



UNIVERSIDADE FEDERAL DE SANTA CATARINA
CENTRO TECNOLÓGICO
PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA MECÂNICA

Leonardo Sales Galvão

**Desenvolvimento e avaliação de um Sistema Estéreo Inercial
baseado em ORB-SLAM3**

Florianópolis
2026

Leonardo Sales Galvão

**Desenvolvimento e avaliação de um Sistema Estéreo Inercial
baseado em ORB-SLAM3**

Dissertação submetida ao Programa de Pós-Graduação em Engenharia Mecânica da Universidade Federal de Santa Catarina como requisito parcial para a obtenção do título de Mestre em Metrologia e Instrumentação.

Orientador(a): Prof. Armando Albertazzi Gonçalves Júnior, *Dr. Eng.*

Coorientador(a): Prof. Tiago Loureiro Figaro da Costa Pinto, *Dr. Eng.*

Florianópolis

2026

Ficha catalográfica gerada por meio de sistema automatizado gerenciado pela BU/UFSC.
Dados inseridos pelo próprio autor.

Galvão, Leonardo Sales

Desenvolvimento e avaliação de um Sistema Estéreo
Inercial baseado em ORB-SLAM3 / Leonardo Sales Galvão ;
orientador, Armando Albertazzi Gonçalves Júnior,
coorientador, Tiago Loureiro Figaro da Costa Pinto, 2026.
173 p.

Dissertação (mestrado) - Universidade Federal de Santa
Catarina, Centro Tecnológico, Programa de Pós-Graduação em
Engenharia Mecânica, Florianópolis, 2026.

Inclui referências.

1. Engenharia Mecânica. 2. Visão Computacional. 3.
Metrologia. 4. SLAM. 5. ORB-SLAM3. I. Gonçalves Júnior,
Armando Albertazzi . II. Pinto, Tiago Loureiro Figaro da
Costa. III. Universidade Federal de Santa Catarina.
Programa de Pós-Graduação em Engenharia Mecânica. IV. Título.

Leonardo Sales Galvão

**Desenvolvimento e avaliação de um Sistema Estéreo Inercial
baseado em ORB-SLAM3**

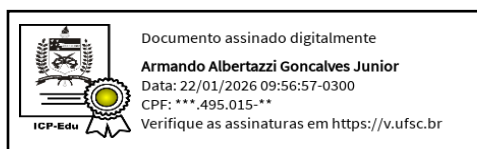
O presente trabalho em nível de Mestrado foi avaliado e aprovado, em 18 de dezembro de 2025, pela banca examinadora composta pelos seguintes membros:

Prof.(a) Armando Albertazzi Gonçalves Júnior, Dr.
Instituição Universidade Federal de Santa Catarina

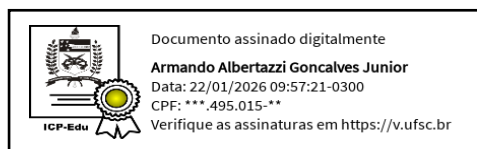
Prof.(a) Henrique Simas, Dr.
Instituição Universidade Federal de Santa Catarina

Prof.(a) Maurício Edgar Stivanello, Dr.
Instituição Instituto Federal de Santa Catarina

Certificamos que esta é a versão original e final do trabalho de conclusão que foi julgado adequado para obtenção do título de Mestre em Mestre em Metrologia e Instrumentação pelo Programa de Pós-Graduação em Engenharia Mecânica.



Coordenação do Programa de Pós-Graduação



Prof.(a) Armando Albertazzi Gonçalves Júnior, Dr.(a)
Orientador(a)

Florianópolis, 2026.

Dedico este trabalho a todos aqueles que me ajudaram a chegar onde estou e para onde ainda estou caminhando.

AGRADECIMENTOS

Agradeço profundamente à minha família pelo apoio incondicional, carinho e suporte em todos os momentos desta jornada. Foram vocês que me deram força nos desafios e celebraram cada conquista, tornando possível a realização deste trabalho.

Aos meus orientadores, Professor Armando, pela valiosa orientação, paciência e conhecimento compartilhado, e ao Professor Tiago, pela essencial coorientação e pelas contribuições que enriqueceram este projeto, meus sinceros agradecimentos.

Ao projeto VORIS, em especial ao Daniel Juchem, meu chefe, pelo constante incentivo e oportunidades, e aos colegas João, Moacir e Gustavo, pela colaboração e pelo ambiente de trabalho motivador.

Aos membros do POSMEC, em especial à Tainara, Igor, Luan e Alice, pela amizade, troca de ideias e pelo apoio durante esta trajetória acadêmica.

Também um agradecimento especial a secretária Rosana e ao professor André por terem me ajudado a conhecer o LABMETRO e pelo suporte a minha entrada ao POSMEC.

Por fim, mas não menos importante, a todos os meus amigos, sejam presenciais ou à distância, que de alguma forma me apoiaram, incentivaram e estiveram ao meu lado ao longo destes anos. Cada palavra de motivação e cada momento compartilhado foram fundamentais para a conclusão desta etapa.

Não se esqueça de sorrir em qualquer situação. Contanto que você esteja vivo, haverá coisas boas mais para frente, e serão muitas!

(ODA, Eiichiro)

RESUMO

O presente trabalho apresenta um sistema de visão estéreo-inercial de grau industrial, projetado para a navegação autônoma de Veículos de Operação Remota (ROVs) em ambientes subaquáticos, e análise metrológica da incerteza de medição resultante. O sistema utiliza um conjunto de câmeras estéreo e uma Unidade de Medição Inercial (IMU) montados rigidamente e adota o algoritmo ORB-SLAM3 no modo estéreo-inercial, reconhecido por melhorar a fusão de dados e a incerteza de medição dos resultados. A metodologia incluiu uma Calibração rigorosa dos parâmetros intrínsecos e extrínsecos, com validação geométrica de alinhamento contra modelo CAD para a transformação Câmera-IMU. A avaliação do sistema foi realizada inteiramente em ambiente terrestre controlado. A avaliação da incerteza de medição da trajetória empregou um *ground truth* micrométrico baseado em laser interferométrico em uma guia linear, analisando o Erro de Pose Absoluta (APE), Erro de Pose Relativa (RPE) e a Repetibilidade de Medição. Adicionalmente, avaliou-se a incerteza de medição geométrica da nuvem de pontos 3D, comparando-a a um modelo de referência improvisado gerado por fotogrametria que possui incerteza de medição conhecida. Os resultados demonstraram que a abordagem de acoplamento rígido e a calibração metrológica conferem ao sistema uma elevada estabilidade operacional e uma incerteza de medição consistente da trajetória em cenários controlados. A análise estatística confirmou que a parametrização otimizada viabiliza uma incerteza de medição do deslocamento coerente com as exigências de aplicações de navegação em superfície terrestre. Atingiu-se também uma reconstrução 3D com incerteza de medição verificável quando comparada ao modelo de referência. O trabalho estabelece uma base validada e confiável para a transposição futura do sistema para aplicações subaquáticas do Projeto VORIS.

Palavras-chave: ORB-SLAM3; Fotogrametria; SLAM.

ABSTRACT

The present work presents an industrial-grade stereo-inertial vision system, designed for the autonomous navigation of Remotely Operated Vehicles (ROVs) in subsea environments, and metrological analysis of the resulting measurement uncertainty. The system uses a set of stereo cameras and an Inertial Measurement Unit (IMU) rigidly mounted and adopts the ORB-SLAM3 algorithm in stereo-inertial mode, recognized for improving data fusion and the measurement accuracy of the results. The methodology included a rigorous Calibration of intrinsic and extrinsic parameters, with geometric alignment validation against a CAD model for the Camera-IMU transformation. The system assessment was performed entirely in a controlled terrestrial environment. The evaluation of trajectory measurement uncertainty employed a micrometric ground truth based on laser interferometry on a linear guide, analyzing the Absolute Pose Error (APE), Relative Pose Error (RPE), and Measurement Repeatability. Additionally, the geometric measurement uncertainty of the 3D point cloud was evaluated, comparing it to an improvised reference model generated by photogrammetry that possesses known measurement uncertainty. The results demonstrated that the rigid coupling approach and the metrological calibration confer high operational stability and consistent trajectory measurement uncertainty to the system in controlled scenarios. Statistical analysis confirmed that the optimized parametrization enables a displacement measurement accuracy coherent with the demands of navigation applications on the terrestrial surface. A 3D reconstruction with verifiable measurement uncertainty was also achieved when compared to the reference model. The work establishes a validated and reliable basis for the future transposition of the system for subsea applications within the VORIS Project.

Keywords: ORB-SLAM3; Photogrammetry; SLAM.

LISTA DE FIGURAS

Figura 1 – ROVs em Ambiente de plataforma <i>offshore</i>	16
Figura 2 – Princípio do funcionamento da câmera <i>pinhole</i>	24
Figura 3 – Sistemas de Coordenação Envolvidos na Projeção de um Ponto do Mundo para o Plano da Imagem, Baseados no Modelo <i>Pinhole</i>	25
Figura 4 – Representação da geometria epipolar de um sistema estéreo	29
Figura 5 – Par de imagens estéreo antes e depois do processo de retificação	30
Figura 6 – Análise detalhada do alinhamento de um ponto correspondente	31
Figura 7 – Par de imagens estéreo antes e depois do processo de retificação	31
Figura 8 – Tipos de distorções de imagem	34
Figura 9 – Diversos padrões de calibração	36
Figura 10 – Representação de uma IMU e seus eixos de medição	38
Figura 11 – Modelo de um giroscópio MEMS vibratório do tipo diapasão	40
Figura 12 – Modelo de um acelerômetro MEMS com detecção capacitiva diferencial	41
Figura 13 – Sistemas de referência definidos para a configuração estéreo inercial do ORB-SLAM3	45
Figura 14 – Diagrama do funcionamento do SLAM	48
Figura 15 – Ilustração do princípio da triangulação para múltiplos pontos	54
Figura 16 – Ilustração do erro de reprojeção	57
Figura 17 – Arquitetura lógica do sistema proposto	65
Figura 18 – Referência espacial do sistema estéreo proposto	66
Figura 19 – Evolução do aparato experimental	75
Figura 20 – Diagrama do funcionamento dos <i>softwares</i>	79
Figura 21 – Padrão de calibração do tipo tabuleiro de xadrez utilizado	82
Figura 22 – Sistemas sendo utilizado para coletar imagens do padrão de calibração	83
Figura 23 – Sistema estéreo-inercial fixado mesa óptica estabilizada	84
Figura 24 – Coleta de dados para calibração extrínseca Câmera-IMU	85
Figura 25 – Cenário de teste em ambiente controlado no LABMETRO mostrando o sistema posicionado na guia linear	87
Figura 26 – Posições do sistema de visão	87

Figura 27 – Ambiente de teste não controlado nos corredores do 4° andar do CCE/UFSC.....	88
Figura 28 – Exemplo de cenário de teste qualitativo realizado em área externa no campus da UFSC.....	89
Figura 29 – Corpo de prova (tubo polimérico) sendo reconstruído pelo GON ATOS	90
Figura 30 – Corpo de prova (tubo polimérico) e arranjo experimental para avaliação da nuvem de pontos no LABMETRO.....	90
Figura 31 – Caixa d'água da UFSC, utilizada como alvo para o teste qualitativo de reconstrução em larga escala.....	91
Figura 32 – Diagrama do procedimento técnico para aquisição dos dados e geração de <i>datasets</i>	92
Figura 33 – Gráfico do erro de reprojeção sistema inicial com resolução cheia.....	100
Figura 34 – Gráfico do erro de reprojeção sistema inicial com resolução cheia após remoção de imagens com alto erro de reprojeção.....	100
Figura 35 – Poses do tabuleiro em relação a câmera do sistema inicial com resolução cheia.....	100
Figura 36 – Gráfico do erro de reprojeção sistema inicial com resolução reduzida.	101
Figura 37 – Poses do tabuleiro em relação a câmera do sistema inicial com a resolução reduzida.....	101
Figura 38 – Gráfico do erro de reprojeção sistema novo.....	102
Figura 39 – Gráfico do erro de reprojeção sistema novo após remoção de imagens com alto erro de reprojeção.....	102
Figura 40 – Poses do tabuleiro em relação a câmera do sistema novo.....	103
Figura 41 – Gráfico de Desvio de Allan para o Acelerômetro.....	105
Figura 42 – Gráfico de Desvio de Allan para o Giroscópio.....	105
Figura 43 – Média e Desvio Padrão dos Erros para Configuração 1.....	109
Figura 44 – Média e Desvio Padrão dos Erros para Configuração 2.....	110
Figura 45 – Média e Desvio Padrão dos Erros para Configuração 3.....	110
Figura 46 – Média e Desvio Padrão dos Erros para Configuração 4.....	111
Figura 47 – Gráfico comparativo de desempenho entre as configurações estéreo.	113
Figura 48 – Gráfico comparativo de desempenho entre a Configuração 2 e a Configuração 4.....	114
Figura 49 – Gráfico polar dos desvios da trajetória para uma amostra da Configuração 2.....	115

Figura 50 – Gráfico polar dos desvios da trajetória para uma amostra da Configuração 4.....	115
Figura 51 – Nuvem de pontos da fotogrametria e posição das imagens.....	117
Figura 52 – Comparativo visual das trajetórias no experimento da guia linear. Esquerda: Trajetória da Fotogrametria. Direita: Trajetória do ORB-SLAM 3.....	118
Figura 53 – Gráfico polar de desvio da retitude fotogrametria.....	118
Figura 54 – Gráfico polar de desvio da retitude ORB-SLAM3.....	119
Figura 55 – Trajetória do SLAM (Exp. 1) colorida pelo Erro de Pose Absoluto (APE) em metros.....	121
Figura 56 – Trajetória do SLAM (Exp. 1) colorida pelo Erro de Pose Relativo (RPE) em metros.....	122
Figura 57 – Visualização do Erro de Pose Absoluto (APE) para a trajetória do Exp. 2.	123
Figura 58 – Visualização do Erro de Pose Relativo (RPE) para a trajetória do Exp. 2.	124
Figura 59 – Vistas da nuvem de pontos esparsa do corredor (Exp. 2).....	127
Figura 60 – Mapa da base da coluna que sustenta caixa d'água. Esquerda: Vista isométrica Nuvem de pontos. Direita: Vista superior da Nuvem de pontos.....	127
Figura 61 – Vistas da nuvem de pontos do tubo. Esquerda: Vista lateral. Direita: Vista superior.....	128
Figura 62 – Nuvem de pontos de referência do tubo obtida com o digitalizador GOM ATOS.....	129
Figura 63 – Nuvens de pontos: Sistema estéreo com câmera de baixa resolução.	130
Figura 64 – Nuvens de pontos: Sistema estéreo inercial com câmera de baixa resolução.....	130
Figura 65 – Nuvens de pontos: Sistema estéreo com câmera de alta resolução....	131
Figura 66 – Nuvens de pontos: Sistema estéreo inercial com câmera de alta resolução.....	131

LISTA DE TABELAS

Tabela 1 – Parâmetros intrínsecos e distorção para as configurações de câmera ...	98
Tabela 2 – Matrizes de Transformação Homogênea entre as câmeras de cada par estéreo.....	99
Tabela 3 – Parâmetros de Ruído da IMU.....	106
Tabela 4 – Resultados de erro para a Configuração 1 nos diferentes ângulos	109
Tabela 5 – Resultados de erro para a Configuração 2 nos diferentes ângulos.....	109
Tabela 6 – Resultados de erro para a Configuração 3 nos diferentes ângulos.....	110
Tabela 7 – Resultados de erro para a Configuração 4 nos diferentes ângulos.....	111
Tabela 8 – Comparativo de desempenho geral entre as quatro configurações.....	112
Tabela 9 – Comparativo quantitativo do deslocamento linear de 3 metros	117
Tabela 10 – Análise ATE/RPE do ORB-SLAM 3 usando Ground-Truth Fotogramétrico (Exp. 1).....	120
Tabela 11 – Análise ATE/RPE do ORB-SLAM 3 usando <i>Ground-truth</i> Fotogramétrico (EXP 2).....	122
Tabela 12 – Resultados quantitativos da medição do diâmetro tubo.....	132

LISTA DE ABREVIATURAS E SIGLAS

ROVs	Remotely Operated Vehicles
ROVs	Veículos de Operação Remota
IMU	Unidade de Medição Inercial
IMU	<i>Inertial Measurement Unit</i>
APE	Erro de Pose Absoluto
APE	<i>Absolute Pose Error</i>
RPE	Erro de Pose Relativo
RPE	<i>Relative Pose Error</i>
P&G	Petróleo e Gás
SLAM	<i>Simultaneous Localization and Mapping</i>
VSLAM	<i>Visual Simultaneous Localization and Mapping</i>
ORB	<i>Oriented FAST and Rotated BRIEF</i>
BRIEF	<i>Binary Robust Independent Elementary Features</i>
FAST	<i>Features from Accelerated Segment Test</i>
ROV	<i>Remotely Operated Vehicle</i>
ROUV	<i>Remotely Operated Underwater Vehicle</i>
GPS	Sistema de Posicionamento Global
FINEP	Financiadora de Estudos e Projetos
VORIS	Veículo Operado Remotamente para Inspeção Submarina
LABMETRO	Laboratório de Metrologia e Automatização
ROS	<i>Robot Operating System</i>
POSMEC	Programa de Pós-Graduação em Engenharia Mecânica
SM	Sistema de Medição
P&D	Pesquisa e Desenvolvimento
RGB-D	<i>Red Green Blue-Depth</i>
DDS	<i>Data Distribution Service</i>
OMG	<i>Object Management Group</i>
MVS	<i>Multi View Stereo</i>
PnP	<i>Perspective-n-Point</i>
RANSAC	<i>Random Sample Consensus</i>
SIFT	<i>Scale-Invariant Feature Transform</i>
SFM	<i>Structure From Motion</i>

EKF	Filtro de Kalman Estendido
MEMS	Sistemas Microeletromecânicos
VIO	Odometria Visual-Inercial
DDS	<i>Data Distribution Service</i>
OpenCV	<i>Open Source Computer Vision Library</i>
G2O	<i>General Graph Optimization</i>
ACM	Material de Alumínio Composto
UV	Ultra Violeta
UFSC	Universidade Federal de Santa Catarina
PLA	Ácido Polilático

SUMÁRIO

1	INTRODUÇÃO	15
1.1	MOTIVAÇÃO	17
1.2	OBJETIVOS GERAIS	19
1.2.1	Objetivos específicos.....	20
1.3	CONTRIBUIÇÕES DO TRABALHO	20
1.4	ESTRUTURA DO DOCUMENTO.....	21
2	REVISÃO BIBLIOGRÁFICA.....	23
2.1	SISTEMA DE VISÃO PASSIVOS.....	23
2.1.1	Modelo da Câmera <i>Pinhole</i>.....	24
2.1.2	Modelo estéreo	27
2.1.3	Calibração de Câmeras.....	32
2.2	NAVEGAÇÃO INERCIAL	37
2.2.1	Unidades de Medição Inercial (IMU).....	37
2.2.2	Princípios de Funcionamento: Giroscópio e Acelerômetro.....	39
2.2.2.1	<i>Giroscópio</i>	39
2.2.2.2	<i>Acelerômetro</i>	40
2.2.3	Modelos de Incerteza e Calibração de IMUs.....	42
2.2.2.1	<i>Erros Determinísticos</i>	43
2.2.2.2	<i>Erros Estocásticos.....</i>	44
2.2.4	Calibração Extrínseca Câmera-IMU.....	44
2.3	ODOMETRIA VISUAL-INERCIAL E SLAM.....	46
2.3.1	O Problema do SLAM (<i>Simultaneous Localization and Mapping</i>).....	46
2.3.2	Abordagens de Fusão de Dados estéreo-Inerciais.....	50
2.4	ALGOTIRMO ORB-SLAM 3.....	52
2.5	RECONSTRUÇÃO 3D COM FOTOGRAMETRIA.....	53
2.5.1	Princípios da Estrutura a partir do Movimento (<i>Structure From Motion</i>)	55
2.5.1.1	<i>Escala e Alinhamento</i>	57
2.5.2	Ferramentas de Referência	59
2.6	ROS 2 (<i>Robot Operating System 2</i>).....	60
2.6.1	Arquitetura e Conceitos Fundamentais	61
2.6.2	Comunicação e Aplicação no Projeto	62

3	MODELO PROPOSTO	64
3.1	CONFIGURAÇÃO DO SISTEMA DE NAVEGAÇÃO VISUAL-INERCIAL...	64
3.1.1	Disposição e Calibração das Câmeras.....	65
3.1.2	Integração e Calibração da IMU	66
3.1.3	Correção de Erros da IMU e Alinhamento com as Câmeras.....	67
3.2	FUSÃO DOS DADOS IMU E ESTÉREO NO ORB-SLAM 3	68
3.3	TRAJETÓRIA E CORPOS DE PROVA	69
3.4	CAPTURA DA TRAJETORIA E MAPEAMENTO.....	70
4	MATERIAS E MÉTODOS	71
4.1	APARELHO EXPERIMENTAL (<i>Hardware Setup</i>).....	71
4.1.1	Câmera Estéreo: Modelo e Especificações	71
4.1.2	Unidade de Medição Inercial (IMU): Modelo e Especificações	73
4.1.3	Plataforma de Montagem e Configuração Física	73
4.1.4	Unidade de Processamento e Armazenamento	75
4.2	AMBIENTE DE DESENVOLVIMENTO (<i>Software</i>).....	76
4.2.1	Sistema Operacional e ROS 2.....	76
4.2.2	Bibliotecas e Dependências.....	77
4.2.3	Implementação e Parametrização do ORB-SLAM 3.....	79
4.3	PROCEDIMENTOS EXPERIMENTAIS.....	81
4.3.1	Procedimento de Calibração Intrínseca e Extrínseca do Sistema Estéreo	81
4.3.2	Procedimento de Calibração da IMU e Alinhamento Câmera-IMU	83
4.3.3	Descrição dos Cenários de Teste e Corpos de Prova.....	86
4.3.4	Metodologia de Coleta de Dados e Geração de <i>Datasets</i>	91
4.3.5	Métricas de Avaliação de Desempenho	93
5	Resultados	97
5.1	RESULTADOS DA CALIBRAÇÃO DO SISTEMA.....	97
5.1.1	Parâmetros Intrínsecos, Extrínsecos e Erro de Reprojeção das Câmeras.....	97
5.1.2	Análise da Retificação das Imagens	103
5.1.3	Calibração Intrínseca da IMU:Análise de Ruído	104
5.1.4	Validação da Calibração Câmera-IMU	106
5.2	AVALIAÇÃO DA ODOMETRIA.....	107
5.2.1	Avaliação do Posicionamento e Configurações	108

5.2.2	Fotogrametria como " <i>ground truth</i> " para Avaliação do ORB-SLAM3	116
5.2.3	Erro de Trajetória Absoluto (APE) e Relativo (RPE)	119
5.2.4	Comparativo de Condições entre os Cenários de Teste	124
5.3	AVALIAÇÃO DO MAPEAMENTO	125
5.3.1	Análise Visual das Nuvens de Pontos	126
5.3.2	Análise dos Mapas de Pontos e Reconstrução 3D dos Tubos	128
5.4	ANÁLISE DE DESEMPENHO COMPUTACIONAL	132
6	DISCUSSÃO	134
6.1	ANÁLISE DOS RESULTADOS: LOCALIZAÇÃO E MAPEAMENTO	134
6.1.1	Análise da Exatidão de Localização (Trajetória)	134
6.1.2	Análise da Precisão de Mapeamento (Nuvem de Pontos)	135
6.2	IMPACTO DA FUSÃO VISUAL-INERCIAL NO DESEMPENHO DO SLAM	136
6.2.1	Vantagens da Fusão Visual-Inercial	136
6.2.2	Limitações	137
6.3	ANÁLISE DAS LIMITAÇÕES E FONTES DE ERRO	138
6.3.1	Desafios e Sensibilidades Operacionais	138
6.3.2	Categorização das Fontes de Erro de Medição	139
6.4	IMPLICAÇÕES PRÁTICAS PARA INSPEÇÃO DE ESTRUTURAS	141
7	CONCLUSÕES	143
7.1	SÍNTESE DO TRABALHO E PRINCIPAIS APRENDIZADOS	143
7.2	TRABALHOS FUTUROS	145
	REFERÊNCIAS	147
	APÊNDICE A – FUNCIONAMENTO DO ORB-SLAM3	156
	APÊNDICE B – RESULTADOS	165
	ANEXO A – LENTE LM3NCM E CÂMERA FLEA3 USB 3	166
	ANEXO B – LENTE EDMUND E CÂMERA BLACKFLY S USB3	167
	ANEXO C – IMU	168
	ANEXO D – UNIDADE DE PROCESSAMENTO E ARMAZENAMENTO	169

1 INTRODUÇÃO

A crescente demanda por recursos energéticos e a complexidade das operações industriais têm impulsionado a exploração e produção em ambientes cada vez mais desafiadores. No setor de Petróleo e Gás (P&G), as infraestruturas, tanto *onshore* quanto *offshore*, exigem monitoramento contínuo e inspeções periódicas para garantir a integridade estrutural, a segurança operacional e a prevenção de acidentes ambientais. Os processos de inspeção são essenciais na indústria de petróleo e gás para garantir a integridade e a segurança dos ativos de infraestrutura (1, p. 600). A inspeção de estruturas submersas, como dutos, dutos flexíveis e equipamentos submarinos, representa um desafio particular devido às condições adversas do ambiente aquático, que incluem baixa visibilidade, fortes correntes e grandes profundidades. Agbakwuru (2), enfatiza que para dutos submarinos onde a baixa visibilidade subaquática é evidente, o processo de inspeção do duto torna-se difícil de implementar. Outros sim, (3), aborda a questão de diversos tipos de métodos ópticos, afirmando que "as soluções de levantamento óptico podem ser classificadas de acordo com vários critérios, incluindo sua natureza ativa ou passiva" (3, p. 7).

Tradicionalmente, muitas dessas inspeções eram realizadas por mergulhadores profissionais, o que envolvia altos custos, riscos à segurança humana e limitações operacionais em termos de profundidade e tempo de submersão. ROVs com tecnologia avançada de sonar oferecem uma alternativa mais segura e eficiente para inspeções de estacas de fundação de píer subaquáticas, alcançando precisão comparável aos métodos tradicionais de mergulhadores, ao mesmo tempo em que reduzem os riscos operacionais e a dependência de mergulhadores humanos (4). Com o avanço da robótica e da visão computacional, os Veículos de Operação Remota (ROVs) tornaram-se ferramentas indispensáveis para a execução dessas tarefas em ambientes hostis (Figura 1). Eles permitem a coleta de dados visuais e instrumentados de forma mais segura, eficiente e em locais de difícil acesso, revolucionando as práticas de inspeção na indústria de P&G. Esse cenário é corroborado por Castillón (5), que destaca que como as operações de inspeção, manutenção e reparo subaquáticas estão sendo cada vez mais robotizadas a fim de reduzir problemas de segurança e custos, assim como esses sistemas robóticos

dependem de sensores de visão para tarefas fundamentais, como navegação, reconhecimento e manipulação de objetos.

Figura 1 – ROVs em ambientes de plataforma *offshore*.



Fonte: ORION S.I.C. (6).

Para que os ROVs possam realizar inspeções autônomas ou semi-autônomas eficazes, a capacidade de localização precisa e a reconstrução tridimensional do ambiente são necessárias. A localização permite que o veículo saiba sua posição e orientação no espaço, enquanto a reconstrução 3D fornece um modelo detalhado da estrutura inspecionada, fundamental para identificar deformações, corrosões ou outros danos. "A localização e o mapeamento são cruciais para ROVs e inspeção autônoma, pois permitem um rastreamento preciso da trajetória e reduzem a dependência de mão de obra" (7, p. 04). Em ambientes subaquáticos, onde sistemas de posicionamento global (GPS) não são viáveis, a dependência de sensores a bordo para estas tarefas é ainda maior.

Nesse contexto, as técnicas de visão computacional estéreo passiva, que se baseiam na luz ambiente para a percepção do ambiente, surgem como uma alternativa confiável e de baixo custo. Além disso os dados provenientes de câmeras estéreo com informações de Unidades de Medida Inercial (IMUs) tem se mostrado particularmente eficaz para superar os desafios de navegação e mapeamento em ambientes complexos. A fusão desses dois sensores tem se mostrado uma excelente

solução para uso, como apresentado nos resultados dos trabalhos de Liu *et al.* (8), Wen *et al.* (9) e Jung *et al.* (10).

Para atender à crescente demanda por sistemas de inspeção subaquática autônomos e de baixa incerteza de medição, uma classe de algoritmos denominados de *Simultaneous Localization and Mapping* (SLAM) emerge como uma solução fundamental. O SLAM é uma técnica que permite a um robô ou veículo construir um mapa 3D de um ambiente desconhecido e, simultaneamente, localizar-se dentro desse mapa (11, p. 3). Dentre os diversos algoritmos de SLAM existentes, o ORB-SLAM3 destaca-se por sua capacidade de operar com diferentes tipos de sensores (monocular, estéreo e *Red-Green-Blue-Depth* (RGB-D), que captura a cor e a informação de profundidade), sua estabilidade operacional e, principalmente, sua elevada precisão na estimativa de trajetória e na construção de mapas esparsos. Sendo descrito pelo trabalho de Campos *et al.* (12), como um sistema confiável e com exatidão, o ORB-SLAM3 alcança resultados de 3,5 cm de erro médio absoluto de trajetória em drones e 9 mm em movimentos manuais em ambientes fechados.

1.1 MOTIVAÇÃO

A crescente demanda por inspeções de infraestruturas *offshore* e subaquáticas exige veículos de operação remota (ROVs) dotados de sistemas de medição (SM) capazes de fornecer dados tridimensionais (3D) com excelente incerteza de medição em ambientes inerentemente desafiadores.

Atualmente, a maioria dos ROVs comerciais utiliza sistemas de navegação baseados em Sonar e/ou câmeras monoculares (13, 14, 15). Tais soluções apresentam limitações significativas em termos de medição e detalhamento da reconstrução 3D quando comparadas a sistemas estéreo ou estéreo-inerciais mais avançados (16, 17). Além disso, a navegação autônoma em ambientes subaquáticos é restrita, visto que o Sistema de Posicionamento Global (GPS) é inviável sob a água, e as alternativas de posicionamento acústico apresentam custo e complexidade operacional elevados.

A literatura especializada (12, 18) aponta que as soluções existentes de SLAM visual frequentemente se concentram apenas nas avaliações da precisão de medição (a componente do erro de medição relacionada à dispersão aleatória) da trajetória

estimada. Tal foco negligencia aspectos críticos para aplicações de inspeção industrial, como a qualidade métrica da nuvem de pontos e a quantificação completa da incerteza de medição do sistema 3D gerado. Fica evidente uma lacuna na validação metrológica rigorosa de sistemas Visual-SLAM aplicados à inspeção de alta qualidade em cenários exigentes.

Diante deste cenário, este trabalho pretende contribuir no avanço do conhecimento na área de inspeção subaquática ao desenvolver e validar um sistema estéreo-inercial acessível e adaptado às necessidades da indústria nacional e de preencher limitações técnicas e lacunas metrológicas. A abordagem proposta diferencia-se por:

- Combinar sensores comerciais com algoritmos de SLAM do estado da arte.
- Fornecer uma avaliação completa tanto da trajetória estimada quanto da precisão métrica da reconstrução 3D.
- Oferecer um protocolo de validação em terra que antecede e facilita a implementação subaquática.
- Estender a aplicação do sistema SLAM para servir como a base de posicionamento e fusão de dados para sensores auxiliares embarcados no ROV.

Este trabalho está inserido no projeto de Pesquisa e Desenvolvimento (P&D) denominado Veículo Operado Remotamente de Inspeção Submarina (VORIS), conduzido pelo Laboratório de Metrologia e Automação (LABMETRO) com financiamento da FINEP (Financiadora de Estudos e Projetos). O projeto VORIS visa “desenvolver e validar em piscina de testes um veículo de inspeção subaquático baseado em um ROV ao qual serão integrados quatro sistemas ópticos de medição 3D com distintas tecnologias e características complementares para extrair e documentar nuvens de pontos contendo texturas de estruturas e equipamentos submarinos, representá-las em 3D e medir parâmetros geométricos a partir das nuvens medidas” (VORIS, 2024).

O projeto VORIS considerou quatro sistemas de medição (SM) ópticos para diferentes distâncias e tecnologias. Ao validar o SM2 (Estéreo Passivo via SLAM), esta pesquisa contribui diretamente para o projeto, mas também para o desenvolvimento de tecnologias autônomas de inspeção subaquática no Brasil. Ao

validar uma solução acessível e adaptável às condições locais, o trabalho abre caminho para a democratização de tecnologias avançadas de inspeção submarina no país, reduzindo a dependência de soluções estrangeiras e promovendo maior segurança nas operações *offshore*.

1.2 OBJETIVOS GERAIS

Este trabalho tem como objetivo principal desenvolver um sistema integrado de SLAM que combine dados visuais estéreo e medidas inerciais através do *software* aberto ORB-SLAM3, visando aplicações em inspeção subaquática.

A implementação deste sistema requer a adaptação dos parâmetros do ORB-SLAM3 para operação em condições adversas. Um aspecto fundamental do trabalho será o desenvolvimento de metodologias para calibração e a unidade de medição inercial, garantindo a qualidade da fusão sensorial.

Como parte integrante do projeto VORIS, esta pesquisa contribuirá para a validação em terra do sistema SM2 antes de sua integração ao ROV. A fase experimental incluirá avaliações abrangentes, tanto da repetibilidade de medição da trajetória estimada, quanto da qualidade métrica das nuvens de pontos geradas, utilizando referências metrológicas estabelecidas utilizando da técnica de fotogrametria. Serão analisados diversos cenários operacionais para caracterizar o desempenho do sistema sob diferentes condições de visibilidade e riqueza de características visuais.

Os resultados esperados permitirão não apenas validar a solução proposta como alternativa viável para localização e mapeamento 3D esparsos visando futuro uso em ROVs, mas também estabelecer procedimentos experimentais para avaliação de sistemas SLAM em aplicações industriais. Este trabalho contribuirá significativamente para o avanço das tecnologias de inspeção subaquática no contexto nacional, oferecendo uma solução acessível e adaptada às condições específicas do projeto VORIS e com potencial aplicação em avaliação de dutos, inspeção de plataformas e outras infraestruturas submarinas críticas.

1.2.1 Objetivos específicos

Para alcançar o objetivo geral proposto, estabelecem-se os seguintes objetivos específicos:

- Desenvolver um sistema estéreo-inercial compatível com o ORB-SLAM3, integrando câmeras comerciais e IMU em uma configuração física apropriada para operação em diversos ambientes.
- Definir e implementar um protocolo de calibração para os parâmetros intrínsecos e excêntricos do par de câmeras *pinhole*, incertezas do sensor IMU e parâmetros extrínsecos entre o sistema estéreo e a IMU.
- Realizar testes sistemáticos em diferentes cenários controlados, com suas determinadas variações, sendo elas: condições de iluminação, texturas de ambientes, padrões de movimento e cenas.
- Avaliar quantitativamente o desempenho do sistema no ar através de análise métrica da trajetória estimada, comparação da nuvem de pontos geradas com dados de referência e comparação da trajetória gerada com dados de referência.
- Otimizar os parâmetros do ORB-SLAM3 para a configuração específica do sistema desenvolvido visando futura aplicação no ROV, incluindo: ajuste dos limiares de extração de características ORB, quantidade máxima extraída de ORBs na cena e valores do método de processamento.
- Investigar o desempenho de diferentes configurações físicas do sistema, avaliando a distância e ângulo entre as câmeras, posicionamento relativo da IMU e abertura das lentes.
- Documentar todo o processo de desenvolvimento e validação, gerando protocolos reproduzíveis para calibração e testes avaliativos do sistema.

Estes objetivos específicos foram estruturados para garantir uma abordagem sistemática e metrologicamente fundamentada no desenvolvimento e validação do sistema proposto.

1.3 CONTRIBUIÇÕES DO TRABALHO

Este trabalho de mestrado visa contribuir diretamente para o projeto VORIS através do desenvolvimento de um sistema de percepção de ambiente baseado em visão computacional. A principal tarefa reside na integração de um sistema de visão estéreo passivo, utilizando câmeras industriais e uma Unidade de Medida Inercial (IMU), com o algoritmo de *Simultaneous Localization and Mapping* (SLAM) ORB-SLAM3. O trabalho contribuirá para futuramente capacitar um ROV para realizar mapeamento tridimensional e gerar uma trajetória precisa durante operações de inspeção.

A contribuição mais específica deste estudo concentra-se na avaliação da incerteza de medição de localização e mapeamento proporcionada pelo ORB-SLAM3 sob a configuração de hardware proposta. Os resultados desta análise fornecerão detalhes críticos sobre a viabilidade e o desempenho de sistemas de visão estéreo-inerciais para futuras implementações em ROVs, especialmente em ambientes desafiadores como o subaquático, onde a navegação autônoma e a inspeção visual de alta qualidade são essenciais para a segurança e eficiência das operações.

1.4 ESTRUTURA DO DOCUMENTO

O presente trabalho está organizado em sete capítulos, os quais apresentam de forma sequencial o desenvolvimento da pesquisa, desde os fundamentos teóricos até os resultados experimentais.

O Capítulo 1 (Introdução) estabelece o contexto do trabalho, destacando a motivação para o desenvolvimento de um sistema de visão estéreo-inercial visando futura aplicação em ROVs comerciais. Além disso, são definidos os objetivos gerais e específicos da pesquisa.

O Capítulo 2 (Revisão Bibliográfica) apresenta a revisão bibliográfica, abordando conceitos essenciais para a compreensão do sistema proposto. Inicialmente, são discutidos os princípios de sistemas estéreo-inerciais passivos, incluindo o modelo da câmera *pinhole*, técnicas de calibração estéreo, geometria epipolar e geração de mapas de disparidade. Em seguida, é apresentada uma revisão sobre SLAM (*Simultaneous Localization and Mapping*), com ênfase no ORB-SLAM, detalhando seu funcionamento, características do descritor ORB e aplicações práticas. Também são explorados os fundamentos da navegação inercial,

abrangendo o princípio de operação de IMUs, teoria de giroscópios e acelerômetros, e métodos de fusão sensorial entre dados visuais e inerciais. Por último, são explicados os conceitos básicos da fotogrametria.

O Capítulo 3 (Modelo Proposto) descreve a arquitetura do sistema desenvolvido. São apresentados os procedimentos para configuração do sistema estéreo, incluindo calibração das câmeras e alinhamento com a IMU. Também é apresentada a integração dos sensores no *framework* do ORB-SLAM 3, assim como as estratégias adotadas para trajetória de inspeção e mapeamento de objetos em ambientes controlados.

O Capítulo 4 (Meios e Métodos) descreve os componentes utilizados no desenvolvimento experimental. São especificadas as câmeras estéreo selecionadas, com ênfase em seu processo de calibração e configuração física para testes. Da mesma forma, são apresentados os detalhes da IMU empregada, incluindo metodologias de calibração e sincronização com o sistema visual. Adicionalmente, são caracterizados o ambiente experimental e os objetos de referência utilizados para validação do sistema.

O Capítulo 5 (Resultados) apresenta os dados obtidos experimentalmente, organizados em quatro eixos principais: desempenho em ambiente controlado do modo monocular e estéreo, do modo estéreo-inercial e avaliação do desempenho dos modos citados anteriormente em ambientes não controlados.

O Capítulo 6 (Discussão) realiza uma análise crítica dos resultados, examinando as limitações identificadas no sistema, a efetividade da integração sensorial e o impacto da IMU na precisão do ORB-SLAM 3.

O Capítulo 7 (Considerações Finais) sintetiza as contribuições do trabalho, apresentando conclusões sobre o sistema desenvolvido e propondo direcionamentos para futuras pesquisas nesta área e, crucialmente, estabelece sugestões e diretrizes sobre como transpor os resultados obtidos na validação em terra para as próximas fases do projeto VORIS que envolvem a implementação e as aplicações subaquáticas.

2 REVISÃO BIBLIOGRÁFICA

Este capítulo apresenta os fundamentos teóricos e as tecnologias que embasam o desenvolvimento e a avaliação do sistema proposto. Serão abordados, inicialmente, os Sistemas de Visão Passivos, detalhando o modelo da câmera *pinhole*, alta angulação, sistema estéreo e a calibração de câmeras, e os conceitos de geometria epipolar e mapas de disparidade. Em seguida, discutem-se os princípios da Navegação Inercial, explicando o funcionamento de um Sistema IMU, incluindo giroscópios, acelerômetros juntamente com sua calibração. Posteriormente, explora-se a fusão de dados estéreo e IMU, destacando os benefícios dessa integração para estimativas de pose e mapeamento mais estáveis. O capítulo também apresenta a técnica de SLAM, com foco no algoritmo ORB-SLAM cobrindo sua definição, funcionamento, as características dos descritores ORB e suas aplicações. Por fim, são detalhados os princípios da Fotogrametria, com ênfase no software COLMAP, e o ROS 2 (*Robot Operating System 2*), percorrendo sobre sua arquitetura e componentes essenciais para sistemas robóticos.

2.1 SISTEMA DE VISÃO PASSIVOS

A visão computacional segundo Milano e Honorato (19, p. 1) é a ciência que capacita máquinas a "enxergarem o meio à sua volta, extraíndo informações significativas a partir de imagens capturadas por câmeras de vídeo, sensores, *scanners*, entre outros dispositivos". Assim buscam extrair informações significativas do mundo físico a partir de imagens, permitindo que máquinas compreendam e interajam com o ambiente de forma análoga à visão humana. Esses sistemas podem ser classificados em duas categorias: ativos e passivos. Os sistemas ativos emitem sua própria forma de energia, como luz estruturada ou pulsos de laser, para iluminar a cena e medir a sua estrutura tridimensional.

Em contraste, os sistemas de visão passivos, que são o foco deste trabalho, operam utilizando apenas a iluminação ambiente disponível na cena. O componente central de um sistema de visão passivo é a câmera, que atua como um sensor para capturar a radiação eletromagnética (luz) refletida ou emitida pelos objetos.

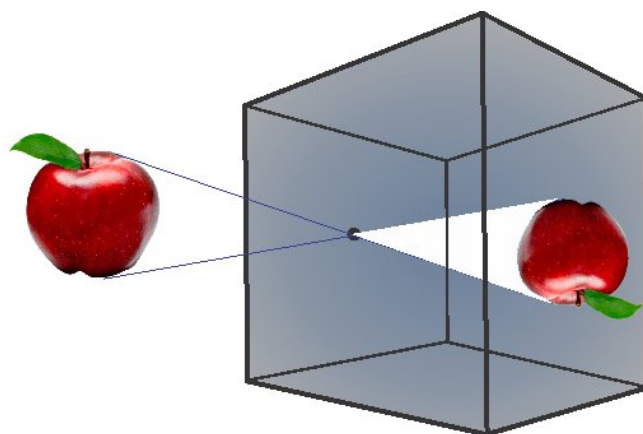
O princípio fundamental por trás da visão passiva é a projeção de um mundo tridimensional (3D) em um plano de imagem bidimensional (2D). Essa projeção, por sua natureza, resulta na perda da informação de profundidade, tornando a tarefa de reconstruir a estrutura 3D original um dos desafios centrais da visão computacional. A grande vantagem dos sistemas passivos reside em sua simplicidade, baixo custo e na riqueza de informações texturais e de cor que as imagens fornecem, sem a necessidade de interferir no ambiente com uma fonte de iluminação própria (20).

2.1.1 Modelo da Câmera *Pinhole*

O modelo da câmera *pinhole* (orifício) é a representação matemática mais fundamental de uma câmera. Ele serve como uma aproximação idealizada que descreve a relação geométrica entre os pontos de um ambiente tridimensional e suas projeções em um plano de imagem bidimensional. O modelo consiste em uma caixa à prova de luz com um pequeno orifício, chamado de centro de projeção ou centro óptico, em uma de suas faces, e um plano de imagem (ou filme) na face oposta(21).

Fisicamente, os raios de luz refletidos pelo objeto na cena viajam em linha reta, passam pelo centro de projeção e atingem o plano de imagem, formando uma imagem invertida tanto na vertical quanto na horizontal. A Figura 2 ilustra este princípio físico.

Figura 2 – Princípio do funcionamento da câmera *pinhole*.

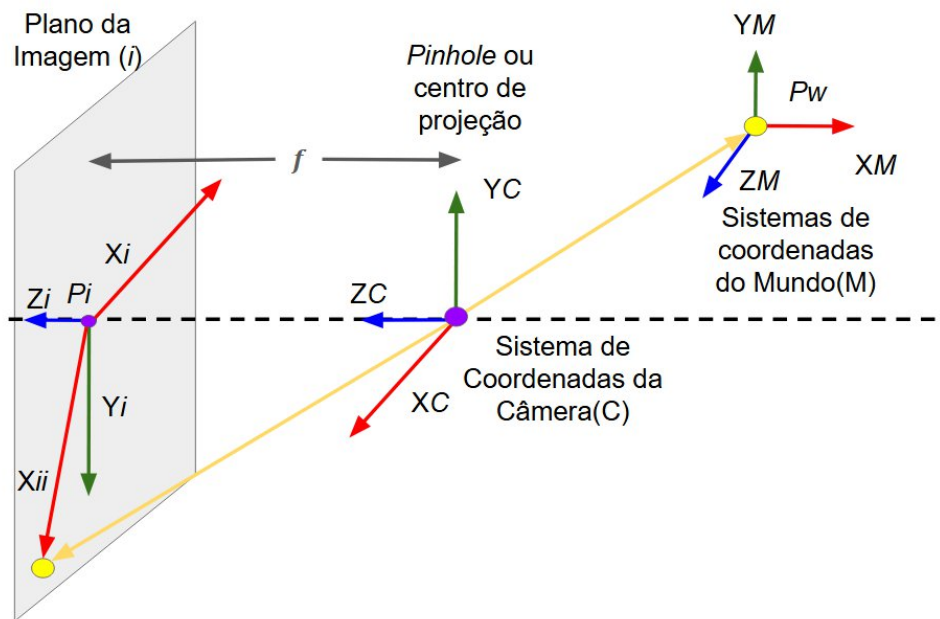


Fonte: Do Autor.

Para simplificar a matemática e evitar o tratamento de imagens invertidas, é comum utilizar um modelo geométrico equivalente. Neste modelo, um plano de imagem virtual é posicionado na frente do centro de projeção, a uma distância

conhecida como distância focal (f). A relação entre as coordenadas de um ponto no mundo 3D e sua projeção no plano de imagem 2D pode ser derivada por semelhança de triângulos. A Figura 3 apresenta o modelo geométrico que utilizaremos para a derivação matemática (22, 23, 24).

Figura 3 – Sistemas de Coordenadas Envolvidos na Projeção de um Ponto do Mundo para o Plano da Imagem, baseados no Modelo *Pinhole*.



Fonte: Do Autor.

Com base na Figura 3, vamos definir as variáveis:

- P_w é um ponto qualquer no espaço 3D
- O Sistema de Coordenadas da Câmera (C) tem sua origem no centro de projeção.
- O X_C , Y_C e Z_C são as coordenadas do ponto P em relação ao sistema C. O eixo Z_C representa a profundidade do ponto.
- O plano de imagem é posicionado a uma distância focal f da origem, ao longo do eixo Z_C .
- X_i e Y_i são as coordenadas da projeção do ponto P no plano de imagem.

Observando a geometria da Figura 3, podemos ver dois pares de triângulos semelhantes. A partir dessa semelhança, obtemos as equações (1 e 2) da projeção perspectiva:

$$\frac{x_i}{f} = \frac{x_c}{z_c} \Rightarrow x_i = f \frac{x_c}{z_c} \quad (1)$$

$$\frac{y_i}{f} = \frac{y_c}{z_c} \Rightarrow y_i = f \frac{y_c}{z_c} \quad (2)$$

estas coordenadas escalares x_i e y_i são as componentes do vetor de posição X_i (a seta vermelha na Figura 3) ao longo dos vetores unitários que definem os eixos do plano da imagem, $\hat{Y}_i$ e $\hat{X}_i$. Matematicamente, a relação entre o vetor e suas componentes é a fórmula apresentada na equação 3:

$$X_i = x_i * \hat{X}_i + y_i * \hat{Y}_i \quad (3)$$

Esta distinção formaliza a conexão entre a geometria vetorial da figura e as grandezas escalares utilizadas nos cálculos subsequentes.

