidade de acesso a dados 3D, como nuvens de pontos e modelos digitais de elevação (DEM, do inglês *Digital Elevation Models*), que fornecem posições e formatos tridimensionais explícitos dos objetos. De acordo com o extenso estudo realizado por Qin *et al.* [17], os métodos de detecção de mudanças em dados 3D geralmente seguem três etapas principais: aquisição ou seleção dos dados, co-registro das diferentes fontes ou períodos de aquisição, e análise e identificação das mudanças ocorridas. Essas etapas são fundamentais para garantir a consistência espacial entre os dados e a acurácia na detecção de alterações. As duas primeiras etapas são consideradas fases de pré-processamento, responsáveis pela geração e pelo alinhamento de dados 3D adquiridos em diferentes momentos, visando à posterior identificação e avaliação das mudanças. A etapa de aquisição dos dados envolve, em geral, dois tipos principais de dados 3D. O primeiro grupo compreende dados com informações tridimensionais explícitas, como nuvens de pontos, modelos 3D e Modelos Digitais de Superfície (DSMs, do inglês *Digital Surface Models*). Esses dados já se encontram em formato 3D e podem ser diretamente utilizados em análises tridimensionais, embora, exijam sensores mais sofisticados e de maior custo. O segundo grupo abrange dados com informações 3D implícitas, como imagens de visão múltipla ou estéreo, que requerem processamento adicional para a obtenção da estrutura tridimensional (Reconstrução 3D), mas que podem ser adquiridos com sensores de menor custo.

Conforme o extensivo estudo de Kharroubi *et al.* [2], os métodos de detecção de mudanças podem ser classificados em três categorias principais: baseados em pontos, baseados em objetos e baseados em *voxels*. Métodos baseados em objetos tendem a ser mais suscetíveis a erros de detecção e classificação, o que os torna pouco adequados ao contexto deste trabalho. Já os métodos baseados em *voxels* são comumente aplicados em tarefas de navegação; entretanto, a definição da granularidade da discretização exerce forte influência sobre a análise, podendo introduzir vieses significativos. Por sua vez, os métodos baseados em pontos, embora conceitualmente mais simples, dependem fortemente da qualidade da nuvem de pontos, ou seja, da fidelidade da reconstrução tridimensional. Diante dessas considerações, este estudo foca na detecção de mudanças a partir de nuvens de pontos, seguindo três etapas principais: remoção de ruídos provenientes do processo de reconstrução, alinhamento das nuvens a um sistema de coordenadas comum e comparação ponto a ponto entre as nuvens.

2.2 Reconstrução 3D

A reconstrução da geometria do espaço 3D a partir de imagens é um problema de visão computacional investigado pela comunidade desde o final do século passado. Com os avanços nas áreas de fotogrametria e visão computacional, surgiram

