

DISSERTAÇÃO DE MESTRADO Nº 944

**ESTUDO DE MÉTODOS DE EXTRAÇÃO DE CARACTERÍSTICAS APLICADOS
AO PROBLEMA DA ESTIMAÇÃO DA POSIÇÃO DE UM VANT EM NAVEGAÇÃO
AUTÔNOMA COM VISÃO COMPUTACIONAL**

Juan Camilo Fonseca Galindo

DATA DA DEFESA: 24/10/2016

Universidade Federal de Minas Gerais

Escola de Engenharia

Programa de Pós-Graduação em Engenharia Elétrica

**ESTUDO DE MÉTODOS DE EXTRAÇÃO DE CARACTERÍSTICAS
APLICADOS AO PROBLEMA DA ESTIMAÇÃO DA POSIÇÃO DE
UM VANT EM NAVEGAÇÃO AUTÔNOMA COM VISÃO
COMPUTACIONAL**

Juan Camilo Fonseca Galindo

Dissertação de Mestrado submetida à Banca Examinadora designada pelo Colegiado do Programa de Pós-Graduação em Engenharia Elétrica da Escola de Engenharia da Universidade Federal de Minas Gerais, como requisito para obtenção do Título de Mestre em Engenharia Elétrica.

Orientador: Prof. André Paim Lemos

Belo Horizonte - MG

Outubro de 2016

G158e

Galindo, Juan Camilo Fonseca.

Estudo de métodos de extração de características aplicados ao problema da estimação da posição de um VANT em navegação autônoma com visão computacional [manuscrito] / Juan Camilo Fonseca Galindo. – 2016.

x, 69 f., enc.: il.

Orientador: André Paim Lemos.

Dissertação (mestrado) Universidade Federal de Minas Gerais, Escola de Engenharia.

Bibliografia: f. 63-69.

1. Engenharia elétrica - Teses. 2. Auxílio a navegação aérea - Teses. 3. Navegação inercial (Aeronáutica) - Teses. 4. Aeronave não tripulada - Teses. I. Lemos, André Paim. II. Universidade Federal de Minas Gerais. Escola de Engenharia. III. Título.

CDU: 621.3(043)

**"Estudo de Métodos de Extração de Características Aplicados
ao Problema da Estimação da Posição de um Vant em
Navegação Autônoma com Visão Computacional"**

Juan Camilo Fonseca Galindo

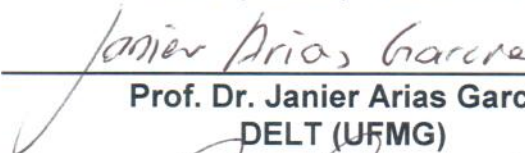
Dissertação de Mestrado submetida à Banca Examinadora designada pelo Colegiado do Programa de Pós-Graduação em Engenharia Elétrica da Escola de Engenharia da Universidade Federal de Minas Gerais, como requisito para obtenção do grau de Mestre em Engenharia Elétrica.

Aprovada em 24 de outubro de 2016.

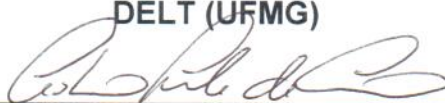
Por:



Prof. Dr. Andre Paim Lemos
DELT (UFMG) - Orientador



Prof. Dr. Janier Arias Garcia
DELT (UFMG)



Prof. Dr. Cristiano Leite de Castro
DEE (UFMG)

Resumo

Este trabalho apresenta um estudo comparativo de métodos de extração de características aplicado à estimativa de posição para a navegação autônoma de VANTs. A partir do estudo inicial dos algoritmos existentes, esse trabalho também apresenta duas propostas de métodos de extração de características invariantes à transformada afim e de baixo custo computacional. A primeira proposta é uma modificação do algoritmo ASIFT. O método utiliza os dados fornecidos pelo INS para reduzir o custo computacional desse algoritmo. A segunda proposta também é baseada no princípio de invariância à transformada afim do método ASIFT, porém utiliza o método SURF para extração de possíveis pontos característicos. Os métodos propostos são comparados com SIFT, SURF e ASIFT para registro de imagens aéreas capturadas a partir de um VANT.

As métricas de avaliação utilizadas no estudo comparativo para medir o desempenho dos extratores de características foram a quantidade de pontos extraídos em cada uma das imagens, a quantidade de pontos casados, o erro do casamento a partir da distância da projeção dos pontos e o custo computacional medido através do tempo de execução. Os experimentos numéricos mostraram que as propostas são capazes de melhorar a precisão e reduzir a complexidade do algoritmo ASIFT, o qual indica que podem ser usados para navegação autônoma em VANT.

Abstract

This work presents a comparative study of feature extraction methods applied to the problem of position estimation for autonomous navigation of UAVs. Based on an initial study of the existing methods, this work also proposes two novel feature extraction algorithms invariant to the affine transformation and with a low computational cost. The first proposal is a modification of ASIFT algorithm. The method utilizes the data provided by the INS to reduce its computational cost. The second proposal is also based on the affine invariance property of the ASIFT algorithm, but uses the SURF method for extracting candidate feature points. The proposed methods are compared with SIFT, SURF and ASIFT to aerial images registration taken from UAVs.

The evaluation metrics used in the comparative study to measure the performance of feature extractors are the amount of extracted points in each image, the amount of matched points, the distance-error of matched points, and the computational cost measured by the time of execution. The numerical experiments showed that the proposals are able to improve the accuracy and reduce the complexity of ASIFT algorithm, which means that can be used for autonomous navigation of UAVs.

Dedico mi título de "Mestre em Engenharia Elétrica" a mi abuelo Fabio Galindo y mi abuelita Felisa Jauregui, dos personas especiales que tuve el placer de disfrutar en vida y por motivos naturales de la vida, partieron de mi lado en el transcurso de mi maestría.

“A picture is worth a thousand words.”

Anonimo

Agradecimentos

A Deus por dar-me a oportunidade de realizar minha maestria no Brasil, especialmente na UFMG.

A minha família, especialmente a meus pais por todo o apoio pessoal ao longo de meu mestrado.

A meu orientador, André Paim Lemos, por sua paciência e seu suporte profissional.