As coordenadas (x_i, y_i) são posições físicas no sensor da câmera. Para uma imagem digital, precisamos convertê-las para coordenadas de *pixel* (u, v) , cuja origem geralmente fica no canto superior esquerdo da imagem. Essa transformação é feita pelos parâmetros intrínsecos da câmera, que são agrupados na matriz de calibração intrínseca K (equação 4).

$$K = \begin{bmatrix} fx & s & cx \\ 0 & fy & cy \\ 0 & 0 & 1 \end{bmatrix} \quad (4)$$

Onde:

- fx e fy são as distâncias focais em *pixels*. Elas são o resultado da multiplicação da distância focal (em metros) pela densidade de *pixels* do sensor (em *pixels*/metro).
- (cx, cy) são as coordenadas do ponto principal em *pixels*.
- s é o coeficiente de inclinação (*skew*), que é zero para a maioria das câmeras modernas.

Para combinar tudo em uma única operação, utilizamos coordenadas homogêneas. A projeção de um ponto 3D $P_c = [x_c, y_c, z_c]^T$ para um ponto 2D em *pixels* $p = [u, v, 1]^T$ é dada pela multiplicação da matriz intrínseca K pelo vetor do ponto 3D (equação 5):

$$\begin{bmatrix} u' \\ v' \\ w' \end{bmatrix} = \begin{bmatrix} fx & s & cx \\ 0 & fy & cy \\ 0 & 0 & 1 \end{bmatrix} * \begin{bmatrix} x_c \\ y_c \\ z_c \end{bmatrix} \quad (5)$$

Realizando a multiplicação de matrizes temos a equação 6 (assumindo $s=0$ para simplificar):

$$\begin{bmatrix} u' \\ v' \\ w' \end{bmatrix} = \begin{bmatrix} fx * xc + cx * zc \\ fy * yc + cy * zc \\ zc \end{bmatrix} \quad (6)$$

Para converter de volta para coordenadas cartesianas 2D (equação 7 e 8), dividimos os dois primeiros elementos pelo terceiro ($w' = zc$):

$$u = \frac{u'}{w'} = \frac{fx * xc + cx * zc}{zc} = fx \frac{xc}{zc} + cx \quad (7)$$

$$v = \frac{v'}{w'} = \frac{fy * yc + cy * zc}{zc} = fy \frac{yc}{zc} + cy \quad (8)$$

Este modelo, embora seja uma simplificação, é a base para a maioria das aplicações em visão computacional. Ele captura a essência da formação de imagem e, quando combinado com modelos de distorção de lente, pode descrever o comportamento de câmeras reais com baixa incerteza de medição.

É fundamental, neste ponto, formalizar a transformação do ponto tridimensional do mundo para o plano bidimensional da imagem. Quando um ponto do mundo PW é transladado e rotacionado para o sistema de coordenadas da câmera PC , essa transformação é descrita por uma matriz de transformação homogênea, composta pela matriz de Rotação (R) e pelo vetor de Translação (t). A projeção final é então realizada utilizando a matriz de calibração intrínseca (K). A combinação dessas transformações resulta na Matriz de Projeção da Câmera (P), que engloba os parâmetros geométricos e intrínsecos do sistema. Essa formalização matricial é o pilar para o cálculo do SLAM, que necessita dessas transformações para integrar múltiplos frames e sensores.

É importante ressaltar que, embora o modelo pinhole seja o mais fundamental, existem outros modelos matemáticos desenvolvidos para câmeras com características ópticas específicas, como as de grande-angular (*wide-angle*) e as olho de peixe (*fish-eye*). Estes modelos são projetados para lidar com as distorções geométricas severas presentes em lentes com um campo de visão muito amplo. Estudos aprofundados sobre estes modelos e suas aplicações em *Structure from Motion* e modelagem 3D podem ser encontrados em trabalhos como os de Schwalbe (25), Van Gool, Tomasi e Kanade (23), e Molinari *et al.* (27).

2.1.2 Modelo estéreo

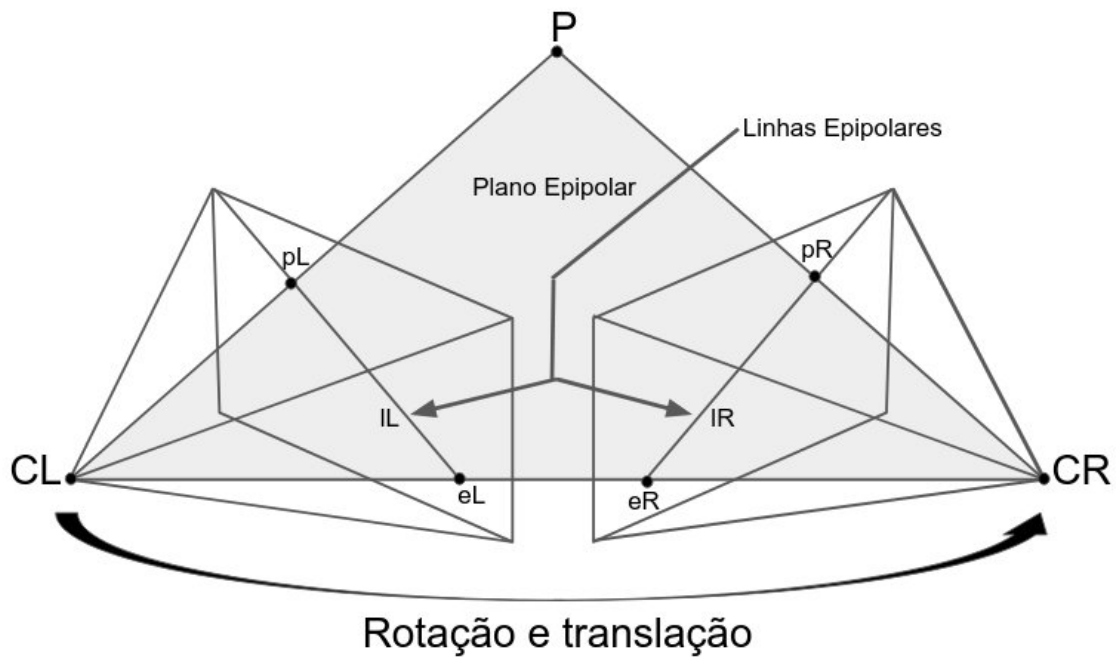
Conforme discutido na seção anterior, o modelo de câmera pinhole projeta o mundo 3D em uma imagem 2D, um processo que inerentemente perde a informação de escala e profundidade. Para superar essa limitação, os sistemas de visão estéreo empregam duas câmeras para observar a mesma cena a partir de pontos de vista ligeiramente diferentes, de forma análoga à visão binocular humana.

O seu princípio fundamental foi descrito de forma pioneira por Sir David Brewster (28, p. 3), que observou que "a criação estereoscópica se deve unicamente à superposição das duas imagens planas pelo aparato óptico empregado, e à percepção distinta e instantânea de distância pela convergência dos eixos ópticos sobre os pontos similares das duas imagens que o estereoscópio uniu". Em termos modernos, a "convergência dos eixos ópticos sobre os pontos similares" é o que chamamos de problema de correspondência, e a "percepção de distância" é a recuperação de profundidade através da triangulação. Em um sistema de visão computacional, este princípio é implementado utilizando um par de câmeras, separadas por uma distância conhecida e fixa, a linha de base (*baseline*), que aqui representaremos por B . O principal desafio da visão estéreo é, portanto, resolver o problema de correspondência, e a solução para isso reside na geometria fundamental que rege o sistema: a geometria epipolar.

Considerando um ponto tridimensional P e os centros ópticos das câmeras esquerda e direita, CL e CR respectivamente, estes três pontos formam um plano único no espaço, conhecido como plano epipolar (20, 23, 24). O epipolo é definido pela intersecção da reta que une os centros ópticos (CL e CR) com o plano de imagem da outra câmera, definindo os epipolos (eL e eR). Finalmente, a intersecção do plano epipolar com cada um dos planos de imagem gera duas linhas, LL e LR , chamadas de linhas epipolares (Figura 4).

A consequência fundamental desta geometria é a restrição epipolar: para qualquer ponto pL na imagem esquerda, que é a projeção do ponto 3D P , seu ponto correspondente pR na imagem direita deve, obrigatoriamente, pertencer à linha epipolar LR . Desta forma, a busca pelo ponto correspondente, que a princípio seria em toda a área 2D da imagem direita, é reduzida a uma busca 1D ao longo de uma única linha. Esta restrição torna o problema de correspondência computacionalmente mais eficiente e menos ambíguo.

Figura 4 – Representação da geometria epipolar de um sistema estéreo.



Fonte: Do Autor.

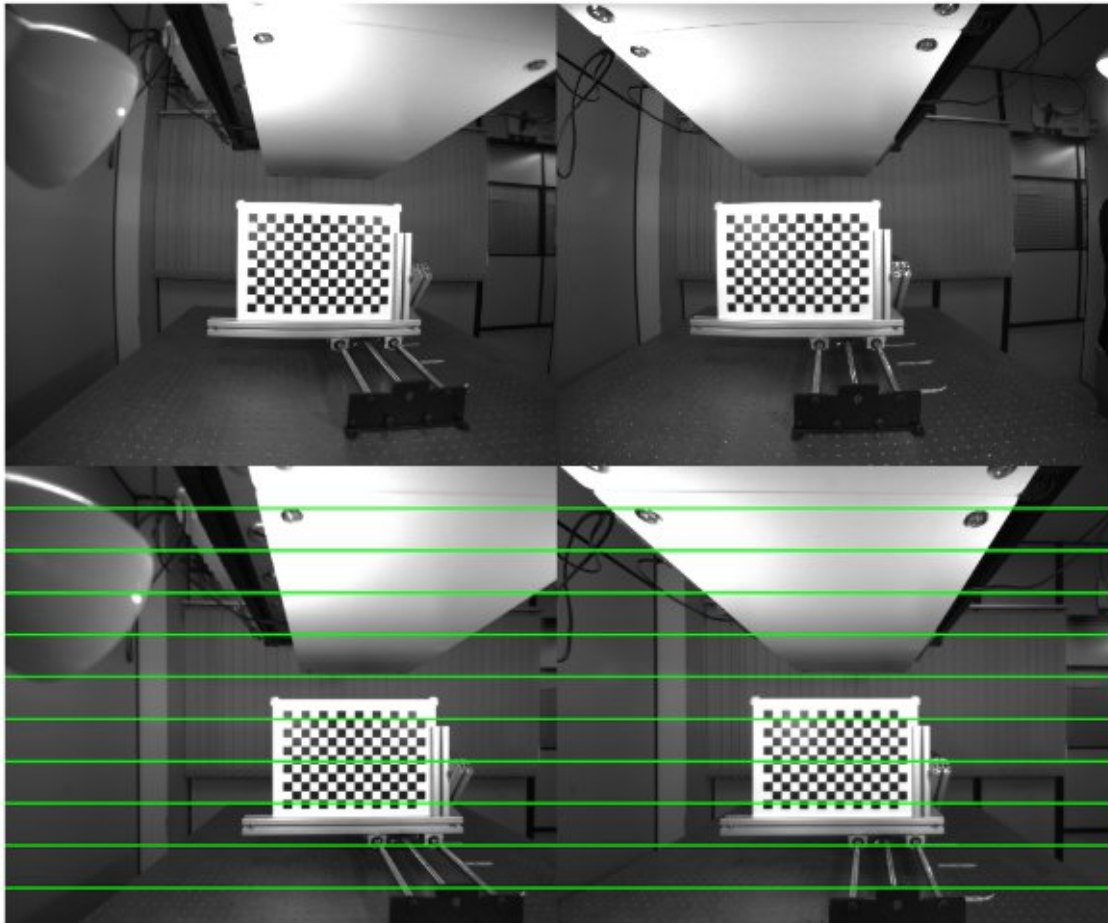
A geometria epipolar, como descrita anteriormente, estabelece que a correspondência para um ponto na imagem esquerda se encontra em uma linha na imagem direita. No caso geral, contudo, essas linhas epipolares podem ter qualquer orientação, o que pode tornar o algoritmo de busca computacionalmente complexo.

Para simplificar drasticamente o problema de correspondência, aplica-se um processo chamado retificação estéreo. A retificação é uma transformação projetiva (ou *warping*) que é aplicada ao par de imagens original. O objetivo deste processo é alinhar virtualmente as duas imagens de forma que os planos de imagem se tornem coplanares e, como consequência, todas as linhas epipolares se tornem horizontais e colineares.

A grande vantagem da retificação é que, no par de imagens retificadas, a busca por um ponto correspondente torna-se extremamente eficiente. Para um ponto (u,v) na imagem esquerda, seu correspondente na imagem direita estará garantidamente na mesma linha horizontal, ou seja, em uma coordenada com a mesma componente vertical v . Isso reduz a busca 1D em uma linha de orientação qualquer para uma varredura simples ao longo de uma mesma linha de *pixels*. A

Figura 5 mostra visualmente o efeito do processo de retificação em um par de imagens estéreo.

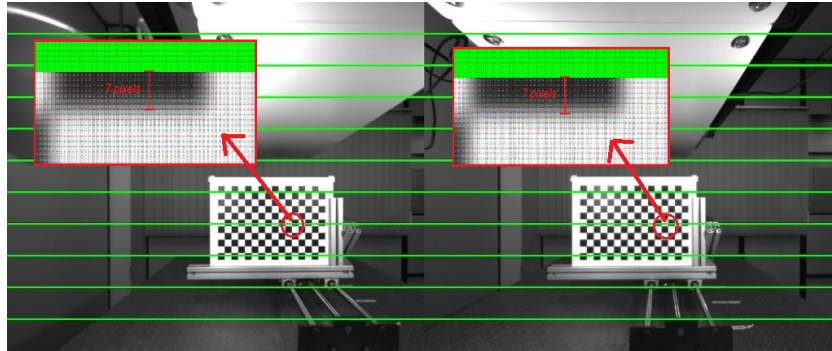
Figura 5 – Par de imagens estéreo antes e depois do processo de retificação.



Fonte: Do Autor.

Para verificar a qualidade da retificação em nível de *pixel*, um ponto de correspondência pode ser analisado em detalhe, como demonstrado na Figura 6. A análise ampliada da vizinhança do ponto destacado em ambas as imagens revela que o centro do ponto correspondente se encontra na mesma coordenada vertical. Neste exemplo, nota-se que o ponto em ambas as imagens esquerda e direita se posiciona a uma distância vertical idêntica (aproximadamente 7 *pixels*) da linha epipolar mais próxima, confirmando o sucesso do alinhamento horizontal com baixa incerteza de medição.

Figura 6 – Análise detalhada do alinhamento de um ponto correspondente.



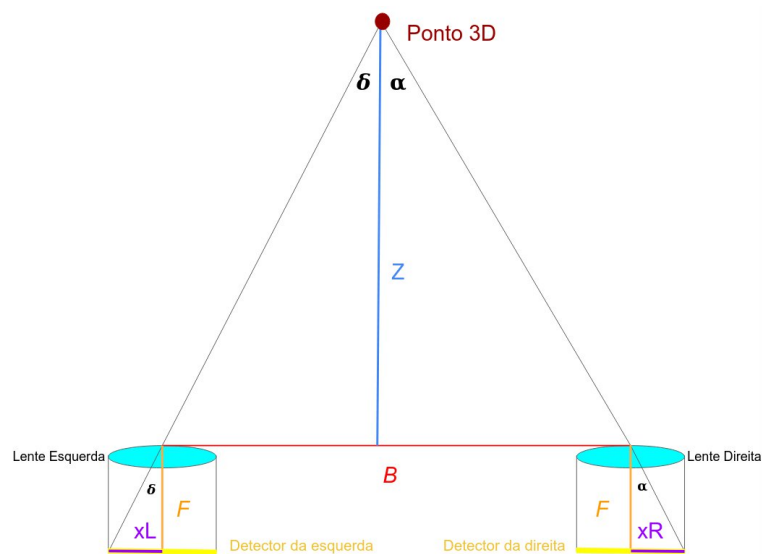
Fonte: Do Autor.

Com o par de imagens estéreo devidamente retificado, o processo de triangulação para encontrar a profundidade de um ponto é simplificado. Como um ponto e seu correspondente agora se encontram na mesma linha horizontal nas duas imagens, a diferença entre suas posições é uma simples medida unidimensional, conhecida como disparidade estéreo (d).

A profundidade (Z) do ponto 3D P pode ser então calculada diretamente a partir da disparidade, da linha de base (B) e da distância focal (F) da câmera, através da relação de semelhança de triângulos, como ilustrado na Figura 7. Para as projeções x_L e x_R mostradas na Figura 7, a disparidade é definida na equação 9:

$$d = x_L - x_R \quad (9)$$

Figura 7 – Par de imagens estéreo antes e depois do processo de retificação.



Fonte: Do Autor.

A partir da semelhança de triângulos na geometria da Figura 7, podemos derivar a relação fundamental entre profundidade e disparidade, chegando à equação 10, a equação fundamental da visão estéreo:

Esta equação é obtida assumindo as condições ideais de um sistema estéreo perfeitamente retificado, onde os eixos ópticos das câmeras são paralelos e perpendiculares à linha de base.

$$Z = \frac{B * F}{d} \quad (10)$$

Esta equação revela a relação da estereoscopia: a profundidade de um ponto (Z) é inversamente proporcional à sua disparidade (d). Isso significa que pontos muito distantes da câmera terão uma disparidade pequena (próxima de zero), enquanto pontos muito próximos terão uma disparidade grande. É este princípio que permite que algoritmos como o ORB-SLAM3 estimem a estrutura tridimensional da cena e a trajetória do sistema a partir de um par de visão estéreo.

2.1.3 Calibração de Câmeras

Nas seções anteriores, estabelecemos os modelos matemáticos ideais para descrever a formação de imagem em câmeras individuais (o modelo *pinhole*) e em pares (o modelo estéreo). Contudo, esses modelos são abstrações geométricas que não representam com total fidelidade o comportamento de câmeras e lentes do mundo real. Câmeras físicas possuem pequenas imperfeições de fabricação e, mais importante, utilizam conjuntos de lentes complexas que causam desvios significativos em relação à projeção linear perfeita assumida pelo modelo *pinhole*. As mais notáveis dessas imperfeições são as distorções geométricas da lente. Conforme descritas na literatura, essas distorções são governadas por um conjunto de coeficientes que modelam desvios radiais e tangenciais, descrevendo matematicamente como o plano de imagem real se difere do plano ideal (29).

Essas distorções fazem com que pontos que deveriam ser projetados em uma determinada coordenada de *pixel* apareçam, na realidade, em locais ligeiramente diferentes, um efeito que se intensifica nas bordas da imagem. Para aplicações de visualização geral, essa imprecisão pode ser imperceptível. No entanto, para aplicações de medição de alta qualidade, como a reconstrução 3D métrica e o

Mapeamento e Localização Simultânea (SLAM), essas distorções não podem ser ignoradas. Utilizar as coordenadas de *pixels* diretamente de uma imagem não calibrada na equação 10 de profundidade estéreo resultaria em erros significativos no cálculo da distância, comprometendo a incerteza de medição da trajetória estimada e do mapa 3D gerado.

É neste ponto que a calibração de câmeras se torna um processo fundamental e indispensável. A calibração é o processo de determinar experimentalmente todos os parâmetros que descrevem o modelo completo de uma câmera – tanto seus parâmetros intrínsecos (distância focal, ponto principal) quanto os coeficientes que modelam as distorções da sua lente. O objetivo final é criar um modelo matemático preciso da câmera que nos permita, a partir de uma coordenada de *pixel* observada e distorcida, encontrar a coordenada do ponto correspondente em um sistema ideal não distorcido, viabilizando assim cálculos geométricos exatos.

O modelo de projeção ideal descreve a geometria de uma câmera pinhole perfeita no espaço 3D (23, 30, 31, 32). Como vimos anteriormente, ele combina os parâmetros extrínsecos, que definem a pose da câmera (sua matriz de Rotação $[R]$ e vetor de Translação $[t]$ em relação a um sistema de coordenadas global), com os parâmetros intrínsecos, que descrevem as propriedades ópticas internas da câmera (agrupados na matriz de calibração K).

A equação completa que mapeia um ponto 3D no mundo, $P_m = [X_m, Y_m, Z_m, 1]^T$ (em coordenadas homogêneas), para um ponto na imagem $p = [u, v, 1]^T$ (também em coordenadas homogêneas) é dada pela equação 11:

$$\lambda \begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \begin{bmatrix} Fx & 0 & Cx \\ 0 & Fy & Cy \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} R11 & R12 & R13 & T1 \\ R21 & R22 & R23 & T2 \\ R31 & R32 & R33 & T3 \end{bmatrix} \begin{bmatrix} X_m \\ Y_m \\ Z_m \\ 1 \end{bmatrix} \quad (11)$$

Onde λ é um fator de escala (profundidade). O resultado desta equação são as coordenadas de *pixel* (u, v) ideais, ou seja, onde o ponto 3D seria projetado se a câmera não tivesse nenhuma distorção de lente.

Na prática, as lentes introduzem distorções não-lineares que fazem com que as coordenadas de *pixel* observadas na imagem real, (u_d, v_d) , sejam diferentes das coordenadas ideais, (u, v) , calculadas acima. Os tipos mais significativos de distorção são:

Distorção Radial: É o efeito mais proeminente, causado pela curvatura da lente. Faz com que linhas retas na cena pareçam curvadas na imagem, especialmente perto das bordas. Pode se manifestar como distorção em barril (*barrel distortion*), comum em lentes grande-angulares, ou distorção em almofada (*pincushion distortion*). É modelada por um conjunto de coeficientes radiais, tipicamente k_1, k_2 e k_3 .

Distorção Tangencial: Ocorre quando a lente não está perfeitamente alinhada e paralela ao plano do sensor da câmera. É modelada por dois coeficientes tangenciais, p_1 e p_2 .

A Figura 8 demonstra o impacto visual da distorção radial, o tipo mais comum, exibindo como uma grade de linhas que são retas no mundo real é mapeada de forma curvilínea pela lente da câmera, gerando os efeitos conhecidos como barril, almofada e complexa.

Figura 8 – Tipos de distorções de imagem.



Fonte: Do Autor

Para corrigir esses efeitos, o modelo completo de distorção é aplicado. Dadas as coordenadas de um ponto ideal normalizado $x' = (u - cx) / Fx$ e $y' = (v - cy) / Fy$, suas coordenadas distorcidas, xd e yd , são calculadas pelas seguintes equações 12, 13 e 14:

$$xd = x'(1 + k_1r^2 + k_2r^4 + k_3r^6) + [2p_1x'y' + p_2(r^2 + 2x'^2)] \quad (12)$$

$$yd = y'(1 + k_1r^2 + k_2r^4 + k_3r^6) + [2p_2x'y' + p_1(r^2 + 2y'^2)] \quad (13)$$

$$r^2 = x'^2 + y'^2 \quad (14)$$

O processo de calibração, portanto, visa encontrar todos esses parâmetros – K, R, t , e os coeficientes k_i, p_i para que se possa mapear com precisão qualquer ponto do mundo 3D para sua localização exata na imagem distorcida. Esse mapeamento bidirecional é fundamental tanto para a projeção de pontos globais na imagem quanto para o processo inverso de correção da distorção, que é a base para a reconstrução 3D.

O processo para determinar os parâmetros intrínsecos e de distorção de uma câmera é tipicamente realizado através da observação de um objeto de calibração, ou padrão, cuja geometria no mundo 3D é conhecida com reduzida incerteza de medição. Ao capturar múltiplas imagens deste padrão a partir de diferentes pontos de vista, um algoritmo estabelece um conjunto de correspondências entre pontos 3D de referência no padrão e suas projeções 2D (em *pixels*) na imagem. A abordagem mais influente e amplamente adotada para essa tarefa foi proposta por Zhang (33), que demonstrou ser possível realizar uma calibração confiável utilizando um padrão planar, como um tabuleiro de xadrez, fotografado em diversas orientações. Ao longo dos anos, diversos tipos de padrões de calibração foram desenvolvidos, cada um com suas vantagens. Os mais comuns na literatura são:

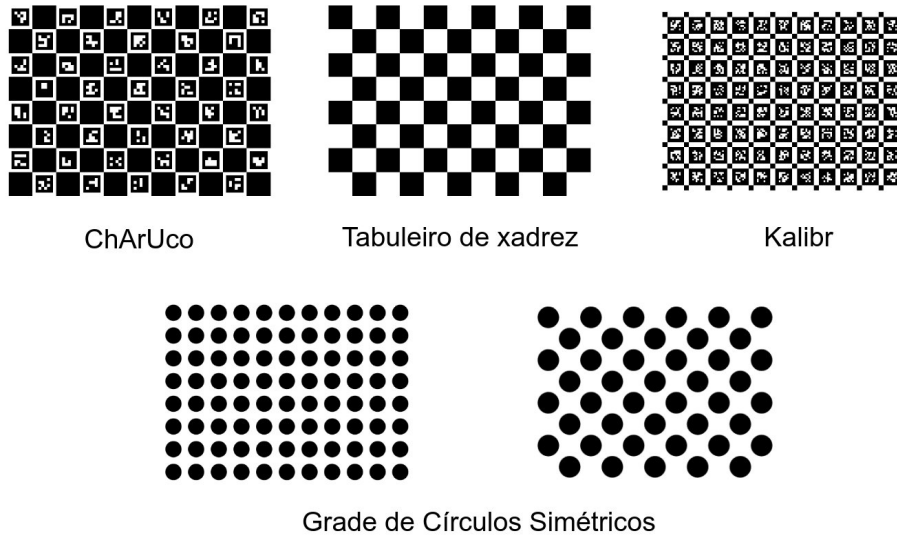
Tabuleiro de Xadrez (*Chessboard*): É o método mais clássico e difundido. O algoritmo busca pelos "pontos de sela" (os cantos internos onde os quadrados pretos e brancos se encontram Figura 9), que podem ser detectados com baixa incerteza de medição. Sua principal vantagem é a simplicidade e a vasta disponibilidade de algoritmos para sua detecção.

Grade de Círculos Simétricos: Este padrão utiliza uma grade de círculos. A detecção é baseada no cálculo do centroide de cada círculo (Figura 9), o que pode ser mais estável a imagens desfocadas e menos sensível a problemas de iluminação do que a detecção de cantos.

Padrões ChArUco: Trata-se de um padrão híbrido, popularizado pela biblioteca OpenCV, que combina as características de um tabuleiro de xadrez com marcadores ArUco (marcadores fiduciais quadrados com códigos binários internos Figura 9). A presença dos marcadores torna a detecção do padrão extremamente resiliente,

permitindo calibrações mesmo com oclusões parciais e fornecendo uma identificação única para cada canto detectado.

Figura 9 – Diversos padrões de calibração.



Fonte: Do Autor

A implementação dos algoritmos de detecção de padrões e da otimização matemática para resolver os parâmetros da câmera está disponível em diversas ferramentas de software e bibliotecas consolidadas, amplamente utilizadas na academia e na indústria. Entre as mais notáveis estão a biblioteca de código aberto OpenCV, o aplicativo *Camera Calibrator* do MATLAB e a clássica *Camera Calibration Toolbox for Matlab* de Jean-Yves Bouquet (34).

Uma vez que o conjunto completo de parâmetros intrínsecos, extrínsecos e de distorção de um sistema de câmeras é determinado através do processo de calibração, dois resultados práticos e fundamentais são alcançados, viabilizando a sua aplicação em tarefas de medição.

O resultado mais direto da calibração é a capacidade de compensar as distorções da lente. Utilizando os coeficientes de distorção (k_i , p_i) encontrados, é possível aplicar uma transformação matemática inversa a uma imagem "crua" capturada pela câmera). Este processo, conhecido como *undistortion* ou correção da imagem, gera uma nova imagem na qual as linhas que são retas na cena são de fato retas na imagem. Em essência, ele transforma a imagem de uma câmera real para

que ela se comporte de acordo com o modelo *pinhole* ideal, um pré-requisito para muitos algoritmos de visão computacional que assumem projeção linear.

Mais importante ainda, a calibração transforma a câmera de um simples dispositivo de imagem em uma ferramenta de medição 3D. Ao fornecer os valores métricos precisos para a distância focal (F) e a linha de base (B), a equação fundamental da visão estéreo (Equação 10) deixa de ser uma relação puramente teórica e se torna uma fórmula funcional para calcular distâncias reais no mundo.

Adicionalmente, o conhecimento dos parâmetros extrínsecos (Rotação R e Translação t) é o pré-requisito matemático para realizar a retificação estéreo (discutida na seção 2.1.2), o processo que alinha as imagens para simplificar a busca por correspondências.

Em resumo, a calibração é a ponte que conecta os modelos geométricos ideais com o hardware físico, garantindo que a reconstrução 3D e a estimativa de trajetória do SLAM sejam não apenas visualmente coerentes, mas com a máxima fidelidade métrica para a instrumentação utilizada.

2.2 NAVEGAÇÃO INERCIAL

Esta seção do trabalho aborda os conceitos fundamentais da navegação inercial. Trata-se de uma tecnologia de estimação de movimento baseada em sensores que medem as acelerações e velocidades angulares de um corpo, permitindo o cálculo de sua trajetória e orientação de forma autônoma. Os dispositivos, princípios de funcionamento e modelos de incerteza associados a esta técnica são detalhados a seguir.

2.2.1 Unidades de Medição Inercial (IMU)

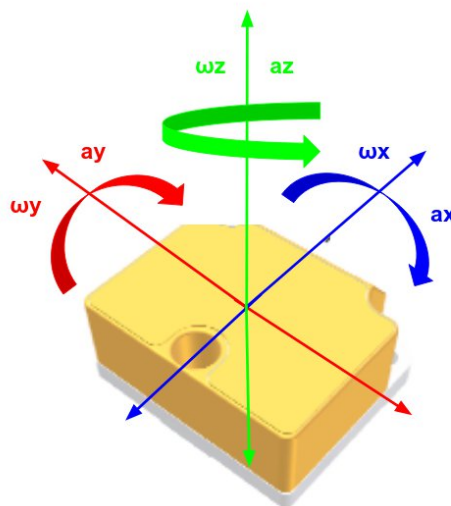
A plataforma de hardware que viabiliza a navegação inercial é a Unidade de Medição Inercial, ou IMU (*Inertial Measurement Unit*). Este dispositivo é o componente central responsável por captar as grandezas de movimento de um corpo. Uma definição abrangente é fornecida por Machado *et al.* (35):

Uma unidade de medição inercial (IMU) é um dispositivo eletrônico que mede e relata as componentes de aceleração resultantes de um corpo, relacionadas as coordenadas

cartesianas, a taxa de variação angular em torno de cada eixo dessas coordenadas e, em alguns casos, o campo magnético que circunda o corpo em questão. Todas essas medidas são feitas por meio dos sensores que compõem o IMU, sendo eles os acelerômetros, giroscópios e os magnetômetros.

Embora a definição cite um conjunto de três sensores, a configuração fundamental e mais comum para aplicações de robótica e navegação consiste na combinação de um acelerômetro de três eixos e um giroscópio de três eixos. Essa dupla de sensores permite medir as três componentes da aceleração linear (a_x , a_y , a_z) e as três componentes da velocidade angular (ω_x , ω_y , ω_z), conforme ilustrado na Figura 10. O conjunto dessas medições caracteriza o que é conhecido como uma IMU de 6 Graus de Liberdade (DoF, *Degrees of Freedom*) (36).

Figura 10 – Representação de uma IMU e seus eixos de medição.



Fonte: STIM300 *Datasheet* (37 Modificado)

O magnetômetro, também mencionado na citação, é um sensor adicional presente em alguns modelos (que os classifica como de 9 DoF), mas sua função de "bússola" é extremamente sensível a interferências magnéticas de motores e metais. Por essa razão, para garantir a estabilidade dos dados de movimento, muitas aplicações de alta performance priorizam o uso de IMUs de 6 DoF.

A vasta disseminação das IMUs nas últimas décadas foi impulsionada principalmente pelo avanço da tecnologia de Sistemas Microeletromecânicos (MEMS). Essa tecnologia permite a fabricação de sensores inerciais em escala microscópica e a baixo custo, integrando componentes mecânicos e eletrônicos em

um único substrato de silício. Antes dos MEMS, os sensores inerciais eram dispositivos grandes e caros, restringindo seu uso a aplicações aeroespaciais e militares. A tecnologia MEMS democratizou o acesso à navegação inercial, tornando-a um componente padrão em robôs e veículos autônomos (38). Dentro da tecnologia MEMS, existe uma gama de sensores, desde os de baixo custo (grau de consumo) até unidades de alta performance (grau tático ou industrial), projetadas para aplicações que exigem maior precisão e estabilidade.

2.2.2 Princípios de Funcionamento: Giroscópio e Acelerômetro

Uma vez definida a Unidade de Medição Inercial como um conjunto de sensores, é fundamental compreender os princípios físicos que governam a operação de seus componentes principais. A capacidade da IMU de quantificar o movimento se origina da conversão de fenômenos mecânicos como os resultantes da aplicação da Segunda Lei de Newton e do Efeito Coriolis em sinais elétricos que podem ser processados e interpretados (39). O entendimento desses princípios é crucial para a correta utilização dos dados e para a modelagem dos erros inerentes a cada sensor (40). As seções seguintes detalham os conceitos por trás do funcionamento dos dois principais componentes de uma IMU: o giroscópio e o acelerômetro.

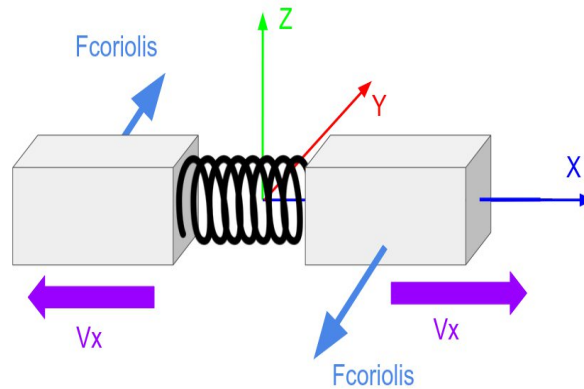
2.2.2.1 *Giroscópio*

O giroscópio é o sensor responsável por medir o vetor de velocidade angular (Ω), ou seja, a taxa de rotação de um corpo em torno de um eixo. Em uma IMU de 6 DoF, um conjunto de três giroscópios ortogonais mede a velocidade angular em torno dos eixos x, y e z do sensor (Ω_x , Ω_y , Ω_z), fornecendo uma descrição completa do movimento rotacional do sistema.

A maioria dos giroscópios modernos baseados em tecnologia MEMS opera com base no Efeito Coriolis. Para aplicar este princípio, o sensor utiliza uma microestrutura, frequentemente na forma de um diapasão vibratório (*tuning fork*), como ilustrado na Figura 11. Nela, duas massas de prova (m) são acionadas eletronicamente para oscilar em direções opostas ao longo de um eixo (o eixo de

acionamento, com velocidade V_x). Isso garante que as massas possuam uma velocidade linear conhecida e constante.

Figura 11 – Modelo de um giroscópio MEMS vibratório do tipo diapasão.



Fonte: Do autor.

Quando o sensor é submetido a uma rotação externa em um eixo perpendicular ao movimento das massas (a velocidade angular Ω_z no eixo z , que se deseja medir), a força de Coriolis atua sobre elas. Essa força, descrita na equação (15), é perpendicular tanto ao eixo de rotação quanto à direção do movimento das massas, fazendo com que elas sejam desviadas e também passem a vibrar em um novo eixo (o eixo de detecção, neste caso, o eixo y).

$$F_{Coriolis} = -2m(\Omega \times V_x) \quad (15)$$

A amplitude desse deslocamento secundário no eixo de detecção é diretamente proporcional à magnitude da velocidade angular de entrada (Ω_z). Este deslocamento é medido por eletrodos que detectam a variação de capacitância causada pela aproximação ou afastamento das massas. Um circuito eletrônico então converte essa variação de capacitância em um sinal de tensão ou em um valor digital que representa a medição da velocidade angular (41).

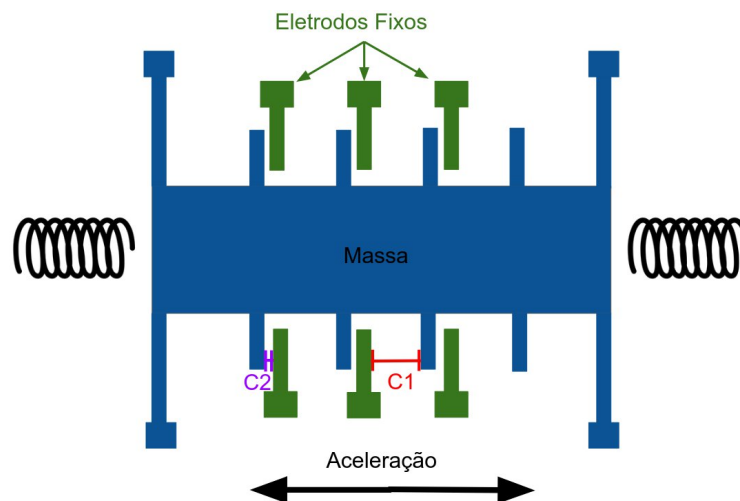
2.2.2.2 Acelerômetro

O acelerômetro é o sensor projetado para medir a aceleração própria (a), que representa a aceleração que um corpo experimenta em seu próprio sistema de

referência, incluindo a influência da gravidade. Em uma IMU, três acelerômetros são dispostos ortogonalmente para medir as componentes a_x , a_y , e a_z do vetor de aceleração.

O princípio de funcionamento de um acelerômetro MEMS moderno baseia-se em um sistema microeletromecânico que pode ser modelado como uma massa de prova (ou massa sísmica) conectada a uma suspensão elástica (molas). A detecção do movimento é realizada, mais comumente, por um método de detecção capacitiva diferencial, conforme ilustrado de forma conceitual na Figura 12 (42).

Figura 12 – Modelo de um acelerômetro MEMS com detecção capacitiva diferencial.



Fonte: SENTURIA (42 Modificado)

Nesta estrutura, a massa de prova é um componente condutor que pode se mover ao longo do eixo de sensibilidade. De cada lado da massa de prova, existem eletrodos fixos. A massa de prova e os eletrodos fixos formam um par de capacitores de placas paralelas, identificados na Figura 12 como C1 e C2.

Quando o sensor está em repouso (aceleração nula), a massa de prova fica centralizada. A distância entre a massa e os eletrodos de cada lado é igual, resultando em capacitâncias iguais: $C1=C2$.

Quando o sensor é submetido a uma aceleração ao longo do seu eixo de sensibilidade, a massa de prova, por sua inércia, se desloca na direção oposta. Por exemplo, se a aceleração for para a direita, a massa se desloca para a esquerda. Este deslocamento altera as distâncias: a distância do capacitor C1 diminui, enquanto a do

capacitor C2 aumenta. Como a capacitância é inversamente proporcional à distância entre as placas ($C=f(1/d)$), a consequência é um aumento na capacitância de C1 e uma diminuição na de C2.

O circuito eletrônico do sensor mede a diferença entre essas duas capacitâncias ($\Delta C=C1-C2$). Essa medição diferencial é extremamente sensível ao deslocamento da massa e, por consequência, à aceleração de entrada. Além disso, essa abordagem ajuda a rejeitar erros e ruídos que afetam ambos os capacitores igualmente. O valor de ΔC é então convertido em um sinal de tensão ou digital que representa a medição da aceleração (40).

É crucial entender que, por medir a aceleração própria, o acelerômetro é sensível à gravidade. Quando um acelerômetro está em repouso sobre uma superfície horizontal, ele mede a aceleração de reação que a superfície exerce sobre ele para anular a força da gravidade, que é de aproximadamente $+9.8 \text{ m/s}^2$ em seu eixo vertical. Portanto, a saída de um acelerômetro é sempre a soma vetorial da aceleração do corpo e da reação à aceleração da gravidade (39).

2.2.3 Modelos de Incerteza e Calibração de IMUs

As seções anteriores descreveram o princípio de funcionamento de um giroscópio e um acelerômetro ideais. No entanto, os sensores físicos do mundo real estão sujeitos a uma variedade de imperfeições e fontes de ruído que fazem com que suas medições se desviem dos valores verdadeiros. Quando estes dados imprecisos são integrados ao longo do tempo para estimar a orientação e a posição, os erros numéricos se acumulam rapidamente, resultando no fenômeno de deriva (*drift*), que torna a navegação inercial pura inviável para longos períodos(40, 43).

Para que uma IMU seja utilizada de forma confiável em aplicações de baixa incerteza de medição, é imperativo modelar, quantificar e compensar esses erros. Este processo envolve duas abordagens teóricas e práticas complementares: a Modelagem de Erros Sistemáticos e a Caracterização. A modelagem de erros sistemáticos foca nos erros determinísticos (sistemáticos), com o objetivo de encontrar parâmetros para a correção matemática do sinal bruto. A caracterização, por sua vez, foca nos erros estocásticos (aleatórios), e seu objetivo é quantificar as propriedades

estatísticas da incerteza para a sintonia de filtros de estimação, como o Filtro de Kalman.

Ambos os processos contribuem para a construção de um modelo matemático abrangente que descreve a relação entre a medição bruta do sensor e a grandeza física verdadeira. Este modelo, cujos termos serão detalhados nas seções seguintes, geralmente assume a seguinte forma apresentada na equações 16 e 17:

Modelo do Giroscópio:

$$\tilde{\omega} = M_g S_g (w_{real} + b_g + n_g) \quad (16)$$

Modelo do Acelerômetro:

$$\tilde{a} = M_g S_g (a_{real} + b_a + n_g) \quad (17)$$

Onde $\tilde{\omega}$ e $\tilde{a}$ são as medições brutas dos sensores, enquanto w_{real} e a_{real} são os valores físicos verdadeiros. Os termos de erro incluem as matrizes de desalinhamento (M_g), as matrizes de fator de escala (S_g), os vetores de vício (b_a) e os vetores de ruído (n_g).

As seções a seguir irão detalhar cada um dos termos de erro presentes nestes modelos, começando pelos erros determinísticos que são o alvo do processo de calibração.

2.2.3.1 Erros Determinísticos

Os erros determinísticos são desvios sistemáticos e previsíveis, e a sua compensação através da calibração é fundamental para a melhoria do desempenho de uma IMU (44). O mais significativo destes erros é o *bias*, uma medição de *offset* constante que, ao ser integrada, causa o principal erro de deriva na navegação. Soma-se a ele o erro de fator de escala, que representa um ganho incorreto no sensor, fazendo com que as medições sejam proporcionalmente maiores ou menores que a realidade, e o erro de desalinhamento, que ocorre quando os eixos do sensor não são perfeitamente ortogonais, causando um "vazamento" de medição entre eles (45).

Em conjunto, estes erros distorcem a percepção do movimento real do sistema. O *bias* introduz um *drift* constante, enquanto o fator de escala e o desalinhamento corrompem a magnitude e a geometria 3D do movimento, respectivamente. O processo de calibração, portanto, é a etapa fundamental que busca estimar e modelar numericamente estes três erros para que suas contribuições possam ser matematicamente removidas das medições brutas, resultando em um fluxo de dados significativamente mais exato e confiável.

2.2.3.2 Erros Estocásticos

Diferentemente dos erros sistemáticos, os erros estocásticos são de natureza aleatória e, portanto, não podem ser eliminados por uma correção fixa. Contudo, a construção de um modelo estocástico preciso é uma etapa necessária, tanto para prever o desempenho do sistema quanto para implementar um sistema de referência de atitude e direção ou um navegador funcional (46). O principal deles é o Ruído Branco (*White Noise*), uma flutuação de alta frequência e sem correlação presente em todas as medições, que pode ser entendida como a "estática" do sinal. Outro crucial é a instabilidade da tendência (*Bias Instability*), que descreve a variação lenta e imprevisível do próprio bias ao longo do tempo, fazendo com que o valor calibrado inicialmente perca sua validade (45).

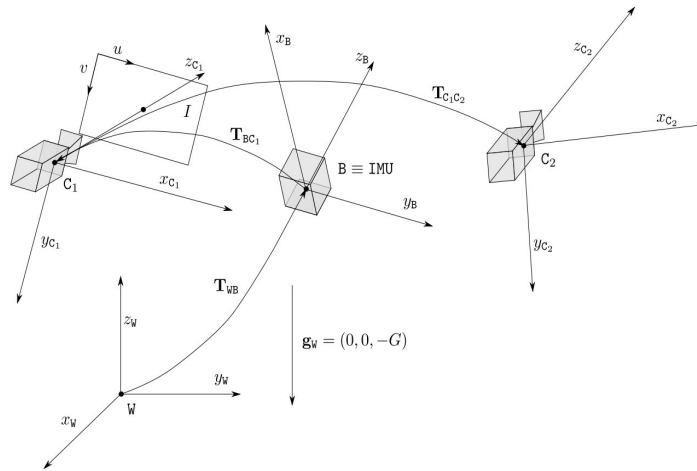
O impacto destes erros é gerenciado através da sua caracterização estatística, geralmente realizada com a análise de Variância de Allan. O Ruído Branco, ao ser integrado, resulta no erro de passeio aleatório (*Angle/Velocity Random Walk*), enquanto a Instabilidade do Vício degrada a performance do sistema a longo prazo. Os parâmetros obtidos na caracterização não corrigem o sinal, mas são essenciais para a sintonia de filtros de estimação, como o Filtro de Kalman, informando ao algoritmo o nível de incerteza de cada medição e permitindo uma fusão de dados mais estável e inteligente.

2.2.4 Calibração Extrínseca Câmera-IMU

Enquanto a calibração intrínseca, discutida na seção anterior, trata dos erros internos de um único sensor, a fusão de dados em um sistema visual-inercial depende

da relação geométrica entre os sensores. O processo para determinar esta relação é a calibração extrínseca. Para entender a transformação, é necessário definir os sistemas de referência utilizados. No caso de uma configuração como a do ORB-SLAM3, estes são definidos como o Global (W), o do Corpo/IMU (B) e o da Câmera (C) como pode se observar na Figura 13 (12).

Figura 13 – Sistemas de referência definidos para a configuração estéreo-inercial do ORB-SLAM3.



Fonte: Gómez Rodríguez *et al.* (12).

A relação entre a pose da câmera no mundo (T_{WC1}) e a pose do corpo/IMU no mundo (T_{WB}) é dada pela transformação extrínseca constante (T_{BC1}), que é o objetivo da calibração. A pose da câmera é, portanto, o resultado da concatenação da pose do corpo com a transformação extrínseca apresentada na equação 18:

$$T_{WC1} = T_{WB} T_{BC1} \quad (18)$$

Onde T_{WB} e T_{WC1} são as transformações homogêneas no Grupo Euclidiano Especial de três dimensões que levam pontos dos referenciais do corpo e da câmera para o referencial mundial, respectivamente. Em um sistema estéreo-inercial como o proposto neste trabalho, a IMU e as duas câmeras (Câmera 1 e Câmera 2) são montadas rigidamente em um mesmo corpo móvel, o que garante que as transformações extrínsecas entre o corpo e cada sensor (T_{BC1} e T_{BC2}) sejam constantes ao longo do tempo. A calibração extrínseca para este sistema consiste em encontrar as matrizes de transformação constante, minimizando a incerteza de medição associada.

A precisão desta calibração é um fator dominante no desempenho global de um sistema Visual-Inercial. Erros em T_{BC1} fazem com que o algoritmo de estimação tente reconciliar duas fontes de informação sobre o movimento que são fundamentalmente inconsistentes, resultando em acúmulo de erros e possível divergência da estimativa da trajetória.

O método geral para encontrar os parâmetros extrínsecos envolve a observação de um alvo de calibração a partir de múltiplas poses, enquanto a câmera e a IMU registram o movimento. Subsequentemente, um processo de otimização é executado para encontrar a transformação (T_{BC1}) e o atraso temporal (td) que minimizam a discrepância entre o movimento observado pela câmera e o medido pela IMU. Ferramentas de código aberto, como o Kalibr, são implementações práticas destes princípios teóricos (47, 48).

2.3 ODOMETRIA VISUAL-INERCIAL E SLAM

Esta seção estabelece o conceito para a compreensão dos modernos sistemas de SLAM visual-inercial. O foco reside na junção entre dados dos sensores visuais e inerciais, uma estratégia de fusão empregada para solucionar um dos problemas centrais da robótica autônoma. O percurso teórico aqui apresentado inicia-se com a investigação da complexa formulação do problema de Localização e Mapeamento Simultâneos e seus desafios inerentes. Na sequência, a discussão aprofunda-se nas principais arquiteturas para a fusão de dados estéreo-inerciais, analisando as diferenças fundamentais entre as abordagens e suas respectivas implicações na precisão e estabilidade do sistema de estimação.