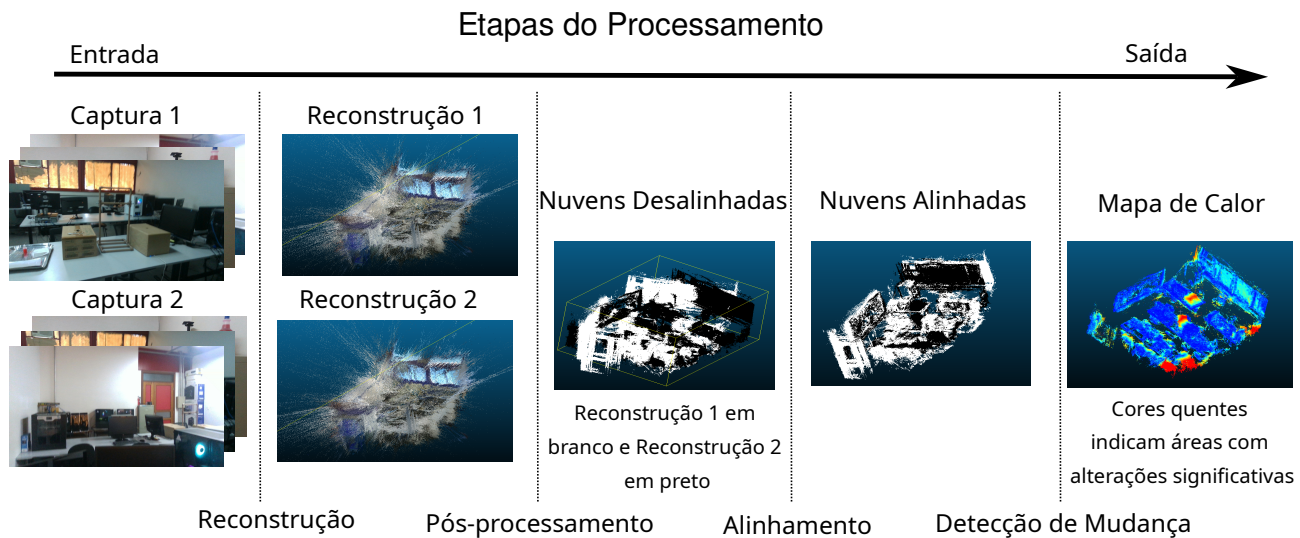


Figura 1. Visão Geral da Metodologia: Duas capturas do ambiente são processadas pelos métodos de reconstrução 3D, limpas, alinhadas globalmente e refinadas com ICP. As mudanças são detectadas comparando-se as nuvens de pontos alinhadas.

métodos automatizados baseados em múltiplas imagens que abordam de diferentes maneiras o desafio da reconstrução 3D [18]. As representações mais comuns da geometria 3D de uma cena, obtida a partir de imagens sobrepostas, incluem mapas de profundidade, nuvens de pontos densas e modelos 3D. Os mapas de profundidade correspondem a imagens 2D em que cada *pixel* codifica a distância em relação ao ponto de captura, enquanto as nuvens de pontos densas consistem em pontos no espaço com coordenadas 3D conhecidas em um sistema de referência. O processo de reconstrução 3D a partir de múltiplas imagens é conhecido como *Multi-View Stereo* (MVS) [11], sendo o software COLMAP [19] uma implementação amplamente utilizada e robusta dessas técnicas.

De forma complementar, o levantamento realizado pelo estudo extensivo de Samavati *et al.* [20] apresenta uma revisão abrangente das abordagens de reconstrução 3D baseadas em aprendizado profundo. Esse estudo classifica os métodos de acordo com critérios diferentes, podendo ser tipo de entrada, imagens únicas e múltiplas, vídeos, etc., também as diferentes representações 3D do modelo como nuvem de pontos, *voxels*, e malhas. Um dos principais destaques do estudo é na capacidade dos métodos recentes de inferirem diretamente a geometria tridimensional a partir de dados visuais, eliminando etapas intermediárias tradicionais como a detecção e correspondência explícita de características [21].

Nesse contexto, o modelo CUT3R [15] representa uma abordagem moderna baseada em aprendizado profundo, capaz de realizar reconstruções 3D contínuas a partir de múltiplas imagens, unificando percepção e reconstrução.

3. Metodologia

Com o objetivo de avaliar a qualidade de técnicas de baixo custo para captura 3D aplicadas à detecção de mudanças, este

trabalho propõe um estudo comparativo entre duas categorias distintas de reconstrução tridimensional. Ambas são submetidas a um processo completo de detecção de mudanças, no qual cada etapa é analisada para identificar as principais características dos métodos de reconstrução utilizados, bem como os desafios envolvidos na aplicação de cada etapa do processo.

O fluxo inicia-se com a aquisição dos dados por meio de uma câmera RGB convencional. As imagens ou vídeos capturados são então processados por um software de reconstrução 3D, resultando em nuvens de pontos representando a geometria da cena. Na etapa seguinte, é realizada a limpeza das nuvens por meio de uma combinação de filtros automáticos para remoção de *outliers* e ajustes manuais visando o refinamento dos dados. Após a filtragem, procede-se ao alinhamento entre diferentes nuvens de pontos. Inicialmente, são aplicadas transformações globais para uma aproximação inicial, seguidas de um refinamento utilizando o algoritmo *Iterative Closest Point* (ICP). Por fim, a detecção de mudanças é realizada com base nas diferenças entre as nuvens alinhadas, gerando como saída um mapa de calor: regiões em tons avermelhados indicam alterações significativas, enquanto áreas em azul indicam pouca ou nenhuma modificação. O fluxo geral do processo está ilustrado na Figura 1.

3.1 Seleção e Pré-processamento dos Dados

O procedimento foi dividido em duas etapas. Na primeira etapa, o operador foi instruído a percorrer todo o ambiente realizando a captura de vídeo de forma contínua, evitando movimentos bruscos ou acelerados com a câmera. Para padronizar a coleta e facilitar a compreensão da tarefa, foi apresentado ao operador um vídeo de exemplo gravado em um ambiente distinto, cuja reconstrução tridimensional resultou em boa qualidade. O objetivo dessa abordagem foi garantir instruções simples e acessíveis, evitando restrições técnicas

complexas que pudessem dificultar a execução por operadores não especializados.

Na segunda etapa, o mesmo ambiente foi modificado com alterações pontuais e controladas, como a remoção ou a realocação de objetos. O operador, então, repetiu o processo de captura com as mesmas instruções da etapa anterior, assegurando consistência entre as duas coletas para posterior comparação.