Ao professor Elcio Shiguemori por todo o tempo e conhecimento dado.

Aos professores e colegas do LITC: Cristiano de Leite Castro, Brayan Rene Acevedo, Carlos Anderson, Ramon Corrêa, Yuri Sousa, Luiz Carlos Bambirra, Honovan Rocha e Frederico Coelho, por todas as contribuições dadas ao longo de meu mestrado. Agradeço especialmente ao professor Antônio Braga por seu compromisso no grupo de pesquisa e por me-permitir desenvolver meu trabalho de mestrado.

A meus amigos Colombianos, especialmente a Jaime Arturo Dulce, Juan José Quiroz, Edna Rocio Triana, Maria Victoria Africano e Diego Camilo Tami.

À CAPES pelo incentivo financeiro concedido, que foi imprescindível para a dedicação integral ao mestrado.

Sumário

Lista de Abreviaturas	ix
1 Introdução	1
1.1 Motivação	1
1.2 Principais Contribuições deste Trabalho	4
1.3 Organização do Texto	5
2 Fundamentação Teórica	6
2.1 Características Visuais	7
2.2 SIFT (Scale Invariant Feature Transformation)	9
2.2.1 Detecção de extremos no espaço/escala	9
2.2.2 Localização exata do ponto de interesse	12
2.2.3 Alocação de Orientação	14
2.2.4 Descritor dos pontos de interesse	15
2.2.5 Casamento SIFT	17
2.3 SURF (Speeded Up Robust Features)	17
2.3.1 Detector Fast-Hessian	18
2.3.2 Descritor SURF	19
2.4 ASIFT (Affine-SIFT)	21
2.4.1 Algoritmo ASIFT	22
2.5 RANSAC (Random Sample Consensus)	23
2.5.1 Distância limite (threshold)	26
2.5.2 Número de iterações	27
2.5.3 Tamanho do conjunto consenso	28
2.6 Resumo do capítulo	28
3 Propostas	29
3.1 ASIFT-V (Affine-SIFT-VANT)	29
3.2 ASURF-V (Affine-SURF-VANT)	32
3.3 Resumo do capítulo	33

4	Experimentos e Resultados Computacionais	37
4.1	Banco de Imagens	37
4.2	Equipamento e Software	40
4.3	Métricas de Avaliação	41
4.4	Análise Estatística	41
4.5	Resultados Experimentais	42
4.5.1	Análise estatística para θ de 0°	44
4.5.2	Análise estatística para θ de 16°	47
4.5.3	Análise estatística para θ de 32°	50
4.5.4	Análise estatística para θ de 48°	54
4.5.5	Análise estatística para θ de 64°	57
4.5.6	Análise estatística para θ de 76°	59
4.6	Resumo do capítulo	60
5	Conclusões e Trabalhos Futuros	61
	Referências Bibliográficas	63

Lista de Figuras

2.1	Sistema de navegação baseado em visão computacional.	6
2.2	Construção de $D(x, y, \sigma)$ para cada oitava de espaço/escala, a imagem inicial é repetidamente convolvida com gaussianas para produzir um conjunto de imagens espaço/escala mostrado na esquerda. Imagens gaussianas adjacentes são subtraídas para produzir diferença-de-Gaussianas, mostradas à direita. Adaptada de Lowe (2004).	11
2.3	26 pixels vizinhos analisados para detectar os máximos e mínimos da diferença-de-Gaussianas. Adaptada de Lowe (2004).	12
2.4	Descritor de ponto de interesse do método SIFT. Adaptada de Lowe (2004). . .	16
2.5	Filtro <i>Box</i> . Derivada parcial de segunda ordem da função gaussiana em imagens discretas na direção y (Figura 2.5a) e na direção xy (Figura 2.5b). A Figura 2.5c e 2.5d representam a aproximação do filtro <i>Box</i> da Figura 2.5a e 2.5b, respectivamente. Adaptada de Bay et al. (2006).	18
2.6	Transformada <i>Haar – wavelet</i> em direção x (Figura 2.6a) e direção y (Figura 2.6b).	19
2.7	Propriedades do descritor em sub-regiões para três imagens com diferentes padrões de intensidade. Adaptada de Bay et al. (2006).	21
2.8	Interpretação geométrica dos parâmetros da transformada afim.	21
2.9	As amostras dos parâmetros $\theta = \arccos(1/t)$ e ϕ realizadas pelo algoritmo ASIFT. As amostragens são representadas pelos pontos pretos da imagem.	23
2.10	Representação do ponto \mathbf{X} que encontra-se dentro de uma superfície planar, no plano da Imagem ₁ (\mathbf{x}) e da Imagem ₂ (\mathbf{x}').	24
2.11	Casamento de imagens, as linhas a cor representam os pontos casados entre as duas imagens. A Figura 2.11a e Figura 2.11b são a implementação do método SIFT com, e sem filtragem de <i>outliers</i> RANSAC, respectivamente. A Figura 2.11c e Figura 2.11d são a implementação do método ASIFT com, e sem filtragem de <i>outliers</i> RANSAC, respectivamente.	25
2.12	Transformação projetiva, estimação 2D. Adaptada de Hartley and Zisserman (2003).	27

