

## AVALIAÇÃO DE MODELOS DE VISUAL SLAM COM ODOMETRIA PARA ROBÓTICA MÓVEL

Celomar Oliveira da Silva Filho<sup>1</sup> (IC), Rafael Francisco dos Santos<sup>2</sup> (PQ)

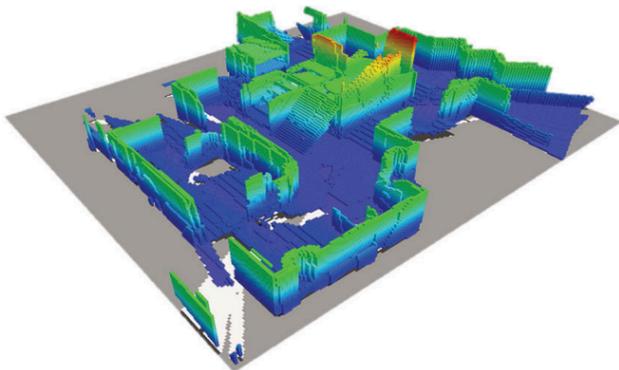
<sup>1,2</sup>Universidade Federal de Itajubá - Campus Itabira

**Palavras-chave:** Visual SLAM, Odometria, Robôs Móveis, ROS.

### Introdução

A robótica móvel é fundamental para sistemas autônomos, com aplicações que vão de fábricas inteligentes à exploração de ambientes desconhecidos, como áreas urbanas ou planetárias. Um dos principais desafios nesse campo é a capacidade dos robôs de se localizar com precisão enquanto mapeiam o ambiente ao redor, o que é conhecido como Simultaneous Localization and Mapping (SLAM). Essa técnica é crucial para que robôs móveis possam operar de forma autônoma em ambientes desconhecidos, criando e atualizando mapas continuamente enquanto se orientam dentro deles, conforme ilustrado na Figura 1.

Figura 1 - Ilustração de mapeamento 3D gerado por técnicas de SLAM.



Fonte: SILVA e YEPES (2015).

A solução tradicional para o SLAM utiliza sensores de alta precisão, como LIDAR ou câmeras monoculares. Porém, essas opções têm limitações: o LIDAR é preciso, mas caro e demanda muito processamento, enquanto câmeras monoculares não fornecem dados de profundidade, o que pode gerar erros em ambientes complexos. Para superar essas limitações, este trabalho foca na odometria visual com câmeras estéreo, que calcula a distância dos objetos ao robô usando a diferença entre duas imagens capturadas simultaneamente por câmeras alinhadas.

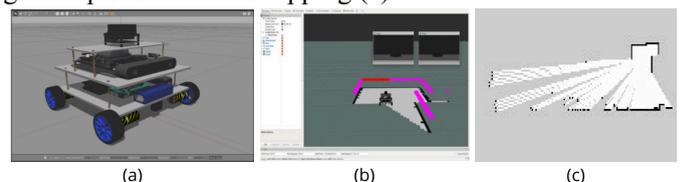
O uso de câmeras estéreo oferece várias vantagens em relação aos sensores tradicionais (BARROS, A. M. et al.). A principal delas é a capacidade de estimar a

profundidade diretamente a partir da disparidade entre as imagens, proporcionando uma percepção espacial mais precisa do ambiente. Além disso, a odometria visual com câmeras estéreo não depende de infraestrutura externa, como GPS, sendo eficaz em ambientes internos ou em locais sem sinal de GPS. A integração da odometria visual com algoritmos de SLAM melhora a navegação, tornando-a mais robusta e precisa, especialmente em cenários dinâmicos ou pouco estruturados, com variações de obstáculos e iluminação.

Para desenvolver e validar os modelos de SLAM com odometria visual, este projeto utilizou o Robot Operating System (ROS), amplamente adotado na pesquisa em robótica. O ROS oferece uma infraestrutura modular que facilita a integração de sensores e algoritmos. Neste trabalho, o ROS foi o núcleo para implementar os algoritmos de SLAM e gerenciar a comunicação entre os módulos de captura de dados visuais e os métodos de estimativa de pose.