3.2 Geração do modelo 3D

Nesta etapa, foram adotadas duas abordagens distintas para a geração de modelos 3D: (i) reconstrução com o COLMAP, baseada em SfM e (ii) reconstrução com a abordagem CUT3R, fundamentada em aprendizado profundo. A seguir, descrevem-se os principais aspectos de cada uma delas.

3.2.1 Structure-from-Motion (SfM)

Para gerar modelos 3D a partir de imagens RGB, empregou-se a ferramenta COLMAP, que utiliza o método SfM. A ferramenta integra técnicas de visão computacional, como extração de características, correspondência de pontos, triangulação e *Bundle Adjustment*, possibilitando a criação de modelos tridimensionais precisos. O processo foi iniciado com a detecção de pontos de interesse nas imagens, que foram comparados entre diferentes vistas. Com base nessas correspondências, o COLMAP estimou as posições relativas das câmeras e reconstruiu a estrutura 3D da cena, refinando os resultados por meio de otimizações. Ao final, gerou-se uma nuvem de pontos densa, representando a geometria da cena com alto nível de detalhamento.

3.2.2 Reconstrução com aprendizado profundo (CUT3R)

Para gerar modelos 3D a partir de imagens RGB utilizando aprendizado profundo, foi empregada a ferramenta CUT3R, que é uma abordagem recente de reconstrução 3D baseada em *Transformers* visuais. Para cada imagem de entrada, o modelo gera uma nuvem de pontos correspondente, que pode ser posteriormente unificada em uma representação única da cena. Como se trata de um método que exige grande quantidade de memória da placa gráfica, a seleção das imagens torna-se essencial para garantir a qualidade do modelo final.

3.3 Pós-processamento das Nuvens de Pontos

O pós-processamento é uma etapa fundamental no *pipeline* de reconstrução 3D, uma vez que os modelos gerados frequentemente contêm ruídos, *outliers* e regiões com densidade de pontos inconsistentes. Para mitigar esses problemas e garantir uma base mais confiável para análises subsequentes, foi utilizada a ferramenta CloudCompare para realizar a limpeza das nuvens de pontos.

O processo de limpeza foi conduzido em duas etapas. Inicialmente, realizou-se a remoção manual dos *outliers* externos ao ambiente utilizando uma caixa delimitadora. Essa caixa foi ajustada de modo a englobar exclusivamente a região de interesse. Na etapa seguinte, foi aplicado o filtro *Statistical Outlier Removal* (SOR), com fator de vizinhança $k = 10$. Esse

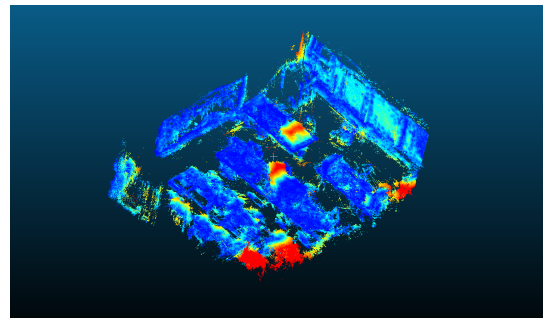


Figura 2. Mapa de calor gerado a partir da comparação C2C entre duas nuvens de pontos completamente alinhadas.

filtro avalia a média das distâncias entre cada ponto e seus k vizinhos mais próximos, removendo aqueles cuja média se desvia significativamente do padrão global. Essa filtragem contribui para uma nuvem de pontos mais coesa, homogênea e com menor nível de ruído.

Embora essas etapas envolvam intervenção manual, elas são essenciais para garantir uma comparação justa entre os métodos analisados e não exigem conhecimentos técnicos avançados para serem executadas de forma adequada.

3.4 Alinhamento tridimensional

Para o alinhamento tridimensional dos modelos 3D, representados como nuvens de pontos, foi adotada uma solução clássica amplamente utilizada na área. Inicialmente, foi realizado um alinhamento global utilizando os descritores *Fast Point Feature Histogram* (FPFH) em conjunto com o algoritmo *RANdom SAMple Consensus* (RANSAC), com o objetivo de estimar uma transformação inicial robusta. Em seguida, essa estimativa será refinada por meio do algoritmo ICP, que requer uma aproximação inicial razoável para garantir uma boa convergência e alinhamento preciso.

3.5 Detecção de Mudança e Avaliação

Nesta etapa, foi utilizado o método *Cloud-to-Cloud* (C2C) para a detecção de mudanças entre as nuvens de pontos. O procedimento consiste em, para cada ponto de uma nuvem, localizar o ponto mais próximo na outra nuvem e calcular a distância euclidiana entre eles. Pontos cuja distância excede um limiar pré-definido são então classificados como alterados.