3.1	Conversão dos ângulos Euler.	30
3.2	Representação das simulações realizadas pelo ASIFT-V.	30
3.3	Diagrama de Fluxo do primeiro passo de ASIFT-V paralelizado	32
3.4	Diagrama de Fluxo do algoritmo ASIFT-V	33
3.5	Diagrama de Fluxo do algoritmo ASURF-V	35
4.1	Imagem Satélite.	38
4.2	Imagens cenários.	39
4.3	Transformações de uma imagem cenário, (a) $\theta = 0^\circ$, $\phi = 0^\circ$, $\psi = 0^\circ$, (b) $\theta = 0^\circ$, $\phi = 100^\circ$, $\psi = 0^\circ$, (c) $\theta = 16^\circ$, $\phi = 100^\circ$, $\psi = 90^\circ$, (d) $\theta = 48^\circ$, $\phi = 100^\circ$, $\psi = 150^\circ$ e (e) $\theta = 64^\circ$, $\phi = 100^\circ$, $\psi = 150^\circ$	40
4.4	Análise gráfica de cada um dos métodos com θ de 0° , os dados são representados por boxplots. A Figura 4.4a apresenta a distância-erro ² em pixels ² obtida pelo casamento final dos pontos de interesse. A Figura 4.4b apresenta o tempo de execução em segundos que utilizou cada um dos métodos em fazer a extração de pontos, o casamento e a filtragem de pontos.	45
4.5	Análise Post-hoc para avaliar a distância-erro ² dos métodos ASIFT, ASIFT-V, ASURF-V, SIFT e SURF, com θ de 0°	46
4.6	Análise Post-hoc para avaliar o tempo de execução dos métodos ASIFT-V, ASURF-V, SIFT e SURF, com θ de 0°	47
4.7	Análise gráfica de cada um dos métodos com θ de 16° , os dados são representados por boxplots. A Figura 4.7a apresenta a distância-erro ² em pixels ² obtida pelo casamento final dos pontos de interesse. A Figura 4.7b apresenta o tempo de execução em segundos que utilizou cada um dos métodos em fazer a extração de pontos, o casamento e a filtragem de pontos.	48
4.8	Análise Post-hoc para avaliar a distância-erro ² dos métodos ASIFT, ASIFT-V, ASURF-V, SIFT e SURF, com θ de 16°	49
4.9	Análise Post-hoc para avaliar o tempo de execução dos métodos ASIFT-V, ASURF-V, SIFT e SURF, com θ de 16°	50
4.10	Análise gráfica de cada um dos métodos com θ de 32° , os dados são representados por boxplots. A Figura 4.10a apresenta a distância-erro ² em pixels ² obtida pelo casamento final dos pontos de interesse. A Figura 4.10b apresenta o tempo de execução em segundos que utilizou cada um dos métodos em fazer a extração de pontos, o casamento e a filtragem de pontos.	51
4.11	Análise Post-hoc para avaliar a distância-erro ² dos métodos ASIFT, ASIFT-V, ASURF-V, SIFT e SURF, com θ de 32°	52
4.12	Análise Post-hoc para avaliar o tempo de execução dos métodos ASIFT, ASIFT-V, ASURF-V, SIFT e SURF, com θ de 32°	53

4.13	Análise gráfica de cada um dos métodos com θ de 48° , os dados são representados por boxplots. A Figura 4.13a apresenta a distância-erro ² em pixels ² obtida pelo casamento final dos pontos de interesse. A Figura 4.13b apresenta o tempo de execução em segundos que utilizou cada um dos métodos em fazer a extração de pontos, o casamento e a filtragem de pontos.	54
4.14	Análise Post-hoc para avaliar a distância-erro ² dos métodos ASIFT, ASIFT-V, ASURF-V, SIFT e SURF, com θ de 48°	55
4.15	Análise Post-hoc para avaliar o tempo de execução dos métodos ASIFT, ASIFT-V, ASURF-V, SIFT e SURF, com θ de 48°	56
4.16	Análise gráfica de cada um dos métodos com θ de 64° , os dados são representados por boxplots. A Figura 4.16a apresenta a distância-erro ² em pixels ² obtida pelo casamento final dos pontos de interesse. A Figura 4.16b apresenta o tempo de execução em segundos que utilizou cada um dos métodos em fazer a extração de pontos, o casamento e a filtragem de pontos.	57
4.17	A Figura 4.17a mostra a análise Post-hoc para avaliar a distância-erro ² dos métodos ASIFT, ASIFT-V e ASURF-V, com θ de 64°	58
4.18	Análise gráfica de cada um dos métodos com θ de 76° , os dados são representados por boxplots. A Figura 4.18a apresenta a distância-erro ² em pixels ² obtida pelo casamento final dos pontos de interesse. A Figura 4.18b apresenta o tempo de execução em segundos que utilizou cada um dos métodos em fazer a extração de pontos, o casamento e a filtragem de pontos.	59

Lista de Tabelas

2.1	Número de N iterações requeridas para garantir, com uma probabilidade $p = 0,99$, uma boa estimativa do modelo matemático. No caso de casamento de imagens $s = 4$. Adaptada de Hartley and Zisserman (2003).	28
3.1	Valores dos ângulos θ e ϕ simulados pelo método ASIFT.	31
4.1	Medida pontual e o desvio padrão da média da quantidade de pontos característicos obtidos pelos métodos em cada uma das imagens, da quantidade dos pontos casados após a filtragem do RANSAC e a porcentagem de casamento.	43
4.2	Análise Post-hoc de superioridade e inferioridade todos contra todos, para avaliar a distância-erro ² dos métodos ASIFT, ASIFT-V, ASURF-V, SIFT e SURF, com θ de 0°	45
4.3	Análise Post-hoc de superioridade e inferioridade todos contra todos, para avaliar o tempo de execução dos métodos ASIFT-V, ASURF-V, SIFT e SURF, com θ de 0°	46
4.4	Análise Post-hoc de superioridade e inferioridade todos contra todos, para avaliar a distância-erro ² dos métodos ASIFT, ASIFT-V, ASURF-V, SIFT e SURF, com θ de 16°	49
4.5	Análise Post-hoc de superioridade e inferioridade todos contra todos, para avaliar o tempo de execução dos métodos ASIFT-V, ASURF-V, SIFT e SURF, com θ de 16°	49
4.6	Análise Post-hoc de superioridade e inferioridade todos contra todos, para avaliar a distância-erro ² dos métodos ASIFT, ASIFT-V, ASURF-V, SIFT e SURF, com θ de 32°	52
4.7	Análise Post-hoc de superioridade e inferioridade todos contra todos, para avaliar o tempo de execução dos métodos ASIFT, ASIFT-V, ASURF-V, SIFT e SURF, com θ de 32°	53
4.8	Análise Post-hoc de superioridade e inferioridade todos contra todos, para avaliar a distância-erro ² dos métodos ASIFT, ASIFT-V, ASURF-V, SIFT e SURF, com θ de 48°	55