Além do ROS, utilizou-se os simuladores Gazebo e RViz para validar modelos em ambientes virtuais. O Gazebo é um simulador 3D que permite criar cenários complexos, testando algoritmos de SLAM em ambientes controlados e dinâmicos. Já o RViz, uma ferramenta de visualização integrada ao ROS, foi usado para monitorar dados sensoriais e verificar a posição do robô em tempo real. Juntos, esses simuladores proporcionaram uma avaliação detalhada do desempenho dos algoritmos antes dos testes práticos com vídeos estéreo em ambientes reais. A Figura 2 ilustra a utilização destes simuladores para realizar a simulação do robô e do ambiente em que ele se encontra, além da aplicação de um método de SLAM para realizar a localização e mapear o ambiente simulado.

Figura 2 - Ilustração de Robô Wally modelado no Gazebo (a), Wally no ambiente de simulação no RViz (b) e mapa gerado por meio do GMapping (c).



Fonte: Silva et al. (2022).

Este trabalho visa otimizar a navegação autônoma de robôs em ambientes complexos por meio da integração de odometria visual e SLAM. Essa abordagem permite que o robô utilize dados visuais para aprimorar a precisão no mapeamento e na localização, oferecendo uma solução de baixo custo e alta eficiência, em comparação com métodos que dependem exclusivamente de sensores de alta precisão, como o LIDAR. O estudo foca na investigação e implementação de diferentes algoritmos de Visual SLAM, utilizando câmeras estéreo e odometria visual, testando seu desempenho em cenários simulados e reais.

A justificativa para este projeto baseia-se na crescente demanda por sistemas de navegação autônoma que operem de forma eficiente em diversos cenários, como ambientes urbanos e industriais, onde obstáculos móveis e condições dinâmicas representam desafios. A implementação de SLAM com odometria visual usando câmeras estéreo oferece uma solução de baixo custo e alta aplicabilidade, já que as câmeras são sensores acessíveis que fornecem informações detalhadas sobre o ambiente.

### Metodologia

Este projeto utilizou-se de três abordagens baseadas em Visual SLAM e, entre eles, foi-se escolhido um algoritmo de cada abordagem para testar a capacidade de desempenho em situações específicas, especialmente no contexto de veículos autônomos, onde a precisão e robustez do mapeamento e localização são fundamentais para garantir a segurança e eficiência do sistema.

#### VISUAL-ONLY SLAM

Os sistemas de SLAM baseados apenas em visão utilizam processamento de imagens 2D. Após a aquisição das imagens de diferentes pontos de vista, o sistema realiza um processo de inicialização para definir um sistema de coordenadas global e reconstruir um mapa inicial. Nos algoritmos baseados em recursos que utilizam filtros (algoritmos baseados em filtragem), a primeira etapa consiste em inicializar os pontos do mapa com alta incerteza, que podem posteriormente convergir para suas posições reais. Após essa etapa, o rastreamento tenta estimar a posição da câmera. Ao mesmo tempo, o processo de mapeamento inclui novos pontos na reconstrução 3D à medida que mais cenas desconhecidas são observadas.

Os sistemas de SLAM baseados apenas em visão podem utilizar câmeras monoculares ou estéreo. O SLAM com câmeras monoculares é amplamente estudado devido ao tamanho reduzido do sensor (o menor entre as abordagens disponíveis), seu baixo custo, facilidade de calibração e menor consumo de energia.

Exemplo de algoritmos: CNN-VSLAM, LSD-SLAM e ORB-SLAM2.

#### VISUAL-INERTIAL SLAM

A abordagem VI-SLAM incorpora medições inerciais para estimar a estrutura e a pose do sensor. Os dados inerciais são obtidos através de uma unidade de medição inercial (IMU), que combina giroscópio, acelerômetro e, adicionalmente, magnetômetro. Dessa forma, a IMU fornece informações sobre a taxa angular (giroscópio) e aceleração (acelerômetro) ao longo dos eixos x, y e z, além do campo magnético ao redor do dispositivo (magnetômetro). Embora a adição de uma IMU enriqueça a quantidade de informações do ambiente e melhore a precisão, ela também aumenta a complexidade do algoritmo, especialmente durante a fase de inicialização, pois, além da estimativa inicial da pose da câmera, o algoritmo também precisa estimar as poses da IMU.