2.3.1 O Problema do SLAM (*Simultaneous Localization and Mapping*)

De acordo com a formulação clássica de Durrant-Whyte & Bailey (49), o problema de Localização e Mapeamento Simultâneos (SLAM) questiona se é possível para um robô móvel, posicionado em um local desconhecido dentro de um ambiente desconhecido, construir incrementalmente um mapa consistente deste ambiente enquanto determina simultaneamente sua localização dentro deste mesmo mapa. Os autores destacam que a solução para o SLAM é vista como o "santo graal" da robótica

móvel, pois ela representa a capacidade de tornar um robô verdadeiramente autônomo (50). A complexidade intrínseca do problema reside na sua natureza de estimação acoplada: um mapa com elevada consistência métrica de medição é necessário para uma localização de alta fidelidade, e essa localização de alta fidelidade é pré-requisito para a construção de um mapa 3D consistente do ambiente. Essa formulação inicial do SLAM e seu papel central para a navegação autônoma permanecem válidos e são confirmados em trabalhos mais recentes na área (51)

Do ponto de vista matemático, o SLAM é formulado como um problema de inferência probabilística. O objetivo é estimar um conjunto de variáveis desconhecidas (o estado do sistema) com base em um conjunto de medições conhecidas. Vamos definir essas variáveis:

O Estado ($X_{1:k}$ e m): Representa tudo aquilo que o sistema deseja estimar e que não é diretamente observável. Ele é composto por duas partes: $X_{1:k} = \{x_1, x_2, \dots, x_k\}$: É a trajetória do agente ao longo do tempo. Cada x_k é a pose do agente (posição 3D e orientação 3D) no instante de tempo k . m : É a representação do mapa do ambiente. No seu caso, será a coleção de todos os pontos 3D (a nuvem de pontos) cuja posição o sistema está estimando.

As Entradas de Controle ($U_{1:k}$): Representam as medições de movimento do próprio agente, que informam como ele se moveu de um instante para o outro. São também chamadas de odometria. $U_{1:k} = \{u_1, u_2, \dots, u_k\}$: Cada u_k corresponde às medições de velocidade angular e aceleração linear fornecidas pela IMU entre os instantes $k - 1$ e k .

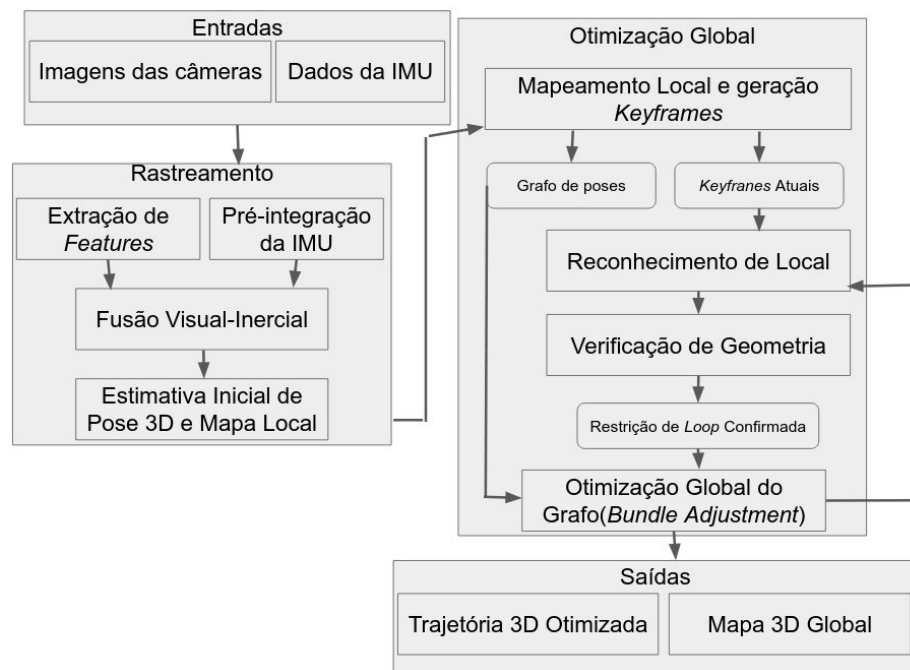
As Observações ($Z_{1:k}$): Correspondem às medições que o agente faz do ambiente externo, ou seja, como ele "vê" o mapa. $Z_{1:k} = \{z_1, z_2, \dots, z_k\}$: No sistema de visão, estas são as imagens capturadas pelas câmeras. Cada observação z_k é a imagem (ou o par de imagens estéreo) no instante k , da qual o sistema extrai as posições dos pontos de interesse (as características visuais) que correspondem aos pontos do mapa m .

O problema geral, conhecido como SLAM Completo (*Full SLAM*), consiste em usar as medições conhecidas (U e Z) para calcular a distribuição de probabilidade posterior sobre as variáveis desconhecidas (X e m) (49). A equação 19 que formaliza este objetivo.

$$P(X_{1:k}, m | Z_{1:k}, U_{1:k}) \quad (19)$$

Em outras palavras, a equação pergunta: "Qual é a probabilidade da trajetória $X_{1:k}$ e do mapa m serem os verdadeiros, dado que eu fiz as medições de movimento $U_{1:k}$ e tive as observações visuais $Z_{1:k}$ " (49, 52, 53). Para resolver este problema, os sistemas modernos implementam um fluxo de processamento complexo, conforme ilustrado no diagrama da Figura 14.

Figura 14 – Diagrama do funcionamento do SLAM.



Fonte: Do autor.

A arquitetura de um sistema SLAM moderno, especialmente um sistema visual-inercial é tipicamente dividida em duas etapas principais que operam em paralelo: uma Etapa de Rastreamento, que opera em tempo real, e uma Etapa de Otimização Global, que processa os dados de forma assíncrona para garantir a consistência do mapa.

Etapa de Rastreamento é a parte do sistema responsável por processar os dados brutos dos sensores em alta frequência para obter uma estimativa rápida, porém local, do movimento do agente. Como o próprio nome sugere, esta etapa "rastrea" a pose do sistema a cada instante. No sistema estéreo-inercial

desenvolvido, esta etapa executa a Odometria Visual-Inercial (VIO) e suas principais funções, conforme o diagrama, são:

Extração de Características (*Features*): Pontos de interesse salientes e rastreáveis nas imagens das câmeras, como cantos e texturas (54). Descritores como ORB (utilizado no ORB-SLAM3) são comumente empregados para esta tarefa.

Pré-integração da IMU: Processa as medições de alta frequência da unidade inercial (aceleração e velocidade angular) para gerar uma estimativa de movimento relativo entre os instantes em que as imagens são capturadas.

Fusão Visual-Inercial: Combina as informações das *features* visuais com o movimento estimado pela IMU para calcular a pose 3D atual do sistema. Esta fusão cria uma odometria mais estável, onde a IMU auxilia em movimentos rápidos ou em ambientes com pouca textura, e a visão corrige a deriva inerente da IMU.

O resultado desta etapa é uma Estimativa Inicial de Pose 3D e Mapa Local, que é localmente precisa, mas acumula erros de deriva ao longo de trajetórias extensas.

A etapa de Otimização Global opera de forma mais lenta e confiável, com o objetivo de refinar as estimativas do rastreamento e gerar um mapa globalmente consistente. A abordagem de estado da arte para esta tarefa é a otimização baseada em grafos, na qual o problema SLAM é modelado como um grafo de restrições esparsas (52, 53). Nesta técnica, a trajetória e o mapa são representados da seguinte forma:

Nós representam as poses do sistema em instantes de tempo chave (*keyframes*) (55) e as posições 3D dos pontos do mapa.

Arestas representam as restrições geométricas entre os nós, derivadas tanto das medições de movimento (odometria) quanto das observações visuais dos pontos do mapa.

O fluxo interno desta etapa, conforme o diagrama, é o apresentado a seguir:

Mapeamento Local e Geração de *Keyframes*: A estimativa inicial da pose alimenta este módulo, que constrói um Grafo de Poses e decide quando inserir um novo *keyframe* (um *Frame* considerado importante para o mapeamento).

Reconhecimento de Local e Verificação: Os *keyframes* atuais são enviados ao módulo de Reconhecimento de Local, que os compara com *keyframes* passados (da sessão atual ou de um mapa previamente carregado) para detectar um possível

fechamento de *loop* (*Loop Closure*) (56). Caso um candidato seja encontrado, a Verificação de Geometria confirma a correspondência em 3D para evitar falsos positivos.

Otimização Global do Grafo: O otimizador recebe o Grafo de Poses e as Restrições de *Loop* Confirmadas. Ele então executa um processo de otimização não linear, como o *Bundle Adjustment* (57, 58), para encontrar a configuração de poses e pontos do mapa que minimize o erro global de todas as restrições.

Realimentação: Um aspecto crucial, destacado no diagrama, é o ciclo de realimentação. O mapa e as poses dos *keyframes*, após serem corrigidos pela Otimização Global, são enviados de volta para o módulo de Reconhecimento de Local e Mapeamento Local. Isso permite que o sistema se localize e feche *loops* futuros com base em um mapa já mais preciso, aumentando a estabilidade geral.

A capacidade de fechar *loops* de forma confiável e otimizar o grafo de restrições é o que permite que um sistema SLAM completo gere mapas 3D precisos e em larga escala, superando as limitações da odometria simples.

Fica claro, portanto, que a arquitetura SLAM moderna é composta por duas etapas principais que trabalham em conjunto: o rastreamento em tempo real e a otimização global. O sucesso de todo o sistema depende diretamente da qualidade da Etapa de Rastreamento, responsável por processar os dados dos sensores. Em sistemas que utilizam tanto câmeras quanto uma IMU, como o deste trabalho, a estratégia para combinar, ou fundir, os dados desses dois sensores é de fundamental importância. Sendo assim, a próxima seção detalhará as principais abordagens para a fusão de dados em sistemas estéreo-inerciais.

2.3.2 Abordagens de Fusão de Dados Estéreo-Inerciais

A eficácia da Etapa de Rastreamento, descrita anteriormente, é ampliada pela fusão de dados de sensores com características complementares. Em um sistema visual-inercial, como o desenvolvido neste trabalho, as câmeras estéreo e a Unidade de Medição Inercial (IMU) oferecem visões distintas, porém sinérgicas, do movimento do agente.

As câmeras estéreo, descritas no item 2.1.3, fornecem informações visuais ricas sobre a geometria e a aparência do ambiente. Elas permitem a extração de

características e a estimativa de profundidade, resultando em informações de movimento em escala métrica absoluta. No entanto, são sensíveis a condições de iluminação desfavoráveis, movimentos rápidos que causam borrão de imagem (*motion blur*) e cenários com pouca ou nenhuma textura.

A IMU, descrita no item 2.2, é um sensor proprioceptivo que mede o movimento do próprio sistema. Ela fornece dados de alta frequência sobre a velocidade angular e a aceleração linear, sendo imune às condições visuais do ambiente. Sua força reside na estimação de movimento em curtos intervalos de tempo, mas suas medições sofrem com ruído e vieses (*bias*) que, quando integrados, resultam em um erro de deriva (*drift*) que cresce rapidamente.

A fusão dessas duas modalidades de sensores (59) visa, portanto, combinar a estabilidade operacional de curto prazo da IMU com a estabilidade de longo prazo e a referência externa da visão, criando um sistema de odometria de maior fidelidade métrica e mais estável do que qualquer um dos sensores operando isoladamente. de maior fidelidade métrica e mais estável

Na literatura de SLAM, existem duas estratégias principais para realizar a fusão de dados visuais e inerciais: o Acoplamento Fraco (*Loosely-Coupled*) e o Acoplamento Forte (*Tightly-Coupled*).

Na abordagem de *Loosely-Coupled*, os sistemas de odometria visual e de navegação inercial operam como módulos independentes. Cada subsistema calcula sua própria estimativa de pose e, em seguida, um filtro de estimação, como o Filtro de Kalman Estendido (EKF), é utilizado para fundir esses resultados já processados. Essa arquitetura se destaca pela sua relativa simplicidade e modularidade. Contudo, essa separação resulta em uma solução considerada sub-ótima, pois o fluxo de informação é limitado, impedindo que os dados brutos de um sensor sejam utilizados para auxiliar ou corrigir o processamento do outro em tempo real (60).

Em contrapartida, o *Tightly-Coupled* representa a abordagem de estado da arte para sistemas de baixa incerteza de medição. Nesta estratégia, as medições brutas de ambos os sensores (por exemplo, as coordenadas das características visuais nas imagens e as leituras de aceleração e velocidade angular da IMU) são integradas em um único e grande problema de otimização conjunta. Ao invés de processar os dados sequencialmente, o sistema busca um vetor de estado unificado que melhor explique todas as medições em um dado intervalo de tempo (61, 62).

A qualidade principal dessa abordagem é um processo de otimização não linear, frequentemente conhecido como *Bundle Adjustment* (63, 64). A complexidade e o alto custo computacional do *Tightly-Coupled* decorrem diretamente desse processo. Ele envolve a criação e a resolução de um problema de otimização massivo, onde o vetor de estado inclui não apenas a pose atual do agente, mas também as poses das câmeras e da IMU para um conjunto de *keyframes*, além das posições 3D de todos os pontos do mapa. O problema se torna ainda maior pela inclusão dos parâmetros estocásticos da IMU, como seus vieses de acelerômetro e giroscópio, que precisam ser estimados em tempo real (62).

Esse otimizador busca minimizar um erro quadrático, ajustando simultaneamente todas essas variáveis para alcançar a maior consistência possível entre as medições visuais (baseadas em erros de reprojeção) e as medições inerciais. A implementação de tal otimizador conjunto, no entanto, representa um desafio de maior complexidade e exige um custo computacional significativamente mais elevado que a abordagem de acoplamento fraco, devido ao grande número de variáveis no vetor de estado e à natureza iterativa e computacionalmente intensiva do processo de minimização.

Para os objetivos de baixa incerteza de medição propostos nesta dissertação, a escolha de um algoritmo como o ORB-SLAM3 foi motivada pelo fato de ele implementar uma sofisticada estratégia de *Tightly-Coupled*. Essa abordagem é fundamental para extrair a mínima incerteza de medição do conjunto de sensores de alta qualidade utilizado neste trabalho, composto pelas câmeras industriais e pela IMU.

2.4 O ALGORITMO ORB-SLAM 3

O ORB-SLAM 3 representa um dos mais avançados e completos sistemas de Localização e Mapeamento Simultâneo (SLAM) de código aberto disponíveis no GitHub. Sua arquitetura versátil oferece suporte a uma vasta gama de configurações de sensores, incluindo câmeras monoculares, estéreo e RGB-D, além de uma integração confiável com dados de Unidades de Medição Inercial (IMU), caracterizando-se como um sistema de SLAM visual-inercial de alta performance (12).

O sistema atual é o resultado de uma contínua evolução. Sua base foi estabelecida pelo pioneiro ORB-SLAM, um sistema versátil e preciso focado em câmeras monoculares (65). Posteriormente, o ORB-SLAM 2 expandiu essas capacidades ao incorporar suporte para câmeras estéreo e RGB-D, consolidando-se como uma solução de código aberto amplamente adotada pela comunidade de robótica e visão computacional (66). Com base nesses predecessores, o ORB-SLAM3 introduz melhorias significativas, como o sistema Atlas que é um gerenciador de múltiplos mapas que permite ao algoritmo operar em sessões distintas, recuperar-se de falhas de rastreamento e fundir mapas de forma coesa, conferindo uma estabilidade sem precedentes (12).

A escolha do ORB-SLAM 3 como ferramenta central para este trabalho de mestrado fundamenta-se precisamente em suas características de baixa incerteza de medição e estabilidade operacional. A capacidade de fusão de dados estéreo-inerciais (*stereo-inertial*) é diretamente compatível com o sistema desenvolvido, que emprega um par de câmeras industriais e uma IMU. A integração de dados inerciais é crucial para mitigar o desvio de trajetória (*drift*) em situações de movimentos rápidos ou em ambientes com pouca textura visual, cenários comuns em aplicações do mundo real.

A incerteza de medição e o desempenho de algoritmos de SLAM são frequentemente avaliados por meio de *benchmarks* públicos, como o disponibilizado pela ETH3D (67). Visto que o objetivo principal desta dissertação é avaliar a incerteza de medição da trajetória e da nuvem de pontos geradas, a renomada incerteza de medição do ORB-SLAM 3 o estabelece como um sistema de referência para validar o desempenho do *hardware* proposto.

Para uma compreensão aprofundada de seu funcionamento, segue mais informações no APÊNDICE A no qual detalha a arquitetura de processamento paralelo do sistema e os princípios do descritor de características ORB, que dá nome e fundamenta o algoritmo.

2.5 RECONSTRUÇÃO 3D COM FOTOGRAMETRIA

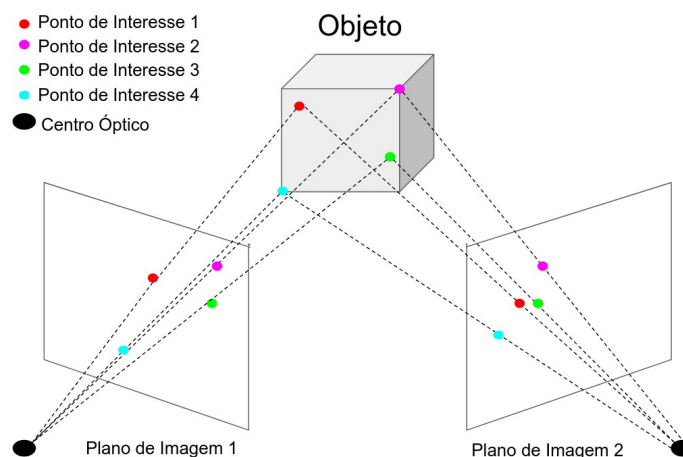
A fotogrametria é formalmente definida como a ciência, arte e tecnologia de obter informações confiáveis sobre objetos físicos e o ambiente por meio de processos de registro, medição e interpretação de imagens fotográficas e outras formas de

sensoriamento remoto (68). Em sua essência, a disciplina busca responder a uma questão fundamental: como extrair métricas e modelos tridimensionais precisos a partir de representações bidimensionais do mundo?

Historicamente, a fotogrametria teve suas origens e maior aplicação na cartografia, com foco em levantamentos aéreos para a criação de mapas topográficos. Essa abordagem clássica, muitas vezes dependente de equipamentos especializados e pontos de controle pré-definidos, evoluiu significativamente com os avanços nas áreas de visão computacional e processamento digital de imagens. A fotogrametria moderna, frequentemente denominada *Structure from Motion* (SfM), flexibilizou muitas dessas restrições, permitindo a reconstrução 3D a partir de conjuntos de imagens menos estruturados, capturados com câmeras convencionais (20, 69, 70).

O princípio fundamental por trás da fotogrametria baseia-se na triangulação. Em vez de determinar a posição de um único ponto, o processo é aplicado a um vasto conjunto de características ou pontos de interesse na superfície de um objeto. A Figura 15 ilustra este conceito de forma clara: múltiplos pontos de interesse em um objeto tridimensional são simultaneamente observados a partir de duas posições distintas (representadas pelos Centros Ópticos 1 e 2). A projeção de cada um desses pontos é registrada nos respectivos Planos de Imagem. Ao estabelecer a correspondência entre as projeções de um mesmo ponto nos dois planos, a intersecção de suas linhas de visada permite calcular sua posição 3D, e a reconstrução conjunta de todos os pontos revela a estrutura geométrica do objeto.

Figura 15 – Ilustração do princípio da triangulação para múltiplos pontos.



Fonte: Do autor.

Graças à sua capacidade de gerar reconstruções 3D geometricamente consistentes por meio de otimizações globais, as técnicas de fotogrametria são amplamente utilizadas na literatura como uma base de comparação de baixa incerteza de medição. Dessa forma, fornecem uma referência de alta fidelidade que permite a avaliação geométrica quantitativa de mapas gerados por outros algoritmos, analisando métricas como a fidelidade da estrutura, a consistência da escala e a deriva da trajetória.

2.5.1 Princípios da Estrutura a partir do Movimento (*Structure from Motion*)

O SfM é uma técnica de fotogrametria e visão computacional que permite reconstruir a estrutura tridimensional de uma cena a partir de uma série de imagens bidimensionais capturadas de diferentes pontos de vista (70). O princípio fundamental do SfM é a capacidade de estimar simultaneamente a geometria da cena (a "Estrutura") e as poses da câmera (o "Movimento") que geraram as imagens. A entrada para um sistema de SfM é tipicamente um conjunto de imagens não ordenadas, e sua saída consiste em dois componentes principais: uma nuvem de pontos 3D esparsa da cena e os parâmetros de pose (posição e orientação) para cada uma das câmeras de entrada.

O processo de SfM, especialmente em sua abordagem incremental, que é a mais comum, é implementado como uma linha de múltiplos estágios, projetado para identificar correspondências confiáveis entre as imagens e otimizar a geometria da reconstrução. As etapas canônicas deste *pipeline* são descritas a seguir:

Detecção e correspondência de características é primeiro passo que consiste em identificar pontos de interesse visualmente distintos em cada imagem. Algoritmos detectores e descritores de características, como SIFT (*Scale-Invariant Feature Transform*) (71, 72), SURF ou ORB, conforme discutido na Seção 2.4.2, são empregados para extrair esses pontos e associá-los a um vetor descritor. Em seguida, os descritores de todas as imagens são comparados para encontrar pares de características correspondentes, ou seja, projeções do mesmo ponto 3D em diferentes imagens.

Com a correspondência de características realizada é feita a Verificação Geométrica que é necessária após correspondência de características que

invariavelmente produzem um número de associações incorretas (*outliers*). Para filtrar esses erros, uma verificação de consistência geométrica é realizada. Utilizando um algoritmo confiável como o RANSAC (*Random Sample Consensus*) (73, 74), estima-se a matriz fundamental ou essencial que descreve a geometria epipolar (conforme Seção 2.1.3) entre pares de imagens. As correspondências que não se ajustam a este modelo geométrico são descartadas.

Com isso *pipeline* de uma reconstrução inicial é iniciado selecionando-se um par de imagens adequado, que geralmente apresenta uma boa sobreposição e uma distância de linha de base suficiente para triangulação precisa. A pose relativa entre essas duas câmeras é calculada, e as correspondências de características validadas são trianguladas para formar a nuvem de pontos 3D inicial. Este primeiro modelo define a origem e a escala arbitrária de todo o sistema de coordenadas da reconstrução.

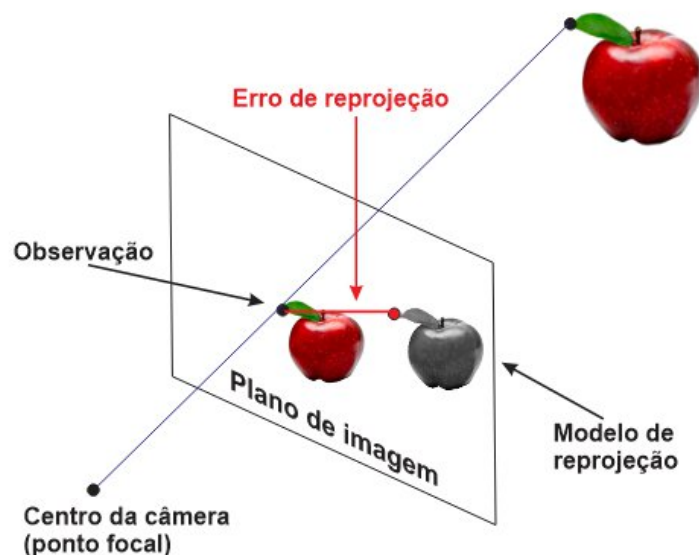
Após a inicialização, o modelo é expandido através de uma reconstrução incremental, que iterativamente adiciona uma nova imagem de cada vez. Para cada nova imagem, suas correspondências 2D com os pontos 3D já existentes no mapa são utilizadas para calcular a pose da câmera, um problema conhecido como *Perspective-n-Point* (PnP) (62). Uma vez que a nova pose da câmera é conhecida, novos pontos 3D podem ser criados por meio da triangulação de características que são vistas nesta nova imagem e em outras imagens já adicionadas ao modelo.

Por fim, Otimização Global com BA é a etapa mais crítica para a obtenção de uma reconstrução com incerteza de medição controlada. O BA é um problema de otimização não linear de larga escala que refina simultaneamente todos os parâmetros estimados: as poses de todas as câmeras, os parâmetros intrínsecos da câmera, e as coordenadas de todos os pontos 3D da cena (63, 64). O objetivo é minimizar o erro de reprojeção total, que é a soma das distâncias euclidianas entre as localizações das características observadas nas imagens e as projeções dos pontos 3D correspondentes nesses mesmos planos de imagem, conforme a pose estimada da câmera. A função de custo E a ser minimizada é dada pela equação 20:

$$E = \sum_{i=1}^n \sum_{j=1}^m w_{ij} \|x_{ij} - \pi(C_i P_i)\|^2 \quad (20)$$

Onde n é o número de câmeras, m é o número de pontos 3D, w_{ij} é um indicador binário (1 se o ponto j é visto pela câmera i , 0 caso contrário), x_{ij} é a coordenada 2D observada do ponto j na imagem i , e $\pi(C_i, P_j)$ é a função de projeção que mapeia o ponto 3D P_j para a imagem, dados os parâmetros da câmera C_i . A Figura 16 ilustra graficamente os componentes centrais desta função de custo: o termo x_{ij} corresponde à "Observação" (a característica medida), enquanto $\pi(C_i, P_j)$ corresponde à "Reprojeção do Modelo". O "Erro de Reprojeção" é a distância vetorial entre esses dois pontos, e o algoritmo de BA ajusta iterativamente a pose da câmera (C_i) e a posição do ponto 3D (P_j) para minimizar essa distância para todas as observações.

Figura 16 – Ilustração do erro de reprojeção



Fonte: Do autor.

O resultado direto do *pipeline* de SfM é uma nuvem de pontos esparsa, consistindo apenas nos locais das características utilizadas para a correspondência, juntamente com as poses de câmera otimizadas.

2.5.1.1 Escala e Alinhamento

É fundamental reconhecer duas propriedades intrínsecas da reconstrução gerada por um sistema de SfM que utiliza uma única câmera (monocular).

Primeiramente, a reconstrução possui uma escala ambígua. Isso ocorre porque, a partir de imagens 2D, é possível inferir a forma da cena, mas não seu tamanho absoluto. Um objeto pequeno visto de perto e um objeto grande visto de longe podem gerar imagens idênticas. Portanto, a geometria da cena é recuperada de forma proporcionalmente correta, mas seu tamanho absoluto permanece arbitrário, pois não há uma referência métrica nas imagens. Em segundo lugar, o modelo 3D e as trajetórias da câmera existem em um sistema de coordenadas arbitrário, tipicamente definido pelo par de imagens que inicializou o processo de reconstrução (22).

Para resolver a ambiguidade de escala e conferir significado métrico ao modelo, é necessário introduzir pelo menos uma medida de distância do mundo real. Isso pode ser alcançado de várias formas, como medir a distância entre dois pontos identificáveis na cena, inserir um objeto de tamanho conhecido ou, de maneira mais confiável, utilizar Pontos de Controle Terrestres (*Ground Control Points* - GCPs) com coordenadas conhecidas, obtidas por GPS ou topografia(68). Uma vez que a distância real entre dois pontos na nuvem de pontos é conhecida, o fator de escala de toda a reconstrução pode ser calculado e aplicado uniformemente.

Após a determinação da escala, a reconstrução precisa ser alinhada a um sistema de coordenadas de referência externo (por exemplo, o sistema de coordenadas de um sensor de *ground-truth* ou um sistema geodésico). Esse processo, conhecido como registro, envolve encontrar a transformação geométrica ótima que sobrepõe o modelo do SfM ao sistema de referência. Para reconstruções monoculares, essa transformação é uma transformação de similaridade ($\text{Sim}(3)$). Ela possui 7 graus de liberdade e pode ser decomposta em um fator de escala (s), uma componente de rotação (R , uma matriz 3×3) e uma componente de translação (t , um vetor 3×1). Existem duas abordagens principais para estimar essa transformação:

Métodos baseados em correspondências explícitas: Quando se dispõe de um pequeno conjunto de pontos de controle (tipicamente 3 ou mais) cujas coordenadas são conhecidas em ambos os sistemas de coordenadas, é possível calcular a transformação de forma direta. Um método clássico para resolver este problema é o algoritmo de Horn (75), que fornece uma solução de forma fechada (*closed-form solution*) para a orientação absoluta utilizando quaternions, encontrando a rotação, translação e escala ótimas que minimizam o erro quadrático médio entre os pontos correspondentes.

Métodos baseados em nuvens de pontos: Quando não há correspondências explícitas, mas as duas nuvens de pontos (a do SfM e a de referência) possuem uma sobreposição significativa, algoritmos iterativos podem ser utilizados. O mais conhecido é o *Iterative Closest Point* (ICP) (76). O ICP funciona em um ciclo de duas etapas: primeiro, associa cada ponto da nuvem de origem ao ponto mais próximo na nuvem de destino; segundo, calcula a transformação que minimiza a distância entre esses pares associados. Esse processo é repetido até que a melhoria entre as iterações caia abaixo de um limiar ou um número máximo de iterações seja atingido.

O alinhamento preciso da trajetória e do mapa reconstruído é um pré-requisito para a avaliação quantitativa do desempenho de sistemas de SLAM e SfM, permitindo o cálculo de métricas de erro padrão como o Erro Absoluto de Trajetória (*Absolute Pose Error* - APE) e o Erro Relativo de Pose (*Relative Pose Error* - RPE) (77).

2.5.2 Ferramentas de Referência

O processo de *Structure from Motion* (SfM), descrito na seção anterior, culmina na geração de uma nuvem de pontos esparsa. Embora essa representação seja geometricamente precisa e fundamental para estimar as poses da câmera, ela possui uma limitação inerente: contém apenas as coordenadas 3D de um conjunto reduzido de pontos de interesse, não representando as superfícies contínuas dos objetos na cena. Para obter uma reconstrução 3D fotorrealista e detalhada, é necessário um passo subsequente de densificação.

Esse processo é realizado por algoritmos de Múltipla Vista Estéreo (*Multi-View Stereo* - MVS). O MVS utiliza as poses de câmera e a calibração, estimadas com baixa incerteza de medição pelo SfM, como entrada. A partir dessa informação, os algoritmos de MVS buscam estabelecer correspondências densas de *pixels* entre múltiplas imagens para estimar a informação de profundidade para uma grande parte da cena. O resultado de um pipeline de MVS é tipicamente uma nuvem de pontos densa, que pode posteriormente ser convertida em uma malha 3D texturizada, fornecendo uma representação completa e visualmente rica da geometria da cena (78, 79).

Diversas soluções de software, tanto de código aberto quanto comerciais, implementam os processos completos de SfM e MVS. Algumas das ferramentas mais proeminentes e utilizadas como referência na literatura são:

COLMAP: É uma solução de *pipeline* de SfM e MVS de código aberto considerada na comunidade acadêmica por sua flexibilidade e, principalmente, pela baixa incerteza de medição de seus resultados. Desenvolvido por Schönberger e Frahm, o COLMAP oferece tanto uma interface gráfica (GUI) para visualização e depuração, quanto uma interface de linha de comando (CLI) que permite a automação completa do processo de reconstrução, sendo ideal para protocolos de avaliação experimental (80).

Agisoft Metashape: Anteriormente conhecido como PhotoScan, é uma das principais ferramentas comerciais de fotogrametria do mercado. É amplamente adotado em aplicações profissionais de topografia, engenharia, arqueologia e inspeção industrial. Seus pontos fortes residem em um fluxo de trabalho altamente otimizado e uma interface de usuário intuitiva, que guia o operador por todas as etapas do processo, desde o alinhamento das fotos até a geração de produtos finais como modelos 3D texturizados, ortomosaicos e modelos digitais de superfície (81).

Outras Ferramentas de Código Aberto: Além do COLMAP, o ecossistema de software livre para fotogrametria é vasto. Ferramentas como o Meshroom, que é um *pipeline* completo baseado no *framework* AliceVision e notável por sua interface de usuário baseada em nós, e o OpenMVG (para a parte de SfM) combinado com o OpenMVS (para a densificação), representam alternativas poderosas e modulares que também são amplamente utilizadas na pesquisa e por entusiastas da área.

2.6 ROS 2 (*Robot Operating System 2*)

A implementação de sistemas robóticos complexos, como o de odometria visual-inercial e reconstrução 3D proposto nesta dissertação, exige uma infraestrutura de *software* que orchestre a comunicação entre múltiplos sensores, algoritmos e processos de forma eficiente e confiável. O *Robot Operating System 2* (ROS 2), sucessor do *framework* ROS original (82), surge como uma solução padrão na comunidade acadêmica e industrial para este desafio, fornecendo um conjunto de bibliotecas, ferramentas e convenções para o desenvolvimento de software robótico

modular (83). Este capítulo apresenta os conceitos fundamentais do ROS 2, sua arquitetura de comunicação e justifica sua adoção como o *middleware* central deste trabalho.

2.6.1 Arquitetura e Conceitos Fundamentais

O paradigma de comunicação do ROS 2 é fundamentado no conceito de um grafo computacional distribuído. Neste modelo, o sistema robótico é decomposto em uma rede de processos independentes que trocam informações entre si. Essa abordagem visa abstrair a complexidade do *hardware* e promover a modularidade, a reutilização de código e a tolerância a falhas, simplificando a tarefa de desenvolver comportamentos robóticos complexos (84). A estrutura deste grafo é composta por três entidades conceituais primárias: os Nós (*Nodes*), os Tópicos (*Topics*) e as Mensagens (*Messages*).

Os nós constituem as unidades computacionais fundamentais dentro da arquitetura do ROS 2. Cada nó é um processo de *software* executável e independente, projetado para realizar uma tarefa singular e bem definida. Por exemplo, um nó pode ser responsável exclusivamente pela interface com um sensor (como uma câmera ou uma IMU), outro pela execução de um algoritmo de planejamento de trajetória, e um terceiro pelo controle de atuadores. No contexto deste trabalho, a arquitetura do sistema é decomposta em múltiplos nós: um nó para a captura e publicação das imagens do par estéreo; um segundo nó dedicado à leitura e publicação dos dados da Unidade Inercial; e o nó principal, onde o algoritmo que é executado. Esta decomposição em processos modulares é uma prática que aumenta a clareza do sistema e facilita a depuração e o desenvolvimento iterativo.

O intercâmbio de mensagens entre os nós é predominantemente realizado por meio de Tópicos, que funcionam como barramentos nomeados que implementam uma semântica de publicação-assinatura (*publish-subscribe*) anônima. Esta semântica desvincula a produção da informação de seu consumo: um nó Publicador (*Publisher*) envia mensagens para um tópico sem ter conhecimento de quais ou quantos nós estão recebendo-as. Similarmente, um nó Assinante (*Subscriber*) recebe mensagens de um tópico sem precisar saber qual nó as originou. Essa característica de baixo acoplamento é fundamental para a escalabilidade e flexibilidade de sistemas

distribuídos. No sistema proposto, por exemplo, o nó da câmera publica continuamente mensagens do tipo imagem no tópico `/stereo/images`, enquanto o nó do SLAM, como assinante, consome essas mensagens para processamento.

As Mensagens são as estruturas de dados, com tipos rigorosamente definidos, que são transportadas através dos tópicos. Cada tópico está associado a um único tipo de mensagem, que funciona como um "contrato" de dados para todos os publicadores e assinantes daquele tópico. O ROS 2 provê um conjunto de tipos de mensagens padrão para dados comuns em robótica, mas também permite a definição de tipos customizados. Para garantir a interoperabilidade, este projeto utiliza tipos de mensagem consolidados, como `sensor_msgs/Image` e `sensor_msgs/Imu`, que são estruturas padronizadas para encapsular dados de imagem e informações inerciais, respectivamente. A utilização de tipos de mensagem bem definidos é crucial para a correta integração entre os diferentes módulos de software do sistema, permitindo que nós desenvolvidos em linguagens de programação distintas (como C++ e Python) comuniquem-se de forma transparente.

2.6.2 Comunicação e Aplicação no Projeto

A fundação da camada de comunicação do ROS 2 repousa sobre uma base de nível industrial, projetada para sistemas de missão crítica. O *framework* delega todo o transporte e gerenciamento de dados ao *Data Distribution Service* (DDS), um padrão aberto de *middleware* mantido pelo *Object Management Group* (OMG). A adoção do DDS como espinha dorsal da comunicação representa a mudança arquitetônica mais significativa em relação ao ROS 1. Essa escolha substitui a topologia de rede centralizada, que dependia de um nó mestre (*roscore*), por um sistema inerentemente distribuído e resiliente. Na prática, isso elimina o ponto único de falha: os nós descobrem-se e estabelecem comunicação de forma automática e descentralizada na rede (*peer-to-peer*), conferindo ao sistema uma maior confiabilidade e escalabilidade (83, 84).

O recurso mais poderoso que o DDS introduz no ecossistema ROS é o controle refinado sobre a comunicação através das políticas de Qualidade de Serviço (QoS). Essencial em sistemas robóticos que operam com múltiplos fluxos de dados com diferentes requisitos temporais e de confiabilidade, o QoS permite que o

desenvolvedor especifique o comportamento desejado para cada publicador e assinante. É possível, por exemplo, configurar a entrega de dados críticos como a estimativa de pose do veículo ou os parâmetros de calibração para ser confiável (*Reliable*). Esta política estabelece um protocolo de confirmação de recebimento para garantir que nenhuma informação seja perdida durante a transmissão, ainda que ao custo de uma maior latência. Em contrapartida, fluxos de dados de alta vazão, como as imagens brutas das câmeras, podem ser configurados para melhor-esforço (*Best Effort*), onde a prioridade é necessário velocidade e a entrega da mensagem mais recente, aceitando a perda eventual de pacotes para manter a baixa latência. Esse nível de configurabilidade é crucial para otimizar o desempenho de redes com diferentes requisitos de comunicação simultâneos.

A adoção do ROS 2 neste trabalho justifica-se, portanto, por sua capacidade de orquestrar de forma desacoplada e modular o complexo fluxo de dados entre os componentes de *hardware* (sensores) e *software* (algoritmos). Essa estrutura não apenas simplifica a integração e a depuração do sistema de odometria visual-inercial, mas também provê a infraestrutura de comunicação de alto desempenho e configurável que é necessária para aplicações de percepção robótica em tempo real.

3 MODELO PROPOSTO

Com base nos fundamentos estabelecidos anteriormente, este capítulo detalha o modelo proposto para o sistema de navegação e mapeamento 3D. A proposta consiste na fusão de dados de um sistema de câmeras estéreo e uma unidade de medição inercial (IMU) utilizando o ORB-SLAM 3 comunicando-se com o ROS2. A seguir, são apresentadas a configuração conceitual do sistema, a lógica da fusão de dados e o plano metodológico para sua validação

3.1 CONFIGURAÇÃO DO SISTEMA DE NAVEGAÇÃO VISUAL-INERCIAL

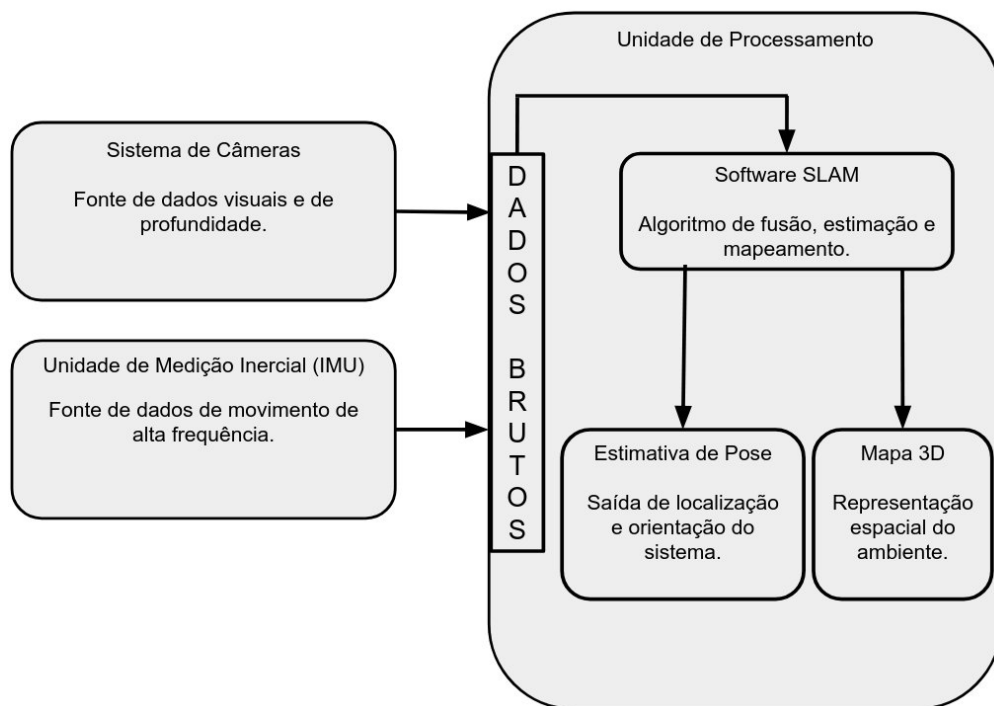
O sistema de navegação proposto neste trabalho foi deliberadamente projetado como uma plataforma estéreo-inercial versátil, com o objetivo de permitir a implementação e avaliação tanto de algoritmos de SLAM puramente estéreo quanto estéreo-inerciais. Para isso, o sistema integra uma unidade de medição inercial (IMU) a um par de câmeras, visando uma fusão de dados para estimação de trajetória em 6 graus de liberdade (6-DoF) e reconstrução tridimensional.

A concepção do aparato físico foi otimizada para uma faixa de operação primária entre 3 e 5 metros. Considerando esta distância, optou-se por uma disposição de câmeras em arranjo paralelo. Esta configuração simplifica o modelo geométrico de projeção e, para a faixa de operação definida, garantindo que ambos os sensores capturem uma visão sobreposta da cena. A disposição física dos componentes, destacando o alinhamento paralelo das câmeras e o posicionamento da IMU, é apresentada na Figura 18.

A escolha pela arquitetura estéreo-inercial é fundamental para os objetivos deste trabalho. A combinação dos sensores resulta em um sistema intrinsecamente complementar: a visão estéreo ancora a estimativa da IMU, pois a linha de base (*baseline*) pré-calibrada estabelece a escala métrica absoluta do ambiente e corrige seu acúmulo de erro (*drift*) a longo prazo. Em contrapartida, a IMU provê a continuidade da navegação em alta frequência, superando as fragilidades do sistema visual em situações de movimentos rápidos, oclusões ou ambientes com baixa textura.

A arquitetura lógica do sistema e o fluxo de dados são ilustrados na Figura 17. Os dados dos sensores (pares de imagens e medições inerciais) alimentam a Unidade de Processamento. Nessa unidade, o *Software* SLAM processa essas informações para gerar as saídas primárias do sistema: a estimativa da trajetória em tempo real (Pose 6-DoF) e o mapa tridimensional do ambiente.

Figura 17 – Arquitetura lógica do sistema proposto



Fonte: Do autor.

Para que este sistema funcione é necessário o equacionamento matemático cada um de seus componentes e a relação entre eles. As subseções seguintes detalham o modelo conceitual adotado para os sensores e os requisitos para sua calibração e alinhamento.

3.1.1 Disposição e Calibração das Câmeras

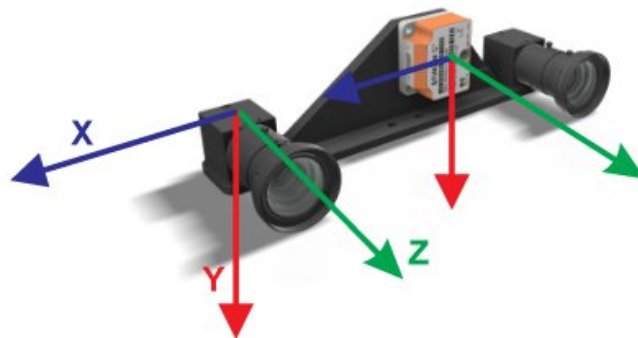
Para o módulo de visão estéreo do sistema proposto, adota-se o modelo de câmera *pinhole* com correção para distorções radiais e tangenciais, cujos fundamentos teóricos foram apresentados na seção 2.1.1. A escolha deste modelo é justificada por sua vasta aceitação na comunidade de visão computacional, seu bom

balanço entre simplicidade e baixa incerteza de medição, e a disponibilidade de métodos de calibração bem estabelecidos.

Para que este modelo represente o sensor físico, é preciso determinar os parâmetros intrínsecos (K , D) e os parâmetros extrínsecos (R , t) do par estéreo, conforme detalhado na Seção 2.1.1.

O processo de calibração de câmera é, conceitualmente, o método utilizado para estimar numericamente todos esses parâmetros (K, D, R, t). A determinação precisa desses valores é um pré-requisito do modelo proposto, pois habilita a retificação das imagens e a triangulação de pontos 3D, permitindo assim a percepção de profundidade. A Figura 18 ilustra a geometria do modelo estéreo adotado, destacando a disposição paralela das câmeras, que alinha os eixos ópticos e simplifica o processo de correspondência de características entre as imagens.

Figura 18 – Referência espacial do sistema estéreo proposto



Fonte: Do autor.

Portanto, o processo de calibração fornece ao modelo os parâmetros estáticos necessários para converter as medições de *pixels* em informação geométrica tridimensional. De maneira análoga, o segundo sensor do sistema, a Unidade de Medição Inercial, também requer a adoção de um modelo matemático para seus erros e um processo de calibração para estimar seus parâmetros dinâmicos. Este componente é detalhado na subseção a seguir.

3.1.2 Integração e Calibração da IMU

A verdadeira força do sistema proposto não reside nos sensores individuais, mas na sinergia resultante de sua fusão. Para alcançar essa sinergia, a primeira etapa conceitual é a integração da Unidade de Medição Inercial (IMU) com o sistema de câmeras. Em essência, a câmera e a IMU observam o mesmo fenômeno, o movimento do aparato no espaço, mas o descrevem em "*linguagens*" fundamentalmente diferentes e a partir de perspectivas distintas. A câmera captura marcos visuais em seu próprio plano de imagem, enquanto a IMU mede forças e velocidades angulares em seu próprio centro de medição. A integração, portanto, é o processo de estabelecer um dicionário matemático que permita traduzir as medições de um sensor para o sistema de coordenadas do outro, criando um único e coerente quadro de referência.

Este é formalizado no modelo como uma transformação rígida espacial constante, que descreve a orientação (rotação) e a posição (translação) da IMU em relação à câmera. Matematicamente, esta relação é representada por uma matriz de transformação homogênea, T_{camimu} . O processo para determinar os valores numéricos desta matriz é conhecido como calibração extrínseca Câmera-IMU. A precisão desta calibração é um fator dominante no desempenho de todo o sistema visual-inercial, pois ela garante a consistência geométrica entre as duas fontes de dados. Um erro nesta transformação faria com que o algoritmo de fusão tentasse reconciliar duas narrativas de movimento conflitantes, resultando em um acúmulo de erros e na potencial divergência da trajetória estimada.

3.1.3 Correção de Erros da IMU e Alinhamento com as Câmeras

A determinação da transformação extrínseca, reconhecida como uma etapa crítica do sistema, motiva uma abordagem que se estende além da otimização puramente algorítmica. Desta forma, o modelo proposto neste trabalho estabelece como premissa de projeto a configuração física do aparato, tratando o alinhamento mecânico entre os sensores como um componente integral do modelo. O objetivo desta premissa é restringir o espaço de busca dos parâmetros de calibração e, conseqüentemente, aumentar a estabilidade operacional e a confiabilidade do processo de estimação.

A abordagem de alinhamento consiste na montagem da Unidade de Medição Inercial (IMU) de tal forma que seus eixos ortogonais sejam manufaturados para serem aproximadamente paralelos aos eixos correspondentes do sistema de coordenadas da câmera de referência. Essa configuração de projeto, que busca a colinearidade dos eixos (X, Y, Z) entre os dois sensores, é ilustrada conceitualmente na Figura 18.