4.9	Análise Post-hoc de superioridade e inferioridade todos contra todos, para avaliar o tempo de execução dos métodos ASIFT, ASIFT-V, ASURF-V, SIFT e SURF, com θ de 48°	56
4.10	Análise Post-hoc de superioridade e inferioridade todos contra todos, para avaliar a distância-erro ² dos métodos ASIFT, ASIFT-V e ASURF-V com θ de 64°	58
4.11	Comparação dos extractores de características SIFT, SURF, ASIFT, ASIFT-V e ASURF-V.	60

Lista de Abreviaturas

AKAZE	<i>Accelerated KAZE</i>
ASIFT	<i>Affine-SIFT</i>
ASIFT-V	<i>Affine-SIFT-VANT</i>
ASURF-V	<i>Affine-SURF-VANT</i>
BREAF	<i>Binary Robust Independent Elementary Features</i>
CTA	Centro Tecnológico Aeroespacial
FAST	<i>Features from Accelerated Segment Test</i>
GNSS	<i>Global Navigation Satellite System</i>
GPS	<i>Global Positioning System</i>
GPU	<i>Graphics Processing Unit</i>
INS	<i>Inertial Navigation System</i>
OpenCV	<i>Open Source Computer Vision Library</i>
OpenMP	<i>Open Multi-Processing</i>
SIFT	<i>Scale Invariant Feature Transformation</i>
SURF	<i>Speed Up Robust Feature</i>
RANSAC	<i>Random Sample Consensus</i>
UAV	<i>Unmanned Aerial Vehicle</i>
VANT	Veículo Aéreo Não Tripulado
VSANT	Veículo de Superfície Aquática Não Tripulado
VTNT	Veículo Terrestres Não Tripulado

Lista de Algoritmos

1	RANSAC	28
2	Paralelização do primeiro passo de ASIFT-V	31
3	Algoritmo ASIFT-V	34
4	Algoritmo ASURF-V	36

Capítulo 1

Introdução

1.1 Motivação

Avanços recentes na tecnologia computacional, desenvolvimento de software, materiais mais leves, sistemas globais de navegação, links de dados, sofisticados sensores e a miniaturização propiciaram o aumento de desenvolvimentos de plataformas autônomas como veículos aéreos (VANTs), terrestres (VTNTs) e de superfície aquática (VSANTs) não-tripulado, sondas de exploração espacial e mísseis. Estas plataformas são utilizadas em tarefas nas quais a intervenção direta de um ser humano não seja necessária ou se mostre inadequada por razões de segurança ou em locais de difícil acesso, representa um grande desafio científico e tecnológico (Jorge et al. (2011)). A robótica, eletrônica e visão computacional são áreas em que tais desafios têm grande impacto.

Atualmente os VANTs têm sido utilizados em aplicações civis e militares, tais como: operações de vigilância, reconhecimento, monitoramento ambiental e agrícola, aerolevantamento, transporte de carga, ataques a alvos no solo, etc (Gonçalves et al. (2015)). No Japão mais de 2000 VANTs são aplicados em pulverização de lavouras e em outras aplicações agrícolas. Já no Brasil, os primeiros relatos de VANT's ocorreram na década de 80, quando o Centro Tecnológico Aeroespacial (CTA) desenvolveu o projeto Acauã para fins militares (Jorge et al. (2011)). Eles são utilizados em operações de vigilância de fronteiras, monitoramento de usinas hidrelétricas e em grandes eventos, e uma empresa anunciou em 2011 operações de monitoramento de gasodutos utilizando VANTs (Chaves (2012)).

Um sistema autônomo de navegação muito aplicado em VANTs é descrito em Goltz and Shiguemori (2008). O sistema é formado pela associação de um sistema inercial de navegação (*Inertial Navigation System* - INS) com dados de um sistema de posicionamento (*Global Navigation Satellite System* - GNSS). O único sistema totalmente operacional em escala mundial é o sistema americano *Global Positioning System*-GPS

(GNS (2016)), mas existem problemas na integridade desse sistema. Se o sinal de GPS, por algum motivo, se torna indisponível ou corrompido, a solução seria estimar o deslocamento do VANT através do INS. Entretanto, os resultados dessa estimativa divergem rapidamente devido a uma acumulação da medida do viés. O sinal de GPS também se torna pouco confiável quando está operando perto de obstáculos devido às reflexões multi-caminho. Outros problemas relacionados à confiabilidade do GPS consistem na exposição a interferências, especialmente quando o GPS opera em frequências civis, além disso, dispositivos bloqueadores de GPS são encontrados com certa facilidade no mercado. Portanto, os VANTs que utilizam somente o sinal de GPS são vulneráveis a ações maliciosas, o que torna necessário a utilização de um sistema de navegação autônomo capaz de lidar com falhas do GPS de curto e longo prazo.

Uma solução é empregar um sistema de navegação baseado em visão computacional o qual fornece a posição do VANT através da captura de imagens do solo. Existem três técnicas principais implementadas na navegação aérea por visão computacional. A primeira consiste em fazer um casamento de marcos ou "*landmarks*" nas imagens capturadas, e desta forma proporcionar o posicionamento do VANT, esta técnica já foi implementada em Silva Filho et al. (2014); Wang et al. (2014); Sineglazov (2014); Pinage et al. (2012). A segunda, consiste no casamento de uma imagem georreferenciada da região a voar, a qual pode ser uma imagem mosaico ou uma imagem satélite, com a imagem capturada pelo VANT, esta técnica foi implementada em Conte and Doherty (2008); Lin and Medioni (2007). Finalmente, a terceira técnica utilizada para navegação autônoma é a odometria, esta técnica faz um casamento de características em imagens capturadas de forma sequencial, deste modo, pode estabelecer o deslocamento espacial do VANT, esta técnica foi implementada em Andert et al. (2014); Chunhui et al. (2014); Roger-Verdeguer et al. (2012)). Além destas técnicas, existem trabalhos como Braga et al. (2015) que propõem outras técnicas de navegação. Esses sistemas utilizam métodos de extração de características e métodos de casamento de imagens para estimar o posicionamento do VANT. Os métodos de extração de características mais comuns são SIFT (*Scale – Invariant Feature Transform*) e SURF (*Speeded Up Robust Features*). O método de extração de características SIFT tem sido amplamente utilizado em aplicações em VANTs (Lyu et al., 2015; Castro et al., 2009; Chen and Meng, 2013; Moranduzzo and Melgani, 2012; Kwon et al., 2012; Li and Aouf, 2012; Suzuki et al., 2010; Cesetti et al., 2010; Canhoto et al., 2009). SIFT foi testado por Goltz and Shiguemori (2008) em aplicações de correspondência entre imagens aéreas que simulam o imageamento de um VANT. Em Canhoto et al. (2009), propõe-se um sistema de navegação baseado em um sistema de visão sem dependência do GPS. Esse trabalho apresenta um método para obter uma estimativa de deslocamento de um VANT usando imagens aéreas sequenciais. O método SURF já foi implementado