A Figura 2 ilustra um mapa de calor das regiões com mudanças detectadas. Cores quentes (vermelho e amarelo) indicam áreas com alterações significativas, enquanto cores frias indicam regiões estáveis ou com pouca variação.

Por fim, as reconstruções foram avaliadas considerando múltiplos critérios: tempo de processamento, qualidade visual da reconstrução, custo computacional e eficácia na detecção de mudanças.

4. Experimentos

Esta seção apresenta o ambiente e o processo de coleta dos dados, bem como as entradas, configurações e parâmetros empregados nas reconstruções

4.1 Aquisição de dados e descrição do ambiente

Para os testes, como ambiente a ser escaneado e reconstruído, foi selecionado um laboratório de computação contendo diversos elementos típicos de ambientes internos, como mesas, armários, cadeiras, computadores, monitores e caixas. Como o objetivo era realizar a reconstrução do ambiente, garantiu-se que não houvesse presença de pessoas e outros elementos dinâmicos durante a aquisição dos dados.

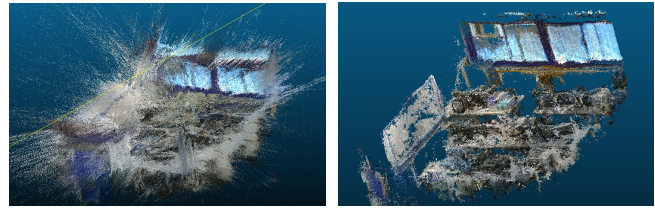
A primeira captura foi realizada com a sala em seu estado original, sem alterações visuais ou estruturais, resultando em aproximadamente 2400 imagens com resolução de 1920×1080 . A segunda captura foi feita no mesmo ambiente, porém com modificações visíveis — principalmente a adição de caixas sobre as mesas — totalizando cerca de 2700 imagens na mesma resolução.

4.2 Reconstruções

Para a reconstrução utilizando SfM, a resolução das imagens foi reduzida de para 1280×720 , mantendo a proporção de aspecto 16 : 9. Essa redução foi necessária devido a limitações de memória VRAM da placa de vídeo utilizada, que resultavam em falhas (*crash*) durante a etapa de MVS. Embora o COLMAP ofereça diversos parâmetros que podem ser ajustados para otimizar a qualidade da reconstrução, permitindo, por exemplo, um melhor equilíbrio entre fidelidade estrutural e estabilidade computacional, neste trabalho optou-se por utilizar os valores padrão. Essa escolha foi motivada pelo objetivo de avaliar a viabilidade da reconstrução em um cenário realista, no qual o operador não necessariamente possui conhecimento técnico aprofundado.

O CUT3R disponibiliza dois *checkpoints* distintos. O *checkpoint* intermediário foi treinado com sequências de apenas 4 imagens e resolução de 224 *pixels*, enquanto o modelo final foi otimizado para processar até 64 imagens com resolução máxima de 512 *pixels*. Conforme demonstrado no artigo original, o modelo final, utilizando 64 imagens e resolução de 512 *pixels*, apresentou os melhores resultados. Por esse motivo, esse cenário foi adotado nos experimentos realizados neste trabalho. As imagens foram amostradas uniformemente a partir da sequência original, de modo a selecionar 64 imagens representativas. A adaptação de resolução foi realizada pelo próprio método, que incorpora internamente as etapas de redimensionamento.

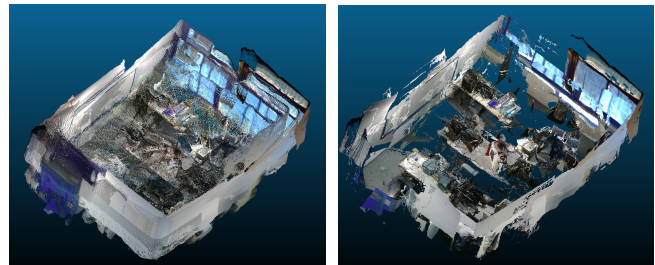
Para a execução dos experimentos, foi utilizada uma estação de trabalho equipada com processador Intel Core i7-12700F, 128 GB de memória RAM e uma GPU NVIDIA RTX 4090 com 24 GB de VRAM, operando sob Ubuntu 22.04 LTS. As imagens foram adquiridas com uma câmera Intel RealSense D435, configurada para capturar quadros RGB-D com resolução de 1920×1080 a 30 FPS, armazenados em modo alinhado para garantir a correspondência entre cor e profundidade. As reconstruções foram realizadas utilizando as ferramentas COLMAP e CUT3R, conforme descrito anteriormente. O COLMAP foi compilado a partir do código-fonte com suporte à arquitetura CUDA 80, permitindo o uso oti-



(a) Reconstrução inicial.