As implicações desta decisão de projeto são tanto matemáticas quanto práticas. Matematicamente, o alinhamento físico impõe uma forte restrição a priori de que a componente de rotação da matriz de transformação, R_{camimu} , deve ser próxima da matriz identidade (I). Isso melhora o condicionamento do problema de otimização e fornece um ponto de partida excelente para os algoritmos de calibração, reduzindo o risco de convergência para mínimos locais incorretos. Na prática, esta premissa serve como um método de verificação: um resultado de calibração que apresente uma rotação significativa pode ser interpretado como um indicador de inconsistências no procedimento experimental, ao invés de um reflexo da realidade física. Desta forma, o alinhamento físico deliberado é uma característica do modelo que visa assegurar uma calibração extrínseca de maior fidelidade métrica, estabilidade operacional e verificabilidade.

3.2 FUSÃO DOS DADOS IMU E ESTÉREO NO ORB-SLAM 3

O sistema adota a arquitetura de processamento e a metodologia de fusão de dados do ORB-SLAM3, detalhada na Seção 2.4. A fusão é realizada de forma fortemente acoplada (*Tightly-Coupled*), o que é um necessário para a estabilidade operacional. Essa abordagem integra as informações da IMU e as observações visuais em um único estimador de estado, garantindo uma estimativa de pose e um mapa tridimensional com reduzida Incerteza de Medição e confiabilidade, em comparação com abordagens de processamento separadas.

O processamento é estruturado em *multi-thread*: a *thread* de rastreamento utiliza a pré-integração da IMU para estimar o movimento inicial, auxiliando a busca visual em tempo real. Em paralelo, a *thread* de mapeamento local refina a geometria do mapa, e a *thread* de fechamento de ciclos corrige erros de deriva acumulados (*drift*) ao longo de grandes trajetórias. A colaboração destas *threads* sobre a estrutura de

dados central (Atlas) viabiliza que o sistema lide com cenários complexos (como perda de rastreamento) e gere um mapa globalmente consistente em 6 graus de liberdade.

3.3 TRAJETÓRIA E CORPOS DE PROVA

Para a validação abrangente do modelo de navegação e mapeamento proposto, é fundamental definir uma estratégia de teste que avalie o sistema em um espectro variado de condições operacionais. Esta estratégia foi desenhada não apenas para quantificar o desempenho em cenários ideais, mas também para verificar a sua estabilidade operacional em condições desafiadoras que simulam aplicações do mundo real, garantindo que o modelo seja aplicável a uma diversidade de ambientes.

As trajetórias propostas para a avaliação do sistema foram planejadas para explorar as capacidades da fusão de dados em diferentes regimes de movimento. Para a avaliação do desempenho de base, foram consideradas trajetórias de percurso aberto e sem fechamento de ciclo, que permitem analisar o acúmulo de erro de deriva (*drift*) em cenários de navegação linear. A principal classe de trajetórias, no entanto, foi desenhada para incluir fechamento de ciclo (*loop closure*), o que permite testar a capacidade do algoritmo em reconhecer locais previamente visitados e aplicar correções globais, um recurso crucial do SLAM que garante a consistência do mapa a longo prazo.

A avaliação do modelo demanda que estas trajetórias sejam executadas em múltiplos cenários ambientais para testar a resiliência do sistema de forma sistemática: (i) cenários controlados, caracterizados por iluminação artificial estável e abundância de características visuais, para estabelecer uma linha de base do desempenho ótimo do sistema; (ii) cenários com baixa textura visual, para avaliar a contribuição e a estabilidade do componente inercial quando o rastreamento visual é propositalmente degradado; e (iii) cenários não controlados, sujeitos a variações de iluminação natural e com menor densidade de pontos de interesse, para testar a estabilidade operacional geral do modelo em condições realistas.

No que tange aos alvos de mapeamento, o modelo propõe a utilização de corpos de prova com características distintas para validar diferentes aspectos da reconstrução 3D. A primeira categoria de alvo consiste em objetos com geometria

métrica conhecida, análogos a estruturas industriais. O propósito destes alvos é permitir uma avaliação quantitativa da incerteza de medição e da escala do mapa gerado. A segunda categoria abrange estruturas de larga escala, como ambientes expandidos e objetos de grandes dimensões. O mapeamento destes alvos visa demonstrar a capacidade do modelo em lidar com a reconstrução de cenas maiores e avaliar seu desempenho em cenários de navegação típicos, validando a escalabilidade e a consistência global do mapa.

3.4 CAPTURA DA TRAJETORIA E MAPEAMENTO

A fase de captura da trajetória e mapeamento é onde o modelo teórico se traduz em um processo operacional. Esta etapa lida com as considerações de aquisição de dados e a geração dos artefatos finais para análise, levando em conta as características do algoritmo ORB-SLAM 3 e os objetivos de avaliação do sistema.

Uma consideração do modelo de captura é o tratamento dos dados de imagem. O ORB-SLAM 3 opera internamente sobre imagens em tons de cinza para a detecção de características, realizando a conversão automaticamente caso as imagens de entrada sejam coloridas. Além disso, o modelo deve considerar o compromisso inerente entre a resolução da imagem e o desempenho computacional. Embora o sistema proposto seja capaz de operar com imagens de alta resolução para capturar um maior nível de detalhe da cena, o algoritmo reconhece que tal configuração impõe uma carga computacional significativa que pode comprometer a operação em tempo real. Portanto, a seleção da resolução de captura é proposta como um parâmetro de projeto a ser balanceado, visando otimizar a densidade de características visuais sem exceder a capacidade de processamento do sistema.

Ao término da execução de uma trajetória, o modelo prevê a persistência dos dados gerados para a subsequente etapa de avaliação quantitativa. O processo de salvamento foi projetado para exportar a trajetória estimada em formatos compatíveis com as ferramentas de *benchmark* padrão da comunidade de SLAM. Da mesma forma, o mapa tridimensional, na forma de uma nuvem de pontos, é salvo em um formato que permite a análise geométrica e integridade estrutural. Esta etapa de exportação é um componente integral do modelo, pois garante que as saídas teóricas do sistema (pose e mapa) se materializem como dados concretos e analisáveis.

4 MATERIAS E MÉTODOS

Este capítulo detalha o aparato experimental e os procedimentos metodológicos empregados para o desenvolvimento e a avaliação do sistema de navegação visual-inercial. Serão apresentados o hardware e o software que compõem a plataforma, bem como a configuração física do sistema. Adicionalmente, são descritos os procedimentos de calibração, a metodologia de coleta de dados e as métricas de desempenho adotadas para validar a precisão da trajetória e a qualidade do mapeamento.

4.1 APARELHO EXPERIMENTAL (*Hardware Setup*)

A incerteza de medição e a estabilidade operacional do sistema de navegação visual-inercial dependem, em grande parte, da qualidade e da confiabilidade dos dados de entrada dos sensores. Por esta razão, a escolha e a configuração do aparato experimental foram passos críticos no desenvolvimento deste trabalho. O sistema foi concebido a partir de um conjunto de sensores de grau industrial e de navegação, selecionados para fornecer informações de alta fidelidade: um par de câmeras estéreo para a percepção visual do ambiente e uma Unidade de Medição Inercial (IMU) para o rastreamento do movimento. A disposição física desses componentes em uma plataforma de montagem rígida é igualmente fundamental, garantindo uma relação espacial fixa entre os sensores e, conseqüentemente, a baixa incerteza de medição das calibrações extrínsecas. Nas subseções seguintes, serão detalhados cada um desses componentes utilizados ao desenvolver esse projeto.

4.1.1 Câmera Estéreo: Modelo e Especificações

A captura de informações visuais, base do sistema de SLAM, foi realizada por meio de um sistema de câmeras estéreo. Ao longo do desenvolvimento do projeto, duas configurações distintas deste sistema foram implementadas e avaliadas. A evolução do primeiro para o segundo sistema foi motivada pela necessidade de aumentar a densidade e a qualidade dos dados para aplicações secundárias, como a fotogrametria, que se beneficiam de maior resolução e informação de cor.

A primeira versão do sistema estéreo, utilizada nas fases iniciais de implementação e validação do ORB-SLAM 3, era composta por um par de câmeras monocromáticas *Point Gray*, modelo FL3-U3-20E4M-C. Estas câmeras, com resolução de 2.0 megapixels (1600x1200) e sensor de obturador global, foram fundamentais para testar o algoritmo e para otimizar a configuração física do aparato. Nesta fase, foram realizados experimentos com diferentes linhas de base (*baseline*) entre as câmeras, especificamente estabilidade na estimação de profundidade (Figura 19 versão 1 e 2), sendo adotada para as versões subsequentes. Os detalhes técnicos completos desta primeira configuração de câmera e da lente Kowa LM3NCM associada estão disponíveis no Anexo A.

Com o avanço do projeto e a intenção de utilizar os dados gerados para reconstrução 3D por meio de fotogrametria, foram identificadas limitações na configuração inicial. A resolução de 1600x1200 *pixels* dificulta a captura dos detalhes finos das cenas, dificultando a localização de pontos de correspondência e o alinhamento preciso das imagens pelos *softwares* de fotogrametria. Adicionalmente, a ausência de informação de cor nos sensores monocromáticos representava outra barreira, uma vez que a cor é um atributo valioso para a segmentação e o reconhecimento de características em nuvens de pontos densas.

Para superar essas limitações, procedeu-se com a evolução do sistema para uma segunda e definitiva versão. Esta nova configuração utiliza um par de câmeras FLIR Blackfly® S, modelo BFS-U3-50S4C-C. Este modelo representa um avanço significativo, sendo uma câmera colorida de visão de máquina com resolução de 5.0 megapixels (2448x2048 *pixels*) e capacidade para operar a 72 quadros por segundo (Hz). Equipada com lentes da Edmund 3.5mm/F2.0 modelo 89410, a câmera mantém a característica indispensável do obturador global (*global shutter*), eliminando distorções de movimento e garantindo a integridade geométrica das imagens. A lente selecionada para esta configuração foi o modelo C-Series de 3.5mm de distância focal (Figura 19 versão 5). A combinação desta câmera de maior resolução e sensor colorido com a lente de amplo campo de visão foi projetada para maximizar a captura de informações texturais e cromáticas, essenciais tanto para a estabilidade do SLAM quanto para a qualidade da reconstrução fotogramétrica. Os detalhes técnicos completos desta segunda configuração de câmera e lente estão disponíveis no Anexo B.

4.1.2 Unidade de medição Inercial (IMU): Modelo e Especificações

O componente inercial do sistema, responsável por fornecer medições de velocidade angular e aceleração linear, é fundamental para a estabilidade do SLAM em situações de movimento rápido ou em ambientes com pouca textura visual. A IMU selecionada para este projeto foi o modelo STIM300 da Safran, um sensor de grau de navegação de alto desempenho. É importante notar que a STIM300 é uma IMU de 6 graus de liberdade, composta estritamente por três eixos de giroscópios e três eixos de acelerômetros, não incluindo um magnetômetro. Diferentemente do sistema de visão, o modelo da IMU foi mantido o mesmo ao longo de todo o desenvolvimento, devido à sua baixa incerteza de medição, baixa deriva (*drift*) e estabilidade de polarização (*bias stability*), características cruciais para a fusão de dados visual-inercial. As especificações técnicas detalhadas do sensor estão disponíveis no Anexo C.

Embora o modelo do sensor tenha permanecido constante, sua integração física no aparato experimental passou por uma importante otimização. Inicialmente, a IMU foi montada diretamente sobre a base horizontal que conecta as duas câmeras, como ilustrado na Figura 19 versão 3. Embora funcional, essa configuração resultava em eixos coordenados desalinhados entre a IMU e o sistema de câmeras estéreo, o que introduzia uma rotação complexa nos parâmetros de calibração extrínseca.

Para simplificar o modelo matemático do sistema e facilitar a convergência dos algoritmos de calibração, a IMU foi reposicionada em um suporte customizado, como mostra a Figura 19 versão 4. O objetivo desta mudança foi alinhar os eixos coordenados da IMU com os do sistema de câmeras, de modo que a matriz de rotação extrínseca entre eles se aproximasse da matriz identidade. Esta orientação alinhada minimiza a complexidade da transformação entre os referenciais dos sensores, contribuindo para uma maior precisão e estabilidade na fusão dos dados. Foi esta configuração final que se utilizou para a coleta de todos os dados experimentais.

4.1.3 Plataforma de Montagem e Configuração Física

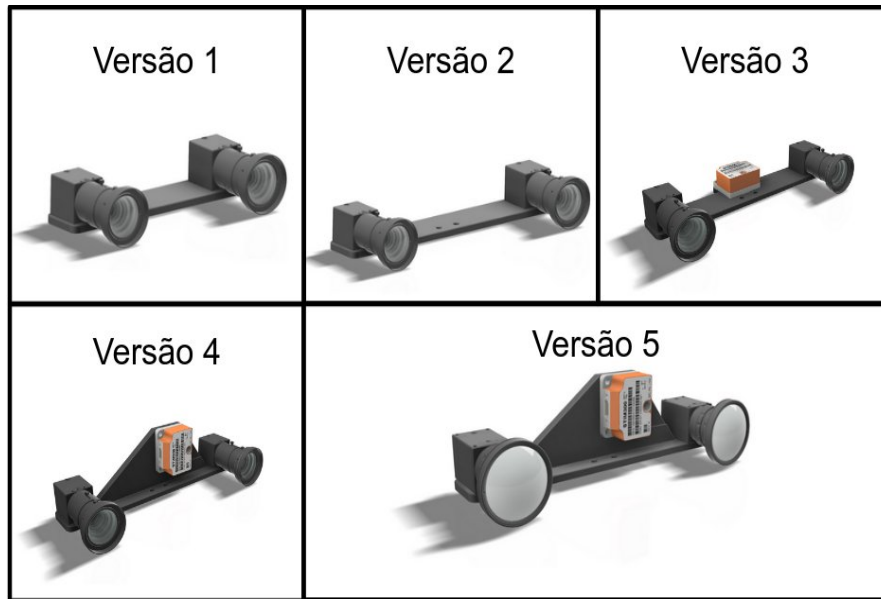
A integração coesa dos sensores em uma única estrutura é um requisito fundamental para a estabilidade de qualquer sistema de navegação visual-inercial. Para este trabalho, foi desenvolvida um suporte, cuja concepção evoluiu em etapas iterativas para otimizar tanto o desempenho do algoritmo quanto a usabilidade durante a coleta de dados. Toda a evolução do aparato experimental está ilustrada na Figura 19.

A primeira iteração da plataforma consistia em um sistema puramente estéreo, com as duas câmeras *Point Gray* (descritas na seção 4.1.1) montadas em paralelo com uma linha de *baseline* de 100 mm. Esta versão inicial, que não possuía a IMU, foi utilizada exclusivamente para os testes preliminares do ORB-SLAM 3 em seu modo estéreo, permitindo validar a funcionalidade básica do sistema de visão.

Numa segunda fase, a IMU STIM300 foi adicionada à estrutura, ainda com a linha de base de 100 mm. Conforme discutido na seção anterior, a montagem inicial da IMU foi realizada de forma desalinhada com os eixos das câmeras. A otimização seguinte focou em corrigir essa questão: a IMU foi reposicionada em um suporte que alinhava suas orientações com as do par estéreo, simplificando a matriz de transformação entre os sensores e, consequentemente, o processo de calibração.

Finalmente, a modificação final e mais impactante foi o aumento da linha de base de 100 mm para 200 mm. Essa alteração foi motivada pela busca de uma maior estabilidade na estimação de profundidade, uma vez que uma base estéreo maior melhora a fidelidade métrica da triangulação de pontos no espaço, especialmente para objetos e superfícies mais distantes. Esta configuração, com *baseline* de 200 mm e a IMU em sua posição otimizada, constitui a versão definitiva do sistema utilizada para a geração dos resultados deste trabalho.

Figura 19 – Evolução do aparato experimental



Fonte: Do autor.

Um aspecto de design ergonômico mantido em todas as versões foi a inclusão de uma manopla na versão 5 apresentada na figura 19 na base inferior da estrutura. Este componente foi projetado para permitir que o sistema fosse segurado e manuseado com apenas uma mão de forma estável, facilitando a coleta de dados com movimentos suaves e controlados. A plataforma foi fabricada utilizando a tecnologia de impressão 3D em uma impressora Bambu Lab X1-Carbon, e o material selecionado foi o polímero PLA (Ácido Polilático), escolhido por sua elevada rigidez para garantir a estabilidade geométrica do aparato.

4.1.4 Unidade de processamento e armazenamento

A execução em tempo real do algoritmo ORB-SLAM 3, simultaneamente à gravação de grandes volumes de dados dos sensores, impõe uma demanda computacional significativa. Para atender a esses requisitos, foi utilizado um hardware de alto desempenho tanto para o processamento quanto para o armazenamento dos dados experimentais.

O processamento central foi realizado em um *notebook* Avell Storm 470, uma estação de trabalho móvel equipada com um processador Intel Core i9-14900HX, 32 GB de memória RAM DDR5 e uma GPU dedicada NVIDIA GeForce RTX 4070. Esta

configuração foi essencial para executar o sistema operacional, o ambiente ROS 2, o algoritmo ORB-SLAM 3, os softwares de interface gráfica e armazenamento das imagens em alta resolução do sistema estéreo e os dados da IMU sem gargalos de processamento, mesmo ao lidar com os fluxos.

Para o armazenamento dos dados brutos, capturados no formato *rosvbag* (MCAP), foi utilizado um SSD externo SanDisk Portable SSD. A escolha por um SSD externo de alto desempenho foi motivada pela necessidade de garantir uma alta taxa de escrita contínua, evitando a perda de pacotes de dados (*dropped frames*) durante a gravação dos experimentos. O dispositivo, que utiliza uma interface USB 3.2 Gen 2, foi conectado a uma das portas USB-C do *notebook* por meio de um cabo USB-C em ambas as extremidades. Essa conexão direta assegura que a máxima velocidade de transferência seja aproveitada, o que é crítico para gravar de forma confiável os arquivos *rosvbag*, que podem atingir dezenas de gigabytes de tamanho em poucos minutos de captura. As especificações técnicas completas tanto do *notebook* de processamento quanto do SSD externo estão detalhadas no Anexo D: Unidade de processamento e armazenamento.

4.2 AMBIENTE DE DESENVOLVIMENTO (*Software*)

Após a detalhada descrição do aparato experimental na seção anterior, esta seção aborda a arquitetura de *software* e as ferramentas computacionais que formam a espinha dorsal do sistema de navegação. A infraestrutura de *software* é responsável por controlar o hardware, adquirir e sincronizar os dados dos sensores, executar o algoritmo de SLAM em tempo real e armazenar os resultados. Serão apresentados o sistema operacional e o *middleware* que compõem a base do sistema, as bibliotecas externas das quais o algoritmo depende e, por fim, a parametrização específica do ORB-SLAM 3 para o *hardware* desenvolvido. A correta configuração e integração desses componentes são determinantes para o desempenho e a precisão do sistema.

4.2.1 Sistema operacional e ROS 2

A escolha do sistema operacional Ubuntu 22.04 LTS (*Jammy Jellyfish*) como base para o projeto foi motivada, primariamente, pelos requisitos do próprio algoritmo

ORB-SLAM 3. Sendo um *software* de código aberto proeminentemente desenvolvido e testado no ecossistema Linux, o Ubuntu é a plataforma para a qual o ORB-SLAM 3 oferece suporte oficial e onde sua compilação e execução são mais estáveis e documentadas. A eficiência do sistema e sua natureza aberta também foram fatores relevantes. Além da compatibilidade com o algoritmo principal, esta escolha está em total alinhamento com o ROS 2 Humble, que também tem o Ubuntu 22.04 LTS como sua plataforma de suporte primária, garantindo uma integração nativa e um ambiente de desenvolvimento estável e com amplo suporte da comunidade de robótica.

Sobre essa base o ROS 2, na sua distribuição Humble, foi empregado como o *middleware* principal do sistema. O ROS 2 atuou como a espinha dorsal de comunicação, gerenciando o fluxo de dados entre os componentes de *hardware* e *software* através de seu sistema de "tópicos" (publicadores e assinantes). Suas funções primordiais no projeto foram:

Aquisição de Dados: Interfacear com os *drivers* dos sensores para receber os fluxos de imagens sincronizadas do par de câmeras estéreo e os dados da IMU.

Distribuição de Dados: Publicar os dados dos sensores em tópicos específicos, que eram assinados pelo nó do ORB-SLAM 3, servindo como a entrada para o algoritmo.

Disseminação dos Resultados: Receber os dados de saída do SLAM, como a pose estimada do sistema, a trajetória percorrida e o mapa de pontos, e publicá-los em outros tópicos para visualização e gravação.

Para a visualização em tempo real e análise dos dados que trafegavam na rede ROS 2, a ferramenta principal utilizada foi o Foxglove Studio. Esta interface gráfica permitiu a introspecção de todos os tópicos ativos, possibilitando a visualização simultânea das imagens das câmeras, dos dados da IMU, da pose estimada e do mapa 3D sendo construído. Além da monitorização em tempo real, o Foxglove Studio foi uma ferramenta importante para a exportação de resultados e para a geração de algumas das análises visuais apresentadas neste trabalho.

4.2.2 Bibliotecas e Dependências

A arquitetura do ORB-SLAM 3 se beneficia de um ecossistema maduro de software de código aberto, apoiando-se em um conjunto de bibliotecas externas

especializadas para realizar tarefas fundamentais. Esta abordagem permite que o algoritmo se concentre em suas contribuições inovadoras para o problema do SLAM, enquanto utiliza soluções confiáveis e altamente otimizadas para operações de base. As dependências mais críticas para o funcionamento do sistema são detalhadas a seguir.

A biblioteca OpenCV (*Open Source Computer Vision Library*) é a espinha dorsal para todas as operações de visão computacional do projeto. Ela é utilizada para o manuseio e pré-processamento das imagens recebidas do sistema estéreo. Sua função mais crucial, no entanto, é na implementação do *frontend* do ORB-SLAM 3, fornecendo as funções para a detecção de cantos FAST e a computação dos descritores binários BRIEF, que, quando combinados e orientados, formam as características ORB (85).

Para todas as operações matemáticas de baixo nível, o sistema utiliza a Eigen, uma biblioteca C++ de alta performance para álgebra linear. A representação de poses da câmera, pontos 3D no mapa, transformações entre sistemas de coordenadas e a formulação dos problemas de otimização são todas dependentes das eficientes estruturas de matrizes e vetores fornecidas pela Eigen (86).

A otimização do mapa e da trajetória, que constitui o backend do SLAM, é gerenciada pela biblioteca g2o (*General Graph Optimization*). Ela é a ferramenta responsável por resolver os complexos problemas de otimização não-linear, como o *Bundle Adjustment*, que refina a estrutura 3D do mapa e as poses da câmera simultaneamente, e a otimização do grafo de poses, essencial para corrigir a trajetória após uma detecção de *loop* (87).

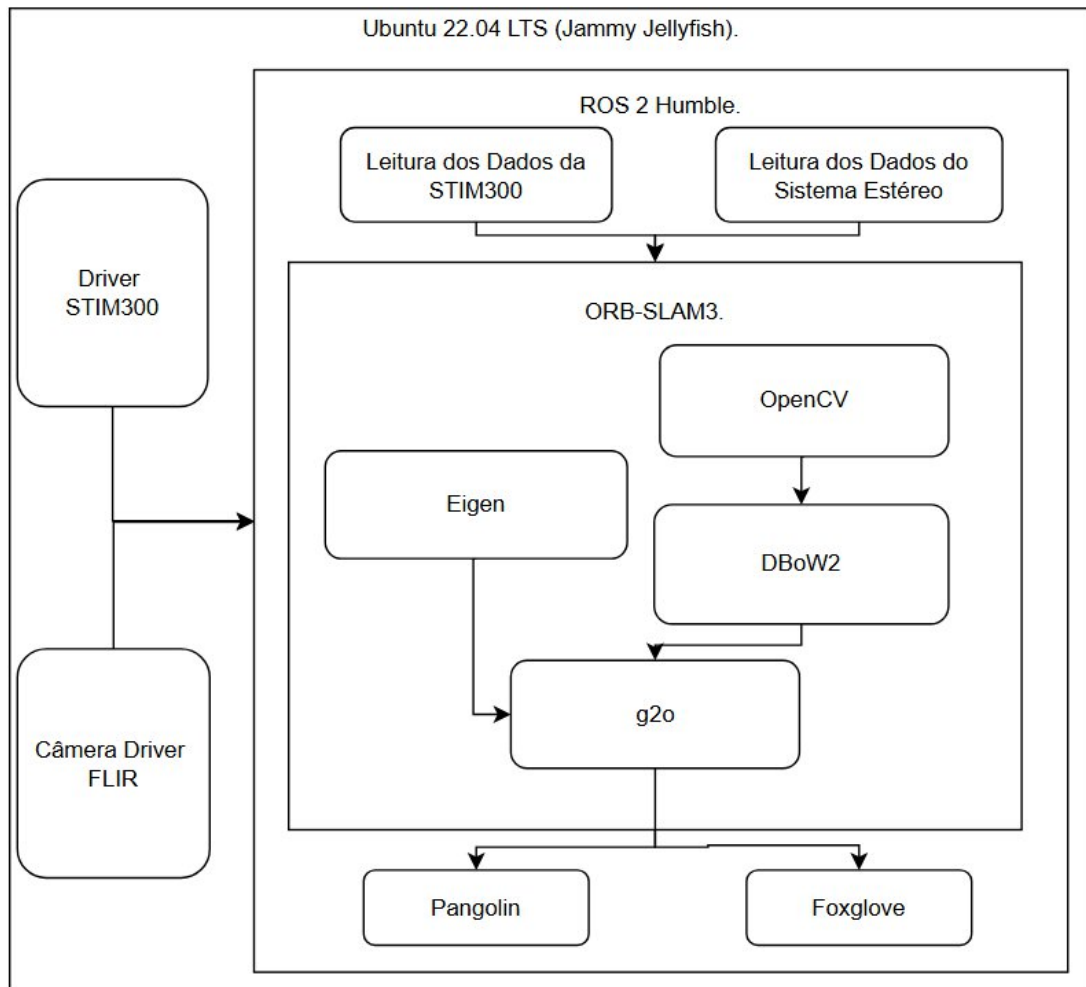
Para a tarefa de reconhecimento de lugar, essencial para a detecção de loop e relocalização, o ORB-SLAM 3 utiliza a biblioteca DBoW2. Esta, implementa um método de *Bag of Words* que converte o conjunto de características visuais de uma imagem em um vetor esparso, permitindo uma comparação extremamente rápida contra um banco de dados de imagens passadas para identificar se o sistema já esteve em um local anteriormente (88).

Finalmente, a visualização 3D nativa e a interface gráfica do ORB-SLAM 3 são implementadas com o auxílio da biblioteca Pangolin. Ela é responsável por criar a janela de visualização em tempo real que renderiza o mapa de pontos esparso, a

trajetória estimada da câmera e a localização dos *Keyframes*, fornecendo um *feedback* visual imediato sobre o estado do sistema (89).

Pode-se visualizar o funcionamento lógico dos softwares descritos anteriormente através do diagrama da figura 20.

Figura 20 – Diagrama do funcionamento dos *softwares*.



Fonte: Do autor.

4.2.3 Implementação e Parametrização do ORB-SLAM 3

A etapa final da configuração do ambiente de software consistiu na compilação e parametrização do algoritmo ORB-SLAM 3. O nó do SLAM foi integrado à arquitetura ROS 2, atuando como um assinante dos tópicos de imagem e IMU e como um publicador da pose, trajetória e mapa estimados, que eram então visualizados em tempo real através do Foxglove Studio, conforme descrito anteriormente. A

customização do comportamento do algoritmo para o *hardware* específico do projeto foi realizada através de um arquivo de configuração no formato YAML, que contém não apenas os parâmetros de calibração dos sensores, mas também os parâmetros operacionais do próprio SLAM.

Dentre os diversos parâmetros ajustáveis, uma atenção especial foi dada às configurações do extrator de características ORB, que é um componente crítico do *frontend* do sistema. O ajuste fino desses parâmetros necessário para otimizar o desempenho do algoritmo para as características das câmeras de alta resolução utilizadas e para os ambientes onde os experimentos foram conduzidos. Para um leitor menos familiarizado, o extrator de características funciona como os "olhos" do algoritmo, buscando pontos de referência únicos em cada imagem para poder se localizar. Os principais parâmetros que controlam esse processo, conforme a documentação do ORB-SLAM 3, são:

ORBextractor.nFeatures: Pense neste parâmetro como a quantidade de "pontos de interesse" que o algoritmo tenta encontrar em cada foto. Esses pontos são cantos ou texturas únicas que servem como "âncoras" para o sistema se localizar. Um número maior de pontos pode tornar o sistema mais confiável (menos propenso a se perder), mas exige mais do processador. O desafio é encontrar a quantidade ideal de pontos de referência, nem poucos a ponto de não ter informação suficiente, nem tantos a ponto de sobrecarregar o sistema.

ORBextractor.nLevels e **ORBextractor.scaleFactor:** Estes dois parâmetros trabalham juntos para permitir que o algoritmo reconheça objetos em diferentes escalas, ou seja, tanto de perto quanto de longe. Imagine que o algoritmo cria várias versões da mesma imagem, como se fossem "níveis de zoom". O **nLevels** define quantos desses níveis de zoom serão criados, e o **scaleFactor** define o "quanto" de zoom é aplicado entre um nível e outro. Isso garante que um ponto de referência seja reconhecido da mesma forma, não importando se a câmera está próxima ou distante dele.

ORBextractor.iniThFAST e **ORBextractor.minThFAST:** Estes funcionam como um "filtro de qualidade" para os pontos de interesse. O **iniThFAST** é o filtro principal e mais exigente; ele define o quão "interessante" ou "nítido" um canto precisa ser para ser considerado um ponto de referência. Se, com essa configuração exigente, o algoritmo não encontrar pontos suficientes (talvez porque a imagem é sem textura),

ele usa o minThFAST, um filtro secundário e menos rigoroso, para garantir que sempre haja um número mínimo de pontos para o sistema não se perder.

O processo de parametrização buscou encontrar uma relação de compromisso ideal entre a estabilidade operacional do rastreamento, a fidelidade métrica da trajetória e a carga computacional. Foi observado que um número excessivamente alto de características (nFeatures) tornava o sistema mais estável à perda de rastreamento, mas ao custo de um maior tempo de processamento e da potencial inclusão de pontos menos estáveis, o que poderia degradar a precisão global da trajetória. Portanto, através de uma série de testes nos ambientes de avaliação, foi buscado um conjunto de valores para estes parâmetros que representasse o melhor balanço entre incerteza de medição e otimização para as condições específicas deste trabalho.

4.3 PROCEDIMENTOS EXPERIMENTAIS

Com o aparato experimental fisicamente montado e o ambiente de software devidamente configurado, esta seção descreve a sequência de procedimentos metodológicos adotados para a calibração, coleta de dados e avaliação do sistema. A execução de um protocolo experimental bem definido é fundamental para garantir a validade e a reprodutibilidade dos resultados. A seguir, são detalhadas as etapas de calibração dos sensores, a descrição dos cenários de teste, a metodologia para a geração dos *datasets* e, por fim, as métricas de desempenho utilizadas para quantificar a precisão do sistema de navegação.

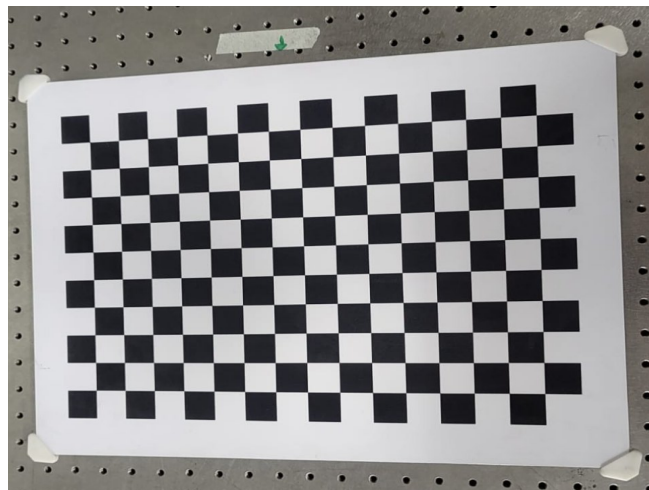
4.3.1 Procedimento de Calibração Intrínseca e Extrínseca do Sistema Estéreo

A calibração do sistema de visão é um pré-requisito indispensável para a reconstrução 3D e a navegação precisa. Este procedimento tem o objetivo duplo de determinar os parâmetros intrínsecos, que modelam as características internas de cada câmera, como a distância focal e as distorções da lente, e os parâmetros extrínsecos, que descrevem a posição e orientação de uma câmera em relação à outra, definindo a geometria do par estéreo.

Para a realização da calibração, foi utilizado o *Camera Calibration Toolbox* do *software* MATLAB. A escolha desta ferramenta se justifica por ser uma solução confiável, amplamente validada na comunidade científica e já empregada com sucesso em outros projetos de visão computacional no laboratório de pesquisa (LABMETRO)(90).

O alvo de calibração utilizado foi um padrão do tipo tabuleiro de xadrez, ilustrado na Figura 21. O padrão, com dimensões de 16x11 quadrados e cada quadrado medindo 25 mm, foi fabricado sobre uma placa de ACM (Material de Alumínio Composto) através de um processo de impressão UV (Ultra Violeta), garantindo uma superfície plana, rígida e com alto contraste. Uma borda branca foi mantida ao redor do padrão para facilitar a detecção dos cantos.

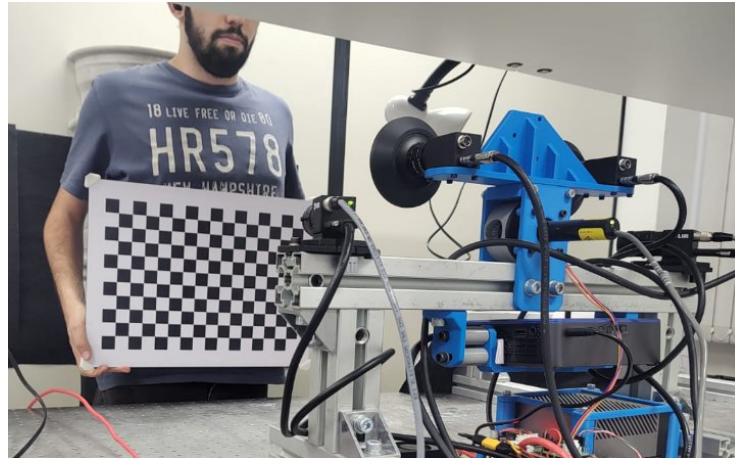
Figura 21 – Padrão de calibração do tipo tabuleiro de xadrez utilizado



Fonte: Do autor.

O procedimento de coleta de dados consistiu na captura de aproximadamente 40 imagens do par estéreo. O padrão foi posicionado em diversas orientações, ângulos e distâncias em relação ao sistema de câmeras, buscando cobrir todo o campo de visão e capturar adequadamente os efeitos de distorção das lentes, especialmente nas bordas da imagem. A Figura 22 ilustra o processo de captura de dados para a calibração. A quantidade de imagens utilizadas segue as recomendações de boas práticas para calibração, que sugerem a captura de um número suficiente de vistas para garantir a convergência e a exatidão dos parâmetros (34).

Figura 22 – Sistema sendo utilizado para coletar imagens do padrão de calibração.



Fonte: Do autor.

A execução da calibração foi realizada utilizando a função de calibração estéreo (*stereo calibration*) do MATLAB. Para o sistema de câmeras mais recente (FLIR Blackfly S), que possui lentes com maior campo de visão, foi utilizado um modelo de alta distorção. Para o sistema mais antigo (Point Gray), um modelo de baixa distorção foi suficiente. Durante o processo, imagens que resultaram em um alto erro de reprojeção foram identificadas como *outliers* e removidas do conjunto de dados para refinar o resultado final. Ao término do processo, os parâmetros intrínsecos e extrínsecos calculados foram transcritos para o arquivo de configuração .yaml do ORB-SLAM 3.

4.3.2 Procedimento de Calibração da IMU e Alinhamento Câmera-IMU

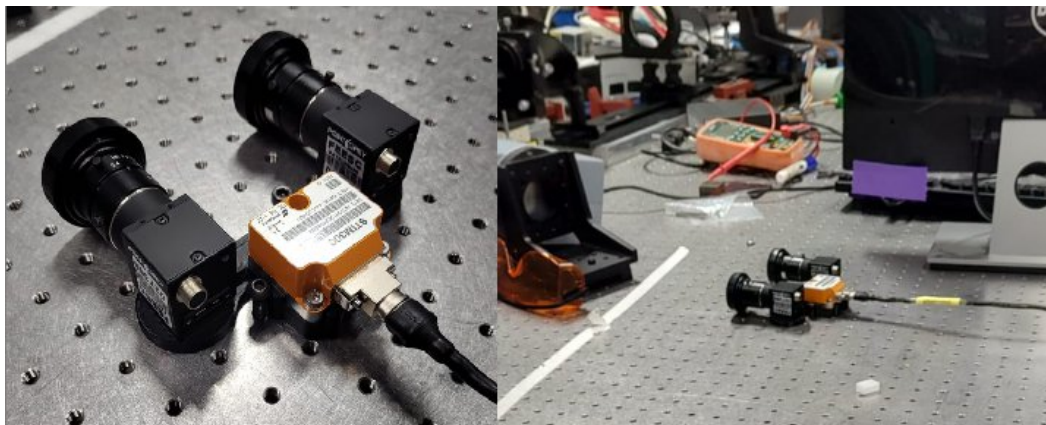
A fusão de dados visual-inercial exige um conhecimento preciso não apenas da geometria do sistema, mas também das características de ruído do sensor inercial. Portanto, o primeiro passo na calibração da IMU foi a determinação de seus parâmetros intrínsecos de ruído: a densidade de ruído (*noise density*) e o passeio aleatório (*random walk*) para o acelerômetro e o giroscópio.

Para esta tarefa, foi empregada a técnica da Variância de Allan (Allan Variance), um método estatístico padrão para a caracterização de ruídos em sensores inerciais. O método analisa a saída do sensor em repouso ao longo de diferentes escalas de tempo para quantificar as fontes de erro estocástico. A ferramenta de software utilizada foi o pacote de código aberto Allan Variance ROS2, uma

implementação que calcula os parâmetros de interesse diretamente a partir de um arquivo rosbag, gerando um arquivo de calibração compatível com outras ferramentas, como o Kalibr (91).

O procedimento experimental para a coleta de dados foi realizado com o máximo de controle para isolar o sensor de quaisquer perturbações externas. O sistema completo foi fixado sobre uma mesa óptica estabilizada (*stabilized optical table*), como mostra a Figura 23. O sistema de estabilização da mesa, que utiliza ar comprimido, foi ativado para anular vibrações do ambiente, garantindo que os dados coletados refletissem apenas o ruído intrínseco da IMU.

Figura 23 – Sistema estéreo-inercial fixado mesa óptica estabilizada.



Fonte: Do autor.

Nestas condições, foi gravado um arquivo rosbag no formato MCAP com os dados da IMU de forma contínua por um período de 10 horas. Para a análise, utilizou-se um segmento de 8 horas de dados, descartando a primeira e a última hora de gravação para assegurar que a análise fosse realizada em um período de máxima estabilidade térmica e operacional do sensor. A execução do pacote Allan Variance ROS2 sobre este *dataset* permitiu o cálculo dos quatro parâmetros de ruído necessários para a configuração do ORB-SLAM 3.

Com os parâmetros intrínsecos da IMU determinados, o passo seguinte foi a calibração extrínseca, que consiste em encontrar a transformação espacial (posição e orientação) entre o sistema de coordenadas da IMU e o da câmera de referência (câmera esquerda). Para esta tarefa, foi utilizada a suíte de ferramentas de código

aberto Kalibr, uma solução de referência na comunidade de robótica para a calibração espaço-temporal de múltiplos sensores (92).

Um requisito do Kalibr é que os arquivos de calibração de entrada (câmera e IMU) sigam um formato .yaml específico. O arquivo de ruídos da IMU, gerado na etapa anterior com a ferramenta Allan Variance ROS2, já estava no formato correto. Contudo, para garantir total compatibilidade e evitar problemas de conversão dos dados de calibração da câmera obtidos no MATLAB, optou-se por recalibrar a câmera esquerda utilizando a própria ferramenta de calibração de câmeras do Kalibr. Este passo intermediário gerou um arquivo *camchain.yaml* no formato nativo esperado pela ferramenta, servindo como uma das entradas para o procedimento final.

O procedimento de coleta de dados para a calibração extrínseca visual-inercial seguiu as recomendações da ferramenta. Um arquivo *rosvbag* foi gravado contendo os fluxos de dados sincronizados da câmera esquerda e da IMU. Durante a gravação, o sistema foi movimentado manualmente em frente a um padrão do tipo *AprilGrid* fixo. Os movimentos foram dinâmicos, incluindo rotações e translações em todos os eixos, de forma a excitar adequadamente a IMU, enquanto se mantinha o padrão sempre visível para a câmera. A Figura 24 ilustra este processo de coleta de dados.

Figura 24 – Coleta de dados para calibração extrínseca Câmera-IMU.



Fonte: Do autor.

A execução da ferramenta *kalibr_calibrate_imu_camera* foi então realizada, utilizando como entrada o *rosvbag* gravado, o arquivo de calibração da câmera (*camchain.yaml*) e o arquivo de ruídos da IMU (*imu.yaml*). O resultado deste processo é um novo arquivo .yaml que contém a matriz de transformação 4x4 que descreve a

pose da câmera em relação à IMU. Por conveniência, a saída do Kalibr já fornece essa transformação no formato requerido pelo ORB-SLAM 3. Desta forma, o arquivo de configuração final do sistema foi montado combinando os parâmetros do sistema estéreo obtidos com o MATLAB, os parâmetros de ruído da IMU obtidos com a Variância de Allan, e a transformação extrínseca câmera-IMU calculada com o Kalibr.

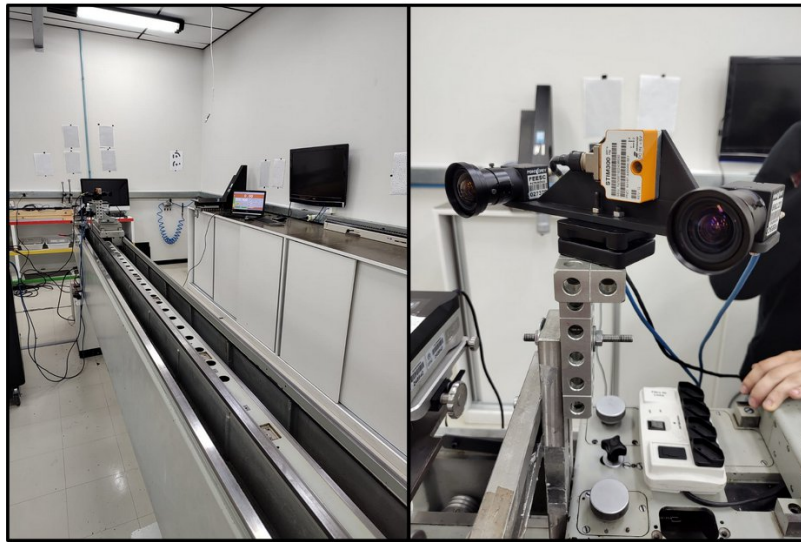
4.3.3 Descrição dos Cenários de Teste e Corpos de Prova

A avaliação do sistema foi dividida em duas frentes complementares: primeiro, a análise do desempenho da trajetória em diferentes ambientes (os cenários de teste); segundo, a avaliação da incerteza de medição métrica do mapa 3D gerado, utilizando objetos com dimensões conhecidas (os corpos de prova). Esta seção detalha os ambientes e objetos utilizados nos experimentos.

Para avaliar a repetibilidade de medição e a estabilidade do sistema na estimação da trajetória, foram definidos três tipos de cenários distintos.

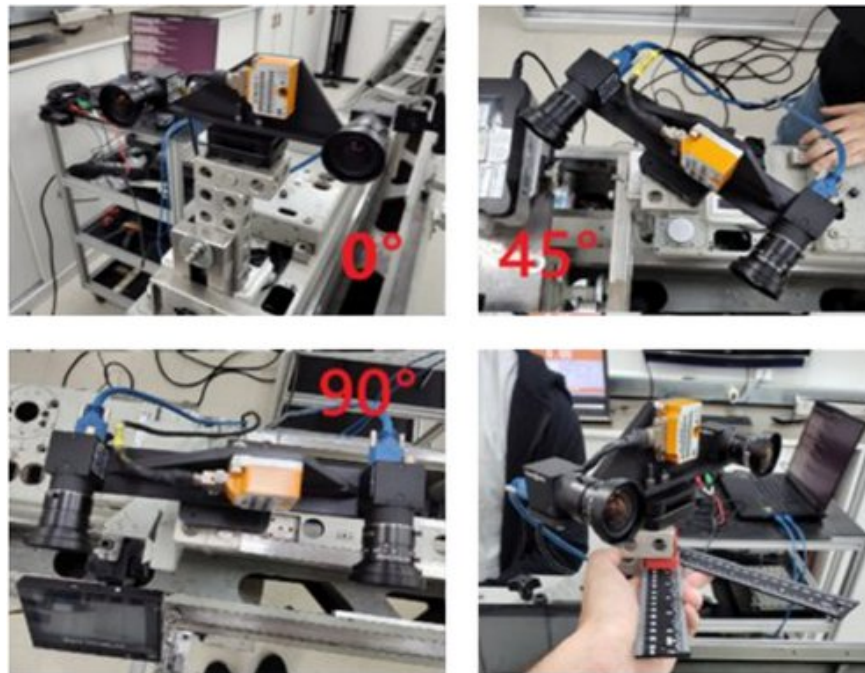
1. Cenário de Validação em Ambiente Controlado: O primeiro conjunto de experimentos foi realizado em uma sala termicamente controlada e sem janelas no LABMETRO, visando uma análise quantitativa de baixa incerteza de medição (Figura 25). Para este teste, o sistema de SLAM foi montado sobre uma guia linear de 3 metros, que também contava com um laser interferométrico para fornecer a medição de deslocamento com reduzida incerteza de medição micrométrica, servindo como *ground truth*. O procedimento experimental consistia em, primeiramente, realizar um pequeno movimento aleatório com o sistema para garantir a inicialização do modo estéreo-inercial. Em seguida, o sistema era posicionado sobre a guia, o laser era zerado, e o conjunto era movido por 3 metros até o final do curso. Após uma pausa de 5 segundos para estabilização, o sistema era retornado à posição inicial, finalizando a coleta de dados daquele ensaio. Este procedimento foi repetido diversas vezes em três orientações distintas na guia (0° , 45° e 90°) para avaliar a repetibilidade de medição do sistema (Figura 26). Para garantir um número suficiente de características visuais, foram afixados nas paredes papéis com pontos pretos que serviram como pontos de interesse para o ORB-SLAM 3.

Figura 25 – Cenário de teste em ambiente controlado no LABMETRO, mostrando o sistema posicionado na guia linear.



Fonte: Do autor.

Figura 26 – Posições do sistema de visão.

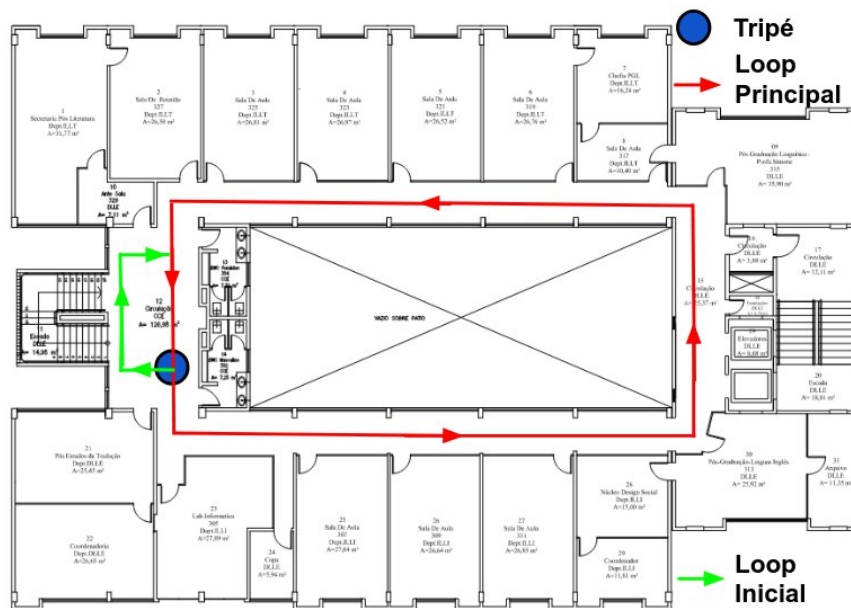


Fonte: Do autor.