em problemas de navegação aérea, como seguimento e detecção de alvos (Sun et al. (2015)), desembarque (Martinez et al. (2013)), criação de mosaicos (Buyukyazi et al. (2013)), e navegação em tempo real utilizando placas gráfica (Guan and Bai (2012)). Li and Aouf (2012) realizou uma análise comparativa entre as duas técnicas com foco na navegação autônoma. Adicionalmente, Li et al. (2012) abordaram os métodos SIFT e SURF para o mapeamento de texturas 3D.

Um problema presente nas aplicações de navegação autônoma são os ângulos entre a imagem capturada pelo VANT e a imagem de referência (normalmente imagem mosaico ou imagem satelital para *landmarks* e casamento de imagens, ou da sequência de imagens na odometria). Existe a necessidade de garantir que as duas imagens estejam nos mesmos ângulos da transformada homográfica para obter um bom casamento e, na prática pode-se não conseguir capturar todas as imagens com os mesmos ângulos. Um método de extração de características, invariante à transformada afim é ASIFT (*Affine-SIFT*). Esse método já foi aplicado por Fernández-Hernandez et al. (2015) em modelagem 3D, Ye et al. (2015) em casamento de imagens sonar e González-Jorge et al. (2014) em fotogrametria. No entanto, o custo computacional é muito alto para navegação autônoma.

Com o objetivo de diminuir o custo computacional para a navegação autônoma em VANTs, uma solução implementada é utilizar algoritmos paralelizados sobre unidade de processamento gráfico (GPU). Benini et al. (2016) implementaram um sistema para pousar um VANT sobre um *marker* usando processamento de imagens em paralelo embarcado em um NVIDIA Jetson TK1. Radecki et al. (2015) aplicaram processamento de imagens SAR (Radar de Abertura Sintética) em um computador portátil baseado em uma placa CARMA DevKit board, equipada com um processador NVIDIA Tegra 3 e uma placa de vídeo NVIDIA Quadro 1000M. Guan and Bai (2012) apresentaram um sistema de navegação em VANT baseado em *landmarks*, embarcado em uma NVIDIA Geforce 240M. Camargo et al. (2010) propuseram um algoritmo para a geração de imagens mosaicos de super-resolução implementado em uma GPU.

Este trabalho propõe dois métodos para extrair características com foco no posicionamento do VANT quando as imagens são capturadas em diferentes ângulos da transformada homográfica. É importante mencionar, que o tempo de processamento dos métodos deve que ser baixo para garantir um fluxo contínuo de imagens na navegação autônoma de VANTs. O primeiro método proposto chamado ASIFT-V (*Affine-SIFT-VANT*) é baseado em ASIFT para a invariância à transformada afim, entretanto, reduz o custo computacional com a implementação dos dados fornecidos pelo INS, além do método ser paralelizado. O segundo método proposto chamado ASURF-V (*Affine-SURF-VANT*) implementa a invariância da transformada afim análoga ao primeiro método proposto e utiliza o extrator de pontos característicos SURF, com o objetivo

de reduzir o custo computacional. Para validar as soluções propostas, esse trabalho apresenta uma análise comparativa dos extratores de características SIFT, ASIFT, SURF, e os dois exatores propostos ASIFT-V e ASURF-V, avaliando a robustez e o custo computacional, quando são apresentadas mudanças de perspectiva entre as duas imagens. Os métodos são utilizados para casar imagens georreferenciadas da região de São José dos Campos - SP, Brasil. Os cenários utilizados, são imagens representativas do setor urbano e setor rural. As imagens do setor urbano contêm muitas características, incluindo quarteirões com diferentes tipos de casas delimitadas por ruas, trechos de avenidas principais e rodovias, praças esportivas, complexos de pontes, campos de futebol, estádios, armazéns, blocos de edifícios, lotes de construção de cimento e de pasto, diferentes tipos de casas e áreas verdes incluindo árvores. As imagens representativas do setor rural contêm poucas características, incluem áreas verdes com pasto, diferentes vegetações e algumas ruas não pavimentadas. As imagens de voo apresentam variações de perspectiva, rotação e translação. Para calcular a robustez dos métodos, analisa-se a quantidade de pontos obtidos, a distância erro dos pontos casados e o custo computacional medido através do tempo de processamento.

1.2 Principais Contribuições deste Trabalho

Dado o problema de não se poder garantir que a imagem capturada pelo VANT esteja nos mesmos ângulos da transformada homográfica em que foi capturada a imagem territorial, as principais contribuições deste trabalho são:

- Implementação e disponibilização de três extratores de características com foco na navegação aérea implementado com bibliotecas OpenCV 3,0 (OCV (2016)) e OpenMP (OMP (2016)), desenvolvidos em C++:
 - ASIFT
 - ASIFT-V: Otimização do método ASIFT aproveitando os ângulos fornecidos pelo INS.
 - ASURF-V: Otimização do método proposto ASIFT-V utilizando o extrator de características SURF para reduzir o tempo computacional.
- Uma análise comparativa dos métodos SIFT, SURF, ASIFT, ASIFT-V e ASURF-V com foco na estimação do posicionamento do VANT em navegação autônoma.