(b) Pós-processamento.

Figura 3. Reconstruções com COLMAP: o ruído inicial é substancialmente reduzido após o pós-processamento.



(a) Reconstrução inicial.

(b) Pós-processamento.

Figura 4. Reconstruções pelo CUT3R: pós-processamento reduz ruído, mas algumas regiões são perdidas por remoção de duplicações.

mizado da GPU. O ambiente para o CUT3R foi configurado com Conda utilizando Python 3.11, CMake 3.14 e PyTorch com suporte a CUDA 12.1.

5. Resultados e Discussão

Nesta seção são apresentados os resultados obtidos a partir dos experimentos, avaliados sob três perspectivas complementares: a qualidade da reconstrução, a eficácia do alinhamento e a eficácia na detecção de mudanças.

5.1 Reconstruções

As Figuras 3a e 3b apresentam a reconstrução gerada pelo COLMAP a partir do conjunto de imagens capturadas. Como mostrado na Figura 3a, a reconstrução inicial contém uma quantidade significativa de ruído. No entanto, após a etapa de pós-processamento, esse ruído é substancialmente reduzido. É importante destacar que a reconstrução apresenta uma boa estrutura espacial, com os objetos posicionados de forma coerente. Por outro lado, regiões com baixa textura, como paredes e o chão, não são reconstruídas adequadamente.

As Figuras 4a e 4b apresentam a reconstrução gerada pelo CUT3R a partir do conjunto de imagens capturadas. As reconstruções apresentam significativamente menos ruído; contudo, exibem efeitos de duplicação dos objetos, como se estes tivessem se movimentado durante a captura. A remoção desse ruído na etapa de pós-processamento resulta na perda de algumas áreas da reconstrução. Esse comportamento pode ser observado na Figura 4b.

A Tabela 1 apresenta uma comparação dos requisitos computacionais entre os métodos avaliados. Observa-se que o

CUT3R possui uma vantagem significativa em termos de tempo de reconstrução, realizando a inferência em aproximadamente 1 minuto, enquanto o COLMAP demanda cerca de 16 horas para completar o processo. Essa diferença ocorre porque o CUT3R realiza inferência direta em GPU, enquanto o COLMAP adota um pipeline clássico baseado em correspondência e reconstrução multivista, fortemente dependente de CPU. Essa discrepância evidencia o potencial do CUT3R para aplicações que exigem reconstruções rápidas. Em relação ao uso de memória, ambos os métodos apresentam consumo semelhante de VRAM, embora o COLMAP possa exigir mais dependendo da resolução e do número de imagens.

Tabela 1. Comparação dos requisitos computacionais dos métodos avaliados. O tempo do CUT3R corresponde à inferência.

Recurso	CUT3R	COLMAP
Tempo	~1 min	~16 h
Uso de VRAM	6.7 GB	6–10 GB
Hardware	GPU	CPU/GPU

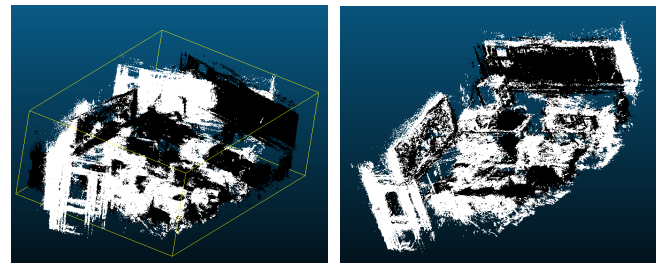
5.2 Alinhamento

Os experimentos de alinhamento evidenciaram que a presença de ruído compromete significativamente a correspondência entre nuvens de pontos. Conforme esperado, os algoritmos de registro falharam quando aplicados diretamente sobre as nuvens brutas, independentemente do método de reconstrução utilizado. Isso ocorre porque o alinhamento espacial, especialmente em abordagens baseadas em ICP, depende de estruturas globais consistentes e pontos de interesse bem definidos, os quais são degradados pelo ruído.

Após a etapa de pós-processamento, as nuvens geradas pelo COLMAP apresentaram alinhamento satisfatório, com sobreposição precisa entre as geometrias, correspondência coerente entre os detalhes estruturais e baixo erro de alinhamento final, conforme ilustrado na Figura 5. Esse resultado indica que, embora o COLMAP produza nuvens ruidosas inicialmente, a estrutura métrica subjacente mantém boa consistência espacial, permitindo um encaixe preciso após a remoção do ruído.