2. Cenário de Avaliação em Ambiente Não Controlado com *Loop Closure*: O segundo cenário foi projetado para avaliar o desempenho do sistema em uma trajetória mais longa e em um ambiente menos estruturado, testando a capacidade

de fechamento de *loop* (*loop closure*) e a deriva acumulada. O teste foi conduzido nos corredores do 4º andar do bloco CCE da UFSC (Figura 27), que permitem a realização de uma trajetória em circuito fechado. Um tripé foi utilizado para marcar fisicamente o ponto de inicial e final da trajetória. O procedimento consistia em iniciar a gravação, posicionar o sistema brevemente no tripé, percorrer a pé todo o circuito do corredor e, ao retornar, reposicionar o sistema no mesmo tripé para garantir um fechamento de *loop* geometricamente preciso. Este ambiente apresentava desafios como iluminação mista (natural das janelas e artificial das lâmpadas). Assim como no ambiente controlado, pontos de interesse artificiais foram adicionados às paredes. Para este cenário, a trajetória de *ground truth* foi gerada utilizando técnicas de fotogrametria, que processaram as imagens coletadas para criar um modelo 3D e uma trajetória de referência.

Figura 27 – Ambiente de teste não controlado nos corredores do 4º andar do CCE/UFSC.



Fonte: Do autor.

3. Cenários de Teste Qualitativo; Por último, foram realizadas diversas coletas de dados em ambientes variados e aleatórios no campus da UFSC (Figura 28). O objetivo destes testes não era uma avaliação quantitativa, mas sim explorar a capacidade geral do sistema de mapear e reconstruir diferentes tipos de locais, como áreas abertas, fachadas de prédios e ambientes com vegetação, avaliando

qualitativamente a estabilidade operacional do algoritmo em cenários de uso mais livres e desafiadores.

Figura 28 – Exemplo de cenário de teste qualitativo realizado em área externa no campus da UFSC.

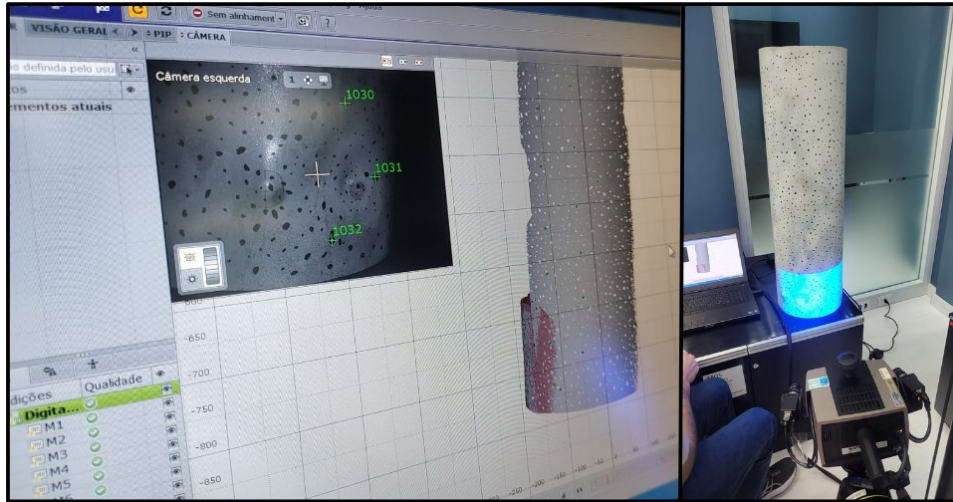


Fonte: Do autor.

Para complementar a análise da trajetória, a avaliação do sistema também focou na incerteza de medição geométrica da nuvem de pontos 3D gerada. Para isso, foram utilizados corpos de prova específicos, permitindo tanto uma avaliação quantitativa contra um modelo de referência quanto uma qualitativa em um cenário de grande escala.

1. Avaliação Quantitativa com Tubo Polimérico; Para a análise quantitativa, foi utilizado um tubo polimérico como corpo de prova. A superfície do tubo foi preparada com diversos pontos pretos para criar uma textura artificial rica em características, facilitando a detecção e o rastreamento pelo ORB-SLAM 3. Um modelo 3D de referência deste tubo foi previamente digitalizado com excelente precisão utilizando o sistema de projeção a laser "GON ATOS" da Certi (Figura 29), servindo como *ground truth* para a geometria do objeto. O experimento, ilustrado na Figura 30, foi conduzido em uma sala de aula do LABMETRO, que foi esvaziada para isolar o objeto. O tubo foi posicionado em uma mesa no centro da sala, e o operador moveu o sistema de SLAM ao seu redor, contornando-o para fechar o *loop* e variando a distância para capturar a geometria de diferentes pontos de vista. Este procedimento foi realizado com ambos os sistemas de câmera, o monocromático e o colorido, para permitir uma comparação direta da qualidade de reconstrução de cada um.

Figura 29 – Corpo de prova (tubo polimérico) sendo reconstruído pelo GON ATOS.



Fonte: Do autor.

Figura 30 – Corpo de prova (tubo polimérico) e arranjo experimental para avaliação da nuvem de pontos no LABMETRO.



Fonte: Do autor.

2. Avaliação Qualitativa de Objeto em Larga Escala: Para avaliar a capacidade do sistema em um cenário mais desafiador e de maior escala, foi realizada a reconstrução da caixa d'água da UFSC (Figura 31), localizada atrás do centro de eventos. Este teste teve um caráter puramente qualitativo, com o objetivo de observar o comportamento do algoritmo na reconstrução de uma estrutura grande e complexa,

sem a intenção de realizar uma comparação métrica, dada a dificuldade em obter um modelo de referência para um objeto de tais dimensões.

Figura 31 – Caixa d'água da UFSC, utilizada como alvo para o teste qualitativo de reconstrução em larga escala.



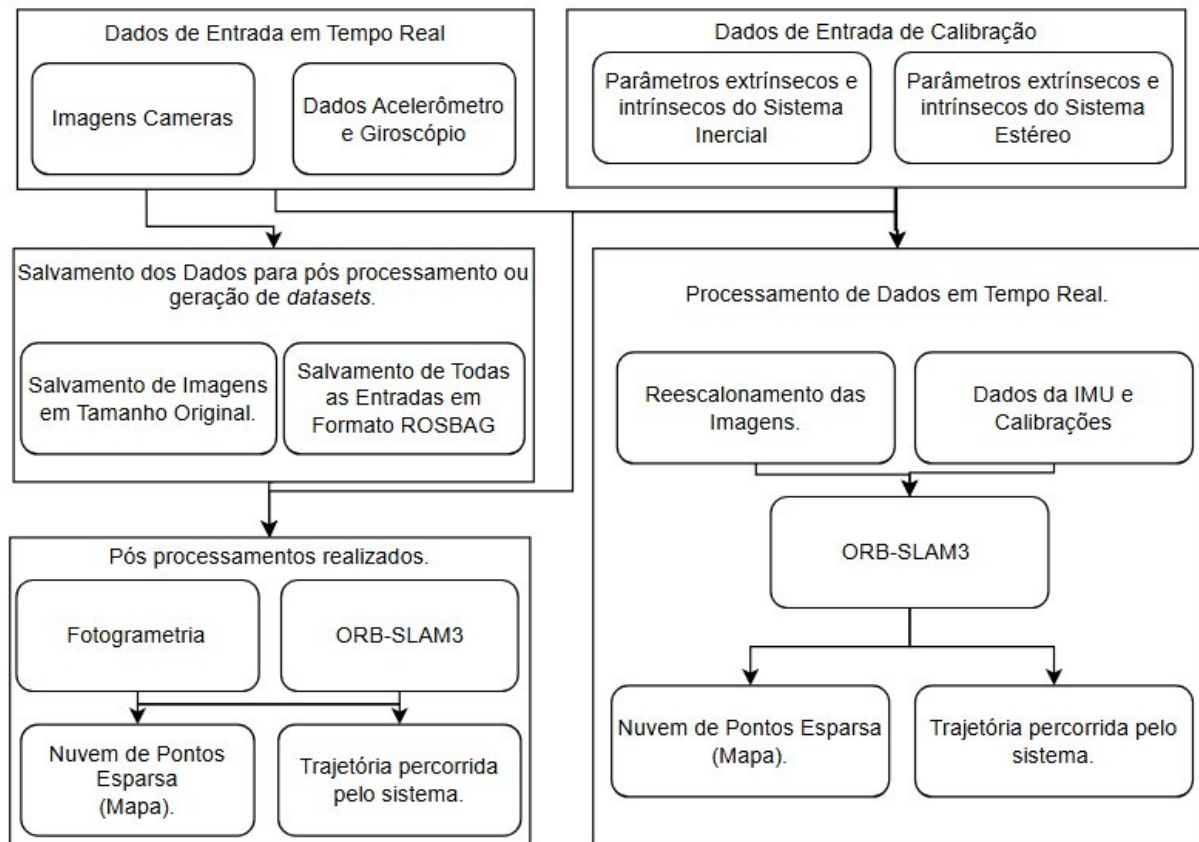
Fonte: Do autor.

É importante notar que diferentes configurações de *hardware* foram utilizadas para testes específicos, de acordo com os requisitos de cada avaliação. Para o teste de trajetória na guia linear, utilizou-se o sistema com as câmeras monocromáticas mais antigas. Já para o teste de trajetória com fechamento de *loop* nos corredores, foi empregado o sistema com as câmeras coloridas de maior resolução, uma vez que a geração do *ground truth* através de fotogrametria se beneficiava da informação de cor e da maior quantidade de detalhes nas imagens. Finalmente, na avaliação do corpo de prova (o tubo polimérico), ambos os sistemas de câmera foram testados para uma análise comparativa do desempenho de reconstrução.

4.3.4 Metodologia de Coleta de Dados e Geração de *Datasets*

Com os cenários de teste definidos, esta seção detalha o procedimento técnico padrão (Figura 32 para a aquisição de dados e a subsequente geração dos *datasets* que foram utilizados para a avaliação do sistema.

Figura 32 – Diagrama do procedimento técnico para aquisição dos dados e geração de *datasets*.



Fonte: Do autor.

O processo de coleta de dados para cada ensaio experimental seguiu um protocolo consistente. Primeiramente, o sistema de *hardware* (câmeras, IMU e notebook) era inicializado. Em seguida, um arquivo de inicialização (*launch file*) do ROS 2 era executado, ativando os *drivers* dos sensores e fazendo com que os dados de imagem e da IMU fossem publicados em seus respectivos tópicos. A gravação era iniciada através de um comando no terminal, que utilizava a ferramenta *ros2 bag record* para salvar todos os dados relevantes em um único arquivo no formato MCAP. O comando de gravação era configurado para registrar os dados por tempo indeterminado, sendo a coleta finalizada manualmente pelo operador com o comando Ctrl+C após a conclusão da trajetória no cenário em questão. O resultado de cada coleta é um arquivo rosbag autocontido, contendo todos os fluxos de sensores com seus respectivos *timestamps* sincronizados.

Para a geração dos *Datasets* de referência (*Ground Truth*), cada rosbag coletado constitui um *dataset* bruto. Para os testes quantitativos, cada um desses

datasets foi pareado com um arquivo de dados de referência (*ground truth*), gerado a partir dos sistemas de medição de baixa incerteza de medição descritos na seção anterior.

Para o Cenário em Ambiente Controlado: A referência foi gerada a partir das leituras do laser interferométrico. Por se tratar de um movimento puramente linear, o *ground truth* consiste em um arquivo de texto simples contendo uma série de poses 3D que descrevem a trajetória retilínea de 3 metros.

Para o Cenário em Ambiente Não Controlado: O *ground truth* da trajetória foi gerado por meio de fotogrametria. As imagens de alta resolução da câmera colorida, extraídas do rosbag, foram processadas em um software de fotogrametria para gerar uma reconstrução 3D densa do ambiente e, a partir dela, a trajetória da câmera com baixa incerteza de medição, que serviu como referência.

Para o Corpo de Prova Quantitativo: O modelo de referência do tubo polimérico foi gerado a partir do *scanner* 3D de baixa incerteza de medição GOM ATOS. O resultado é uma nuvem de pontos densa e metricamente exata do objeto, que serve como *ground truth* para a avaliação da qualidade da reconstrução geométrica do SLAM.

Ao final deste processo, cada ensaio quantitativo resultou em um par de dados: o rosbag com os dados brutos dos sensores do sistema e o arquivo correspondente com o *ground truth*, prontos para a etapa de avaliação de desempenho.

4.3.5 Métricas de Avaliação de Desempenho

A metodologia de avaliação de desempenho do sistema foi um processo iterativo, aprimorado ao longo do desenvolvimento do projeto e documentado em publicações (93). Esta abordagem permitiu uma análise progressivamente mais completa e rigorosa, culminando nas métricas finais detalhadas nesta seção.

A avaliação inicial do sistema, publicada no congresso PDPETRO (93), utilizou uma configuração de *baseline* e resolução de 1600x1200. Neste estudo, a avaliação foi conduzida em oito posições distintas dentro do ambiente controlado, com o objetivo de analisar como a variação natural da cena, resultando em um número maior ou menor de características ORB detectadas, interferiria no desempenho do sistema. Os testes na guia linear demonstraram um desempenho coerente com o esperado para

o ORB-SLAM 3, apresentando incerteza de medição semelhante a encontrada na literatura referente a medição do deslocamento. A análise dos resultados também revelou limitações na metodologia de avaliação da época, como o uso de uma resolução computacionalmente custosa e, principalmente, a não avaliação do desvio no eixo Z, marcando este estudo como uma validação preliminar.

Posteriormente, uma avaliação mais aprofundada e metodologicamente rigorosa foi conduzida, cujos detalhes são aprofundados neste mestrado e foram publicados no congresso Metrologia 2025. Utilizando o mesmo sistema de câmeras monocromáticas, este estudo realizou uma extensa análise paramétrica dos modos estéreo e estéreo-inercial. Para mitigar a falta de características visuais, a sala foi preparada com diversas folhas de papel contendo pontos pretos, garantindo uma densa e uniforme distribuição de pontos de interesse. A avaliação foi realizada em três posições distintas, com o objetivo não apenas de testar em cenas diferentes, mas de incorporar a variabilidade induzida pela mudança de cena como uma componente na análise estatística dos erros de medição. Uma contribuição chave deste trabalho foi a introdução da métrica de retitude para a trajetória na guia linear, quantificando o quão reta a trajetória gerada se manteve. Os resultados obtidos demonstraram que, com a parametrização otimizada, o sistema era capaz de atingir tanto uma incerteza de medição correta quanto uma boa repetibilidade de medição.

Definida a parametrização otimizada do sistema, o passo metodológico seguinte foi estabelecer e validar o processo de geração de *ground truth* para os ensaios em larga escala. Antes que a fotogrametria, processada via COLMAP, pudesse ser empregada como uma referência confiável, sua própria fidelidade métrica precisava ser quantificada. Para este fim, foi conduzido um experimento de validação metrológica em ambiente laboratorial (temperatura controlada a 20°C), utilizando a guia linear de precisão e tendo como padrão de referência absoluto o interferômetro a laser Renishaw XL-80. O procedimento metodológico envolveu a captura de imagens com o novo sistema de câmeras coloridas de alta resolução, enquanto este era deslocado ao longo da guia. A trajetória resultante do processamento fotogramétrico foi então avaliada, comparando-se seu deslocamento medido e seu desvio de retitude contra a referência de baixa incerteza de medição fornecida pelo laser. Este ensaio permitiu, portanto, quantificar a qualidade geométrica da trajetória

fotogramétrica, atestando sua alta fidelidade e validando seu uso como *ground truth* para a avaliação de desempenho do ORB-SLAM 3.

Com base nos aprendizados e na configuração de parâmetros otimizada validada, procedeu-se com a avaliação final do sistema utilizando as métricas detalhadas a seguir, focando tanto na Incerteza de Medição da trajetória em larga escala quanto na Incerteza de Medição da reconstrução 3D. Assim, tivemos a análise de trajetória em ambiente não controlado e a avaliação da nuvem de pontos.

A limitação do sistema antigo, com câmeras monocromáticas de baixa resolução, para gerar um *ground truth* confiável via fotogrametria em ambientes complexos motivou a realização de um teste específico com o novo sistema. Utilizando as câmeras coloridas de alta resolução ("Configuração 2"), foi realizada a trajetória em circuito fechado nos corredores do CCE, conforme descrito na seção 4.3.3.

Para a avaliação quantitativa, a trajetória estimada pelo ORB-SLAM 3 foi comparada com a trajetória de referência gerada pela fotogrametria utilizando o pacote de software EVO (*Evaluation of Odometry and SLAM*). Esta ferramenta, padrão na comunidade de robótica, alinha as duas trajetórias e calcula as métricas de erro, das quais as duas principais foram utilizadas:

Erro de Pose Absoluto (APE - *Absolute Pose Error*): Mede a consistência global da trajetória e o desvio espacial total acumulado. Após o alinhamento, calcula-se a distância euclidiana direta entre cada pose estimada pelo SLAM e a pose correspondente no *ground truth*. O resultado agregado é tipicamente apresentado como o Erro Quadrático Médio (RMSE), fornecendo uma única medida do desempenho global do sistema, que representa o erro final de *drift* após o percurso completo.

Erro de Pose Relativo (RPE - *Relative Pose Error*): Mede a deriva (*drift*) local do sistema. Esta métrica avalia o erro no movimento relativo entre pares de poses consecutivas ao longo de um intervalo de tempo ou distância fixo. É uma métrica eficaz para quantificar a taxa de acúmulo de erro de translação e rotação (*drift rate*) ao longo do percurso.

A avaliação final focou em uma análise comparativa da qualidade da reconstrução 3D do corpo de prova (o tubo polimérico), cujo modelo de referência foi obtido com o *scanner* GOM ATOS. Este teste foi projetado para comparar exhaustivamente o desempenho de ambos os sistemas de hardware (antigo

monocromático e novo colorido) e ambos os modos de operação do ORB-SLAM 3 (estéreo e estéreo-inercial).

Para a comparação, foi utilizado o *software* de código aberto CloudCompare. A nuvem de pontos gerada por cada configuração do sistema SLAM foi alinhada com a nuvem de pontos de referência do *scanner*. A avaliação consistiu na realização de medições de dimensões chave do tubo (como diâmetro e comprimento) diretamente na nuvem de pontos reconstruída. Os valores medidos em cada um dos quatro cenários de teste (sistema antigo/estéreo, sistema antigo/estéreo-inercial, sistema novo/estéreo, sistema novo/estéreo-inercial) foram então comparados com as dimensões nominais do GOM para determinar o erro de medição geométrica de cada configuração.

5 RESULTADOS

5.1 RESULTADOS DA CALIBRAÇÃO DO SISTEMA

Este capítulo apresenta os resultados experimentais obtidos para validar o desempenho do sistema de navegação visual-inercial proposto utilizando o algoritmo ORB-SLAM 3. Serão expostos os dados referentes à calibração do sistema, seguidos pela avaliação quantitativa da precisão e incerteza de precisão da odometria e pela análise da qualidade dos mapas 3D reconstruídos. Por fim, será discutido o desempenho computacional da solução para aferir sua viabilidade prática.

5.1.1 Parâmetros Intrínsecos, Extrínsecos e Erro de Reprojeção das Câmeras

A primeira etapa na validação do sistema sensorial consiste na análise da calibração intrínseca de cada câmera e extrínseca do sistema, que modela suas características ópticas internas (intrínseca). Conforme o procedimento detalhado no Capítulo 4, foram realizados testes utilizando padrões de tabuleiros de xadrez com quadrados de 50 mm e 25 mm. Embora ambos os padrões tenham sido avaliados em termos de erro de retroprojeção na calibração das imagens, a configuração com o tabuleiro de 25 mm apresentou incerteza menor nos resultados da calibração e resultou em um desempenho superior do sistema ORB-SLAM 3 na estimação da trajetória. Portanto, os resultados e parâmetros discutidos nesta seção correspondem à calibração obtida com o padrão de 25 mm, adotado neste trabalho.

A calibração foi executada para três configurações distintas do sistema de visão, permitindo uma análise comparativa de resultados de calibração: o sistema inicial com câmeras monocromáticas em resolução total (1600x1200); o mesmo sistema operando com resolução de imagem reduzida (800x600); e o novo sistema composto por câmeras coloridas, também com resolução reduzida (800x600). Os parâmetros intrínsecos e os coeficientes de distorção resultantes, sendo distância focal (*pixels*), ponto principal (*pixels*), distorção radial e tangencial para cada uma das seis câmeras apresentadas como C_e para câmera da esquerda e C_d para câmera da direita estão consolidados na Tabela 1.

Tabela 1 – Parâmetros intrínsecos e de distorção para as configurações de câmera.

Parâmetros		Sistema inicial total		Sistema inicial red		Sistema novo	
		Ce	Cd	Ce	Cd	Ce	Cd
Distância	f_x	809.423	805.458	405.714	403.969	417.699	417.317
Focal	f_y	808.183	804.166	405.188	403.518	374.574	373.903
Ponto	c_x	779.603	780.385	389.052	389.263	408.626	398.645
Principal	c_y	600.768	601.102	300.099	300.462	306.899	306.225
Distorção Radial	k_1	-0.050	-0.053	-0.041	-0.049	-0.200	-0.202
	k_2	0.106	0.112	0.088	0.109	0.079	0.083
	k_3	-0.052	-0.055	-0.041	-0.056	-0.013	-0.015
Distorção	p_1	1.5×10^{-4}	1.8×10^{-5}	6.5×10^{-5}	-3.2×10^{-4}	-6.8×10^{-5}	-3.0×10^{-4}
Tangencial	p_2	7.3×10^{-5}	-3.0×10^{-4}	-1.0×10^{-3}	-1.3×10^{-3}	5.1×10^{-4}	-6.3×10^{-4}

Fonte: Do autor.

Além dos parâmetros intrínsecos que caracterizam cada câmera individualmente, a calibração estéreo também determina os parâmetros extrínsecos. Eles definem a transformação espacial entre as câmeras esquerda e direita de cada par, o que é fundamental para a retificação das imagens e o cálculo de profundidade. Para uma análise detalhada, esta transformação é decomposta em seus dois componentes fundamentais: Rotação (R) e Translação (t), apresentados a seguir.

A Tabela 2 exibe a matriz de transformação homogênea de 4x4 para cada par estéreo. A matriz R descreve a orientação do sistema de coordenadas da Câmera Direita em relação à Câmera Esquerda. O vetor t define o deslocamento da origem da Câmera Direita em relação à origem da Câmera Esquerda. Os componentes do da ultima coluna da matriz de Transformação Homogênea (t_x, t_y, t_z) são expressos em metros e representam a linha de base (*baseline*) do sistema estéreo, sendo t o componente principal.

Tabela 2 – Matrizes de Transformação Homogênea entre as câmeras de cada par estéreo.

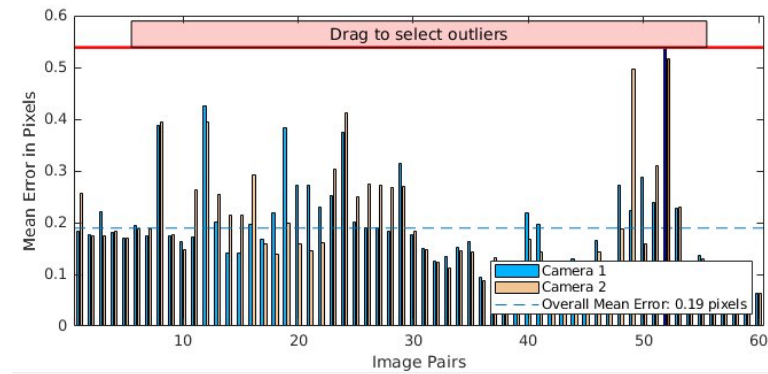
Configuração do Sistema	Matriz de Rotação (R) [Esquerda para Direita]
Sistema Inicial Total	$\begin{bmatrix} 1.0000 & -0.0081 & 0.0034 & -0.1996 \\ 0.0081 & 1.0000 & -0.0042 & 0.0011 \\ -0.0034 & 0.0042 & 1.0000 & -0.0007 \end{bmatrix}$
Sistema Inicial Reduzido	$\begin{bmatrix} 1.0000 & -0.0084 & 0.0043 & -0.1997 \\ 0.0084 & 0.9999 & -0.0059 & 0.0011 \\ -0.0043 & 0.0059 & 1.0000 & -0.0004 \end{bmatrix}$
Sistema Novo Reduzido	$\begin{bmatrix} 0.9997 & 0.0229 & -0.0030 & -0.2003 \\ -0.0229 & 0.9997 & 0.0035 & 0.0014 \\ 0.0029 & -0.0036 & 1.0000 & -0.0002 \end{bmatrix}$

Fonte: Do autor.

A qualidade final da calibração de cada sistema foi quantificada através do erro médio de reprojeção, que corresponde à distância euclidiana média (em *pixels*) entre os cantos detectados no padrão de xadrez e os pontos 3D correspondentes reprojados na imagem, utilizando os parâmetros calculados. Para garantir a máxima exatidão, foi adotado um procedimento manual de identificação e remoção de *outliers* – pares de imagens que contribuíam com erros desproporcionalmente altos. Após a remoção, a calibração foi otimizada. A análise para cada uma das três configurações é detalhada a seguir.

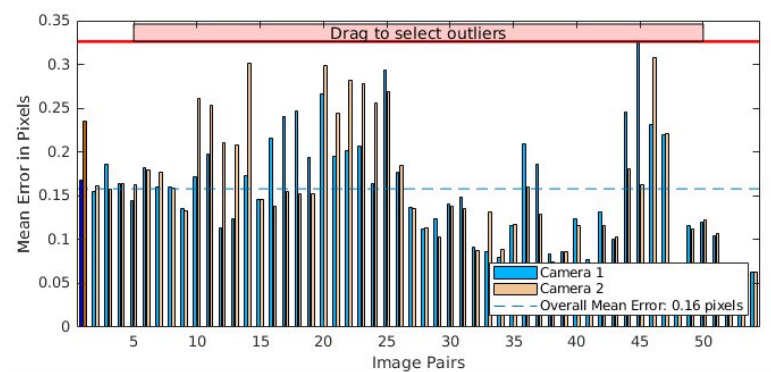
Para o sistema inicial (1600x1200) em sua resolução total, a análise do erro de reprojeção revelou a existência de alguns pares de imagens com erros elevados, que se afastavam da média geral, conforme pode ser visto no gráfico inicial da Figura 33. Com isso obteve-se uma distribuição de erro mais estável e de maior fidelidade métrica, apresentada na Figura 34. Com este refinamento, o erro médio de reprojeção final para esta configuração foi reduzido para 0.16 *pixels*, um valor que atesta a alta fidelidade do modelo de câmera obtido. A Figura 35 ilustra a distribuição espacial das poses da câmera em relação ao tabuleiro, confirmando que a calibração se baseou em uma variedade de ângulos e distâncias, garantindo sua estabilidade operacional.

Figura 33 – Gráfico do erro de reprojeção sistema inicial com resolução cheia.



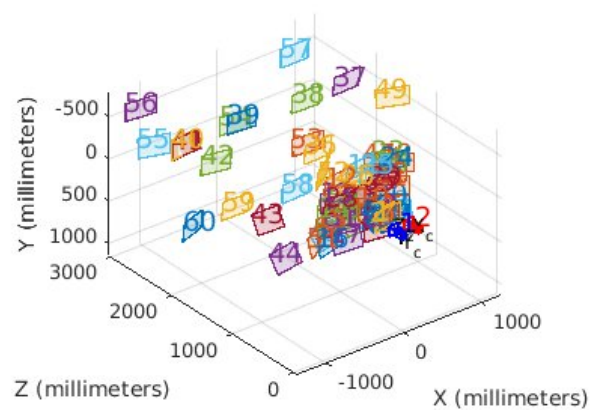
Fonte: Do autor.

Figura 34 – Gráfico do erro de reprojeção sistema inicial com resolução cheia após remoção de imagens com alto erro de reprojeção.



Fonte: Do autor.

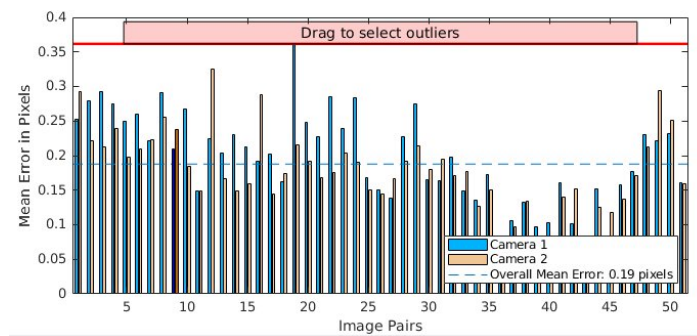
Figura 35 – Poses do tabuleiro em relação a câmera do sistema inicial com resolução cheia.



Fonte: Do autor.

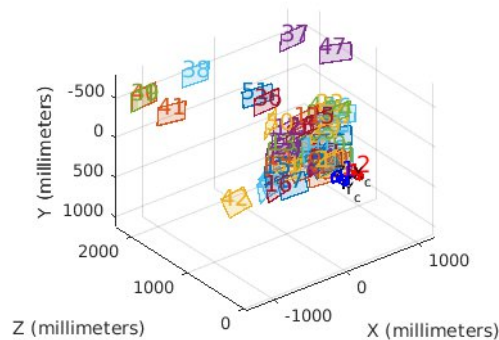
Na calibração do sistema com resolução reduzida, o gráfico de erro (Figura 36) apresentou uma distribuição mais homogênea, sem a presença de *outliers* que impactassem significativamente o resultado. Embora algumas imagens tenham sido descartadas durante a captura, as restantes formaram um conjunto de dados consistente. Por essa razão, não foi necessária a remoção de amostras adicionais. O erro médio final obtido foi de 0.19 *pixels*. As poses utilizadas nesta calibração são visualizadas na Figura 37.

Figura 36 – Gráfico do erro de reprojeção sistema inicial com resolução reduzida.



Fonte: Do autor.

Figura 37 – Poses do tabuleiro em relação a câmera do sistema inicial com a resolução reduzida.

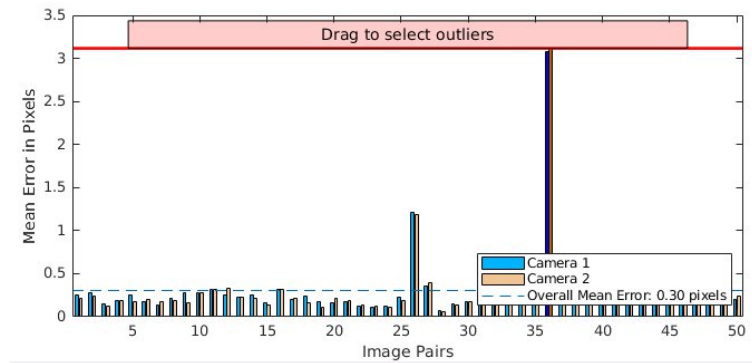


Fonte: Do autor.

O processo de calibração para o novo sistema de câmeras coloridas demonstrou a importância da etapa de refinamento. O gráfico de calibração inicial (Figura 38) apresentou *outliers* acentuados, que elevavam o erro médio geral para 0.30 *pixels*. Após a identificação e remoção desses pares de imagens com erros

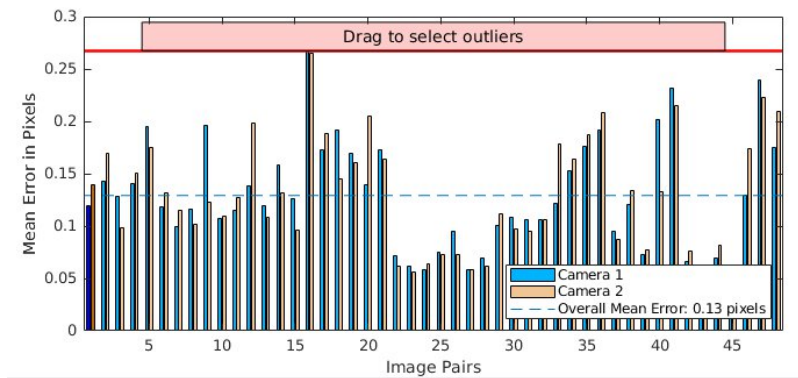
elevados (*outliers*), o modelo foi otimizado. Como resultado, o erro médio final foi reduzido para 0.13 *pixels*, o menor valor entre as três configurações avaliadas (Figura 39). A diversidade de poses que contribuíram para este resultado preciso é mostrada na Figura 40.

Figura 38 – Gráfico do erro de reprojeção sistema novo.



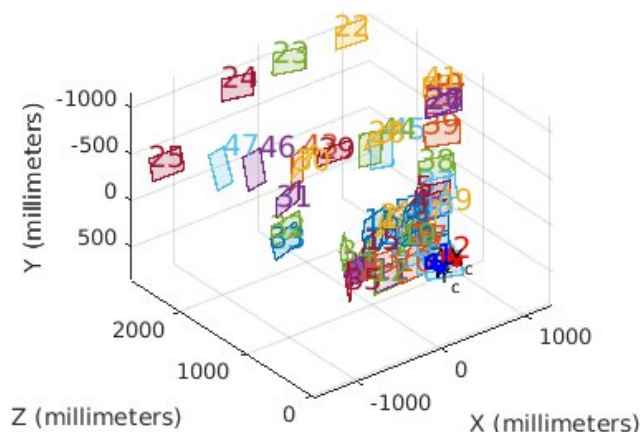
Fonte: Do autor.

Figura 39 – Gráfico do erro de reprojeção sistema novo após remoção de imagens com alto erro de reprojeção.



Fonte: Do autor.

Figura 40 – Poses do tabuleiro em relação a câmera do sistema novo.



Fonte: Do autor.

A calibração das três configurações de sistema foi concluída com sucesso, resultando em parâmetros de alta fidelidade métrica. Estes resultados validam a baixa incerteza de medição dos parâmetros intrínsecos e extrínsecos obtidos, estabelecendo uma base sensorial confiável. A qualidade desta calibração será visualmente inspecionada na próxima seção, através da análise da retificação das imagens.

5.1.2 Análise da Retificação das Imagens

Uma vez validados os parâmetros intrínsecos e extrínsecos do sistema, o próximo passo consiste em avaliar a aplicação prática desses parâmetros no processo de retificação estéreo. A retificação alinha as imagens do par de câmeras de modo que as linhas epipolares se tornem colineares e paralelas ao eixo horizontal da imagem. Este alinhamento aumenta a eficiência dos algoritmos de correspondência estéreo, pois restringe a busca por pontos correspondentes a uma única linha, otimizando significativamente o processo de triangulação e o cálculo de mapas de disparidade.

Para validar visualmente a qualidade da calibração, o resultado do processo de retificação foi analisado para cada uma das três configurações de sistema (Sistema Inicial, Sistema Inicial Reduzido e Sistema Novo). Em cada par de imagens, o par

superior exibia a captura original, enquanto o par inferior exibia as mesmas imagens após a retificação.

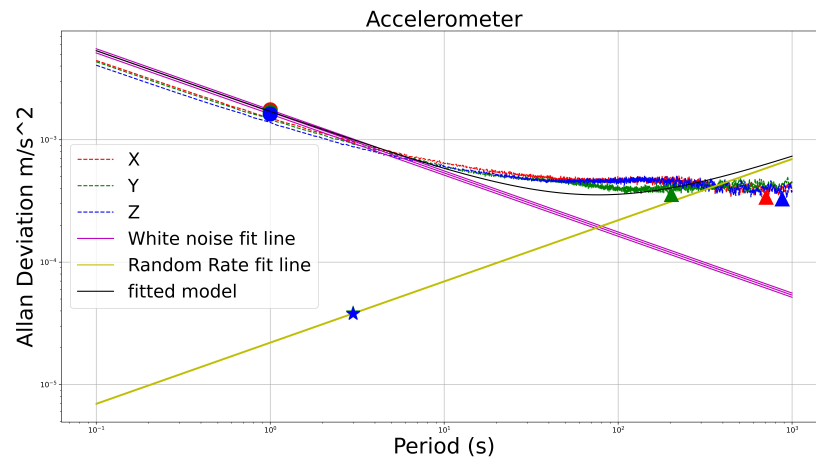
Como observado, após a aplicação da transformação de retificação, os cantos do tabuleiro de xadrez e outros pontos de características correspondentes entre os pares de imagens alinham-se ao longo das linhas horizontais traçadas. Este resultado confirma de forma visual e inequívoca a incerteza de medição e a elevada repetibilidade de medição da calibração extrínseca para os três sistemas, validando que eles estão aptos para a execução de tarefas que dependem do cálculo de profundidade, como a reconstrução 3D e a odometria visual-inercial.

5.1.3 Calibração Intrínseca da IMU: Análise de Ruído

Antes de determinar a relação extrínseca entre a IMU e as câmeras, é fundamental caracterizar as propriedades de ruído estocástico do sensor inercial. Esses parâmetros são entradas essenciais para os filtros de estimação, como o Filtro de Kalman, utilizados pelo ORB-SLAM 3, pois permitem modelar a incerteza das medições e melhorar significativamente a precisão da fusão de dados visual-inercial.

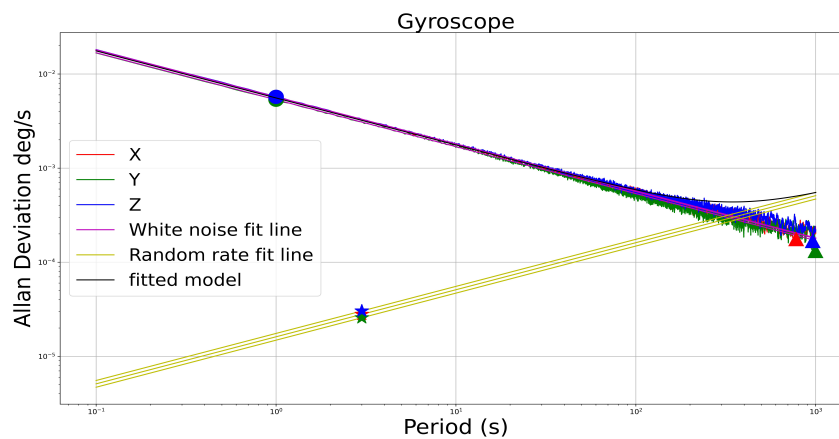
Para determinar os parâmetros de ruído, a IMU foi mantida em uma posição estática por um período prolongado para coletar dados do giroscópio e do acelerômetro, publicados no tópico ROS `/imu/data_raw` a uma frequência de 250 Hz. A análise foi realizada utilizando o método da Variância de Allan (91), uma técnica padrão para identificar diferentes fontes de ruído em sensores inerciais. As Figuras 41 e 42 apresentam os gráficos de desvio de Allan para o acelerômetro e o giroscópio, respectivamente.

Figura 41 – Gráfico de Desvio de Allan para o Acelerômetro.



Fonte: Do autor.

Figura 42 – Gráfico de Desvio de Allan para o Giroscópio.



Fonte: Do autor.

A partir da análise dos gráficos de Desvio de Allan, é possível extrair os parâmetros de ruído para cada sensor: a Densidade de Ruído (*Noise Density*), que representa o ruído branco de alta frequência, e o Passeio Aleatório (*Random Walk*), que caracteriza o desvio do bias em baixas frequências. Os valores numéricos obtidos para a IMU utilizada neste trabalho estão sumarizados na Tabela 3.

Tabela 3 – Parâmetros de Ruído da IMU.

Sensor	Parâmetro	Unidade	Valor
Acelerômetro	Ruído Branco	$m/s^2/\sqrt{Hz}$	1.79×10^{-3}
	Instabilidade de Bias	$m/s^2.\sqrt{Hz}$	3.75×10^{-5}
Giroscópio	Ruído Branco	$rad/s/\sqrt{Hz}$	1.00×10^{-4}
	Instabilidade de Bias	$rad/s.\sqrt{Hz}$	5.01×10^{-7}

Fonte: Do autor.

Estes valores foram utilizados para parametrizar adequadamente o modelo de ruído no arquivo de configuração do ORB-SLAM 3, garantindo que o algoritmo operasse com a máxima fidelidade às características do sensor. Com as propriedades intrínsecas da IMU devidamente modeladas, a próxima seção aborda a calibração extrínseca entre o sensor inercial e o sistema de câmeras estéreo.

5.1.4 Validação da Calibração Câmera-IMU

A calibração extrínseca entre o sistema de câmeras e a IMU exigiu uma atenção especial ao posicionamento físico dos componentes para garantir a fusão ótima dos dados. Conforme detalhado no Capítulo 4, a montagem inicial do sensor apresentou limitações. Para solucionar esta questão, a IMU foi reposicionada para uma nova configuração, buscando um alinhamento mais favorável, como idealizado no modelo teórico. Como a estrutura mecânica que une as câmeras e a IMU se manteve a mesma para os três sistemas de visão testados, os resultados da calibração desta configuração final, apresentados a seguir, são válidos para todos os experimentos subsequentes.

O processo de calibração para a configuração final resultou na matriz de transformação homogênea (T_{CI}) apresentada na matriz abaixo. Esta matriz descreve a rotação e a translação necessárias para mapear um ponto do sistema de coordenadas da IMU para o sistema de coordenadas da câmera esquerda.

$$\begin{bmatrix} 0.9999 & 0.0033 & 0.0008 & -0.0746 \\ -0.0033 & 0.9999 & -0.0062 & -0.0074 \\ -0.0008 & 0.0062 & -0.9999 & -0.0381 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

A validação desta transformação foi realizada através da análise de seus componentes de rotação e translação.

Análise da Rotação: Observa-se que a submatriz de rotação 3x3 (canto superior esquerdo de T_{CI}) é extremamente próxima de uma matriz identidade. Os elementos da diagonal principal são quase 1, e os demais elementos são próximos de zero. Isso indica que os eixos da câmera e da IMU foram montados fisicamente com um alinhamento, com um erro rotacional mínimo entre eles, o que valida a qualidade da montagem mecânica.

Análise da Translação: O vetor de translação extraído da matriz (última coluna) define o deslocamento da IMU em relação à câmera como sendo de -74.6 mm no eixo X, -7.4 mm no eixo Y e -38.1 mm no eixo Z. A forte correspondência entre os valores calibrados e as dimensões do projeto físico confirma a exatidão da calibração translacional.

Com isso, validou todo o aparato sensorial. Tendo sido demonstrada a incerteza de medição dos parâmetros intrínsecos das câmeras, da configuração extrínseca do sistema estéreo e da transformação final Câmera-IMU, estabelece-se uma base de medição confiável e confiável. A partir deste ponto, o foco do capítulo se volta para a avaliação dos resultados práticos do sistema, iniciando pela análise de desempenho da odometria na próxima seção.

5.2 AVALIAÇÃO DA ODOMETRIA

Após o sensor ter sido calibrado, esta seção apresenta os resultados: a avaliação de desempenho da odometria. Aqui, o sistema é avaliado com experimentos práticos definidos no Capítulo 4. A análise vai além de uma simples validação, investigando também as configurações de parâmetros que produzem os resultados de melhor qualidade. Para isso, será apresentada uma avaliação completa que combina métricas de erro, uma inspeção visual detalhada das trajetórias geradas e

um comparativo de desempenho que destila as forças e limitações de cada abordagem testada.

5.2.1 Avaliação do Posicionamento e Configurações

A primeira avaliação de desempenho do sistema de odometria foi conduzida em um ambiente controlado, com o objetivo de isolar variáveis e determinar a configuração ótima dos parâmetros de software. Para isso, foi utilizado um percurso retilíneo de 3 metros sobre uma guia linear de baixa incerteza de medição, cuja trajetória de referência (*ground-truth*) foi aferida pelo sistema a laser Renishaw XL-80, conforme detalhado na metodologia do Capítulo 4. A análise nesta seção se desdobra em duas etapas fundamentais: primeiramente, investiga-se a influência do ângulo de montagem do sistema na guia (0° , 45° e 90°), para verificar a existência de erro sistemático. Em seguida, com base nessa análise preliminar, realiza-se um comparativo de desempenho entre quatro configurações distintas do sistema, visando eleger a mais maior fidelidade métrica e consistente para os testes subsequentes.

Para investigar a influência do ângulo de montagem do sistema na guia, foram avaliadas quatro configurações distintas, cujas especificações são detalhadas abaixo para referência:

Configuração 1: Estéreo, 1600x1200 pixels, 1500 ORBs.

Configuração 2: Estéreo, 800x600 pixels, 1500 ORBs.

Configuração 3: Estéreo, 800x600 pixels, 3000 ORBs.

Configuração 4: Estéreo-Inercial, 800x600 pixels, 1500 ORBs.

A análise se baseia em três métricas de erro, calculadas a partir de um mínimo de cinco repetições para cada ângulo. O Erro de Deslocamento mede a diferença entre a distância total percorrida estimada pelo sistema e os 3 metros de referência. O Erro de Retitude representa a média do desvio perpendicular de todos os pontos da trajetória em relação a uma reta ajustada por mínimos quadrados com os pontos. Por fim, o Erro Máximo de Retitude indica o ponto de maior desvio perpendicular observado durante o percurso.

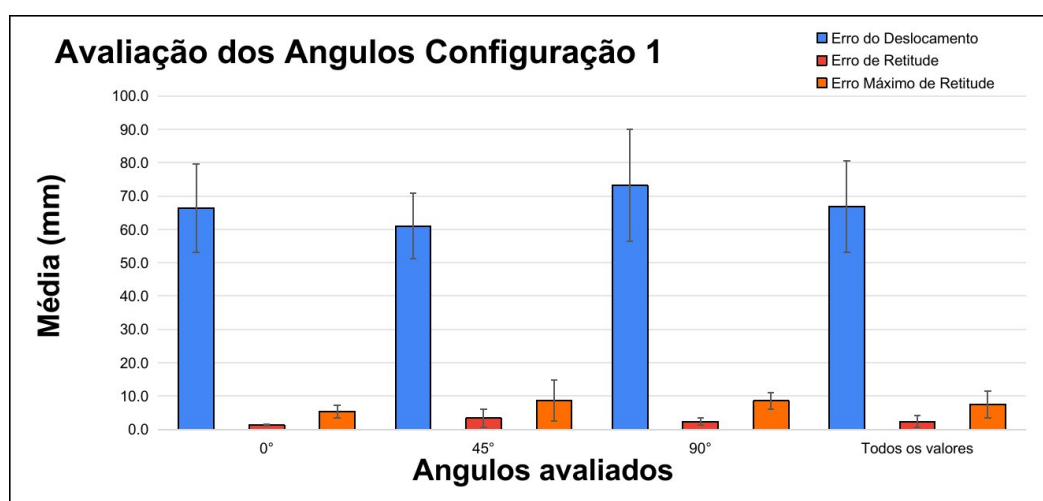
A Tabela 4 e Figura 43 sumariza os resultados de média e desvio padrão para a Configuração 1.

Tabela 4 – Resultados de erro para a Configuração 1 nos diferentes ângulos.

CONFIGURAÇÃO 1		0°	45°	90°	Todos os casos
Erro do Deslocamento	Média(mm)	66.4	61.0	73.2	66.9
	Desvio Padrão(mm)	13.3	9.8	16.8	13.8
Erro de Retitude	Média(mm)	1.4	3.3	2.4	2.4
	Desvio Padrão(mm)	0.2	2.8	1.1	1.8
Erro Maximo de Retitude	Média(mm)	5.4	8.7	8.5	7.5
	Desvio Padrão(mm)	1.8	6.1	2.4	4.0

Fonte: Do autor.

Figura 43 – Média e Desvio Padrão dos Erros para Configuração 1.



Fonte: Do autor.

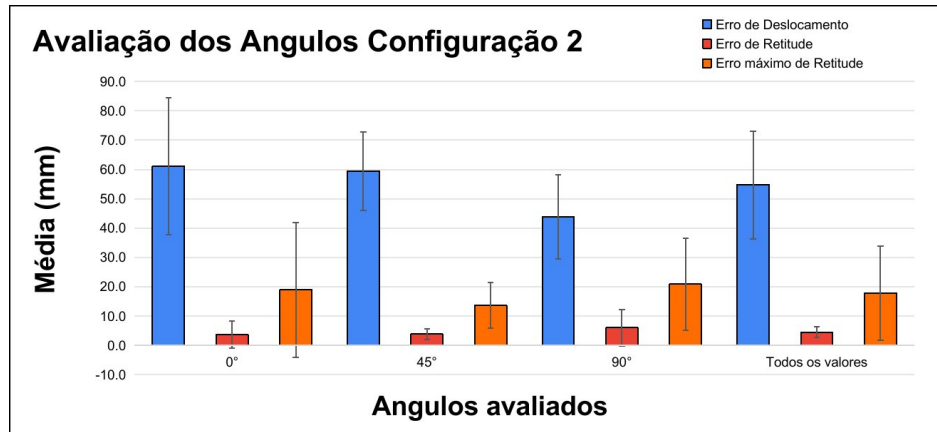
Os resultados para a Configuração 2, que utiliza resolução reduzida, são apresentados na Tabela 5 e visualizados na Figura 44.