Exemplo de algoritmos: ROVIO, VIORB e ORB-SLAM3.

#### VISUAL-RGB-D

Os sistemas de SLAM baseados em dados RGB-D começaram a ganhar mais atenção com o lançamento do Microsoft Kinect em 2010. Sensores RGB-D consistem em uma câmera monocular RGB e um sensor de profundidade, permitindo que os sistemas de SLAM adquiram diretamente informações de profundidade com precisão viável em tempo real, utilizando hardware de baixo custo. Como os dispositivos RGB-D fornecem diretamente o mapa de profundidade para os sistemas de SLAM, a estrutura geral desse tipo de SLAM difere das abordagens apresentadas anteriormente.

Exemplos de algoritmos: RGBDSLAMv2, SLAM++ e ORB-SLAM2.

#### DATASETS E BENCHMARKING

Entre todos os algoritmos de SLAM descritos na literatura, é crucial realizar uma comparação justa entre eles para determinar qual apresenta o melhor desempenho em determinadas situações. Diversos conjuntos de dados de benchmark com diferentes características são propostos na literatura para explorar as capacidades e a robustez dos algoritmos de SLAM. Aqui, apresenta-se os conjuntos de dados de benchmark disponíveis publicamente, utilizados para avaliar os algoritmos de SLAM.

O conjunto de dados KITTI (Geiger; Lenz; Stiller and Urtasun) contém sequências de ambientes externos gravadas por câmeras estéreo em cores e em tons de cinza. Além disso, o KITTI também inclui dados de um scanner a laser 3D e a referência de ground truth fornecida por um sistema INS/GPS. O sistema de sensores é sincronizado e montado em um carro. Os autores do KITTI também disponibilizam tracklets para a classificação de objetos dinâmicos e benchmarks para avaliar tarefas de robótica,

como odometria visual e SLAM.

O conjunto de dados de referência EuRoC (Burri; Nikolic; Gohl; Schneider; Rehder; Omari; Achtelik and Siegwart, R) é amplamente utilizado para avaliar algoritmos de SLAM baseados apenas em visão e em visão-inercial, além de algoritmos de odometria. Os dados foram coletados em dois ambientes internos por meio de um veículo aéreo micro (MAV) e consistem em onze sequências de imagens estéreo e dados de uma unidade de medição inercial (IMU). A verdade de referência (ground truth) é obtida através de uma estação total e um sistema de captura de movimento.

A integração dos algoritmos com o ROS foi crucial para aplicar os métodos em um cenário de tempo-real em veículos autônomos. A criação de dois nós, um para captar as imagens em tempo-real e outro para rodar os algoritmos de visual SLAM, permitiu que o sistema realizasse as tarefas de mapeamento e localização dos métodos e desempenhar bem com qualquer abordagem.

### Resultados e discussão

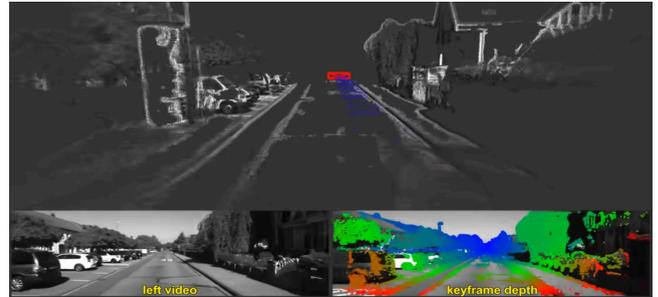
Apesar dessas vantagens dos sistemas monoculares, apresentam maior complexidade na inicialização, ao ser necessário obter pelo menos duas vistas diferentes para determinar a profundidade inicial, além de enfrentar desafios como a estimativa de pose, problemas de deriva e dificuldade na estimativa de escala. Esse último problema foi compensado com o uso de câmeras estéreo, que oferecem a vantagem de capturar a visão estéreo em um único quadro. No entanto, o sensor das câmeras estéreo é maior do que o de uma câmera monocular e exige maior processamento por quadro, principalmente devido à necessidade de retificação das imagens na etapa de correspondência estéreo.