Em contraste, o alinhamento das nuvens geradas pelo CUT3R falhou mesmo após o pós-processamento. A principal causa observada é a inconsistência de escala entre as reconstruções, uma vez que o CUT3R não preserva uma escala absoluta entre diferentes sessões de reconstrução como pode ser visto na Figura 6. Essa limitação inviabiliza o uso direto de algoritmos de registro rígido baseados em transformações isométricas.

Considerando que o objetivo deste trabalho é avaliar os métodos de reconstrução no contexto de detecção de mudanças, a constatação da incompatibilidade de escala nas nuvens geradas pelo CUT3R representa um achado relevante. Essa limitação pode impactar diretamente a aplicabilidade do método em cenários que exigem registro preciso entre reconstru-



(a) Nuvens desalinhadas.

(b) Nuvens alinhadas.

Figura 5. Alinhamento entre nuvens sem mudanças (brancas) e com mudanças (pretas) usando COLMAP. Em (a) estão desalinhadas; em (b), sobreposição precisa evidencia o sucesso do ajuste.

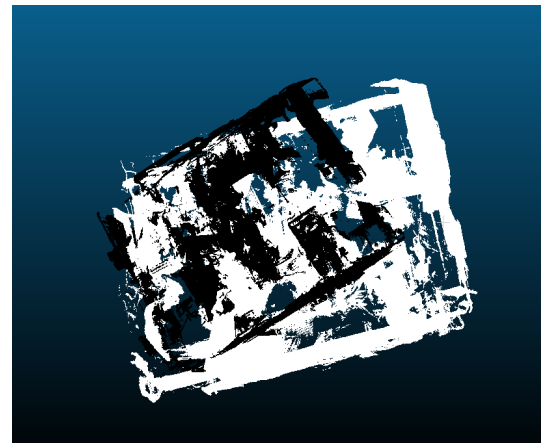


Figura 6. Resultados de reconstrução geradas pelo CUT3R não preservam a escala entre diferentes sessões de reconstrução.

ções. Para viabilizar a continuidade da avaliação, assumiu-se um cenário em que a discrepância de escala foi corrigida manualmente. Com essa normalização aplicada, o processo de alinhamento entre as nuvens geradas pelo CUT3R apresentou resultados satisfatórios, permitindo a análise subsequente de mudanças espaciais.

5.3 Detecção de Mudanças

Apesar da presença de ruídos e da dificuldade em reconstruir regiões com baixa textura, o modelo gerado pelo COLMAP não apresentou duplicação de estruturas. Essa consistência espacial resultou em melhor desempenho na detecção. Como ilustrado na primeira linha da Figura 7, um ajuste adequado do limiar de detecção permitiu que a comparação entre os modelos reconstruídos apresentasse baixos índices de falsos positivos, identificando corretamente as alterações nas caixas posicionadas sobre as mesas no centro da cena.

Em contraste, embora o modelo CUT3R tenha produzido reconstruções visualmente mais completas, observou-se maior ocorrência de duplicação de elementos, como a replicação incorreta de paredes e portas, o que compromete a interpretação espacial da nuvem de pontos. Consequentemente, a detecção de mudanças baseada exclusivamente em diferenças