Tabela 5 – Resultados de erro para a Configuração 2 nos diferentes ângulos.

CONFIGURAÇÃO 2		0°	45°	90°	Todos os casos
Erro do Deslocamento	Média(mm)	61.1	59.4	43.8	54.8
	Desvio Padrão(mm)	23.3	13.4	14.4	18.4
Erro de Retitude	Média(mm)	3.7	3.8	6.0	4.5
	Desvio Padrão(mm)	4.7	1.8	6.2	1.7
Erro Maximo de Retitude	Média(mm)	19.0	13.7	20.9	17.8
	Desvio Padrão(mm)	23.0	7.9	15.8	16.0

Fonte: Do autor.

Figura 44 – Média e Desvio Padrão dos Erros para Configuração 2.



Fonte: Do autor.

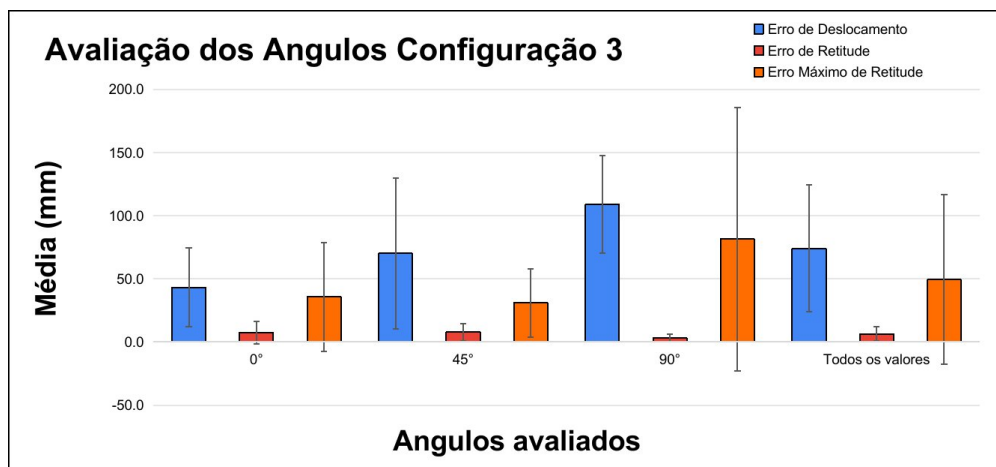
A Tabela 6 e a Figura 45 detalham os resultados da Configuração 3, que utiliza o dobro de *features* ORB.

Tabela 6 – Resultados de erro para a Configuração 3 nos diferentes ângulos.

CONFIGURAÇÃO 3		0°	45°	90°	Todos os casos
Erro do Deslocamento	Média(mm)	43.3	70.1	109.0	74.1
	Desvio Padrão(mm)	31.3	59.6	38.5	50.4
Erro de Retitude	Média(mm)	7.5	7.7	3.1	6.3
	Desvio Padrão(mm)	8.8	7.0	2.8	5.7
Erro Maximo de Retitude	Média(mm)	35.7	31.0	81.5	49.4
	Desvio Padrão(mm)	43.1	27.2	104.4	67.2

Fonte: Do autor.

Figura 45 – Média e Desvio Padrão dos Erros para Configuração 3.



Fonte: Do autor.

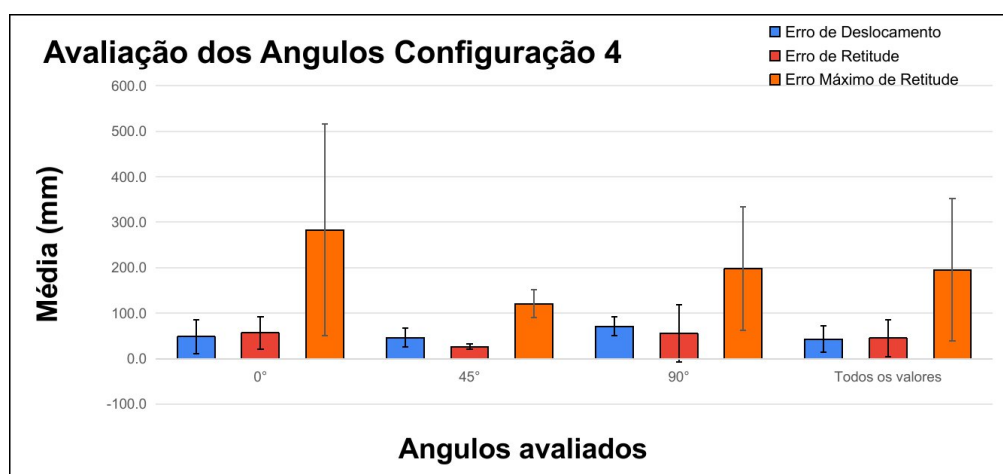
Finalmente, a análise para a Configuração 4 (Estéreo-Inercial) é apresentada na Tabela 7 e nas Figuras 46.

Tabela 7 – Resultados de erro para a Configuração 4 nos diferentes ângulos.

CONFIGURAÇÃO 4		0°	45°	90°	Todos os valores
Erro do Deslocamento	Média(mm)	48.2	46.0	71.3	43.1
	Desvio Padrão(mm)	36.5	20.5	20.4	29.3
Erro de Retitude	Média(m)	56.5	26.4	55.8	45.0
	Desvio Padrão(mm)	35.9	5.5	63.6	40.6
Erro Maximo de Retitude	Média(mm)	283.0	120.8	198.1	195.6
	Desvio Padrão(mm)	232.5	30.6	135.8	156.3

Fonte: Do autor.

Figura 46 – Média e Desvio Padrão dos Erros para Configuração 4.



Fonte: Do autor.

Após a análise individual de cada uma das quatro configurações, observa-se um comportamento consistente: não há evidências de que um ângulo de montagem específico (0°, 45° ou 90°) produza resultados sistematicamente melhores ou piores. Embora os valores médios flutuem, a alta dispersão dos dados, evidenciada pelas barras de desvio padrão, resulta em uma sobreposição significativa das faixas de variação. Isso indica que as diferenças observadas não são estatisticamente relevantes. Portanto, conclui-se que o ângulo de montagem não é uma variável de influência predominante neste experimento, dado o cenário de baixa riqueza de características visuais utilizado. É, no entanto, pertinente considerar que em cenários ricos em detalhes visuais, a diferença de ângulos poderia se tornar estatisticamente significativa, influenciando a visibilidade de *features* e, consequentemente, a

fidelidade da triangulação. Essa limitação está fora do escopo principal deste trabalho, permitindo que os dados de todos os ângulos sejam agrupados para a análise comparativa entre as configurações, que será realizada na próxima etapa.

Tendo estabelecido que o ângulo de montagem não é uma variável de influência significativa, os dados de todos os ângulos foram agrupados para permitir uma comparação direta do desempenho geral entre as quatro configurações. A Tabela 8 sumariza os valores médios e de desvio padrão para cada caso, utilizando um código de cores para facilitar a interpretação visual, onde verde representa os resultados mais favoráveis e vermelho, os mais desfavoráveis.

Tabela 8 – Comparativo de desempenho geral entre as quatro configurações.

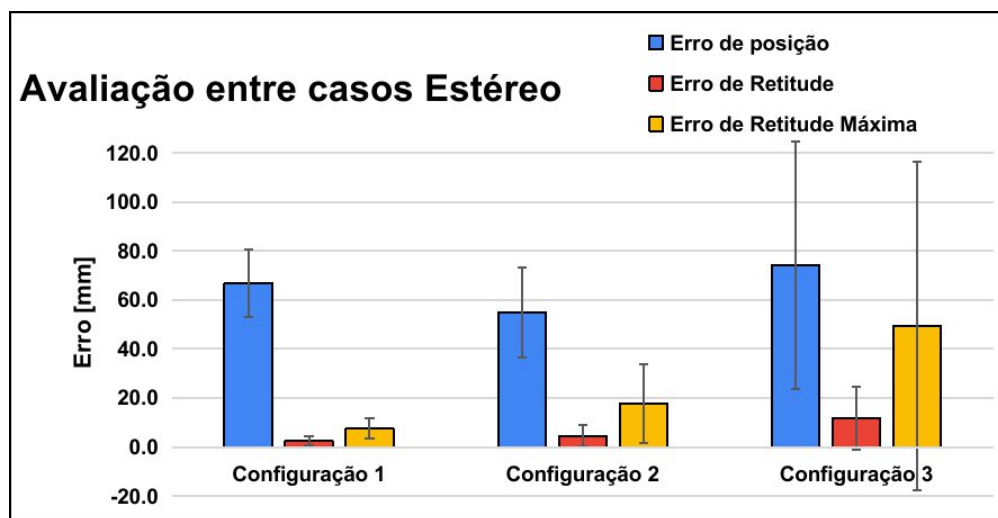
Comparando configurações	Configuração 1	Configuração 2	Configuração 3	Configuração 4
Média Erro de Posição	66,9	54,8	74,1	43,1
Desvio Padrão Erro de Posição	13,8	18,4	50,4	29,3
Média Dist. Pontos-Reta	2,4	4,5	6,3	45,0
Desvio Padrão Pontos-Reta	1,8	1,7	5,7	40,6
Média Dist. Máxima Reta	7,5	17,8	49,4	195,6
Desvio Padrão Dist. Máxima Reta	4,0	16,0	67,2	156,3

Fonte: Do autor.

A análise comparativa será realizada em duas etapas: primeiramente, uma avaliação entre as configurações puramente estéreo (Configurações 1, 2 e 3); em seguida, um confronto direto entre a melhor configuração estéreo e a estéreo-inercial (Configuração 4).

Ao comparar as três configurações puramente estéreo, a Figura 47 ilustra seus desempenhos relativos. A Configuração 1 (alta resolução) apresenta os melhores resultados nas métricas de retitude. Contudo, considerando que o Erro de Posição final é a métrica prioritária para a avaliação da odometria, a Configuração 2 se destaca com o menor erro médio de deslocamento (54,8 mm). Mesmo com um desvio padrão um pouco superior (18,4 mm) ao da Configuração 1, a utilização da resolução menor tem a grande vantagem de diminuir a latência de processamento, o que é um benefício crucial para sistemas em tempo real. Portanto, a combinação do baixo erro médio com a otimização computacional a define como a de melhor desempenho entre as abordagens puramente estéreo.

Figura 47 – Gráfico comparativo de desempenho entre as configurações estéreo.

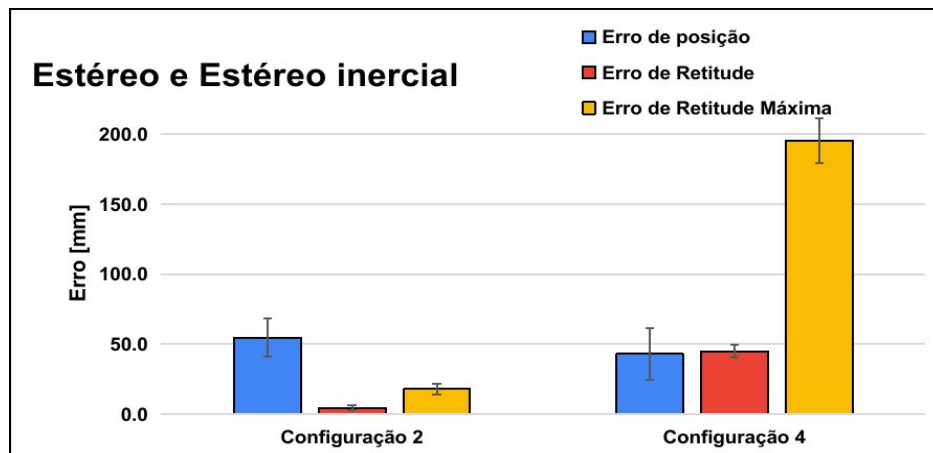


Fonte: Do autor.

A comparação central do estudo é ilustrada na Figura 48, que confronta o desempenho da melhor configuração estéreo (Configuração 2) com a configuração estéreo-inercial (Configuração 4). Os dados demonstram que a configuração estéreo-inercial alcança uma precisão de posicionamento, com um erro médio inferior (43,1 mm contra 54,8 mm). Em contrapartida, sua precisão é menor, evidenciada por um desvio padrão mais elevado (29,3 mm contra 18,4 mm).

A discrepância mais notável, entretanto, ocorre nas métricas de retitude. A configuração estéreo-inercial apresenta resultados drasticamente desfavoráveis, com um desvio máximo médio da reta de 195,6 mm. Uma hipótese para este comportamento reside na alta sensibilidade da IMU a perturbações não captadas pelo sistema visual. Supõe-se que as vibrações de alta frequência da guia linear durante o movimento, embora pequenas, introduziram acelerações espúrias que foram integradas pelo sensor inercial, corrompendo a estimativa de retitude da trajetória. O sistema puramente estéreo, por ser menos suscetível a essas vibrações, manteve uma trajetória mais retilínea, ainda que com menor precisão no deslocamento final.

Figura 48 – Gráfico comparativo de desempenho entre a Configuração 2 e a Configuração 4.

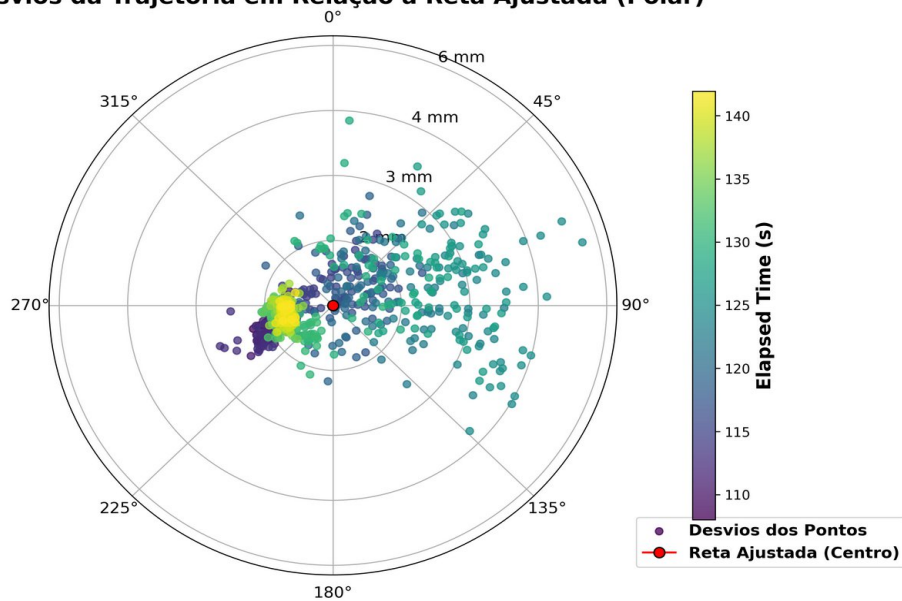


Fonte: Do autor.

A magnitude desta discrepância na retitude é visualizada de forma ainda mais clara na Figura 49 e Figura 50, que apresenta um gráfico polar dos desvios da trajetória para uma amostra de teste de cada configuração. No caso da Configuração 2 (Estéreo), observa-se que a grande maioria dos pontos da trajetória se concentra em um raio de até 6 mm da reta ideal. Em contraste, a Configuração 4 (Estéreo-Inercial) exibe uma dispersão muito maior, com múltiplos pontos se desviando a mais de 120 mm e alguns chegando a quase 160 mm da reta, confirmando o impacto das perturbações no sensor inercial.

Figura 49 – Gráfico polar dos desvios da trajetória para uma amostra da Configuração 2.

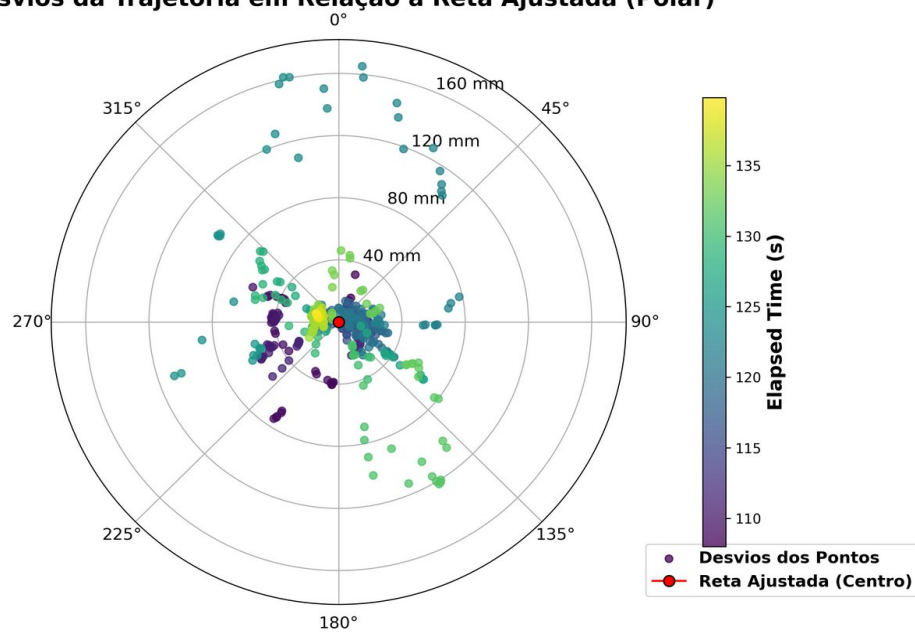
Desvios da Trajetória em Relação à Reta Ajustada (Polar)



Fonte: Do autor.

Figura 50 – Gráfico polar dos desvios da trajetória para uma amostra da Configuração 4.

Desvios da Trajetória em Relação à Reta Ajustada (Polar)



Fonte: Do autor.

Esta análise revela desempenhos com características complementares no ambiente controlado. A fusão inercial (Configuração 4) melhora a estimativa de deslocamento final, mas se mostra extremamente sensível às vibrações da guia, o que degrada severamente a retitude da trajetória. Em contrapartida, a configuração puramente estéreo otimizada (Configuração 2) oferece um resultado mais equilibrado e est a essas perturbações específicas. Apesar desta sensibilidade, a superioridade da Configuração 4 na estimativa de posição é um fator decisivo. Considerando que os próximos experimentos não envolvem a vibração de alta frequência característica da guia linear, espera-se que este comportamento de instabilidade na retitude não se repita. Desta forma, a Configuração 4 (Estéreo-Inercial) é adotada como a principal abordagem para as avaliações em cenários mais complexos que serão apresentados a seguir, onde se espera que os benefícios da fusão inercial se manifestem de forma mais completa. Os dados detalhados de cada um dos testes realizados, bem como as informações para acesso aos *datasets* (*rosbags*) correspondentes, estão disponíveis no APÊNDICE B.

5.2.2 Fotogrametria como “ground truth” para avaliação do ORB-SLAM3

Antes de utilizar a trajetória gerada por fotogrametria como “referência” (*ground-truth*) para os experimentos em cenários complexos, é fundamental validar a incerteza de medição e a qualidade desta técnica. Os dados do experimento da guia linear de 3 metros (apresentado na Seção 5.2.1) fornecem a oportunidade ideal para esta validação, pois contam com uma referência de exatidão: o laser interferômetro Renishaw XL-80.

Nesta análise, a trajetória estimada pelo ORB-SLAM 3 (na Configuração 2, para fins de comparação) e a trajetória gerada pela fotogrametria com o *software* COLMAP(Figura 51), são comparadas diretamente com a medição do laser. A Tabela 9 sumariza os resultados, comparando o deslocamento total, o erro percentual e as métricas de retitude para ambas as abordagens.

Figura 51 – Nuvem de pontos da fotogrametria e posição das imagens.



Fonte: Do autor

Tabela 9 – Comparativo quantitativo do deslocamento linear de 3 metros.

	Referência	SLAM	Fotogrametria
Distância percorrida (mm)	3000,0	2902,7	2950,5
Erro de Posição (%)	0,0	-3,2	-1,7
Desvio Médio de Retitude (mm)	0,0	6,1	0,5
Desvio Máximo de Retitude (mm)	0,0	45,5	1,5
Desvio Padrão de Retitude (mm)	0,0	5,6	0,3

Fonte: Do autor

Conforme apresentado na tabela, o laser apresenta o deslocamento de 3000,0 mm. A trajetória do ORB-SLAM 3 apresentou um erro de -3,2%, enquanto o método de fotogrametria registrou este deslocamento com um erro significativamente menor, de apenas -1,65%.

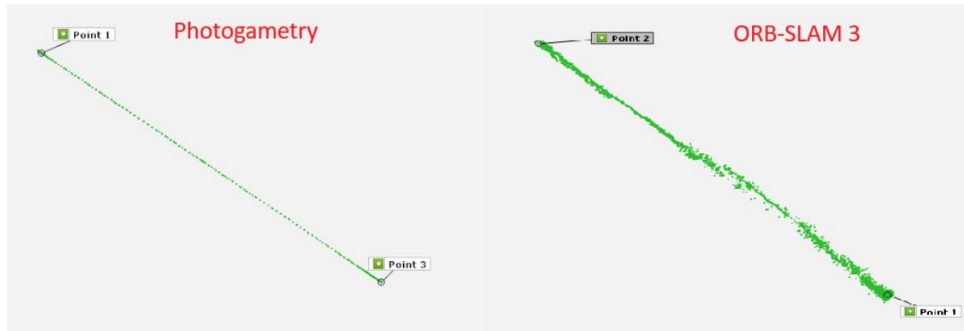
Mais importante, a tabela quantifica a retitude das trajetórias. Os resultados da fotogrametria apresentam qualidade geométrica superior, com um desvio médio da reta de apenas 0,5 mm e desvio máximo de 1,5 mm. Em contraste, a trajetória em tempo real do SLAM, otimizada para velocidade, exibiu um desvio médio de 6,1 mm e um máximo de 45,5 mm.

Essa diferença quantitativa é também claramente visível na Figura 52. A trajetória da fotogrametria, embora mais esparsa devido à taxa de captura de 4 Hz ,

forma uma linha limpa e reta. A trajetória do ORB-SLAM 3 é muito mais densa (60 Hz), mas exibe um ruído local considerável e um claro desvio do caminho linear.

Figura 52 – Comparativo visual das trajetórias no experimento da guia linear.

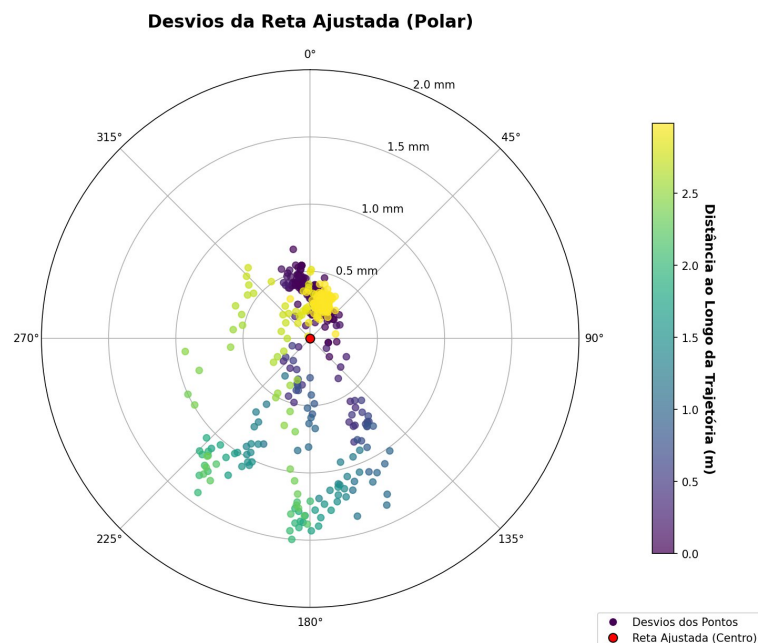
Esquerda: Trajetória da Fotogrametria. Direita: Trajetória do ORB-SLAM 3.



Fonte: Do autor

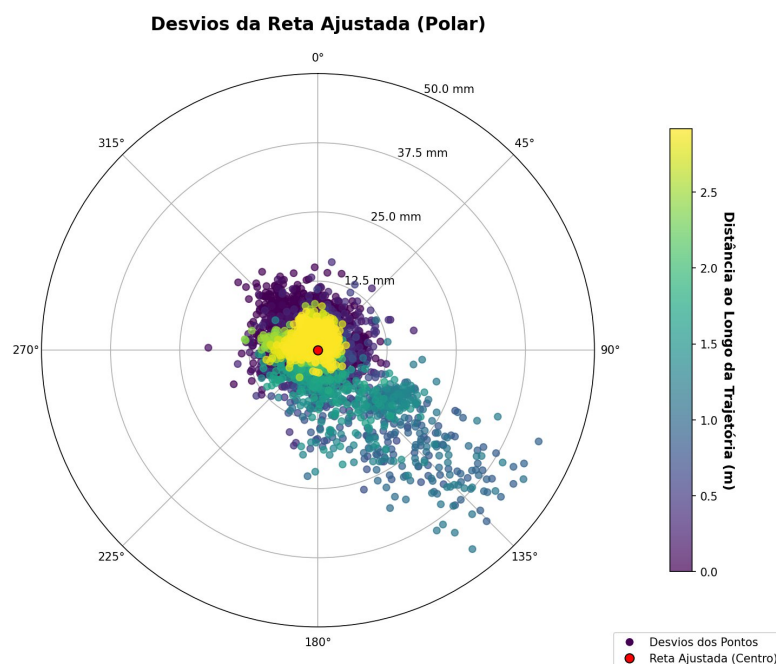
Esta análise é complementada pelos gráficos de desvio polar. A Figura 53 mostra os desvios da trajetória de fotogrametria, confirmando visualmente que todos os pontos estão contidos dentro do raio de 1,5 mm de desvio máximo. Em contrapartida, a Figura 54 mostra a ampla dispersão dos pontos do SLAM, que reflete seu desvio máximo de 45,5 mm. Note que as escalas dessas figuras são diferentes.

Figura 53 – Gráfico polar de desvio da retitude fotogrametria.



Fonte: Do autor

Figura 54 – Gráfico polar de desvio da retitude ORB-SLAM3.



Fonte: Do autor

Os resultados desta seção comprovam que a trajetória gerada pela fotogrametria (utilizando COLMAP e *Bundle Adjustment*) apresenta uma fidelidade métrica e consistência geométrica superior à estimativa em tempo real do SLAM. Embora seja reconhecida uma limitação na medição do deslocamento (devido à identificação de pontos de início/fim e janelas de tempo), sua baixa incerteza de medição confirma a adequação desta trajetória como “referência”. A comparação de desempenho favorável com o laser é restrita ao contexto específico do teste experimental e não à tecnologia laser interferométrica em geral. Desta forma, a trajetória fotogramétrica está validada para a avaliação do ORB-SLAM 3 em cenários mais complexos mesmo sendo uma técnica que geram uma trajetória menos densa.

5.2.3 Erro de Trajetória Absoluto (APE) e Relativo (RPE)

Com a metodologia de *ground-truth* fotogramétrico validada na seção anterior (5.2.2), é possível agora utilizá-la para uma avaliação quantitativa do desempenho do ORB-SLAM 3. Esta análise é dividida em duas partes: primeiro, uma avaliação

APE/RPE do desempenho do SLAM no próprio experimento da guia linear (Exp. 1) e, em seguida, a aplicação no cenário mais complexo do corredor (Exp. 2).

Para completar a avaliação, foi realizada análise APE/RPE no experimento da Guia Linear (Exp. 1) uma comparação quantitativa final utilizando a ferramenta EVO. Desta vez, a trajetória fotogramétrica validada foi tratada como o *ground-truth* e a trajetória do ORB-SLAM 3 (do mesmo Experimento 1) foi tratada como a estimativa. Os resultados para o Erro de Pose Absoluto (APE) e o Erro de Pose Relativo (RPE), incluindo suas estatísticas de média e Raiz do Erro Quadrático Médio (RMSE), estão sumarizados na Tabela 10.

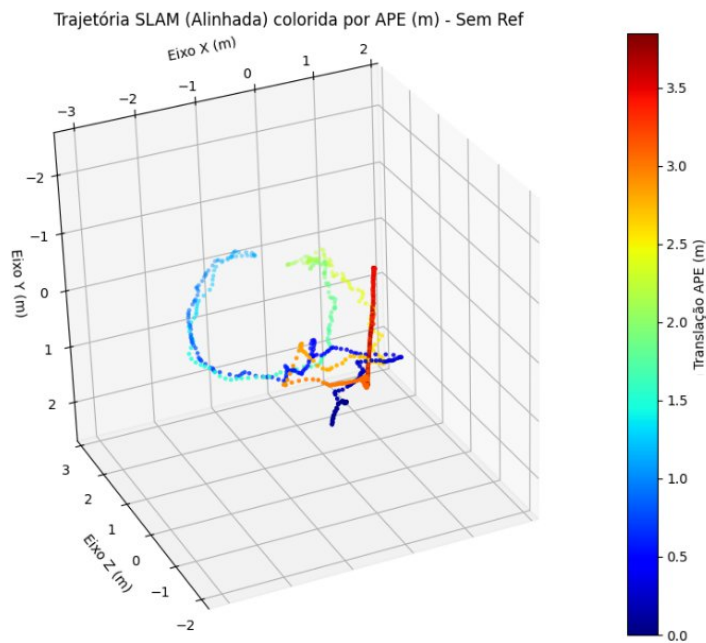
Tabela 10 – Análise APE/RPE do ORB-SLAM 3 usando Ground-Truth Fotogramétrico (Exp. 1).

Métrica	RMSE (m)	Média (m)	Máx (m)
APE	1,4050	1,0788	3,8231
RPE	0,0694	0,0531	0,2659

Fonte: Do autor

A trajetória do SLAM exibiu um Erro de Pose Absoluto (APE) significativo, com um RMSE de 1,405 m e um erro máximo de 3,823 m. Nesta Figura 55, as linhas em 3D representam a trajetória total estimada pelo sensor, que inclui o percurso de 3 metros (ida e volta) realizado na Guia Linear e as movimentações adicionais para inicialização e transição. Este APE elevado é resultado da acumulação de erro (*drift*) inerente ao algoritmo em um percurso complexo, mas é também influenciado pela baixa densidade de pontos na trajetória de referência (*“Ground-Truth”* Fotogramétrico), o que impõe um desafio no processo de alinhamento e comparação. O mapa de cores ilustra esta dinâmica: o erro (azul) é baixo no início da trajetória e cresce progressivamente, atingindo o máximo (vermelho) ao final do percurso.

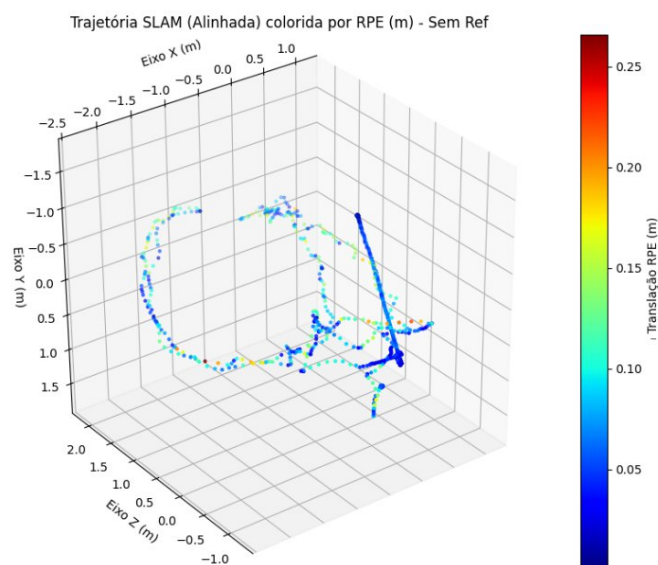
Figura 55 – Trajetória do SLAM (Exp. 1) colorida pelo Erro de Pose Absoluto (APE) em metros.



Fonte: Do autor.

A análise do Erro de Pose Relativo (RPE), mostrada na Figura 56, revela a origem desse *drift*. O RPE, que mede o erro local, é visivelmente maior (pontos em amarelo/vermelho) no exato início e fim da trajetória. Isso corresponde às pausas estacionárias de 5 segundos do procedimento experimental (Seção 4.3.5), um cenário desafiador onde o sistema SLAM acumulou erro de *drift* mesmo estando parado. A capacidade da metodologia em capturar e visualizar esse comportamento específico do SLAM demonstra sua alta sensibilidade e eficácia como ferramenta de avaliação.

Figura 56 – Trajetória do SLAM (Exp. 1) colorida pelo Erro de Pose Relativo (RPE) em metros.



Fonte: Do autor.

Apesar do elevado Erro Absoluto (APE), a análise do Erro Relativo (RPE) demonstrou que o sistema mantém uma consistência local considerável, mesmo em um percurso complexo. O valor baixo de RMSE do RPE, de 0,0694 m, confirma que a taxa de acúmulo de erro é controlada durante o movimento.

Tendo estabelecido a baixa incerteza de medição e a qualidade do “*ground-truth*” fotogramétrico na seção anterior, esta metodologia foi então aplicada para avaliar a trajetória do ORB-SLAM 3 no modo estéreo-inercial (Configuração 4) durante o experimento do corredor através de uma análise APE/RPE no experimento do corredor (Exp. 2). Os resultados quantitativos para o “*Loop Principal*” maior que 50 metros estão apresentados na Tabela 11, resumizando as métricas de erro chave computadas pelo EVO.

Tabela 11 – Análise APE/RPE do ORB-SLAM 3 usando *Ground-Truth* Fotogramétrico (Exp. 2).

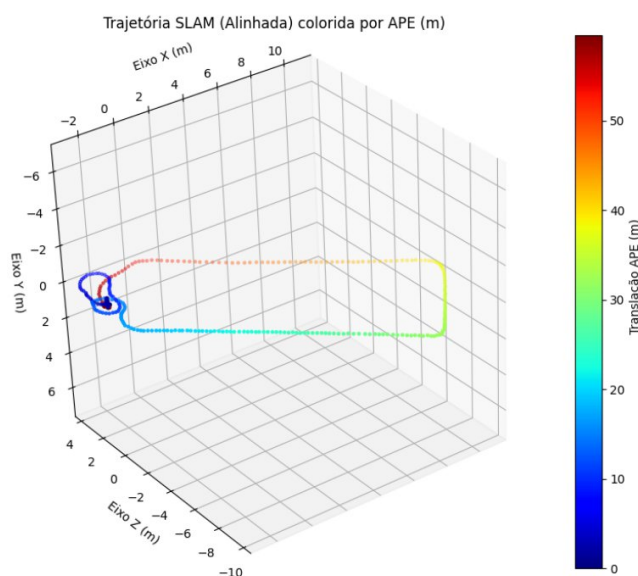
Métrica	RMSE (m)	Média (m)	Máx (m)
APE	9,9907	9,5761	15,2803
RPE	0,1362	0,1153	0,3242

Fonte: Do autor

Os resultados para o "Loop Principal", uma trajetória de mais de 50 metros, mostram um Erro de Pose Absoluto (APE) final de 9,991 m (RMSE). Essa métrica quantifica o *drift* global total acumulado ao longo de todo o percurso. Este valor elevado deve ser interpretado levando em consideração a limitação da trajetória de referência, uma vez que a baixa densidade de pontos da "referência" fotogramétrico pode influenciar o alinhamento global e contribuir para a magnitude final do APE. O Erro de Pose Relativo (RPE), que mede o *drift* translacional local (quadro a quadro), foi de 0,136 m (RMSE).

Além das métricas quantitativas, a análise qualitativa dos gráficos de trajetória fornece uma introspecção adicional sobre o comportamento do sistema SLAM. A Figura 57 visualiza o Erro de Pose Absoluto (APE) ao longo da trajetória do "Loop Principal". O gradiente de cor, de azul (erro baixo) para vermelho (erro alto), mostra claramente como o *drift* global se acumula ao longo do caminho. O erro é mais baixo no início do percurso e aumenta progressivamente, atingindo seu valor máximo quando o operador retorna ao ponto de partida, o que se alinha com os dados quantitativos da Tabela 11.

Figura 57 – Visualização do Erro de Pose Absoluto (APE) para a trajetória do Exp. 2.

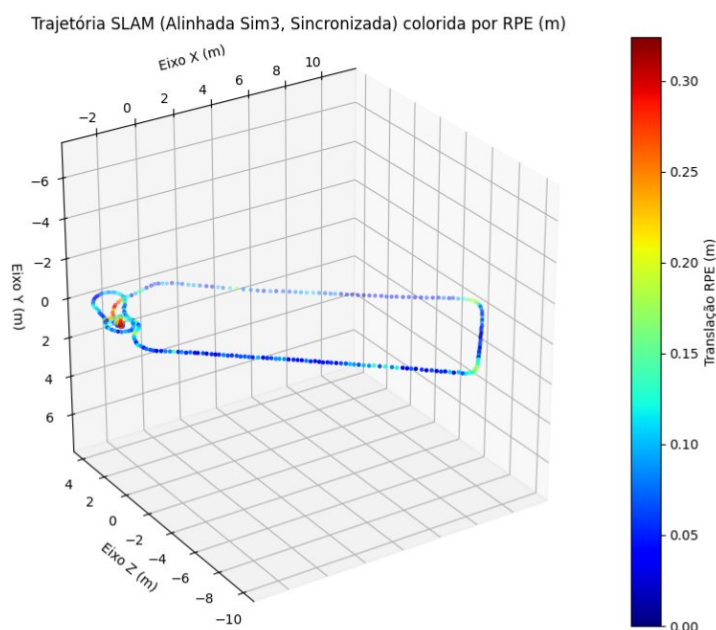


Fonte: Do autor.

A Figura 58 visualiza o Erro de Pose Relativo (RPE), que mede o drift local. Este gráfico mostra que a maior parte da trajetória possui um erro local muito baixo

(azul/verde). Contudo, ele também destaca áreas específicas, particularmente durante rotações bruscas, onde o RPE tem picos (pontos vermelhos, máx. 0,32 m). Isso demonstra a capacidade da metodologia não apenas de fornecer um número de erro global único, mas também de identificar momentos específicos de alto *drift* translacional durante o experimento.

Figura 58 – Visualização do Erro de Pose Relativo (RPE) para a trajetória do Exp. 2.



Fonte: Do autor.

5.2.4 Comparativo de Condições entre os Cenários de Teste

Após a análise dos resultados dos experimentos da guia linear (Seção 5.2.1) e do corredor (Seções 5.2.2 e 5.2.3), esta seção realiza um comparativo direto entre o desempenho do sistema no ambiente controlado do laboratório e no cenário de aplicação realista dos corredores.

O experimento da guia linear, realizado em laboratório, representou um cenário controlado para a avaliação do SLAM. Este ambiente possui características favoráveis, como iluminação artificial constante e difusa, ausência de janelas ou sombras abruptas, e um ambiente totalmente estático, sem o movimento de pessoas ou objetos. Sob essas condições "perfeitas", a principal fonte de erro é a própria incerteza do algoritmo e dos sensores.

Em contrapartida, o experimento nos corredores do *loop* principal (Exp. 2) introduziu um conjunto significativo de desafios do mundo real, transformando-o em um ambiente dinâmico e visualmente complexo. As principais dificuldades encontradas neste cenário foram:

- A) Variações de Iluminação: Ao contrário do laboratório, o corredor possui longos trechos com iluminação artificial direta, intercalados com áreas de sombra e a presença de janelas que causam grandes mudanças na luminosidade. Isso exigiu o uso do ajuste automático do tempo de exposição e ganho das câmeras. Embora necessária, essa funcionalidade pode introduzir flutuações rápidas de brilho que dificultam o rastreamento consistente de *features* (pontos de interesse) pelo algoritmo.
- B) Elementos Dinâmicos: Por ser uma área de circulação, o ambiente não era estático. A passagem de pessoas durante a captura é particularmente problemática para algoritmos de SLAM, que, em sua maioria, operam sob a premissa de um "mundo estático". Pontos de interesse detectados em objetos móveis (como uma pessoa caminhando) são, na verdade, *outliers* geométricos e, se não forem filtrados corretamente, podem corromper a estimação da pose e do mapa, levando a um aumento no *drift* da trajetória.

Portanto, enquanto o experimento da guia foi fundamental para a otimização de parâmetros e validação metrológica (provando, por exemplo, a sensibilidade da IMU a vibrações), a avaliação no corredor foi o teste da estabilidade do sistema. Ela validou a capacidade da Configuração 4 (Estéreo-Inercial) de operar e gerar uma trajetória consistente mesmo fora de condições laboratoriais ideais, enfrentando os desafios de iluminação e de ambientes dinâmicos de uma aplicação real.

5.3 AVALIAÇÃO DO MAPEAMENTO

Esta seção aborda a avaliação do mapa de pontos gerado pelo ORB-SLAM 3. O objetivo é analisar a qualidade e a fidelidade métrica da reconstrução 3D do ambiente, que é uma saída importante do algoritmo. A avaliação será conduzida em duas frentes: inicialmente, através de uma Análise Visual da Nuvem de Pontos, onde

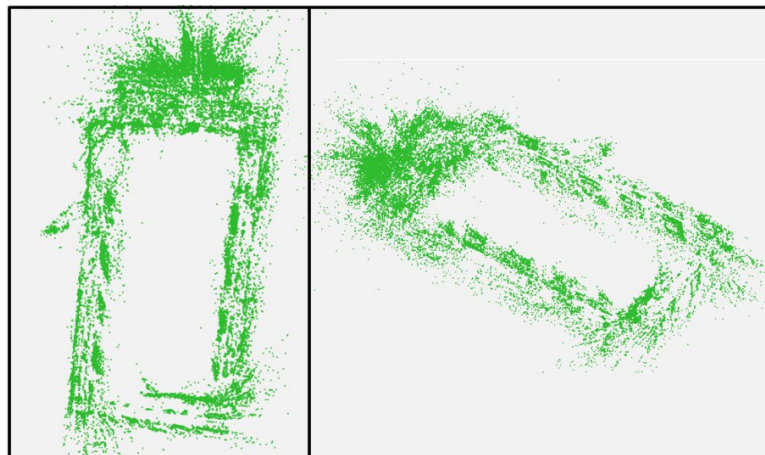
a coerência geométrica do mapa é inspecionada de forma descritiva, verificando se objetos com formas conhecidas (como caixas ou tubos) são representados corretamente. Posteriormente, será introduzida a metodologia para uma validação quantitativa, que utilizará como referência uma nuvem de pontos de baixa incerteza de medição obtida pelo sistema de digitalização industrial GOM ATOS.

5.3.1 Análise Visual das Nuvens de Pontos

A primeira etapa da avaliação do mapeamento é uma análise visual e qualitativa. O objetivo é verificar se a nuvem de pontos esparsa gerada em tempo real pelo ORB-SLAM 3 é geometricamente coerente e se representa de forma fidedigna as estruturas do mundo real. Para isso, foram analisados os mapas de três cenários distintos, comparando a forma da nuvem de pontos resultante com a geometria conhecida do objeto ou ambiente. Para avaliar isso serão analisados 3 casos:

Caso 1: Reconstrução dos Corredores é o mapa gerado durante o "*Loop Principal*" do Experimento 2 (descrito na Seção 5.2.3). A Figura 59 apresenta duas perspectivas da nuvem de pontos esparsa resultante. Observa-se que o mapa forma um contorno retangular nítido. Esta geometria é visualmente condizente com a planta baixa do corredor do edifício, que foi apresentada anteriormente na Figura 27. É importante ressaltar, contudo, que esta nuvem de pontos apresenta ruído e pontos espúrios inerentes ao mapeamento em tempo real, necessitando de um processo posterior de limpeza e filtragem. Apesar do *drift* quantificado na odometria, o sistema foi capaz de manter a coerência estrutural global do ambiente e fechar o *loop* corretamente.

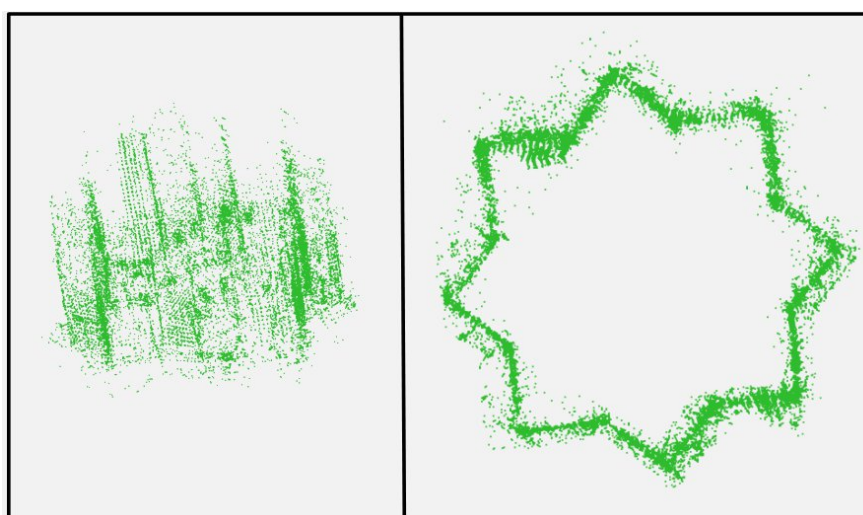
Figura 59 – Vistas da nuvem de pontos esparsa do corredor (Exp. 2).



Fonte: Do autor.

Caso 2: Reconstrução da base da coluna que sustenta Caixa d'Água foi a reconstrução de uma estrutura externa de geometria complexa: a caixa d'água da UFSC. A Figura 60 apresenta a nuvem de pontos resultante (vista superior e isométrica). A correspondência visual entre o mapa gerado, com seu formato estrelado característico, e a planta baixa é evidente. Este resultado valida a capacidade do sistema em mapear geometrias mais complexas em larga escala.

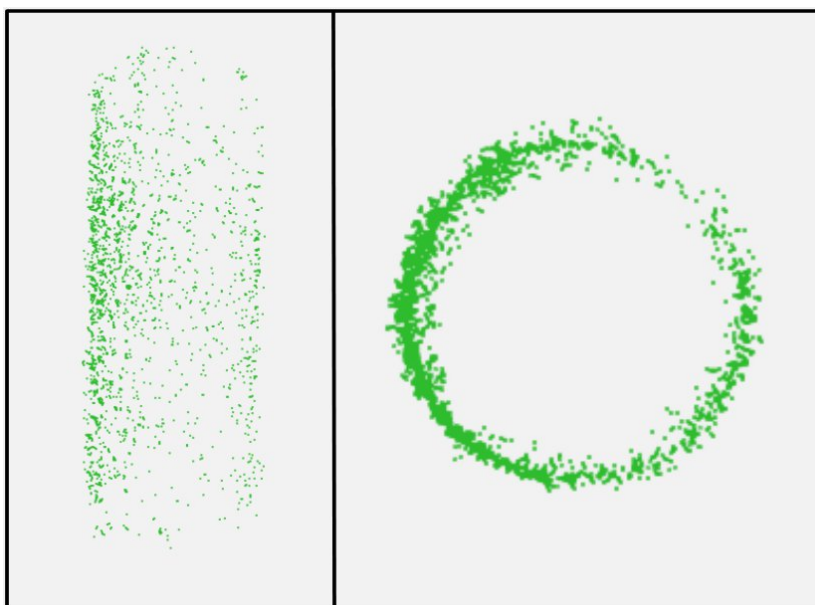
Figura 60 – Mapa da base da coluna que sustenta caixa d'água. Esquerda: Vista isométrica Nuvem de pontos. Direita: Vista superior da Nuvem de pontos.



Fonte: Do autor.

Caso 3: Reconstrução do Tubo cilíndrico. Foi avaliada a reconstrução de um corpo de prova de formato conhecido: um tubo cilíndrico. A Figura 61 apresenta duas vistas ortogonais da nuvem de pontos deste objeto. A vista superior (direita) revela um formato circular claro, enquanto a vista lateral (esquerda) mostra o perfil retilíneo e alongado do tubo. A combinação dessas duas perspectivas é condizente com a geometria de um cilindro, demonstrando a fidelidade do mapeamento em nível de objeto.