- Publicação do artigo: Fonseca-Galindo, J.C. and Lemos, A.P. "Proposta de um método de extração de características aplicado ao problema da estimação da posição de um VANT em navegação autônoma" **XXI Congresso brasileiro de automática**, 2016.
- Publicação do artigo: Fonseca-Galindo, J.C.; Lemos, A.P. and Shiguemori, E.H. "Novo extrator de características aplicado à estimação da posição de um VANT em navegação autônoma com visão computacional" **XIX Encontro Nacional de Modelagem Computacional**, 2016.

1.3 Organização do Texto

No capítulo 2 é apresentada uma visão geral sobre o problema de navegação baseado em visão computacional. Adicionalmente, são detalhadas as técnicas de extração de características SIFT, SURF e ASIFT, assim como o método RANSAC (*RANdon SAmple Consensus*) para a filtragem de *outliers* utilizado nos extratores.

Apresenta-se duas propostas de extração de características para o posicionamento de um VANT em navegação aérea no capítulo 3. A primeira proposta chamada ASIFT-V é uma variação do método ASIFT utilizando os dados fornecidos pelo INS, a segunda proposta chamada ASURFT-V é uma variação da primeira proposta utilizando como base o algoritmo SURF.

Usando os dois métodos propostos, mas os métodos SIFT, SURF e ASIFT, os resultados dos experimentos computacionais e as análises das métricas de avaliação são apresentados no capítulo 4.

Finalmente, o capítulo 5 faz uma discussão sobre os resultados, as conclusões desta dissertação, suas contribuições e indica trabalhos futuros.

Capítulo 2

Fundamentação Teórica

Com o objetivo de se fornecer uma visão geral de um sistema de navegação baseado em métodos de visão computacional pode-se tomar como base o sistema de navegação absoluta ilustrado pela Figura 2.1, descrito em Sanfourche et al. (2012). O sistema tem dois módulos principais: o módulo de processamento de imagens e o módulo de filtragem probabilístico. Complementarmente o sistema possui uma câmera que é utilizada para capturar imagens do solo em tempo real e alguns sensores embarcados, como por exemplo, giroscópio, acelerômetro e magnetômetro que entregam medidas inerciais para determinar o estado em termos de posição, velocidade e altitude.

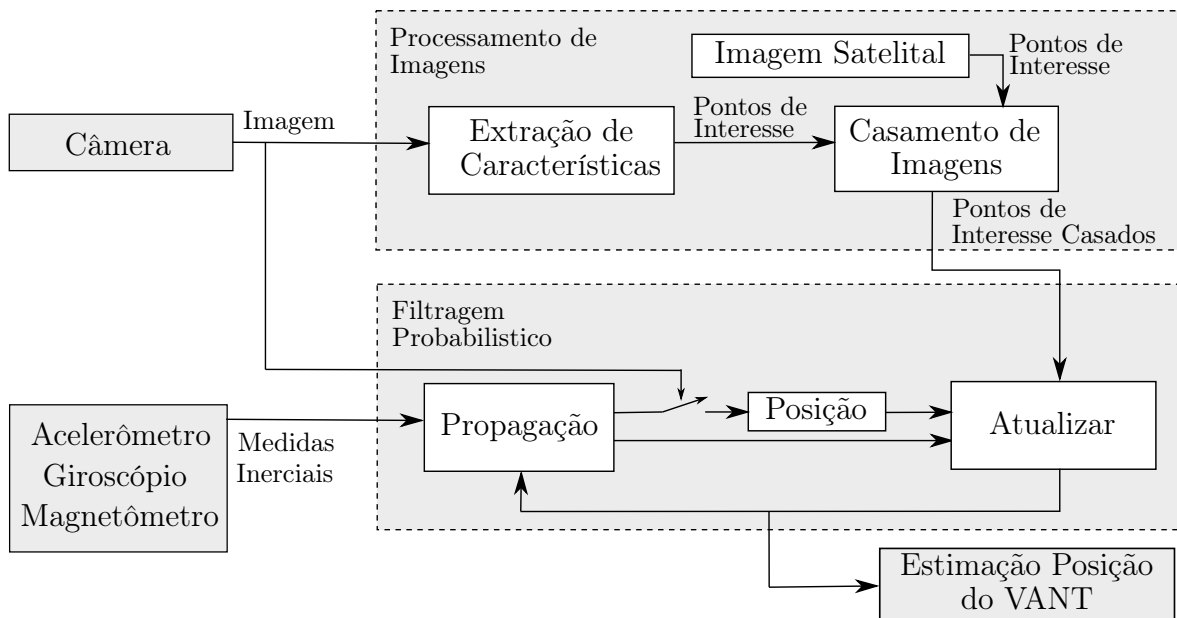


Figura 2.1: Sistema de navegação baseado em visão computacional.

O módulo de processamento de imagens tem como entrada a imagem capturada pela câmera. A imagem é processada utilizando métodos de extração de características ou pontos de interesse, como por exemplo SIFT (Lowe (2004)), SURF (Bay et al. (2006)),

FAST (*Features from Accelerated Segment Test*, Rosten and Drummond (2006)), ASIFT (Morel and Yu (2009)), BRIEF (*Binary Robust Independent Elementary Features*, Calonder et al. (2010)), ORB (*Oriented FAST and Rotated BRIEF*, Rublee et al. (2011)), KAZE (Alcantarilla et al. (2012)) e AKAZE (*Accelerated KAZE*, Alcantarilla et al. (2013)), em seguida, é casada com uma imagem satélite georreferenciada do território voado armazenada no VANT, para indicar seu posicionamento. Uma técnica como RANSAC (Fischler and Bolles (1981)) é utilizada para estimar a matriz homográfica entre as duas imagens e dessa maneira remover *outliers*. Adicionalmente, métodos de navegação como a odometria são utilizados para obter o deslocamento do VANT através de uma sequência de imagens. O módulo de filtragem probabilístico utiliza técnicas como filtro Kalman estendido (Anderson and Moore (1979)) ou filtro de partículas (Del Moral (1996)) junto com as medidas do INS e a posição obtida pelo módulo de processamento de imagens, para calcular a posição final da VANT.

O foco deste trabalho é a extração de características no módulo de processamento de imagens. Em seguida será apresentada uma introdução geral às características visuais de uma imagem e os diferentes tipos de características existentes. Posteriormente, serão detalhados os métodos de extração de pontos de interesse SIFT, SURF, e ASIFT, do mesmo modo que o método de filtragem de *outliers*, RANSAC, que são utilizados no estudo comparativo.