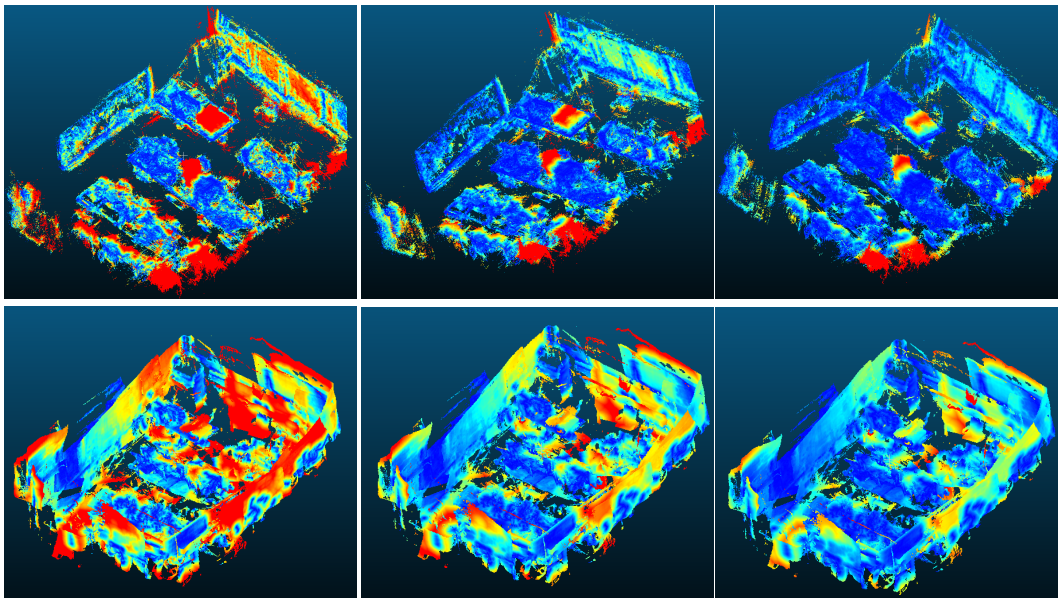


Figura 7. Mapas de calor gerados pelos métodos COLMAP (linha superior) e CUT3R (linha inferior) com três limiares de detecção de mudanças: 0.3, 0.5 e 0.7 (da esquerda para a direita). Limiares mais baixos resultam em maior sensibilidade à detecção de alterações, ampliando a cobertura mas também aumentando a ocorrência de falsos positivos.

resultou em mapas com informações imprecisas, mesmo após um ajuste cuidadoso do limiar, como evidenciado na segunda linha da Figura 7.

Por fim, foram utilizadas métricas de distância, incluindo o Erro Quadrático Médio (MSE, do inglês *Mean Square Error*), a Distância de Chamfer e a média das distâncias ponto a ponto entre as nuvens, para quantificar a diferença entre as reconstruções. Embora as nuvens não sejam idênticas, considera-se que as mudanças entre elas representam apenas uma pequena fração da nuvem total. Assim, valores reduzidos nessas métricas podem ser interpretados como indicativos de reconstruções que preservam melhor a consistência espacial e, conseqüentemente, são mais adequadas para a análise de mudanças. Conforme apresentado na Tabela 2, o método COLMAP obteve o melhor desempenho, apresentando valores menores tanto na comparação da primeira (original) com a segunda nuvem (com mudanças) ($1 \rightarrow 2$) quanto na comparação inversa ($2 \rightarrow 1$).

Tabela 2. Comparação quantitativa entre as reconstruções geradas pelos métodos CUT3R e COLMAP com base nas distâncias entre as nuvens antes e depois das mudanças (valores $\times 10^{-3}$).

Métrica	CUT3R	COLMAP
Chamfer Distance (bidirecional)	111.3	54.9
Média das distâncias ($1 \rightarrow 2$)	124.3	96.5
Média das distâncias ($2 \rightarrow 1$)	197.5	98.2
MSE ($1 \rightarrow 2$)	28.4	27.2
MSE ($2 \rightarrow 1$)	82.9	27.7

6. Conclusão

A análise comparativa evidencia que métodos de aprendizado profundo, como o CUT3R, apresentam grande potencial para reconstruções rápidas e visualmente detalhadas, sendo vantajosos em aplicações em tempo real ou em cenários com grande volume de imagens. No entanto, a ausência de preservação de escala entre diferentes reconstruções limita sua aplicação direta em tarefas de detecção de mudanças, especialmente quando o alinhamento rígido das nuvens é necessário, e a presença de objetos duplicados compromete a identificação precisa de alterações. Por outro lado, métodos tradicionais baseados em SfM demonstram robustez espacial, tornando-os mais adequados para detecção precisa de mudanças, embora exijam maior tempo computacional. Esses resultados sugerem que uma estratégia híbrida, combinando reconstruções rápidas baseadas em aprendizado profundo com ajustes de escala ou refinamento métrico via métodos clássicos, pode otimizar tanto a qualidade quanto a eficiência na análise de alterações tridimensionais. Como limitação deste trabalho, destaca-se que apenas uma metodologia de detecção de mudanças foi testada; abordagens mais robustas, incluindo métodos baseados em redes profundas, podem superar as limitações observadas ou revelar novas restrições. Como trabalho futuro, propõe-se o uso de modelos de referência, como CADs, para permitir uma avaliação mais aprofundada das métricas quantitativas, oferecendo uma análise mais precisa da acurácia das reconstruções e da detecção de mudanças.

Agradecimentos

Os autores agradecem à CAPES, ao CNPq e à FAPEMIG pelo apoio financeiro. Agradecem também ao Instituto de

Inteligência Artificial e Computacional (Idata-UFV), à FINEP (IA-AD-UFV #0284/22) e ao Cluster/UFV pelo fornecimento dos recursos de GPU utilizados neste trabalho.

Contribuição dos Autores

Walter Bueno de Brito Neto foi responsável por todo o desenvolvimento do projeto, experimentos, análises e redação do manuscrito; Michel M. Silva contribuiu com orientação metodológica, discussões técnicas e revisão do manuscrito; Thiago L. Gomes atuou como orientador e contribuiu por meio de supervisão contínua, orientação metodológica, discussões técnicas e revisão do manuscrito.