Figura 61 – Vistas da nuvem de pontos do tubo. Esquerda: Vista lateral. Direita: Vista superior.



Fonte: Do autor.

Em todos os três casos, a análise visual demonstrou que a nuvem de pontos esparsa do ORB-SLAM 3, apesar de não ser densa e ter elevado ruído, captura com sucesso a geometria estrutural e as formas dos ambientes e objetos. Esta validação qualitativa é um passo importante antes da análise quantitativa com o sistema GOM ATOS, que será detalhada na próxima seção.

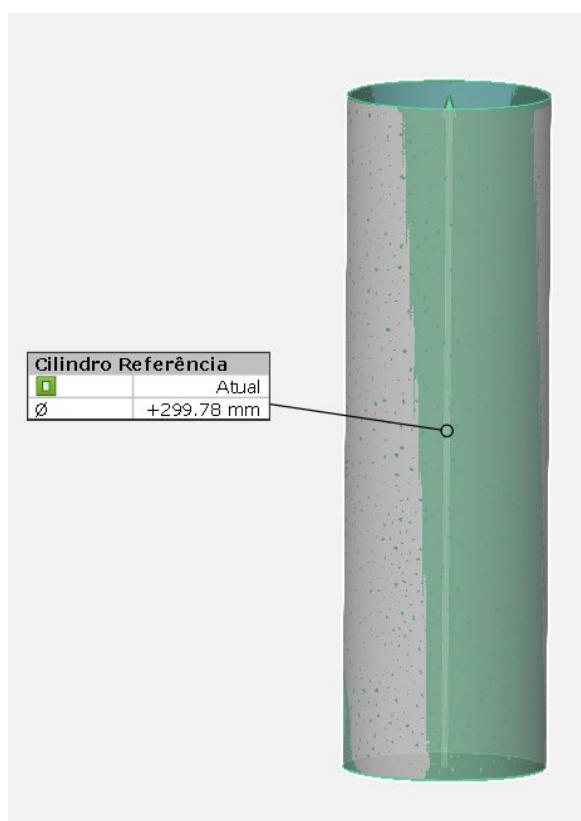
5.3.2 Análise dos Mapas de Pontos e Reconstrução 3D dos Tubos

Após a análise qualitativa da geometria das nuvens de pontos, esta seção apresenta uma avaliação quantitativa da fidelidade métrica do mapeamento. O foco

é a reconstrução 3D do tubo cilíndrico, que permite uma comparação direta de suas dimensões estimadas contra uma referência metrológica de baixa incerteza de medição.

Para estabelecer este *ground-truth*, o tubo foi digitalizado utilizando o sistema de luz estruturada GOM ATOS, conforme descrito na metodologia (Capítulo 4). Este sistema fornece uma nuvem de pontos densa e com baixa incerteza de medição, cujo diâmetro foi utilizado como o valor de referência para esta análise. A Figura 62 exibe a digitalização de referência do tubo obtida com o GOM ATOS.

Figura 62 – Nuvem de pontos de referência do tubo obtida com o digitalizador GOM ATOS.



Fonte: Do autor.

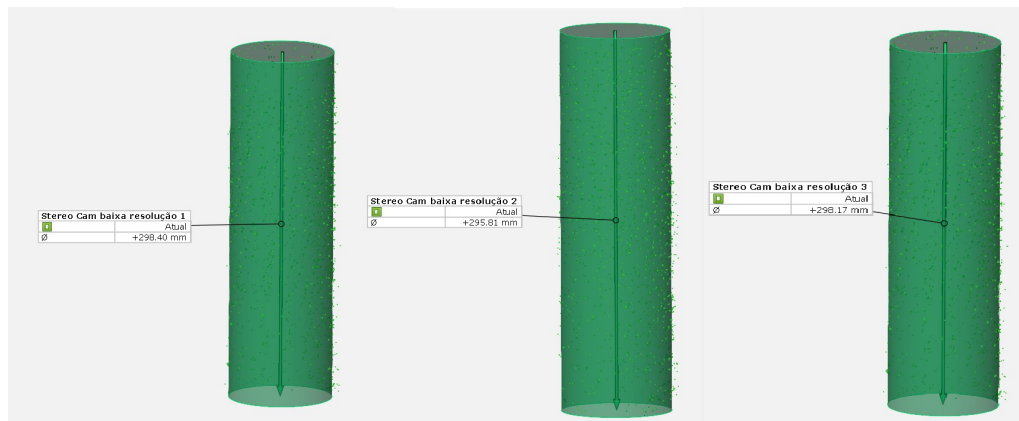
Para a avaliação do ORB-SLAM 3, o tubo foi mapeado em quatro configurações distintas, com três repetições para cada, totalizando 12 nuvens de pontos para análise. As configurações avaliadas foram:

1. Sistema estéreo com câmera de baixa resolução (Figura 63);
2. Sistema estéreo inercial com câmera de baixa resolução (Figura 64);

3. Sistema estéreo com câmera de alta resolução (Figura 65);
4. Sistema estéreo inercial com câmera de alta resolução (Figura 66).

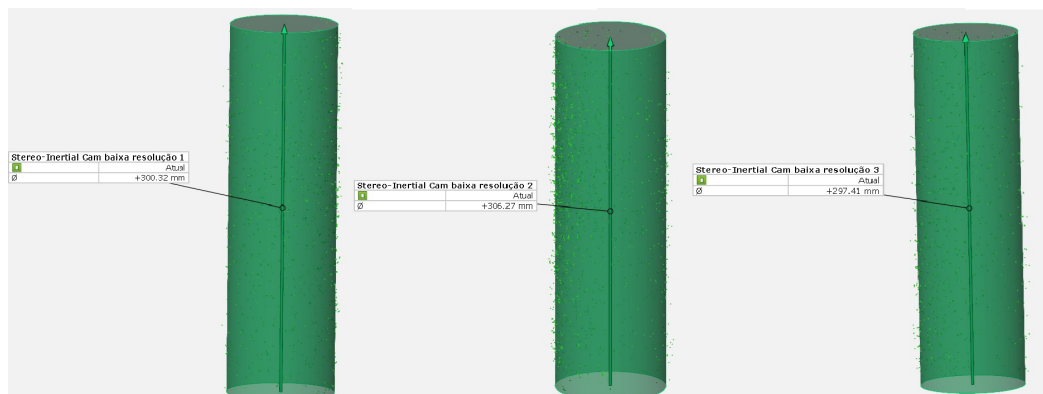
Embora tenham sido usadas câmeras coloridas, o ORB-SLAM 3 foi configurado para utilizar imagens monocromáticas. Esta decisão garante que a comparação de desempenho entre os sistemas seja justa e focada no impacto da fusão inercial e nas diferenças de hardware do sensor, e não na informação de cor.

Figura 63 – Nuvens de pontos: Sistema estéreo com câmera de baixa resolução.



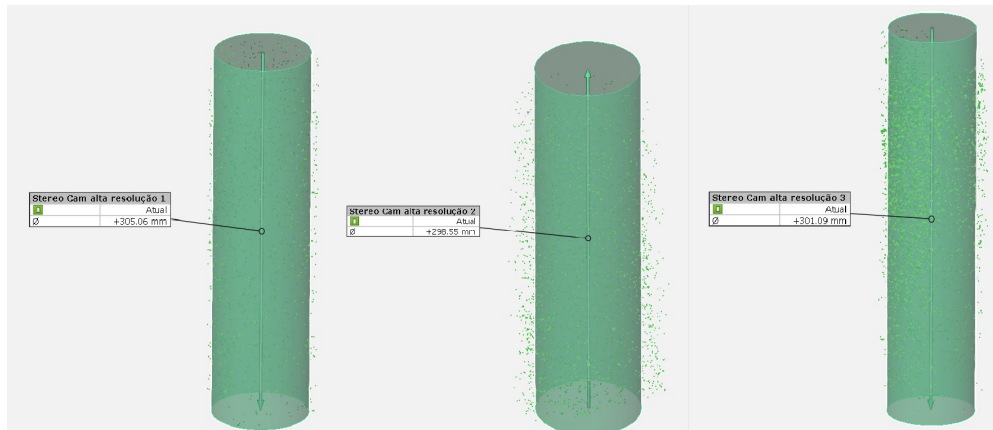
Fonte: Do autor.

Figura 64 – Nuvens de pontos: Sistema estéreo inercial com câmera de baixa resolução.



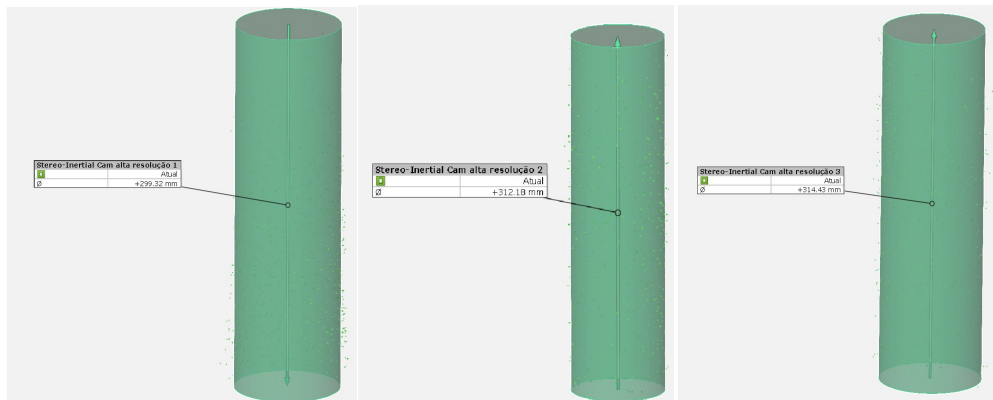
Fonte: Do autor.

Figura 65 – Nuvens de pontos: Sistema estéreo com câmera de alta resolução.



Fonte: Do autor.

Figura 66 – Nuvens de pontos: Sistema estéreo inercial com câmera de alta resolução.



Fonte: Do autor.

A Tabela 12 compila os resultados quantitativos, apresentando a média e o desvio padrão do diâmetro medida nas três repetições de cada configuração. O erro absoluto (em mm) e o erro percentual (%) em relação à referência GOM também são apresentados. As Figuras 63 a 66 oferecem uma visualização de todas as 12 nuvens de pontos geradas pelo ORB-SLAM 3.

Tabela 12 – Resultados quantitativos da medição do diâmetro tubo.

	Estéreo Cam. baixa resolução	Estéreo Inercial Cam. baixa resolução	Estéreo Cam. alta resolução	Estéreo Inercial Cam. alta resolução
Experimento 1(mm)	298,4	300,32	305,06	299,32
Experimento 2(mm)	295,81	306,27	298,55	312,18
Experimento 3(mm)	298,17	297,41	301,09	314,43
Média (mm)	297,46	301,33	301,57	308,64
Desvio padrão(mm)	1,43	4,52	3,28	8,15
Tendência (mm)	-2,32	1,55	1,79	8,86
Tendência (%)	-0,77	0,52	0,60	4,34

Fonte: Do autor.

A Tabela 12 compila os resultados quantitativos, apresentando a média e o desvio padrão do diâmetro medido nas três repetições de cada configuração. É importante salientar que estes resultados numéricos foram obtidos de nuvens de pontos esparsas que passaram por um processo de filtragem e limpeza (como a remoção de *outliers*), sendo, portanto, uma análise do mapa pós-processado. O erro absoluto (em mm) e o erro percentual (%) em relação à referência GOM também são apresentados. As Figuras 63 a 66 oferecem uma visualização de todas as 12 nuvens de pontos geradas pelo ORB-SLAM 3.

Em suma, embora as nuvens de pontos esparsas geradas pelo ORB-SLAM 3 não possuam a fidelidade métrica exigida para medições finas, elas oferecem consistência estrutural suficiente para a avaliação visual da forma dos objetos. Tais mapas são, portanto, viáveis para aplicações como inspeção de integridade estrutural e identificação de defeitos ou deformidades “óbvias”.

5.4 ANÁLISE DE DESEMPENHO COMPUTACIONAL

Esta seção relata as observações sobre o desempenho computacional e a estabilidade operacional do ORB-SLAM 3, obtidas durante a execução dos testes no hardware especificado no Capítulo 4. A análise foca nas limitações práticas encontradas e no comportamento do sistema sob diferentes condições de operação.

Em termos de carga de processamento, observou-se que a configuração Estéreo-Inercial (Configuração 4) apresentou um desempenho computacional geral

equivalente ao da configuração puramente Estéreo (Configuração 2). A fusão dos dados da IMU não representou um aumento de custo computacional que inviabilizasse a execução em tempo real. No entanto, foi notada uma leve queda de desempenho em ambas as configurações após períodos prolongados de uso contínuo.

A análise revelou duas limitações de software claras no ORB-SLAM 3 que impactam diretamente sua estabilidade e viabilidade:

Resolução da Imagem: O algoritmo mostrou-se altamente sensível a imagens de resolução muito alta. Em testes com resoluções a partir de 1600x1200 *pixels*, o sistema apresentou instabilidade, travamentos frequentes, dificultando a conclusão dos experimentos.

Número de *Features* ORB: De forma similar, a definição de um número muito elevado de *features* ORB (pontos de interesse) a serem detectados, como valores acima de 5000, também levou à sobrecarga do sistema e consequente instabilidade.

Por fim, foi identificada uma limitação operacional crítica relacionada ao hardware de testes. A execução do ORB-SLAM 3 no notebook especificado demonstrou um alto consumo de energia, resultando em um rápido esgotamento da bateria. Foi observado de forma consistente que, caso o notebook entrasse em modo de economia de energia ou atingisse um nível de bateria fraca, o sistema operacional reduzia o desempenho da CPU. Isso causava uma queda drástica na performance do SLAM, levando à perda de quadros e, invariavelmente, à falha no rastreamento da trajetória.

6 DISCUSSÃO

O presente capítulo dedica-se à análise crítica dos resultados obtidos com o sistema de visão estéreo inercial desenvolvido, integrando as câmeras industriais FLIR e a unidade de medição inercial (IMU) STIM300. O foco central recai sobre a avaliação da exatidão do algoritmo ORB-SLAM3, tanto na estimativa da trajetória quanto na reconstrução da nuvem de pontos. A discussão busca não apenas validar os dados quantitativos apresentados anteriormente, mas também compreender o comportamento do sistema sob a ótica da fusão sensorial e suas limitações intrínsecas.

6.1 ANÁLISE DOS RESULTADOS: LOCALIZAÇÃO E MAPEAMENTO

A interpretação dos resultados de localização e mapeamento é realizada à luz do rigor metrológico estabelecido pelas etapas de calibração (Seção 5.1), que resultaram em um erro médio de reprojeção de no máximo 0,19 *pixels*, um valor que atesta a alta fidelidade do modelo de câmera. A qualidade da calibração estabelece a base para que os desvios observados na trajetória e no mapa sejam atribuídos primariamente a fatores operacionais e ao comportamento do algoritmo SLAM.

6.1.1 Análise da exatidão de Localização (Trajetória)

A exatidão da trajetória foi avaliada em dois cenários distintos — o ambiente controlado (guia linear) e o ambiente complexo —, ambos confirmando a factibilidade do sistema para a estimativa de pose.

Desempenho em Cenário Controlado (Guia Linear): No experimento da guia linear (Exp. 1), a Configuração 4 (Estéreo-Inercial) alcançou o menor erro médio de deslocamento final (43,1 mm em 3 metros), validando a superioridade da fusão visual-inercial na estimativa de posição.

Contudo, esta mesma configuração exibiu uma perda drástica na retitude da trajetória, com um desvio máximo médio de 195,6 mm. Esta anomalia é justificada pela alta sensibilidade da IMU STIM300 às vibrações de alta frequência inerentes ao movimento do sistema sobre a guia linear. A IMU, ao integrar essas acelerações

espúrias, introduziu um ruído significativo na estimativa de pose, corrompendo a geometria retilínea, embora o algoritmo de otimização *tightly-coupled* tenha conseguido compensar esse erro para a posição final.

Desempenho em Cenário Complexo (Corredor): No cenário mais desafiador do corredor (Exp. 2), o sistema demonstrou capacidade de operar e gerar uma trajetória consistente fora de condições laboratoriais ideais, lidando com elementos dinâmicos e variações de iluminação.

O Erro de Pose Absoluto (APE) com RMSE de 9,991 m em um *loop* de 50 metros quantifica o *drift* global acumulado. A acumulação progressiva de erro, evidenciada pelo gradiente de cor (azul para vermelho), é um comportamento esperado em sistemas SLAM que enfrentam desafios visuais como o ajuste automático de exposição das câmeras FLIR e o rastreamento de elementos dinâmicos (pessoas) e principalmente o uso da fotogrametria com uma trajetória de baixa densidade para avaliação do APE. Tais elementos violam a premissa de um mundo estático, introduzindo *outliers* geométricos que corrompem a estimativa da pose.

O Erro de Pose Relativo (RPE) de 0,136 m (RMSE) demonstra que o desempenho local é aceitável, mas picos de erro (máx. 0,32 m) durante rotações bruscas indicam que a estabilidade operacional local do sistema é momentaneamente desafiada.

Em suma, a trajetória gerada pelo ORB-SLAM3, mesmo no ambiente complexo, se mostrou factível e consistente com os percursos movimentados. O sistema demonstrou uma incerteza de medição global (APE) de 9,9907 m e uma incerteza de odometria local (RPE/RMSE) de 0,1362 m. Essa fidelidade métrica, aliada à estabilidade operacional do rastreamento, valida a utilização do ORB-SLAM3 para fornecer a posição do ROV em cenários de dificuldade similar e servir como referência para a fusão de pontos de outros sensores.

6.1.2 Análise da Precisão de Mapeamento (Nuvem de Pontos)

A avaliação da nuvem de pontos revelou uma distinção clara entre a qualidade estrutural e a incerteza de medição.

Fidelidade Estrutural: O mapeamento demonstrou capacidade de reconstruir a geometria estrutural dos três ambientes avaliados (corredores retangulares) e objetos

(cilindro e base caixa d'água de formato estrelado). O mapa esparso gerado pelo ORB-SLAM3 é visualmente coerente e permite identificar o que está sendo reconstruído, sendo capaz, inclusive, de detectar falhas ou deformidades visíveis.

Incerteza de medição: A avaliação quantitativa da medição do diâmetro do tubo cilíndrico (Tabela 12) mostrou que a incerteza de medição típica é de 4,5 mm, o que não é suficiente para aplicações típicas na área de engenharia.

Todas as nuvens de pontos foram geradas utilizando a mesma configuração de ORBextractor e a resolução de 800x600 *pixels*, garantindo uma comparação direta do impacto do hardware e da fusão.

A Configuração Estéreo-Inercial de alta resolução apresentou o pior desempenho (Erro: 8,86 mm, Desvio Padrão: 8,15 mm) , superando o erro das outras configurações (mais que o dobro) em mais de 5,7 vezes (Erro: 1,55 mm) em relação ao erro avaliado. Esta elevada variabilidade, evidenciada pelo alto Desvio Padrão, demonstra que a tentativa de extrair medições finas (milimétricas) de um mapa esparso e otimizado para a pose (localização) resulta em repetibilidade limitada e alta incerteza.

Conclui-se que, para os três experimentos realizados o ORB-SLAM3 cumpriu sua função primária de localização e construção esparsa do mapa inerentemente limitou sua aplicabilidade para inspeção metrológica, onde a reconstrução densa com boa exatidão é indispensável.

6.2 IMPACTO DA FUSÃO VISUAL-INERCIAL NO DESEMPENHO DO SLAM

A fusão de dados visuais estéreo das câmeras FLIR com os dados de alta frequência da IMU STIM300 foi o pilar do sistema e se manifestou em diferentes aspectos do desempenho, revelando tanto os benefícios esperados desta arquitetura *tightly-coupled* quanto as limitações oriundas do componente inercial.

6.2.1 Vantagens da Fusão Visual-Inercial

O principal impacto positivo da fusão reside no refinamento da estimativa de posição e escala. Conforme demonstrado no experimento da guia linear (Exp. 1), a

configuração Estéreo-Inercial apresentou a menor tendência de posição total (43,1 mm), superando o sistema puramente estéreo, que tem uma maior dificuldade em manter a escala em movimentos longos e não controlados. Este resultado é uma característica positiva da arquitetura *tightly-coupled* escolhida.

A literatura confirma que a integração restringe o acúmulo de erro (*drift*) translacional puro da odometria visual, o que é crucial para aplicações de inspeção: A abordagem de SLAM visual-inercial de acoplamento rígido (*tightly-coupled*) que utiliza estimativa de Máximo a Posteriori (MAP) resulta em um sistema que opera de forma estável e é duas a dez vezes mais fiel metricamente do que abordagens anteriores, particularmente na estimativa de trajetória e escala em ambientes internos e externos (12).

A contribuição da IMU STIM300, uma unidade de grau de navegação, foi particularmente importante na inicialização estável do sistema, fornecendo uma estimativa de escala métrica imediata e a orientação de gravidade, mitigando a ambiguidade inicial típica da visão pura. Esta estabilidade na escala é o que permite que a trajetória gerada pelo ORB-SLAM3 sirva como a referência para a fusão de outros sensores no contexto do ROV.

6.2.2 Limitações

Apesar dos ganhos em incerteza de posicionamento, o modo estéreo-inercial introduziu desafios operacionais e sensibilidades que devem ser considerados na aplicação prática:

Requisito de Inicialização (*Bootstrapping*): Para iniciar a estimação de pose no modo Estéreo-Inercial, o ORB-SLAM3 exige que o operador realize um pequeno movimento inicial (*bootstrapping*), que é um requisito fundamental para que o algoritmo determine os parâmetros de escala e orientação da IMU e construa o mapa inicial. Embora essa manobra seja vital para garantir a fidelidade métrica do sistema, ela representa uma dificuldade operacional para o operador, que precisa realizar um movimento controlado e não abrupto antes de iniciar a missão de inspeção de fato, diferentemente de um sistema puramente visual que pode iniciar imediatamente.

Suscetibilidade a Movimentos Abruptos e Vibrações: O sistema se demonstrou mais sensível a movimentações abruptas, colisões ou vibrações de alta frequência.

Esta sensibilidade se deve ao fato de o acelerômetro de alta sensibilidade da STIM300 captar instantaneamente picos de aceleração decorrentes de vibrações ou impactos.

Conforme justificado no experimento da guia linear (Seção 6.1.1), se houver um movimento brusco ou uma batida, a aceleração alta capturada pelo acelerômetro é integrada pelo modelo. Quando essa integração inercial não é devidamente restrita por features visuais válidas (por exemplo, em uma perda momentânea de *tracking* ou em cenas com pouca textura), ela pode fazer com que o sistema se desloque de forma incoerente na estimativa da posição, introduzindo um grande erro translacional que é difícil de corrigir e pode exigir a reinicialização do sistema.

6.3 ANÁLISE DAS LIMITAÇÕES E FONTES DE ERRO

Esta seção enumera e discute as fontes de incerteza do sistema de visão estéreo-inercial relacionadas ao *hardware* (câmeras FLIR e IMU STIM300) e ao *software* (ORB-SLAM3). O foco é estabelecer o patamar de desempenho e suas limitações métricas para uma análise franca dos resultados.

6.3.1 Desafios e Sensibilidades Operacionais

As limitações de um sistema SLAM baseado em features visuais são amplamente influenciadas pela qualidade da cena e pela dinâmica de movimento, sendo as seguintes as principais restrições observadas:

Escassez de Textura (*Feature Scarcity*): O ORB-SLAM3 depende da detecção e descrição de pontos de interesse (ORB *features*) com padrões de textura distintos. Ambientes com pouca textura ou superfícies uniformemente coloridas (e.g., paredes brancas, superfícies metálicas lisas, ou água com baixa visibilidade) limitam drasticamente o número de *features* rastreáveis. A insuficiência de pontos de correspondência impede a triangulação 3D, degradando o desempenho da odometria visual e, por consequência, o *tracking* da posição.

Condições de Iluminação: Apesar da capacidade das câmeras industriais FLIR de realizarem alteração automática do tempo de exposição (AEC) e do ganho, a baixa luminosidade extrema e a ocorrência de reflexos comprometem a qualidade das imagens. Em ambientes pouco iluminados, o ruído eletrônico inerente ao sensor

óptico aumenta, a capacidade de distinção dos *features* diminui e o algoritmo não consegue manter a estabilidade do rastreamento.

Ambientes Dinâmicos: O pressuposto fundamental da otimização *Bundle Adjustment* no *backend* do SLAM é que o ambiente seja estático. A presença de elementos em movimento (pedestres, veículos, objetos flutuantes, sombras variáveis) é tratada pelo ORB-SLAM3 como *outliers* que precisam ser rejeitados. No entanto, em cenas com grande proporção de elementos dinâmicos, a taxa de rejeição de *outliers* pode ser excessiva, degradando a capacidade de estimação da posição e, criticamente, dificultando o reconhecimento de lugares e o fechamento de *loop* (*loop closure*), que é essencial para corrigir o *drift* acumulado.

Suscetibilidade a Movimentos Abruptos e Vibrações: A grande vantagem do modo Estéreo-Inercial — sua sensibilidade a mudanças de movimento — é também uma limitação operacional. A IMU STIM300 de grau de navegação, ao registrar acelerações muito abruptas (e.g., um impacto acidental, uma batida ou uma trepidação de alta frequência como observado na guia linear), introduz picos de ruído que, uma vez integrados, causam deslocamentos de pose significativos e levam a resultados incorretos. Se esse evento ocorrer em um momento de escassez visual, isto é, com textura insuficiente, o erro pode ser fatal para a sessão de SLAM, exigindo a reinicialização do sistema.

6.3.2 Categorização das Fontes de Erro de Medição

As fontes de erro de medição do sistema podem ser categorizadas em três grupos principais que contribuem para o erro final na trajetória e na nuvem de pontos:

Erro de Calibração (Sistemático e Aleatório): A incerteza de precisão de todo o sistema de fusão sensorial é limitada pela exatidão da calibração:

Componente Sistemática: Erros sistemáticos decorrem de imprecisões dos parâmetros experimentalmente determinados utilizado (e.g., o modelo *pin-hole* para câmeras, ou o modelo de ruído da IMU, como a Variância de Allan). A calibração *inter-sensor* (extrínseca) entre as câmeras e a IMU, embora realizada com métodos rigorosos (e.g., Furgale et al.), também contém um erro residual na determinação das transformações rígidas que se propaga ao longo de todas as medições.

Componente Aleatória: Resulta do ruído de medição durante a calibração, como a incerteza na localização dos cantos do tabuleiro de calibração ou a variação na temperatura dos sensores, que afeta a estabilidade dos parâmetros intrínsecos.

Erro do Algoritmo SLAM: O algoritmo ORB-SLAM3, sendo um sistema baseado em otimização, introduz seus próprios erros:

Acúmulo de Erro (*Drift*): Mesmo com o *Bundle Adjustment* e a restrição inercial, o erro na estimação de pose se acumula ao longo do tempo (principalmente em ambientes grandes e sem *loop closure*), o que é inerente à natureza da odometria.

Erro Residual: Após a otimização, sempre resta um erro residual (erro de reprojeção) nos features visuais e nas medições inerciais, que não pode ser completamente eliminado pelo processo de Mínimos Quadrados Não Lineares (LM).

Erro por Dinâmica de Captura (*Motion Blur*) Diferente da escassez de textura ou ruído eletrônico, a desfocagem por movimento (*Motion Blur*) constitui uma fonte de erro de medição direta. Durante movimentos angulares rápidos ou vibrações da estrutura, o tempo de exposição do sensor — mesmo que curto — captura o deslocamento do cenário, resultando em "arrasto" nos *pixels*. Isso degrada a precisão na localização do centroide dos *features* ORB, introduzindo uma incerteza na triangulação que não pode ser mitigada apenas pelo aumento de ganho ou contraste.

Variação da Resolução Espacial Embora os parâmetros intrínsecos de captura sejam constantes (resolução de imagem fixa em 800x600 *pixels*), a resolução espacial (*Spatial Resolution*) do sistema varia em função da profundidade. Devido à geometria projetiva, a área física representada por um único pixel aumenta com a distância. Portanto, objetos próximos são reconstruídos com alta densidade de pontos, enquanto estruturas distantes sofrem com a discretização da textura, reduzindo a precisão da nuvem de pontos gerada pelo SLAM nessas regiões, independentemente da resolução do sensor.

Uma consideração crítica para a interpretação dos valores de APE (*Absolute Pose Error*) no percurso avaliado é a diferença na densidade de geração de poses entre os métodos. Enquanto o ORB-SLAM3 produz uma estimativa de trajetória contínua e de alta frequência, a reconstrução por fotogrametria (utilizada como referência) fornece um número significativamente menor de poses discretas. Essa escassez de pontos de referência na fotogrametria limita a granularidade da comparação, contribuindo para o aumento do erro acumulado observado, uma vez

que a validação depende de interpolações entre pontos distantes que não capturam as micro-correções de trajetória realizadas pelo sistema inercial.

6.4 IMPLICAÇÕES PRÁTICAS PARA A INSPEÇÃO DE ESTRUTURAS

A avaliação integrada do desempenho do sistema de visão estéreo-inercial, utilizando o ORB-SLAM3 com as câmeras FLIR e a IMU STIM300, estabelece um conjunto claro de implicações práticas para o seu uso em inspeção de estruturas, a partir de um ROV.

A principal contribuição operacional do sistema reside na sua capacidade de fornecer uma estimativa de pose com estabilidade dinâmica para a malha de controle do ROV. A fusão Visual-Inercial demonstrou uma estabilidade operacional, mantendo a continuidade do rastreamento mesmo sob as condições do Experimento 2. Embora a magnitude do erro de posição final (APE) tenha sido inflacionada pela baixa densidade de pontos da referência fotogramétrica — o que penaliza a validação em longas distâncias —, a análise do comportamento local revela a verdadeira aplicabilidade do sistema. Quantitativamente, a consistência entre quadros consecutivos, medida pelo RPE, manteve-se na ordem de 0,32 m por metro percorrido. Este baixo índice de erro relativo atesta que, apesar do acúmulo de deriva (*drift*) em escala global, o sistema oferece a exatidão incremental e a suavidade de trajetória necessárias para a navegação autônoma segura em ambientes confinados, mitigando riscos de colisão decorrentes de instabilidades na estimação

Neste contexto, a trajetória do ORB-SLAM3 não funciona apenas como um sistema de navegação, mas também como uma referência temporal e espacial para a fusão de dados provenientes de sensores auxiliares. O mapa e as poses geradas pelo SLAM atuam como um *framework* para integrar a nuvem de pontos de equipamentos de medição mais lentos, garantindo que todos os dados coletados ao longo do tempo sejam mapeados em um único sistema de coordenadas consistente. O seu valor máximo, portanto, é alcançado quando a nuvem de pontos esparsa do SLAM é utilizada como o mapa base para integrar e registrar a nuvem de pontos densa proveniente de outros sensores de medição (como escâneres a laser, sonares ou fotogrametria de Múltipla Vista Estéreo) com maior exatidão metrológica.

Apesar dos resultados obtidos na localização e navegação, a análise quantitativa dos resultados de mapeamento (Seção 5.3.2) estabelece uma limitação crítica para a aplicação do sistema em metrologia de baixa incerteza de medição. Os resultados de medição de diâmetro do corpo de prova (Tabela 13) demonstraram que a nuvem de pontos esparsa gerada pelo ORB-SLAM3 apresenta erro de medição e desvio padrão (repetibilidade) significativos, não sendo adequada para medições críticas ou para a avaliação detalhada de deformidades com alta exatidão. A natureza esparsa do mapeamento e sua otimização primária pela pose (localização), e não pela densidade geométrica, impõem esta restrição.

Em contrapartida, o sistema apresenta bom desempenho na reconstrução e descrição de formas e estruturas (fidelidade estrutural). Para os casos avaliados, o mapeamento visual qualitativo (Seção 5.3.1) confirmou que o ORB-SLAM3 é capaz de reconstruir a geometria e o contorno de objetos e ambientes de forma coerente (cilindros, estruturas complexas).

Esta capacidade de reprodução estrutural é valiosa para a identificação de macro-falhas, como rupturas, desvios de alinhamento ou deformações bem visíveis. Desta forma, o sistema deve ser estrategicamente posicionado: ele é o sistema de navegação estável e o pilar de fusão de dados, mas as medições dimensionais extraídas de sua nuvem de pontos devem ser utilizadas apenas para fins de referência ou estimativa inicial, sendo necessários sensores complementares para a inspeção metrológica.

7 CONCLUSÕES

O presente capítulo reúne as principais conclusões da pesquisa desenvolvida, retomando o objetivo central de avaliar o desempenho de um sistema de visão estéreo passivo integrado a sensores inerciais de classe tática. Após a análise detalhada dos experimentos realizados com o algoritmo ORB-SLAM3, utilizando o conjunto de hardware composto pelas câmeras industriais FLIR e pela IMU STIM300, são sintetizadas as contribuições técnicas e os aprendizados obtidos durante o processo de validação da odometria visual-inercial. Inicia com uma síntese conclusiva sobre o desempenho obtido na estimativa de trajetória e da qualidade da nuvem de pontos gerada frente aos desafios de iluminação e textura. Em seguida, na seção de trabalhos futuros, encerra-se o documento com a identificação das lacunas remanescentes e a proposição de novas abordagens tecnológicas e metodológicas visando a continuidade e o aprimoramento da solução de inspeção de SLAM.

7.1 SÍNTESE DO TRABALHO E PRINCIPAIS APRENDIZADOS

A presente pesquisa atingiu seu objetivo principal ao desenvolver, integrar e avaliar dois sistemas de visão estéreo-inercial, baseados no algoritmo ORB-SLAM3, voltado para a navegação autônoma em ambientes sem GNSS. A abordagem metodológica, que partiu da construção de um hardware estável com câmeras industriais FLIR e uma IMU STIM300 assim como os métodos de avaliação, provaram ser eficazes.

A síntese dos resultados obtidos permite estabelecer as seguintes conclusões fundamentais:

Quanto à Integridade do Aparato e Calibração: O protocolo de calibração desenvolvido mostrou-se suficiente para o sucesso da fusão sensorial. Os baixos erros de reprojeção obtidos (na ordem de 0,13 a 0,19 *pixels*) e o alinhamento físico preciso entre os eixos da câmera e da IMU validaram o *setup* experimental. Conclui-se que a qualidade da calibração extrínseca é um dos pilares que sustenta a estabilidade do sistema, confirmando que o cuidado com a rigidez mecânica e o

sincronismo temporal dos sensores é pré-requisito mandatório para aplicações de SLAM visual-inercial.

Quanto à incerteza de medição da trajetória (Odometria): A fusão estéreo-inercial (modo *tightly-coupled*) demonstrou superioridade na manutenção da escala métrica e na exatidão de posicionamento final em comparação ao modo puramente visual. Nos testes em ambiente controlado (guia linear), o sistema foi capaz de estimar o deslocamento de 3 metros com um erro médio de apenas 43,1 mm (cerca de 1,4%). No entanto, observou-se uma sensibilidade crítica da IMU a vibrações de alta frequência.

A degradação da retitude da trajetória observada na guia linear evidência que, embora a IMU corrija o *drift* de escala, ela pode introduzir ruídos de posição se o sistema for submetido a trepidações mecânicas não filtradas. Em contrapartida, nos testes em cenário de corredores (mais representativo da dinâmica de um ROV), o sistema mostrou-se estável a movimentos complexos e variações de iluminação, mantendo a coerência global da trajetória em percursos longos (*loop* de 50 m) com uma média do erro relativo aceitável (0,1362 m) para fins de navegação.

Quanto à Qualidade do Mapeamento: A avaliação da nuvem de pontos gerada pelo ORB-SLAM3 revelou uma dicotomia importante entre fidelidade estrutural e incerteza de medição. O sistema foi capaz de reconstruir a geometria global de ambientes e objetos complexos (como o corredor e a caixa d'água) de forma visualmente coerente, permitindo a identificação de estruturas e obstáculos. Entretanto, a análise quantitativa no corpo de prova (tubo polimérico) indicou incertezas dimensionais e uma variabilidade (desvio padrão) que restringem o uso direto da nuvem de pontos esparsa para metrologia dimensional de precisão. Conclui-se que o mapa gerado é adequado para a localização e navegação do robô, mas deve ser utilizado como uma referência estrutural (um "mapa base") sobre o qual dados de sensores mais densos e precisos (como *scanners* a laser ou fotogrametria densa offline) devem ser fundidos para tarefas de inspeção detalhada.

Conclusão Geral: O sistema desenvolvido foi bem sucedido no uso do ORB-SLAM3 com hardware industrial como uma solução viável e estável para fornecer a pose (posição e orientação) de veículos autônomos. O sistema cumpre o papel de "espinha dorsal" de navegação, capaz de operar em ambientes onde o GPS é negado, fornecendo a estabilidade necessária para que o ROV execute suas missões de

inspeção. Também foi criado uma *playlist* com vídeos os testes e resultados obtidos ao decorrer deste desenvolvimento (102). As limitações mapeadas, especialmente quanto à sensibilidade vibracional e à densidade do mapa, não invalidam a solução, mas definem claramente o escopo de sua aplicação e direcionam os aprimoramentos futuros.

7.2 TRABALHOS FUTUROS

A avaliação experimental do sistema de visão estéreo-inercial realizada neste trabalho, bem como a confirmação da fidelidade métrica do algoritmo ORB-SLAM3 em ambiente controlado (ar), encorajam a continuidade da pesquisa em direção à sua aplicação final: a robótica subaquática. Com base nos resultados obtidos e nas metodologias de avaliação estabelecidas, vislumbram-se as seguintes frentes de pesquisa para a evolução do sistema:

Integração ao Veículo VORIS: O passo natural subsequente é a integração do hardware desenvolvido (câmeras FLIR e IMU STIM300) ao Veículo Operado Remotamente (ROV) do projeto VORIS. Embora testes preliminares com o sistema monocular, utilizando a câmera nativa do submarino, já tenham sido realizados, faz-se necessária a implementação física do conjunto estéreo-inercial dedicado. Isso envolve o projeto de encapsulamento estanque que respeite a calibração extrínseca rígida entre as câmeras e a IMU, além da calibração intrínseca específica para o meio aquático, considerando os efeitos de refração.

Validação Metrológica Subaquática: Um dos maiores desafios mapeados é a validação da odometria em ambiente subaquático com o mesmo rigor metrológico aplicado neste trabalho. Propõe-se a replicação da metodologia de validação aqui utilizada, adaptando o sistema de referência para a água. Isso pode ser realizado através de duas abordagens:

A construção de uma infraestrutura submersa análoga à guia linear com interferômetro a laser utilizada nos testes em terra, garantindo um *ground truth* de trajetória com incerteza de precisão.

A utilização de técnicas de fotogrametria subaquática para a reconstrução de corpos de prova de referência, comparando a nuvem de pontos gerada pelo SLAM com modelos obtidos por sistemas comerciais de alta exatidão. O objetivo é verificar

se o desempenho do ORB-SLAM3 se mantém consistente sob a influência de turbidez e absorção de luz características do meio.

Benchmarking de Algoritmos de SLAM: Aproveitando o *dataset* e a infraestrutura de validação desenvolvidos, sugere-se a avaliação comparativa de outros algoritmos de SLAM do estado da arte, como o VINS-Fusion ou o S-PTAM, utilizando as mesmas técnicas de *front-end* e *back-end* parametrizadas neste trabalho. A intenção é investigar se outras abordagens de fusão sensorial apresentam melhor desempenho ou eficiência computacional frente à dinâmica específica do ROV.

Otimização Algorítmica e Computacional: Por fim, identifica-se a oportunidade de realizar otimizações específicas no código fonte do ORB-SLAM3. Tais otimizações visam o aprimoramento do *loop closing* e da seleção de *keyframes* em cenários com partículas em suspensão ou baixa textura, além da adequação do custo computacional para processamento embarcado em tempo real no hardware do ROV, garantindo a autonomia necessária para missões de inspeção ou intervenção.

REFERÊNCIAS

- 1 ODILI, P. O. *et al.* Integrating Advanced Technologies in Corrosion and Inspection Management for Oil and Gas Operations. **Engineering Science & Technology Journal**, v. 5, n. 2, p. 597-611, fev. 2024. Disponível em: <http://www.fepbl.com/index.php/estj>. Acesso em: 19 jun. 2025.
- 2 AGBAKWURU, J. Oil/Gas Pipeline Leak Inspection and Repair in Underwater Poor Visibility Conditions: Challenges and Perspectives. **Journal of Environmental Protection**, v. 3, n. 5, p. 394-399, maio 2012. DOI: 10.4236/jep.2012.35049.
- 3 CHEMISKY, B. *et al.* Underwater Survey for Oil and Gas Industry: A Review of Close Range Optical Methods. **Remote Sensing**, v. 13, n. 14, p. 2789, jul. 2021. DOI: 10.3390/rs13142789.
- 4 HUANG, H. *et al.* Using ROVs for Tilt Angle Measurement of Pier Foundation Piles at Pier-Type Docks. **2025 IEEE Underwater Technology (UT)**, p. 1-6, 2025. Disponível em: <https://doi.org/10.1109/UT61067.2025.10947429>. Acesso em: 19 jun. 2025.
- 5 CASTILLÓN, M. *et al.* State of the art of underwater active optical 3D scanners. **MDPI Sensors Journal**, Basel, v. 19, n. 5161. Nov. 2019. DOI: 10.3390/s19235161. Disponível em: <https://www.mdpi.com/1424-8220/19/23/5161>. Acesso em: 19 jun. 2025.
- 6 ORION S.I.C. **ROVs para inspeções em FPSOs**. [S. l.], 2023. Disponível em: <https://orionsic.com.br/rovs-para-inspecoes-em-fpsos/>. Acesso em: 19 jun. 2025.
- 7 WEI, P.; YING, Y.; LU, W.; GE, Z.; YANG, F.; PENG, C. A Navigation System for ROV's inspection on Fish Net Cage. **arXiv**, abs/2503.00482, 2025. Disponível em: <https://doi.org/10.48550/arXiv.2503.00482>. Acesso em: 19 jun. 2025.
- 8 LIU, H. *et al.* SLAM Method of Mine Inspection Robot based on Stereo Vision and IMU. In: **CHINA AUTOMATION CONGRESS (CAC), 2024**, Qingdao, China. [S. l.]: IEEE, 2024. p. 1356-1361. DOI: 10.1109/CAC63892.2024.10864764.
- 9 WEN, S. *et al.* Joint optimization based on direct sparse stereo visual-inertial odometry. **Autonomous Robots**, v. 44, p. 791-809, 2020. DOI: 10.1007/s10514-019-09897-6.
- 10 JUNG, J.; HEO, S.; PARK, C. Observability Analysis of IMU Intrinsic Parameters in Stereo Visual-Inertial Odometry. **IEEE Transactions on Instrumentation and Measurement**, v. 69, p. 7530-7541, 2020. DOI: 10.1109/TIM.2020.2985174.
- 11 ZHENG, Shuran *et al.* Simultaneous Localization and Mapping (SLAM) for Autonomous Driving: Concept and Analysis. **Remote Sensing**, v. 15, n. 4, p. 1156, fev. 2023. DOI: 10.3390/rs15041156.

- 12 CAMPOS, C. *et al.* ORB-SLAM3: An Accurate Open-Source Library for Visual, Visual-Inertial, and Multimap SLAM. **IEEE Transactions on Robotics**, [S. l.], v. 37, n. 6, p. 1874-1890, dez. 2021. DOI: 10.1109/TRO.2021.3075644.
- 13 DEEP TREKKER. **Deep Trekker**. [S. l.], 2025. Disponível em: <https://www.deeptrekker.com>. Acesso em: 10 jul. 2025.
- 14 VIDEORAY. **VideoRay**. [S. l.], 2025. Disponível em: <https://videoray.com>. Acesso em: 11 jul. 2025.
- 15 BLUEYE ROBOTICS. **Blueye Robotics**. [S. l.], 2025. Disponível em: <https://www.blueyerobotics.com>. Acesso em: 19 jul. 2025.
- 16 HUGHES SUBSEA. **Hughes Subsea**. [S. l.], 2025. Disponível em: <https://www.hughes-subsea.com>. Acesso em: 12 jul. 2025.
- 17 VOYIS. **Voyis**. [S. l.], 2025. Disponível em: <https://voyis.com>. Acesso em: 19 jul. 2025.
- 18 SHARAFUTDINOV, Dinar *et al.* Comparison of modern open-source Visual SLAM approaches. **arXiv**, [S. l.], abs/2108.01654v2, 2023. Disponível em: <https://arxiv.org/abs/2108.01654>. Acesso em: 14 jul. 2025.
- 19 MILANO, Danilo de; HONORATO, Luciano Barrozo. **Visão Computacional**. Campinas: UNICAMP - Universidade Estadual de Campinas, Faculdade de Tecnologia, 2014.
- 20 SZELISKI, Richard. **Computer Vision: Algorithms and Applications**. London: Springer, 2010.
- 21 DANIILIDIS, Kostas; KLETTE, Reinhard. Imaging beyond the pinhole camera. In: **Imaging Beyond the Pinhole Camera**. Dordrecht: Springer Netherlands, 2006. p. 1-2. DOI: 10.1007/978-1-4020-4894-4. Disponível em: <https://doi.org/10.1007/978-1-4020-4894-4>. Acesso em: 24 jul. 2025.
- 22 HARTLEY, Richard; ZISSERMAN, Andrew. **Multiple View Geometry in Computer Vision**. 2. ed. Cambridge: Cambridge University Press, 2003.
- 23 NAYAR, S. K. **First Principles of Computer Vision: Image Formation – Pinhole & Perspective Projection**. Columbia University, 2021. Disponível em: <https://fpcv.cs.columbia.edu/>. Acesso em: julho de 2025.
- 24 TRUCCO, Emanuele; VERRI, Alessandro. **Introductory Techniques for 3-D Computer Vision**. Upper Saddle River, NJ: Prentice Hall, 1998.
- 25 SCHWALBE, E. Geometric Analysis of the PanoCam-System for 3D City Modelling. In: INTERNATIONAL WORKSHOP ON VISION TECHNIQUES FOR SPATIAL MEASUREMENT, MODELLING AND RECONSTRUCTION (VT-SMMR), 2005, Berlin, Germany. **Anais [...]**. Berlin, Germany: TU Berlin, 2005. p. 81-88.

26 VAN GOOL, Luc; TOMASI, Carlo; KANADE, Takeo. Structure from Motion: A Tutorial. **Foundations and Trends® in Computer Graphics and Vision**, v. 4, n. 1, p. 1-195, 2009.

27 MOLINARI, M. E.; D'ANNUNZIO, L.; PEPE, M. The Role of Stereo Vision System and Structure from Motion for 3D Reconstruction and Mapping. **ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences**, v. IV-2/W7, p. 79-85, 2019.

28 BREWSTER, David. **The Stereoscope: Its History, Theory, and Construction**. London: John Murray, 1856. Disponível em: <https://archive.org/details/b21302200/page/n111/mode/2up>. Acesso em: 28 jun. 2025.

29 WANG, J. *et al.* A new calibration model of camera lens distortion. **Pattern Recognition**, v. 41, n. 2, p. 607-615, 2008. DOI: 10.1016/j.patcog.2007.06.012.

30 ZHANG, Zhengyou. Camera calibration. In: **Encyclopedia of computer vision**. [S. l.: s. n.], 2003. p. 5-40.

31 BENEVENUTI, Fabio. **Calibração de um conjunto binocular para uso em estereoscopia e robótica móvel**. 2001. 112 f. Dissertação (Mestrado em Engenharia Elétrica) - Universidade Federal de Santa Catarina, Programa de Pós-Graduação em Engenharia Elétrica, Florianópolis, 2001.

32 SMITH, W. J. **Modern optical engineering**. 3. ed. New York: McGraw-Hill, 2000. 603 p.

33 Zhang, Z. (2000). A flexible new technique for camera calibration. **IEEE Transactions on Pattern Analysis and Machine Intelligence**, 22(11), 1330–1334. <https://doi.org/10.1109/34.888718>

34 BOUGUET, Jean-Yves. **Visual methods for three-dimensional modeling**. 1999. Tese (Ph.D.) - California Institute of Technology, Pasadena, 1999. Disponível em: <https://thesis.library.caltech.edu/541>. Acesso em: 1 ago. 2025.

35 MACHADO, H. V. R. *et al.* SISTEMA DE NAVEGAÇÃO INERCIAL BASEADO EM IMU INERTIAL NAVIGARION SYSTEM BASED ON IMU-INERTIAL MEASUREMENT UNIT. Anais do **XX Congresso Brasileiro de Automática**, João Pessoa, 2018.