2.1 Características Visuais

Em Farinella et al. (2013), uma característica visual descreve uma propriedade pela qual elementos reais ou abstratos podem ser distinguidos, ou seja, descreve uma propriedade especial de uma imagem como um todo ou um objeto dentro da imagem e pode ser tanto uma propriedade local ou uma característica global da imagem. Assim, por exemplo, uma característica visual pode ser a cor, no caso de imagens coloridas, ou no caso de imagens com diferentes escalas de cinzas, forma, tamanho, orientação, textura. Porém, percepção visual envolve não só as características da imagem locais, tais como linhas, ângulos, cores ou forma, também envolve processos de agrupamento para o que foi detectado com os sentidos e a respectiva interpretação.

Uma característica representa uma peça de informação que possa ser relevante para a resolução de uma tarefa especial. Um elemento visual, em relação a um produto, inclui a forma, configuração, e ornamentação padrão do produto. As características visuais de uma imagem são de grande importância em visão computacional e reconhecimento de padrões. Tais características são utilizadas em diversas aplicações, como por exemplo, o registo e recuperação de imagens, reconstrução 3D, navegação, reconhecimento e

rastreamento de objetos.

De acordo com Farinella et al. (2013) e Tuytelaars and Mikolajczyk (2007), uma característica local é um padrão de imagem variando de sua vizinhança espacial. A variação tipicamente surge a partir de uma alteração de uma ou mais propriedades de imagem, como a cor, a intensidade ou a textura. Em geral, as características locais podem ser cantos, gotas, bordas ou pequenos *patches* da imagem. Se apenas a localização da característica é de interesse, o ponto de interesse é o termo comumente usado. Para a maioria das aplicações, no entanto, uma descrição de cada recurso é derivado das propriedades locais da imagem a fim de identificar e combinar características semelhantes. Ao localizar ou descrever as características, uma região ou vizinhança local deve ser analisada e assim, determinar não apenas a localização, mas o tamanho e forma da vizinhança.

Diferentes estratégias para extração de características tem sido propostas na literatura que visam encontrar os respectivos pontos de interesse da imagem. Farinella et al. (2013) define que um característica local ou ponto de interesse em uma imagem precisa possuir as seguintes propriedades:

- **Distintividade/Informatividade:** Os pontos de interesse devem diferir significativamente de sua vizinhança espacial dentro da imagem, assim, sendo altamente distintivos.
- **Repetibilidade (Invariância/Robustez):** Os pontos de interesse devem ser capazes de produzir características semelhantes para imagens semelhantes.
- **Localidade:** Os pontos de interesse devem somente depender de uma pequena vizinhança espacial na imagem.
- **Quantidade:** O número total de pontos de interesse extraídos de uma imagem deve ser suficientemente grande.
- **Precisão:** O ponto de interesse deve ser localizado com precisão. Isto inclui uma localização espacial dentro da imagem, tal como escala e possível forma.
- **Eficiência:** A extração de pontos de interesse deve ter baixo custo computacional, assim, ser adequada para aplicação em tempo real para reconhecimento de objetos em navegação.

Uma vez em que a característica local é detectada, é atribuída uma descrição adequada, a fim de permitir encontrar novamente em diferentes imagens da mesma cena.

2.2 SIFT (Scale Invariant Feature Transformation)

O método SIFT desenvolvido por Lowe (2004) extrai características locais ou pontos de interesse de uma imagem. Os pontos de interesse são invariantes à escala, translação e rotação, além de parcialmente invariantes às mudanças de iluminação, afinidade ou projeção em 3D. Em Lowe (1999) defini-se que os pontos de interesse são criados para permitir deformações geométricas locais por representação de gradientes em múltiplos planos de orientação com imagens suavizadas em múltiplas escalas. Esses pontos são detectados de forma eficiente através de uma abordagem por etapas de filtragem que identifica os pontos estáveis no espaço. O método SIFT é composto pelas seguintes etapas:

1. Detecção de extremos no espaço/escala: essa etapa realiza uma busca sobre todas as escalas e localizações mediante o uso da função Diferença de Gaussianas Lowe (1999), para identificar candidatos a pontos de interesse invariantes à escala e orientação.
2. Localização dos pontos de interesse: para cada candidato, um modelo é ajustado para determinar a localização e escala. Os pontos de interesse são selecionados baseados medidas de estabilidade da relação das curvaturas principais (Harris and Stephens (1988)) , rejeitando candidatos com baixo contraste ou localizados ao longo de uma borda.
3. Alocação de orientação: nesta etapa é atribuído a cada ponto de interesse uma direção de acordo com as direções do gradiente e a região que os rodeia. Todas as operações futuras são realizadas sobre os dados da imagem que foi transformada em relação à escala, localização e a orientação atribuída de cada ponto de interesse, proporcionando assim invariância a essas transformações.
4. Descritor dos pontos de interesse: na última etapa, os gradientes locais são medidos na escala selecionada, na região em torno ao ponto de interesse, permitindo níveis significativos de distorção de forma local e as variações de luminosidade.

2.2.1 Detecção de extremos no espaço/escala

A primeira etapa é identificar a localização e escala que possam ser atribuídas a pontos sob diferentes vistas do mesmo objeto. A detecção invariante à mudança de escala, pode ser realizada através da procura de características estáveis em todas as escalas possíveis, utilizando uma função contínua conhecida como o espaço/escala (Witkin (1984)).