Referências

- [1] QIN, R. 3d change detection in an urban environment with multi-temporal data. *ETH Zurich*, 2015.
- [2] KHARROUBI, A. et al. Three dimensional change detection using point clouds: A review. *Geomatics*, v. 2, n. 4, p. 457–485, 2022. ISSN 2673-7418. Disponível em: <<https://www.mdpi.com/2673-7418/2/4/25>>.
- [3] XIAO, W. et al. 3d urban object change detection from aerial and terrestrial point clouds: A review. *International Journal of Applied Earth Observation and Geoinformation*, v. 118, p. 103258, 2023. ISSN 1569-8432. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S1569843223000808>>.
- [4] TANEJA, A.; BALLAN, L.; POLLEFEYS, M. Geometric change detection in urban environments using images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, v. 37, p. 2193–2206, 2015.
- [5] VIEIRA, A.; DREWS-JR, P.; CAMPOS, M. Spatial density patterns for efficient change detection in 3d environment for autonomous surveillance robots. *IEEE Transactions on Automation Science and Engineering*, v. 11, p. 766–774, 2014.
- [6] TINOCO, C. R.; TRAVENCOLO, B. A. N.; MARTINS, L. G. A. Evaluation of a coordination model for robot swarms in collective environment mapping. In: *2025 IEEE Congress on Evolutionary Computation (CEC)*. [S.l.: s.n.], 2025. p. 1–8.
- [7] GÖNCZ, L.; MAJDIK, A. Object-based change detection algorithm with a spatial ai stereo camera. *Sensors (Basel, Switzerland)*, v. 22, 2022.
- [8] COLLIS, R. Lidar. *Applied optics*, v. 9, p. 1782–1788, 2019.
- [9] KESELMAN, L. et al. Intel realsense stereoscopic depth cameras. *CoRR*, abs/1705.05548, 2017. Disponível em: <<http://arxiv.org/abs/1705.05548>>.
- [10] ULLMAN, S. The interpretation of structure from motion. *Proceedings of the Royal Society of London. Series B. Biological Sciences*, v. 203, p. 405–426, 1979.
- [11] FURUKAWA, Y.; HERNÁNDEZ, C. Multi-view stereo: A tutorial. *Found. Trends Comput. Graph. Vis.*, v. 9, p. 1–148, 2015.
- [12] GOMES, L.; Regina Pereira Bellon, O.; SILVA, L. 3d reconstruction methods for digital preservation of cultural heritage: A survey. *Pattern Recognition Letters*, v. 50, p. 3–14, 2014. ISSN 0167-8655. Depth Image Analysis. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S0167865514001032>>.
- [13] LUO, Z. et al. The nerf signature: Codebook-aided watermarking for neural radiance fields. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2025.
- [14] DOSOVITSKIY, A. et al. *An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale*. 2021. Disponível em: <<https://arxiv.org/abs/2010.11929>>.
- [15] WANG, Q. et al. *Continuous 3D Perception Model with Persistent State*. 2025. Disponível em: <<https://arxiv.org/abs/2501.12387>>.
- [16] SINGH, A. Review article digital change detection techniques using remotely-sensed data. *International Journal of Remote Sensing*, Taylor & Francis, v. 10, n. 6, p. 989–1003, 1989. Disponível em: <<https://doi.org/10.1080/01431168908903939>>.
- [17] QIN, R.; TIAN, J.; REINARTZ, P. 3d change detection – approaches and applications. *ISPRS Journal of Photogrammetry and Remote Sensing*, v. 122, p. 41–56, 2016. ISSN 0924-2716. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S0924271616304026>>.
- [18] VERYKOKOU, S.; IOANNIDIS, C. An overview on image-based and scanner-based 3d modeling technologies. *Sensors*, Basel, Switzerland, v. 23, n. 2, p. 596, jan 2023.
- [19] SCHÖNBERGER, J. L. et al. A vote-and-verify strategy for fast spatial verification in image retrieval. In: *Asian Conference on Computer Vision (ACCV)*. [S.l.: s.n.], 2016.
- [20] SAMAVATI, T.; SORYANI, M. Deep learning-based 3d reconstruction: a survey. *Artif. Intell. Rev.*, Kluwer Academic Publishers, USA, v. 56, n. 9, p. 9175–9219, jan. 2023. ISSN 0269-2821. Disponível em: <<https://doi.org/10.1007/s10462-023-10399-2>>.
- [21] MANDIKAL, P. et al. 3d-lmnet: Latent embedding matching for accurate and diverse 3d point cloud reconstruction from a single image. *CoRR*, abs/1807.07796, 2018. Disponível em: <<http://arxiv.org/abs/1807.07796>>.