36 EL-FATATRY, A. **Inertial Measurement Units – IMU**. 2003.

37 SAFRAN. **STIM300 Datasheet**. TS1524 rev.30. Novembro, 2023.

38 LI, X.; XU, F.; XU, G. Redundant Inertial Measurement Unit Reconfiguration and Trajectory Replanning of Launch Vehicle. **Sensors**, v. 23, n. 13, 2023.

- 39 LAWRENCE, A. **Modern Inertial Technology: Navigation, Guidance, and Control**. Nova York:Springer, 1998.
- 40 WOODMAN, O. J. **An introduction to inertial navigation**. Cambridge: University of Cambridge, Computer Laboratory, 2007. 67 p. (Technical Report, n. 696). Disponível em: <http://www.cl.cam.ac.uk/techreports/>. Acesso em: 25 jul. 2025.
- 41 APOSTOLYUK, V. **Coriolis Vibratory Gyroscopes: Theory and Design**. 1. ed., 2016.
- 42 SENTURIA, S. D. **Microsystem Design**. Boston: Kluwer Academic Publishers, 2001.
- 43 MIEZAL, M.; TAETZ, B.; BLESER, G. On inertial body tracking in the presence of model calibration errors. **Sensors**, v. 16, n. 7, 2016. DOI: 10.3390/s16071132.
- 44 ÜNSAL, D. **Estimation of deterministic and stochastic IMU error parameters**. 2012. 129 f. Tese (Mestrado em Ciências) - Middle East Technical University, Ancara, 2012.
- 45 NARASIMHAPPA, M. *et al.* MEMS-Based IMU Drift Minimization: Sage Husa Adaptive Robust Kalman Filtering. **IEEE Sensors Journal**, v. 20, n. 1, p. 250–260, 2020. DOI: 10.1109/JSEN.2019.2941273.
- 46 HEMERLY, E. M. MEMS IMU stochastic error modelling. **Systems Science and Control Engineering**, v. 5, n. 1, p. 1–8, 2017. DOI: 10.1080/21642583.2016.1262801.
- 47 REHDER, J.; SIEGWART, R. Camera/IMU Calibration Revisited. **IEEE Sensors Journal**, v. 17, n. 11, p. 3257–3268, 2017. DOI: 10.1109/JSEN.2017.2674307.
- 48 FU, B. *et al.* High-Precision Multicamera-Assisted Camera-IMU Calibration: Theory and Method. **IEEE Transactions on Instrumentation and Measurement**, v. 70, p. 1–13, 2021. DOI: 10.1109/TIM.2021.3051726.
- 49 DURRANT-WHYTE, H.; BAILEY, T. Simultaneous Localization and Mapping: Part I History of the SLAM Problem. **IEEE Robotics & Automation Magazine**, v. 13, n. 2, p. 99-110, 2006.
- 50 CADENA, C. *et al.* Past, present, and future of simultaneous localization and mapping: Toward the robust-perception age. **IEEE Transactions on Robotics**, v. 32, n. 6, p. 1309–1332, 2016. DOI: 10.1109/TRO.2016.2624754.
- 51 CHONG, T. J. *et al.* Sensor Technologies and Simultaneous Localization and Mapping (SLAM). **Procedia Computer Science**, v. 76, p. 174–179, 2015. DOI: 10.1016/j.procs.2015.12.336.
- 52 THRUN, S. Simultaneous Localization and Mapping. In: JEFFERIES, M. E.; YEAP, W. K. (ed.). **Robotics and Cognitive Approaches to Spatial Mapping**.

Berlin: Springer, 2007. p. 137-161. (Springer Tracts in Advanced Robotics, v. 38). DOI: 10.1007/978-3-540-75388-9_3.

53 WANG, Z.; HUANG, S.; DISSANAYAKE, G. **Simultaneous Localization and Mapping**: Exactly Sparse Information Filters. [S. l.]: World Scientific Publishing, 2011. DOI: 10.1142/8145.

54 MUTLAG, W. K. *et al.* Feature Extraction Methods: A Review. **Journal of Physics: Conference Series**, v. 1591, n. 1, p. 012028, 2020. DOI: 10.1088/1742-6596/1591/1/012028.

55 GUAN, G. *et al.* Keypoint-based keyframe selection. **IEEE Transactions on Circuits and Systems for Video Technology**, v. 23, n. 4, p. 729–734, 2013. DOI: 10.1109/TCSVT.2012.2214871.

56 HO, K. L.; NEWMAN, P. Detecting loop closure with scene sequences. **International Journal of Computer Vision**, v. 74, n. 3, p. 261–286, 2007. DOI: 10.1007/s11263-006-0020-1.

57 LIU, Z.; ZHANG, F. BALM: Bundle Adjustment for Lidar Mapping. **IEEE Robotics and Automation Letters**, v. 6, n. 2, p. 3184–3191, 2021. DOI: 10.1109/LRA.2021.3062815.

58 PARRA, Á. *et al.* Visual SLAM: Why Bundle Adjust?. In: **IEEE International Conference on Robotics and Automation (ICRA)**, 2019, Montreal. Anais... Montreal: IEEE, 2019. p. 8793749. DOI: 10.1109/ICRA.2019.8793749.

59 LING, Y.; KUSE, M.; SHEN, S. Edge alignment-based visual–inertial fusion for tracking of aggressive motions. **Autonomous Robots**, v. 42, n. 3, p. 513–528, 2018. DOI: 10.1007/s10514-017-9642-0.

60 ISHIHARA, J. *et al.* An IMU/Magnetometer/GPS-based localization system using correlated Kalman filtering. In: **Anuário do Seminário de Iniciação Científica e Tecnológica da UTFPR**, 2008. Disponível em: <https://www.researchgate.net/publication/228865673>. Acesso em: 19 ago. 2025.

61 SHEN, S.; M. e N., and K., V. Tightly-Coupled Monocular Visual-Inertial Fusion for Autonomous Flight of Rotorcraft MAVs. In: **IEEE International Conference on Robotics and Automation (ICRA)**, 2015, Seattle. Anais... Seattle: IEEE, 2015.

62 CAO, S.; LU, X.; SHEN, S. GVINS: Tightly Coupled GNSS-Visual-Inertial Fusion for Smooth and Consistent State Estimation. **IEEE Transactions on Robotics**, v. 38, n. 4, p. 2004–2021, 2022. DOI: 10.1109/TRO.2021.3133730.

63 SONG, J.; RICHARD, A.; OLIVARES-MENDEZ, M. Improving Monocular Visual-Inertial Initialization with Structureless Visual-Inertial Bundle Adjustment. **arXiv**, [S. l.], abs/2502.16598, 2025. Disponível em: <http://arxiv.org/abs/2502.16598>. Acesso em: 19 ago. 2025.

- 64 CUCCI, D. A.; REHAK, M.; SKALOUD, J. Bundle adjustment with raw inertial observations in UAV applications. **ISPRS Journal of Photogrammetry and Remote Sensing**, v. 130, p. 1–12, 2017. DOI: 10.1016/j.isprsjprs.2017.05.008.
- 65 MUR-ARTAL, R.; MONTIEL, J. M. M.; TARDOS, J. D. ORB-SLAM: A Versatile and Accurate Monocular SLAM System. **IEEE Transactions on Robotics**, v. 31, n. 5, p. 1147–1163, 2015. DOI: 10.1109/TRO.2015.2463671.
- 66 MUR-ARTAL, R.; TARDOS, J. D. ORB-SLAM2: An Open-Source SLAM System for Monocular, Stereo, and RGB-D Cameras. **IEEE Transactions on Robotics**, v. 33, n. 5, p. 1255–1262, 2017. DOI: 10.1109/TRO.2017.2705103.
- 67 BENCHMARK for SLAM. In: **ETH3D**. [S. l.]. Disponível em: https://www.eth3d.net/slam_benchmark?metric=time. Acesso em: 13 de ago. de 2025.
- 68 COELHO, L. C. Teixeira. **Fotogrametria Digital**. Rio de Janeiro: EdUERJ, 2007.
- 69 FRANCO, G. G.; NAIME, A. F. Structure from Motion (SFM) – Uma Breve Revisão Histórica, Aplicações nas Geociências e Perspectivas Futuras. **Anuário do Instituto de Geociências - UFRJ**, v. 44, p. 1-15, 2021. DOI: 10.11137/1982-3908_2021_44_40853.
- 70 ULLMAN, S. The interpretation of structure from motion. **Proceedings of the Royal Society of London. Series B, Biological Sciences**, v. 203, n. 1153, p. 405–426, 1979. DOI: 10.1098/rspb.1979.0006.
- 71 LOWE, D. G. Distinctive Image Features from Scale-Invariant Keypoints. **International Journal of Computer Vision**, v. 60, n. 2, p. 91–110, 2004. DOI: 10.1023/B:VISI.0000029664.99615.94.
- 72 LINDBERG, T. Scale Invariant Feature Transform. **Scholarpedia**, v. 7, n. 5, p. 10491, 2012. DOI: 10.4249/scholarpedia.10491.
- 73 FISCHLER, Martin A.; BOLLES, Robert C. Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. **Communications of the ACM**, v. 24, n. 6, p. 381–395, 1981. DOI: 10.1145/358669.358692.
- 74 RAGURAM, R. *et al.* A Comparative Analysis of RANSAC Techniques Leading to Adaptive Real-Time Random Sample Consensus. In: FORSYTH, D.; TORR, P.; ZISSERMAN, A. (ed.). **COMPUTER VISION – ECCV 2008**. Berlin, Heidelberg: Springer, 2008. p. 500-513. (Lecture Notes in Computer Science, v. 5303). DOI: 10.1007/978-3-540-88688-4_37.
- 75 HORN, B. K. P. Closed-form solution of absolute orientation using unit quaternions. **Journal of the Optical Society of America A**, v. 4, n. 4, p. 629–642, 1987.

76 BESL, P. J.; MCKAY, N. D. A method for registration of 3-D shapes. **IEEE Transactions on Pattern Analysis and Machine Intelligence**, v. 14, n. 2, p. 239–256, 1992. DOI: 10.1109/34.121791.

77 STURM, J. et al. A benchmark for the evaluation of RGB-D SLAM systems. In: **IEEE/RSJ INTERNATIONAL CONFERENCE ON INTELLIGENT ROBOTS AND SYSTEMS (IROS)**, 2012, Vilamoura. Proceedings [...]. [S. l.]: IEEE, 2012. p. 573-580. DOI: 10.1109/IROS.2012.6385773.

78 SEITZ, S. M. et al. A Comparison and Evaluation of Multi-View Stereo Reconstruction Algorithms. In: **IEEE COMPUTER SOCIETY CONFERENCE ON COMPUTER VISION AND PATTERN RECOGNITION (CVPR)**, 2006, New York. Anais... New York: IEEE, 2006. p. 519–528. DOI: 10.1109/CVPR.2006.19.

79 FURUKAWA, Y.; HERNÁNDEZ, C. **Multi-View Stereo: A Tutorial**. [S. l.]: Now Publishers, 2015. DOI: 10.1561/06000000052.

80 SCHÖNBERGER, Johannes Lutz *et al.* A Vote-and-Verify Strategy for Fast Spatial Verification in Image Retrieval. In: **ASIAN CONFERENCE ON COMPUTER VISION (ACCV)**, 2016. Proceedings [...]. [S. l.: s. n.], 2016.

81 AGISOFT. **Agisoft Metashape User Manual: Professional Edition, Version 2.1**. [S. l.], 2024. Disponível em: <https://www.agisoft.com>. Acesso em: 13 ago. 2025.

82 QUIGLEY, Morgan et al. ROS: an open-source Robot Operating System. In: **ICRA WORKSHOP ON OPEN SOURCE SOFTWARE**, 2009, Kobe. Anais [...]. [S. l.: s. n.], 2009. p. 1-5.

83 MACENSKI, Steven; FOOTE, Tully; GERKEY, Brian; LALANCETTE, Christophe; WOODALL, William. Robot Operating System 2: Design, architecture, and uses in the wild. **Science Robotics**, v. 7, n. 66, p. eabm6074, 2022. DOI: 10.1126/scirobotics.abm6074.

84 JOSEPH, Lentin. **Robot Operating System (ROS) for Absolute Beginners**. 1. ed. New York: Apress, 2018.

85 BRADSKI, G. The OpenCV Library. **Dr. Dobb's Journal of Software Tools**, v. 25, n. 11, p. 120-125, 2000.

86 GUENNEBAUD, G.; JACOB, B. et al. **Eigen v3**. 2010. Disponível em: <http://eigen.tuxfamily.org>. Acesso em: 11, ago. 2025.

87 KÜMMERLE, R.; GRISETTI, G.; STRASDAT, H.; KONOLIGE, K.; BURGARD, W. g2o: A general framework for graph optimization. In: **IEEE International Conference on Robotics and Automation (ICRA)**, 2011, Shanghai. Anais [...] p. 3607-3613. DOI: 10.1109/ICRA.2011.5979542.

88 GÁLVEZ-LÓPEZ, D.; TARDÓS, J. D. Bags of Binary Words for Fast Place Recognition in Image Sequences. **IEEE Transactions on Robotics**, v. 28, n. 5, p. 1188-1197, 2012. DOI: 10.1109/TRO.2012.2197158.

89 LOVEGROVE, S.; KAEHLER, A.; SELA, G. et al. **Pangolin**: A lightweight and flexible library for managing OpenGL display and user interaction. 2013. Disponível em: <https://github.com/stevenlovegrove/Pangolin>. Acesso em: 21, ago. 2025.

90 THE MATHWORKS INC. **Camera Calibration Toolbox**. Natick, Massachusetts: The MathWorks Inc., 2022. Disponível em: <https://www.mathworks.com/products/computer-vision/camera-calibration.html>. Acesso em: 29 ago. 2025.

91 BUCHANAN, R. **Allan Variance ROS**. Oxford Robotics Institute, DRS Lab, nov. 2021. Software. Disponível em: https://github.com/ori-drs/allan_variance_ros. Acesso em: 29 ago. 2025.

92 FURGALE, P.; REHDER, J.; SIEGWART, R. Unified Temporal and Spatial Calibration for Multi-Sensor Systems. In: **IEEE/RSJ INTERNATIONAL CONFERENCE ON INTELLIGENT ROBOTS AND SYSTEMS (IROS)**, 2013, Tóquio. Anais [...]. Tóquio: IEEE, 2013. p. 1280-1286.

93 GALVÃO, Leonardo Sales *et al.* Desempenho de ORB-SLAM 3 em mapeamento 3D subaquático com ROV: um estudo de viabilidade. In: **12º CONGRESSO BRASILEIRO DE PESQUISA E DESENVOLVIMENTO EM PETRÓLEO E GÁS (PDPETRO)**, 2024,. Anais [...].I: ABPG, 2024.

94 RUBLEE, Ethan; RABAUD, Vincent; KONOLIGE, Kurt; BRADSKI, Gary. ORB: An efficient alternative to SIFT or SURF. In: **IEEE International Conference on Computer Vision (ICCV)**, 2011, Barcelona. Anais... Barcelona: IEEE, 2011. p. 2564-2571. DOI: 10.1109/ICCV.2011.6126544.

95 ROSTEN, E.; DRUMMOND, T. Machine Learning for High-Speed Corner Detection. In: LEONARDIS, A.; BISCHOF, H.; PINZ, A. (ed.). **Computer Vision – ECCV 2006**. Heidelberg: Springer, 2006. p. 430-443. (Lecture Notes in Computer Science, v. 3951). DOI: 10.1007/11744023_34.

96 NEUBECK, A.; VAN GOOL, L. Efficient Non-Maximum Suppression. In: **INTERNATIONAL CONFERENCE ON PATTERN RECOGNITION (ICPR)**, 18., 2006, Hong Kong. Anais... Hong Kong: IEEE, 2006. p. 850–855. DOI: 10.1109/ICPR.2006.479.

97 HU, M. K. Visual pattern recognition by moment invariants. **IRE Transactions on Information Theory**, v. 8, n. 2, p. 179–187, 1962. DOI: 10.1109/TIT.1962.1057692.

98 THOMAS, S. *et al.* Comparison of centroid computation algorithms in a Shack–Hartmann sensor. **Monthly Notices of the Royal Astronomical Society**, v. 371, n. 1, p. 323–336, 2006. DOI: 10.1111/j.1365-2966.2006.10661.x.

99 CALONDER, M.; LEPETIT, V.; STRECHA, C.; FUA, P. BRIEF: Binary Robust Independent Elementary Features. In: DANIILIDIS, K.; MARAGOS, P.; PARAGIOS, N. (ed.). **Computer Vision – ECCV 2010**. Berlim: Springer, 2010. p. 778–792. (Lecture Notes in Computer Science, v. 6314). DOI: 10.1007/978-3-642-15561-1_56.

100 HAMMING, R. W. Error detecting and error correcting codes. **The Bell System Technical Journal**, v. 29, n. 2, p. 147-160, abr. 1950. DOI: 10.1002/j.1538-7305.1950.tb00463.x.

101 BURT, P.; ADELSON, E. The Laplacian Pyramid as a Compact Image Code. **IEEE Transactions on Communications**, v. 31, n. 4, p. 532–540, 1983. DOI: 10.1109/TCOM.1983.1095851.

102 GALVÃO, L. S. Desenvolvimento e avaliação de um Sistema Estéreo Inercial baseado em ORB-SLAM3. **YouTube**, 2024. Disponível em: https://www.youtube.com/playlist?list=PLA_zm5EGc4sh-vxIJubQ-NPP6CWi4xgxz. Acesso em: 6 jan. 2026.

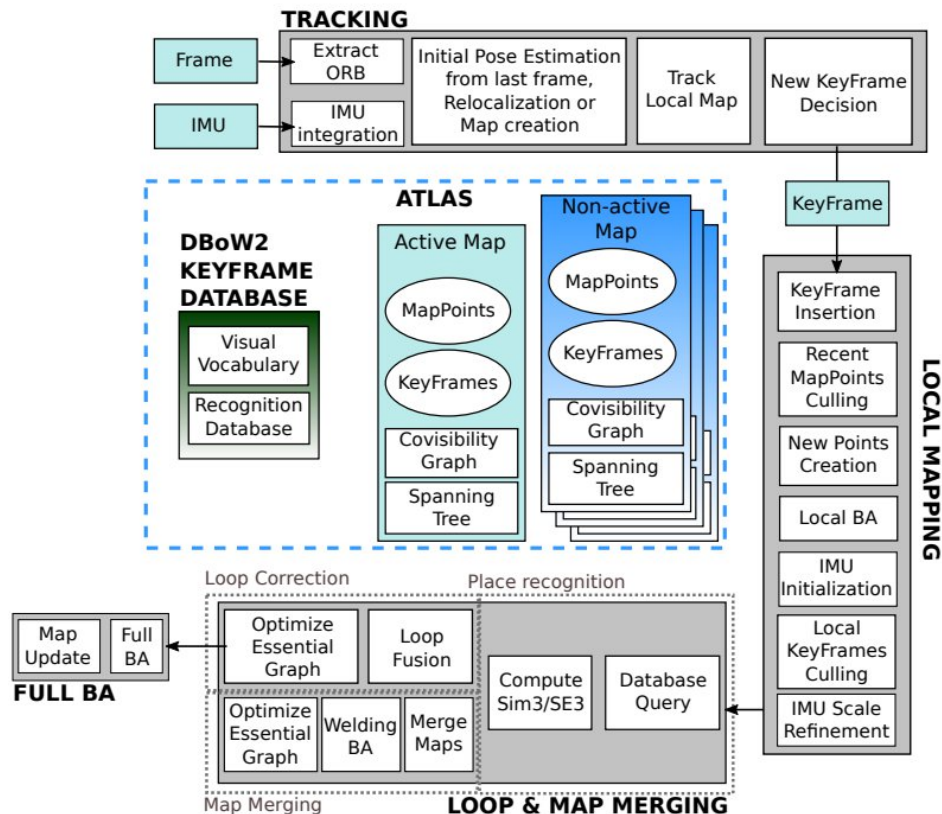
APÊNDICE A – FUNCIONAMENTO DO ORB-SLAM3

Este documento serve como uma extensão detalhada da fundamentação teórica introduzida na Seção 2.4, fornecendo uma análise aprofundada sobre a operação interna do sistema ORB-SLAM 3. Este apêndice está estruturado para apresentar, primeiramente, a Arquitetura Geral e *Threads* do ORB-SLAM 3, detalhando o processamento paralelo (Rastreamento, Mapeamento Local e Fechamento de Ciclos) e a estrutura de gerenciamento de mapas (Atlas). Em seguida, são minuciosamente explicados os princípios do Descritor de Características ORB (*Oriented* FAST and *Rotated* BRIEF), cobrindo o método de detecção FAST, a atribuição de orientação via Centroide de Intensidade, a criação do descritor *steered* BRIEF e a técnica de Pirâmide de Imagens para invariância de escala. O objetivo é fornecer a base técnica necessária para a compreensão dos resultados de performance apresentados nos capítulos seguintes.

ARQUITETURA GERAL E THREADS DO ORB-SLAM3

A alta performance e a capacidade de operação em tempo real do ORB-SLAM 3 são fundamentadas em uma arquitetura de processamento paralelo. O sistema divide as tarefas computacionais em três *threads* principais que rodam simultaneamente: Rastreamento (*Tracking*), Mapeamento Local (*Local Mapping*) e Fechamento de Ciclos e Fusão de Mapas (*Loop & Map Merging*). Estas *threads* operam sobre uma estrutura de dados central e inovadora chamada Atlas, que gerencia um ou mais mapas. A Figura 1 ilustra esquematicamente os componentes do sistema e o fluxo de informações entre eles. A seguir, cada um desses módulos é detalhado.

Figura 1 – Diagrama do funcionamento do SLAM.



Fonte: CAMPOS, 2021

Analisando o fluxo de dados apresentado no diagrama, o processo se inicia com a entrada de dados dos sensores: um *Frame* de imagem e, no modo visual-inercial, as medições da IMU. Esses dados alimentam a *thread* de Rastreamento (*Tracking*), que extrai características visuais e estima a pose atual do sensor. Uma das principais saídas desta *thread* é a decisão de gerar um novo *KeyFrame*, um *frame* considerado relevante para a construção do mapa. Uma vez criado, o *KeyFrame* é enviado simultaneamente para as outras duas *threads* principais. A *thread* de Mapeamento Local (*Local Mapping*) utiliza o *KeyFrame* para construir e refinar a geometria do mapa ativo, inserindo novos pontos e otimizando a área local através de um *Bundle Adjustment* (*Local BA*). Ao mesmo tempo, a *thread* de Fechamento de Ciclos e Fusão de Mapas (*Loop & Map Merging*) utiliza este mesmo *KeyFrame* para realizar uma busca no banco de dados de reconhecimento de lugar, o DBoW2 (*Database of Binary Words*) (GÁLVEZ-LÓPEZ; TARDÓS, 2012), procurando por similaridades que possam indicar um fechamento de ciclo ou uma oportunidade de fundir mapas. Todas essas operações interagem com o Atlas, que armazena tanto o

mapa que está sendo ativamente construído (*Active Map*) quanto mapas previamente criados (*Non-active Map*), garantindo a estabilidade operacional e a capacidade de reutilização de informações do sistema (CAMPOS et al., 2021).

Com a visão geral do fluxo de dados estabelecida, é possível agora detalhar as responsabilidades e os processos específicos de cada um dos módulos principais que compõem a arquitetura do sistema em quatro funções:

A *thread* de Rastreamento (*Tracking*) funciona como o *front-end* do sistema, processando as informações dos sensores para calcular a pose do *Frame* atual em tempo real. No modo visual-inercial, a velocidade do corpo e os vieses (*biases*) da IMU também são estimados ao se incluir os resíduos inerciais na otimização. Suas principais responsabilidades incluem a extração de características ORB da imagem e a decisão de quando um novo *Frame* deve ser promovido a um *KeyFrame* (*KeyFrame*). Caso o rastreamento seja perdido, a *thread* tenta se relocalizar em todos os mapas contidos no Atlas.

A *thread* de Mapeamento Local (*Local Mapping*) é responsável por adicionar novos *KeyFrame* e pontos ao mapa ativo, remover os redundantes e refinar a geometria do mapa. Esse refinamento é realizado por meio de uma otimização chamada *Bundle Adjustment* (ou BA), que opera em uma janela local de *KeyFrames* próximos à pose atual da câmera. No caso da operação visual-inercial, é também nesta *thread* que os parâmetros da IMU são inicializados e refinados (SONG, 2025; CUCCI, 2017).

A cada novo *KeyFrame*, a terceira *thread* detecta regiões comuns entre o mapa ativo e o Atlas inteiro. Se a área comum pertence ao próprio mapa ativo, o sistema realiza uma Correção de Ciclo (*Loop Correction*). Caso a área pertença a um mapa diferente, os dois mapas são unidos de forma consistente em uma operação de Fusão de Mapas (*Map Merging*), e o mapa resultante se torna o novo mapa ativo. Após uma Correção de Ciclo, um BA Global (*Full BA*), que é uma otimização completa e mais custosa, é lançado em uma *thread* independente para refinar todo o mapa sem afetar o desempenho em tempo real do sistema.

O Atlas é a estrutura de dados central e a principal inovação do ORB-SLAM 3, permitindo o verdadeiro SLAM multi-mapa e multi-sessão. Ele consiste em um conjunto de mapas, onde um é definido como o mapa ativo e os demais são armazenados como mapas não ativos. Essa estrutura permite ao sistema ser estável a longos períodos de perda de rastreamento, simplesmente iniciando um novo mapa

que pode ser fundido posteriormente. O reconhecimento de lugar em todo o Atlas é possibilitado por um banco de dados DBoW2 unificado, que contém uma representação de "banco de elementos visuais" de todos os *KeyFrame* e permite consultas eficientes para relocalização, fechamento de ciclos e fusão de mapas(ZHANG, 2019).

O DESCRITOR DE CARACTERÍSTICAS ORB (*ORIENTED FAST AND ROTATED BRIEF*)

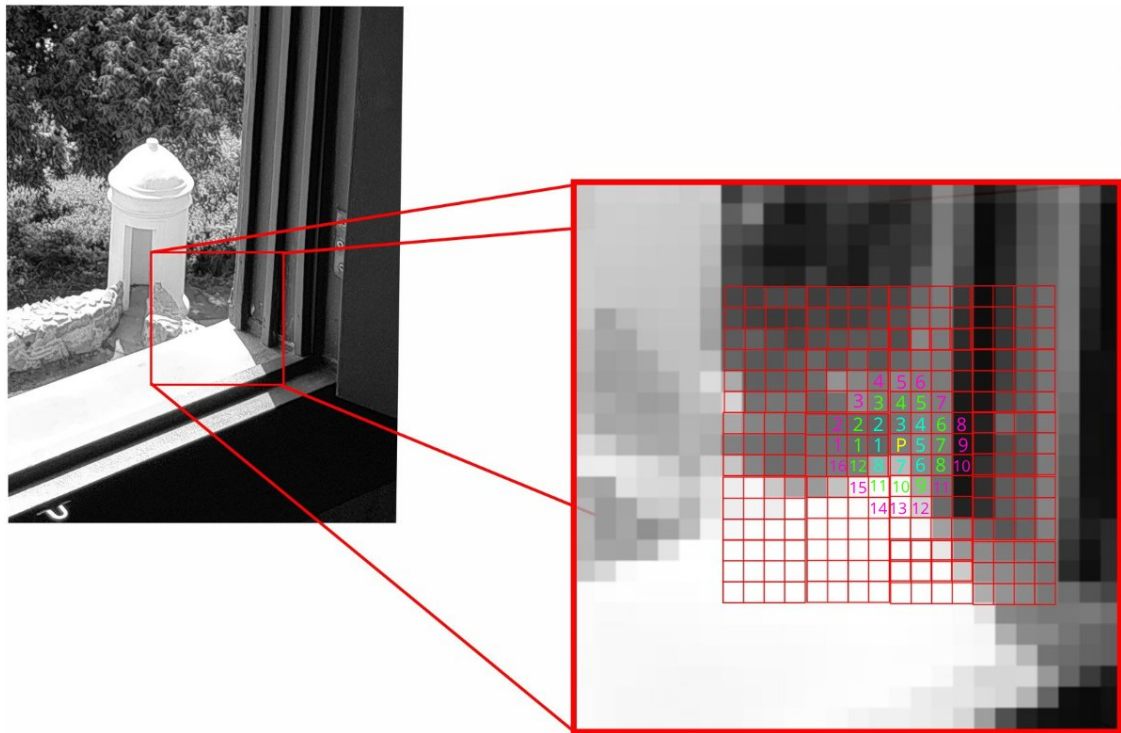
Em sistemas de SLAM que operam com base em características visuais (*feature-based*), a capacidade de detectar pontos de interesse (ou *features*) de forma consistente e de descrevê-los unicamente é um pilar fundamental. Esses pontos funcionam como marcos ou âncoras no ambiente, permitindo que o sistema se localize e construa o mapa. O sistema ORB-SLAM, em todas as suas versões, fundamenta-se no descritor de características que lhe dá o nome: o ORB, acrônimo para *Oriented FAST (Features from Accelerated Segment Test) and Rotated BRIEF (Binary Robust Independent Elementary Features)*.

A popularidade e o sucesso deste descritor em aplicações de robótica em tempo real se devem a uma combinação de vantagens críticas: ele é extremamente rápido de computar e de comparar, é estável a mudanças na rotação da câmera e possui invariância à escala de escala. Conforme o nome sugere, o ORB não é um algoritmo completamente novo, mas sim uma fusão aprimorada e inteligente de duas técnicas já existentes: o detector de cantos *FAST* e o descritor binário *BRIEF* (94).

A primeira etapa do processo ORB é a detecção dos pontos de interesse na imagem, uma tarefa para a qual o ORB utiliza como base o algoritmo FAST. A utilização do FAST se deve à sua altíssima velocidade de processamento, que é um requisito essencial para sistemas de SLAM em tempo real (95).

O princípio do algoritmo, ilustrado na Figura 2, consiste em avaliar se um pixel candidato p é um canto, analisando um círculo de 16 pixels em sua vizinhança, normalmente com um raio de 3 pixels. O FAST então executa um teste de alta velocidade para descartar rapidamente os não-cantos, verificando um subconjunto de pixels opostos no círculo. Se o pixel candidato passa nesse teste inicial, uma verificação completa é feita para confirmar se um arco de N pixels contíguos é significativamente mais claro ou mais escuro que o pixel central(96).

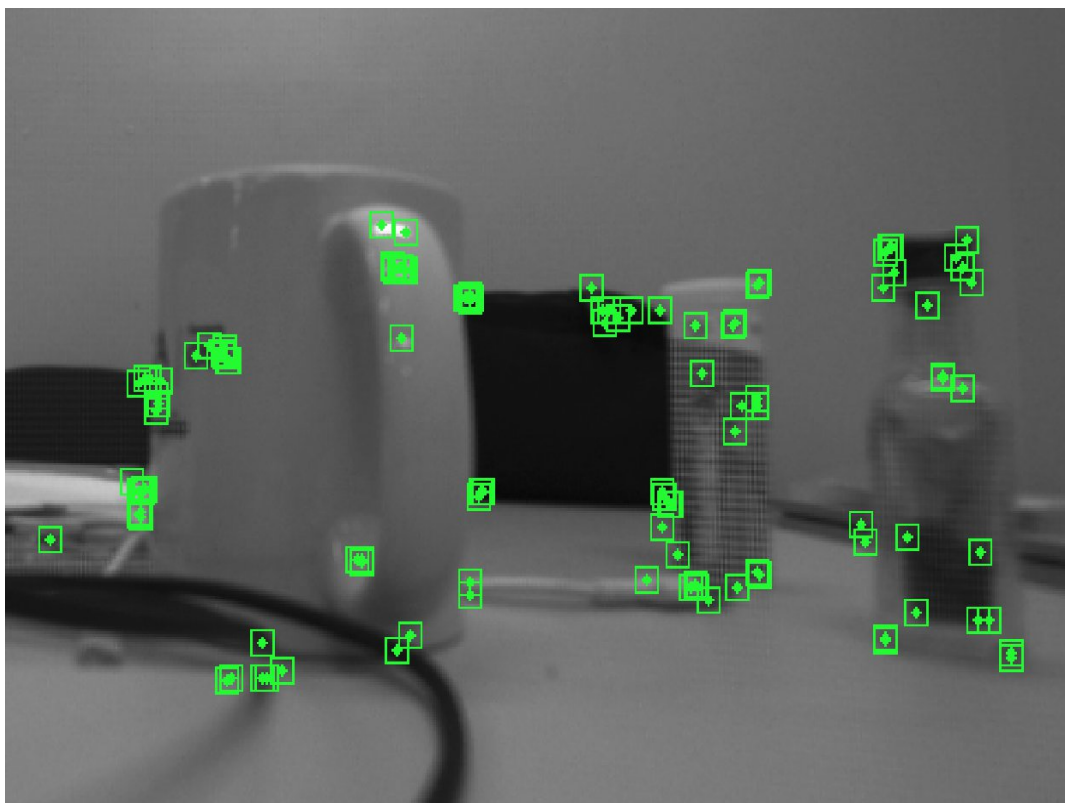
Figura 2 – Diagrama do funcionamento do SLAM.



Fonte: Galvão, 2024 (Modificado)

Embora a implementação padrão no ORB-SLAM utilize parâmetros bem definidos, o algoritmo *FAST* é inerentemente configurável. O raio do círculo de análise e o número de *pixels* contíguos (N) podem ser ajustados (criando variações como FAST-9, FAST-12, etc.) para otimizar a detecção para diferentes tipos de cenas ou resoluções de imagem. Após a detecção inicial, uma etapa de Supressão de Não-Máximos (*Non-Maximal Suppression*) é aplicada para garantir que apenas os cantos mais proeminentes em cada região sejam selecionados(97), resultando em um conjunto de pontos de interesse bem distribuídos, como mostra a Figura 3.

Figura 3 – Exemplo de características visuais detectadas em uma imagem, resultantes da aplicação do detector FAST.



Fonte: Do autor.

É fundamental ressaltar que o FAST atua apenas como um detector. Ele fornece com eficiência as coordenadas dos pontos de interesse, mas não lhes atribui uma orientação nem produz um descritor para identificá-los unicamente. A solução para essa limitação é o próximo passo no *pipeline* do ORB.

Após o algoritmo FAST ter detectado eficientemente a localização dos pontos de interesse, surge uma limitação fundamental: esses pontos não possuem uma orientação definida. Isso os torna vulneráveis a rotações da câmera, pois a aparência da vizinhança de um ponto muda conforme o ângulo de visão, dificultando sua correspondência em imagens futuras.

Para resolver essa deficiência, o ORB introduz sua primeira grande contribuição: uma etapa de orientação, que corresponde ao 'O' (de *Oriented*) em seu acrônimo. O objetivo é atribuir um ângulo canônico para cada característica, de modo que sua descrição futura seja sempre normalizada em relação a essa direção.

Este cálculo de orientação é realizado através do método do Centróide de Intensidade (*Intensity Centroid*). O método analisa uma área circular de *pixels*

(conhecida como *patch*) ao redor do canto detectado(98). A intensidade (brilho) de cada *pixel* nessa área é tratada como um "peso". O centroide é, efetivamente, o "centro de massa" dessa distribuição de intensidades, onde as regiões mais brilhantes "puxam" o centro para mais perto delas. A orientação da característica é então definida pelo ângulo do vetor que liga o centro geométrico do canto a este "centro de brilho". Com um ângulo bem definido para cada ponto, o sistema está pronto para a próxima etapa: criar uma descrição que seja, de fato, confiável à rotação(97).

Uma vez que os pontos de interesse foram localizados pelo FAST e suas orientações determinadas pelo Centroide de Intensidade, a próxima etapa é criar um descritor para cada ponto. Um descritor é uma "impressão digital" ou assinatura vetorial que representa unicamente a aparência da vizinhança de um ponto, permitindo que ele seja comparado e correspondido com pontos em outras imagens. Para esta tarefa, o ORB utiliza uma versão modificada do descritor *BRIEF*(99).

A principal característica e vantagem do *BRIEF* é que ele produz um descritor binário, ou seja, uma sequência de zeros e uns (ex: 101100...). Para construir essa sequência, o BRIEF realiza uma série de testes de comparação de intensidade entre pares de *pixels*, amostrados de locais pré-definidos em uma área ao redor do ponto de interesse. Para cada par de *pixels* (p_1, p_2) o teste definido pela equação 1.

$$\tau(p_1, p_2) = \begin{cases} 1 & \text{se } I(p_1) < I(p_2) \\ 0 & \text{se } I(p_1) \geq I(p_2) \end{cases} \quad (1)$$

Onde $I(p)$ é a intensidade do *pixel*. A repetição desse teste para centenas de pares de *pixels* gera o vetor binário que serve como descritor. A grande vantagem dessa abordagem é que a comparação entre dois descritores para encontrar correspondências pode ser feita de forma extremamente eficiente utilizando a distância de Hamming(100), que conta o número de bits diferentes entre duas sequências), em vez de cálculos mais lentos como a distância Euclidiana, usada por descritores baseados em números reais.

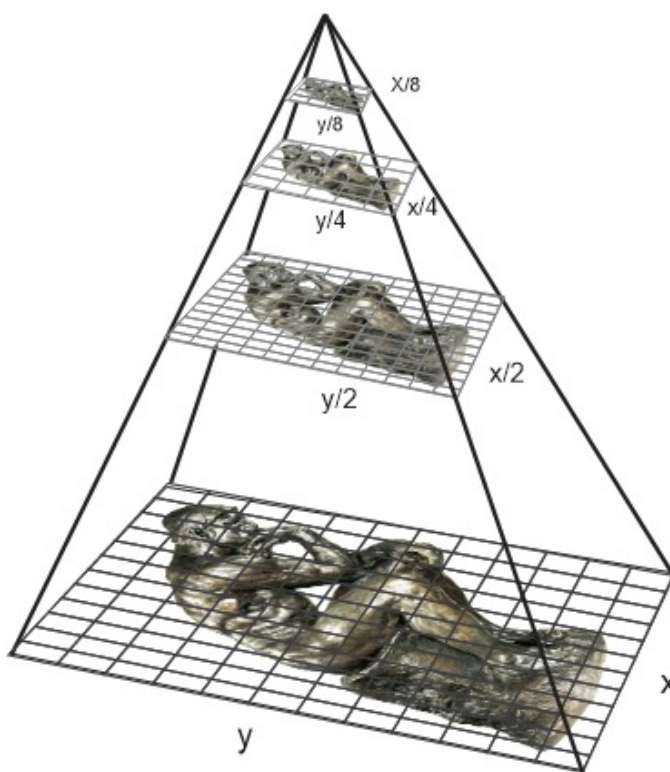
Contudo, o algoritmo BRIEF original é altamente sensível à rotação da câmera, pois o padrão de amostragem dos pares de *pixels* é fixo. A grande contribuição do ORB, que dá origem ao 'R' (de *Rotated*) em seu nome, é resolver essa limitação. Utilizando o ângulo de orientação (θ) calculado na etapa anterior, o ORB rotaciona o padrão de amostragem dos pares de *pixels* por esse mesmo ângulo antes de construir o descritor. Esse processo, conhecido como *steered BRIEF*, garante que o descritor

seja sempre computado em relação à orientação da própria característica, e não à orientação da imagem, conferindo ao ORB sua confiabilidade à rotação(94).

Após resolver o problema da rotação, resta uma última grande invariância a ser alcançada: a de escala. Um ponto de interesse detectado em um objeto parecerá muito maior quando a câmera está próxima e muito menor quando está distante. O descritor *BRIEF*, mesmo rotacionado, não lida nativamente com essa variação de tamanho.

Para conferir estabilidade a mudanças de escala, o ORB emprega uma técnica chamada Pirâmide de Imagens (*Image Pyramid*)(101). Em vez de processar apenas a imagem em sua resolução original, o sistema primeiro constrói uma pilha de versões da mesma imagem, cada uma com uma resolução progressivamente menor, com um fator de escala constante entre os níveis. A estrutura de uma pirâmide de imagens é ilustrada na Figura 4.

Figura 4 – Representação de uma Pirâmide de Imagens.



Fonte: Do autor.

O processo de detecção e descrição de características é então executado de forma independente em cada um dos níveis dessa pirâmide. As características

detectadas nos níveis mais altos da pirâmide (as imagens menores) correspondem a características de maior escala na cena real. Ao buscar por correspondências entre imagens, o sistema compara os descritores de todos os níveis, tornando-se capaz de associar pontos de interesse que são vistos em diferentes distâncias da câmera.

Em conclusão, o descritor ORB estabelece-se como uma solução completa e de alta performance por combinar de forma inteligente várias técnicas poderosas. Ele utiliza o FAST para uma detecção de pontos extremamente rápida, o centroide de Intensidade para atribuir uma orientação confiável, uma versão rotacionada do BRIEF para uma descrição binária eficiente e invariante à rotação, e a Pirâmide de Imagens para alcançar a invariância à escala. É essa combinação de velocidade e estabilidade operacional que torna o ORB a espinha dorsal do algoritmo ORB-SLAM e uma escolha ideal para aplicações de robótica e realidade aumentada em tempo real (102).

APÊNDICE B – RESULTADOS

Este documento complementa a análise quantitativa apresentada no Capítulo 5, servindo como um repositório de dados dos experimentos realizados. Este apêndice disponibiliza o detalhamento de todos os resultados brutos e agregados obtidos nos cenários de teste, incluindo o experimento na guia linear (Configurações de 1 a 4).

Configuração	n° do teste	X (m)	Y (m)	Z (m)	X1 (m)	X2 (m)	Y1 (m)	Y2 (m)	Z1 (m)	Z2 (m)	DESLOCAMENTO (m)	Distancia Percorrida (m)	Ero em mm	Desvio médio de Retitude (m)	Desvio Máximo de Retitude(m)	Ângulo do sistema	
Estéreo, 1600x1200 pixels, 1500 ORBs.	1	2.314	1.712	-0.462	-0.147	2.167	-0.385	-0.097	-0.021	0.441	2.915	3	84.7	0.00114	0.00314	0°	
	2	-0.186	2.895	-0.384	-0.507	-0.321	-0.214	-3.109	-0.234	0.15	2.926	3	73.7	0.00132	0.00561		
	3	-2.909	-0.259	-0.151	0.173	3.082	-0.548	-0.289	-0.136	0.015	2.924	3	75.6	0.00155	0.00764		
	4	-1.962	2.148	-0.44	-0.191	1.771	-0.234	-2.382	-0.067	0.373	2.942	3	57.7	0.00167	0.00724	45°	
	5	-1.484	2.541	0.16	-0.337	1.147	-0.628	-3.169	-0.121	-0.281	2.947	3	53.0	0.00113	0.00360		
	6	-1.796	2.312	0.33	-0.501	1.295	-0.65	-2.962	-0.437	-0.767	2.946	3	53.8	0.00158	0.00503		
	7	0.474	2.915	-0.026	-0.503	-0.029	-0.584	-0.449	-0.249	-0.223	2.953	3	46.6	0.00144	0.00502	90°	
	8	-1.106	2.714	0.308	-0.611	0.495	-0.569	-3.283	-0.287	-0.595	2.947	3	53.2	0.00689	0.02097		
	9	-0.973	2.759	-0.065	-0.468	0.505	-0.329	-3.088	-0.248	-0.183	2.926	3	73.7	0.00374	0.00815		
	Estéreo, 800x600 pixels, 1500 ORBs.	10	-0.369	2.906	0.214	-0.594	-0.225	-0.379	-3.285	-0.129	-0.343	2.937	3	62.9	0.00152	0.00566	45°
		11	-2.047	2.049	0.457	-0.497	1.55	-0.592	-2.641	-0.339	-0.796	2.932	3	67.9	0.00170	0.00628	
		12	-0.021	2.924	0.288	-0.306	-0.285	-0.496	-3.42	-0.241	-0.529	2.938	3	61.8	0.00298	0.00608	
13		0.511	2.87	0.187	-0.181	0.33	-0.633	-3.703	-0.276	-0.463	2.921	3	78.9	0.00251	0.01009	90°	
14		0.527	2.871	0.301	-0.656	-1.183	-0.604	-3.475	-0.198	-0.499	2.934	3	65.6	0.00455	0.01282		
15		-0.191	2.895	-0.042	-0.24	-0.049	-0.864	-3.759	-0.16	-0.118	2.902	3	98.4	0.00159	0.00689		
16		-0.387	2.897	0.061	-0.036	0.351	-0.69	-3.587	-0.251	-0.312	2.923	3	76.6	0.00219	0.00688	0°	
17		-0.123	2.95	0.021	-0.015	0.108	-0.663	-3.613	-0.224	-0.245	2.953	3	47.4	0.00148	0.00708		
18		-0.046	2.927	0.02	-0.623	-0.577	-0.072	-2.999	-0.302	-0.232	2.927	3	72.6	0.00189	0.00744		
Estéreo, 800x600 pixels, 3000 ORBs.		19	2.321	1.722	-0.468	-0.155	2.166	-0.397	-2.119	-0.023	0.445	2.928	3	72.3	0.00184	0.00889	45°
		20	-0.578	2.968	-0.105	-0.407	0.171	-0.229	-3.097	-0.227	-0.122	2.928	3	72.5	0.00180	0.00956	
		21	-2.899	-0.268	-0.137	0.16	3.079	-0.555	-0.287	-0.143	-0.006	2.915	3	85.4	0.00223	0.01273	
	22	-1.99	2.138	-0.442	-0.193	1.797	-0.244	-2.382	-0.069	0.373	2.954	3	45.9	0.00137	0.00967	90°	
	23	-0.685	2.898	-0.065	-0.485	0.2	-0.501	-3.399	-0.002	0.063	2.979	3	21.4	0.00184	0.00724		
	24	-1.401	2.569	0.169	-0.589	0.812	-0.767	-3.336	-0.147	-0.316	2.931	3	68.9	0.01324	0.06583		
	25	0.286	2.944	0.016	-0.37	-0.084	-0.442	-3.386	-0.192	-0.208	2.958	3	42.1	0.00700	0.02970	0°	
	26	-0.465	2.894	0.124	-0.669	-0.204	-0.608	-3.502	-0.23	-0.354	2.934	3	66.3	0.00229	0.01192		
	27	-0.997	2.783	-0.057	-0.443	0.554	-0.299	-3.082	-0.239	-0.162	2.957	3	43.3	0.00220	0.01090		
	28	-0.393	2.899	0.224	-0.615	-0.222	-0.422	-3.321	-0.151	-0.375	2.934	3	65.9	0.00354	0.01007	45°	
	29	-0.16	2.921	-0.051	-0.734	-0.401	-0.574	-3.495	-0.134	-0.083	2.926	3	74.2	0.00335	0.00941		
	30	-0.033	2.919	0.305	-0.299	-0.266	-0.517	-3.436	-0.213	-0.518	2.935	3	64.9	0.00456	0.01004		

ANEXO A – LENTE LM3NCM e CÂMERA FLEA3 USB3



Link: https://www.kowa-lenses.com/LM3NCM-1-1.8-3.5mm-2MP-C-Mount-Lens/10558?srsId=AfmBOoqTO69JxFYEp5T8-1ctvSu1ojJ2CBjs_2epn49zsdwv4Nwo9k-q



Link: <https://www.teledynevisionsolutions.com/products/flea3-usb3/?model=FL3-U3-13E4M-C&vertical=machine%20vision&segment=iis>

ANEXO B – LENTE EDMUND E CÂMERA BLACKFLY S USB3

Link: https://www.edmundoptics.com/p/35mm-c-series-fixed-focal-length-lens/31933/?srsltid=AfmBOordolGWw_jfIXI31zS33W16VAPR-kWPaNyKNiEvgLagdv8j15Lo



Link: <https://www.teledynevisionsolutions.com/products/blackfly-s-usb3/?model=BFS-U3-50S4C-C&vertical=machine%20vision&segment=iis>

ANEXO C – IMU



Link: <https://safran-navigation-timing.com/product/stim300/>

ANEXO D – UNIDADE DE PROCESSAMENTO E ARMAZENAMENTO

Link: <https://avell.com.br/storm-470-blue-avnb25210139?srsItid=AfmBOorEOKZ2IN6v82c6HshQIJ8R6fGwZAFmVZ7qhvoNsKV4Hht2-Ffv>



Link: <https://shop.sandisk.com/products/ssd/external-ssd/sandisk-usb-3-2-ssd?sku=SDSSDE30-1T00-G26>