Atualmente, um dos desafios enfrentados nos sistemas de SLAM é lidar com objetos temporariamente presentes no ambiente, mas que podem desaparecer em momentos subsequentes, como ocorre com veículos gravados em uma sequência de vídeo que podem não estar mais na cena em instantes futuros. Esse fenômeno é especialmente relevante em ambientes urbanos dinâmicos, onde elementos como carros e pedestres podem causar inconsistências na reconstrução do ambiente.

No contexto dessa pesquisa, o algoritmo LSD-SLAM, integrado ao ROS, foi testado utilizando uma sequência do conjunto de dados KITTI, que inclui dados de uma câmera estéreo (imagem 2D combinada com dados de profundidade). Como essa sequência foi capturada durante o dia, em um ambiente rico em detalhes, a reconstrução semi-densa gerada reflete esses elementos detalhados, sendo utilizada para fins de localização e mapeamento. Isso pode ser observado claramente na Figura 3, onde os objetos e estruturas temporários capturados na cena são

reconstruídos com uma densidade suficiente para permitir uma navegação precisa do sistema autônomo, mesmo em meio à movimentação de elementos dinâmicos.

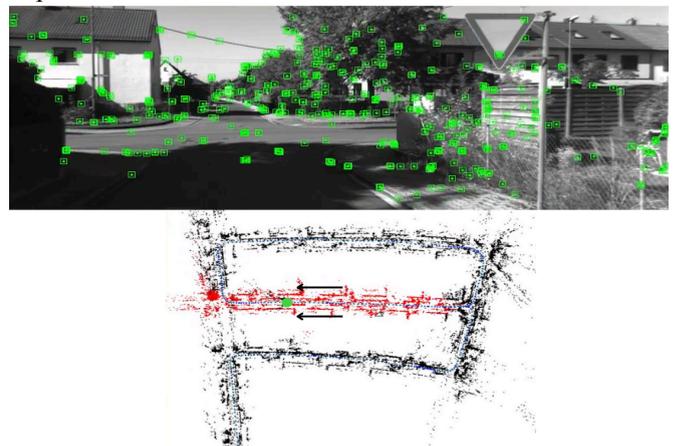
Figura 3 - Imagem do mapeamento e localização realizado pelo LSD\_SLAM, utilizando o dataset Sequence KITTI00.



O algoritmo ORB-SLAM3, que integra dados visuais e inerciais, foi testado utilizando uma sequência do conjunto de dados KITTI, que inclui imagens 2D da câmera estéreo combinadas com dados de uma unidade de medição inercial (IMU).

A utilização do ORB-SLAM3 permitiu não apenas a reconstrução precisa da trajetória do veículo, mas também a criação de um mapa esparsos dos arredores, suficiente para uma navegação robusta em um cenário dinâmico. A inclusão dos dados da IMU ajudou a corrigir erros de escala e drift, comuns em SLAM visual puro. Assim, o sistema conseguiu lidar com a variação de elementos dinâmicos capturados na sequência eficientemente, corrigindo a pose do veículo em tempo real e mantendo uma correspondência mais próxima com os dados de ground truth (INS/GPS) fornecidos pelo KITTI. Isso demonstra a capacidade do ORB-SLAM3 em lidar com cenas complexas, proporcionando uma localização precisa e consistente para sistemas autônomos, mesmo em ambientes externos e com mudanças dinâmicas significativas.

Figura 4 - Algoritmo ORB-SLAM3 identificando os features da figura, localizando-se e criando o mapeamento disperso.



A Tabela 1 complementa as diferenças de ambos os algoritmos do visual SLAM, onde o ORB\_SLAM3 oferece maior precisão, robustez e correção de erros ao longo do tempo, tornando-o mais adequado para trajetos longos e ambientes dinâmicos. Já, o LSD-SLAM, embora eficiente em cenários controlados, é mais limitado em termos de robustez em trajetos longos e em lidar com variações no ambiente.

Tabela 1 - Comparação entre os métodos de Visual SLAM LSD-SLAM e o ORB-SLAM3

Critério	LSD_SLAM	ORB_SLAM3
Precisão da posição	Média/Alta	Alta
Fechamento de loop	Presente, mas menos eficiente para grandes trajetórias	Muito eficiente, com otimização global precisa
Mapeamento	Semi-denso	Esparso
Dependência de iluminação	Alta	Alta, mas a IMU ajuda em cenários difíceis
Tempo de processamento	Mais rápido (processamento direto)	Mais lento (fusão de dados e otimizações)

Desse modo, o ORB-SLAM3 se destaca em cenários complexos, apresentando melhor desempenho em condições dinâmicas e em ambientes com pouca luz. Sua flexibilidade para trabalhar com diferentes tipos de câmeras (monoculares, estéreo e até mesmo RGB-D) torna-o uma escolha popular para aplicações de SLAM, oferecendo um desempenho otimizado e uma integração mais fácil em diversos sistemas robóticos.

### Conclusões

As técnicas de SLAM baseadas em visão são um campo de pesquisa vasto e promissor devido à sua capacidade de oferecer robustez e precisão com o uso de sensores baratos e compactos. No decorrer desta pesquisa, o objetivo principal foi a integração do Visual SLAM com o ROS em um robô móvel terrestre, explorando o potencial de localização e mapeamento em tempo real. A aplicação do Robotic Operating System (ROS) foi essencial para facilitar a integração do sistema robótico, e a pesquisa validou a viabilidade e eficácia do Visual SLAM na estimativa da localização e mapeamento do ambiente, fornecendo insights importantes para futuras aplicações em

robótica e automação.

Em trabalhos futuros, pode-se explorar a utilização de outros algoritmos de Visual SLAM, que permitam o robô aumentar sua eficácia em relação à criação do mapeamento e localização de sua posição, otimizando a sua navegação em tempo real e, o teste e validação destes em um robô móvel no ambiente externo, garantindo o desempenho e eficácia no cenário específico da utilização.

Em suma, esta pesquisa demonstrou que a abordagem de diferentes algoritmos de Visual SLAM apresenta resultados promissores para a estimativa de localização e mapeamento em tempo real, utilizando câmeras estéreo e sensores inerciais (IMU). A integração bem-sucedida desses algoritmos ao robô, combinada com a análise dos resultados obtidos, contribui significativamente para o avanço do conhecimento científico na área. Esses resultados podem ser aplicados em diversos contextos da robótica e automação, como em veículos autônomos, drones, e robôs móveis, que dependem de uma navegação precisa e confiável em ambientes dinâmicos.

### Agradecimentos

O autor expressa sua gratidão à UNIFEI pelo apoio financeiro que viabilizou a realização desta pesquisa e ao CNPq pelo incentivo ao desenvolvimento de novos talentos e pela valorização das diversas áreas do conhecimento. Agradecimentos especiais também ao orientador Rafael Francisco dos Santos por sua orientação e suporte ao longo deste trabalho.

### Referências

- BARROS, A. M. et al.** A comprehensive survey of visual slam algorithms. *Robotics*, v. 11, n. 1, 2022. ISSN 2218-6581.
- Burri, M.; Nikolic, J.; Gohl, P.; Schneider, T.; Rehder, J.; Omari, S.; Ahtelik, M.; Siegwart, R.** The EuRoC micro aerial vehicle datasets. *Int. J. Robot. Res.* 2016, 35, 1157–1163.
- Geiger, A.; Lenz, P.; Stiller, C.; Urtasun, R.** Vision meets robotics: The KITTI dataset. *Int. J. Robot. Res.* 2013, 32, 1231–1237.
- SILVA, C. S. E.; YEPES, I.** Desenvolvimento de sistema SLAM com odometria visual para Vant de inspeção em ambientes internos. 2015. 1
- SILVA, G. Felix da et al.** Desenvolvimento de um modelo de simulação de um robô a quatro rodas utilizando ROS e Gazebo. *Revista Principia - Divulgação Científica e Tecnológica do IFPB*, v. 61, 07 2022. 2