Koenderink (1984) e Lindeberg (1994) demonstraram que o único possível *kernel* espaço/escala é a função gaussiana. Portanto, o espaço/escala é definido como a convolução de uma máscara gaussiana $G(x, y, \sigma)$ com uma imagem de entrada $I(x, y)$, representado na Equação (2.1). Onde $*$ é a operação de convolução em x e y , e $G(x, y, \sigma)$ é a equação (2.2).

$$L(x, y, \sigma) = G(x, y, \sigma) * I(x, y) \quad (2.1)$$

$$G(x, y, \sigma) = \frac{1}{2\pi\sigma^2} e^{-(x^2+y^2)/2\sigma^2} \quad (2.2)$$

Lowe (1999) propõe para uma detecção eficiente da localização de pontos de interesse estáveis em espaço/escala, utilizar os valores extremos da função diferença-de-gaussianas convoluido com a imagem $D(x, y, \sigma)$ (Equação (2.3)), os quais podem ser calculados a partir da diferença de duas escalas próximas separadas por um fator multiplicativo constante (k). Portanto, a função D pode ser calculada pela simples subtração de imagens (Equação (2.4))

$$D(x, y, \sigma) = (G(x, y, k\sigma) - G(x, y, \sigma)) * I(x, y) \quad (2.3)$$

$$D(x, y, \sigma) = L(x, y, k\sigma) - L(x, y, \sigma) \quad (2.4)$$

Lowe (2004) demonstrou que, a função da diferença-de-gaussianas tem escalas diferentes por um fator constante que já incorpora a normalização de escala σ^2 , necessária para invariância a escala. O fator k na Equação (2.4) é uma constante para todas as escalas, portanto, não influencia os extremos locais. Lowe verificou que com $k = \sqrt{2}$ a aproximação não tem impacto sobre a estabilidade da detecção extremos ou mesmo de localização para diferenças significativas na escala.

Uma abordagem eficiente para a construção de $D(x, y, \sigma)$ é mostrado na Figura 2.2. A imagem inicial é incrementalmente convoluida com gaussianas para produzir imagens separadas por um fator constante k no espaço/escala, mostrado no empilhamento na imagem à esquerda da Figura 2.2. Lowe (2004) escolhe dividir cada oitava de espaço/escala em q intervalos, assim $k = 2^{(1/q)}$. Portanto, deve-se produzir $q + 3$ imagens na pilha de imagens filtradas para cada oitava, de modo que a detecção de extrema final cobre uma oitava completa. Escalas de imagem adjacentes são subtraídas para produzir imagens diferença-de-Gaussianas, mostradas à direita da Figura 2.2. Uma vez que uma oitava completa tenha sido processada, a imagem tem duas vezes o valor inicial de σ . A precisão da amostragem em relação ao σ não é diferente do início da oitava anterior, portanto, o custo computacional é reduzido.

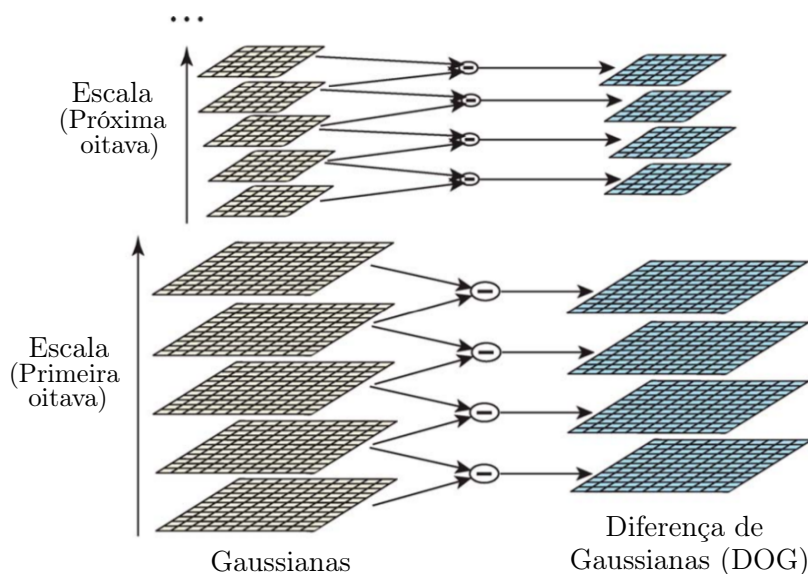


Figura 2.2: Construção de $D(x, y, \sigma)$ para cada oitava de espaço/escala, a imagem inicial é repetidamente convolvida com gaussianas para produzir um conjunto de imagens espaço/escala mostrado na esquerda. Imagens gaussianas adjacentes são subtraídas para produzir diferença-de-Gaussianas, mostradas à direita. Adaptada de Lowe (2004).

2.2.1.1 Frequência de Amostragem na Escala

A determinação da frequência de amostragem maximiza a estabilidade dos extremos. Lowe (2004), demonstrou que mais alta repetibilidade é obtida quando a amostragem é de 3 escalas por oitava. A repetibilidade não continua melhorando à medida que mais escalas são amostradas. A razão é que isto resulta em muitos mais extremo locais a serem detectados, mas estes extremos são em média, menos estáveis e, por conseguinte, são menos susceptíveis de serem detectados na imagem transformada.

O número de pontos de interesse e o número total de combinações corretas cresce com o aumento da amostragem das escalas. Para muitas aplicações, seria ótimo utilizar um número maior de amostras das escalas. No entanto, o custo de computação também cresce com este número.

As experiências de Lowe (2004) mostraram que a função de espaço-escala diferença-de-Gaussianas tem um grande número de extremos e que seria muito caro para detectar todos eles. Felizmente, pode-se detectar o subconjunto mais estável e útil, mesmo com uma pequena amostra de escalas.

2.2.1.2 Frequência de Amostragem no Domínio espacial

Da mesma forma como determinou-se a frequência de amostragem por oitava de espaço-escala, Lowe (2004) determina a frequência da amostragem no domínio da imagem em relação à escala de suavização da imagem. Lowe também mostra uma determinação

experimental do valor de suavização (σ), que é aplicado a cada nível de imagem antes da construção da representação espacial de escala para uma oitava. Ele propôs que existe um custo de utilização de um alto valor para σ em termos de eficiência, por isso, optou-se por utilizar $\sigma = 1.6$, que fornece um valor perto de repetibilidade ótima.

2.2.1.3 Localização dos pontos de interesse

A fim de detectar os máximos e os mínimos locais de $D(x, y, \sigma)$, cada ponto da amostra é comparado com seus oito vizinhos na imagem atual e nove vizinhos na escala acima e abaixo apresentados com os círculos verdes na Figura 2.3. O ponto de interesse é selecionado somente se é maior ou menor do que todos os vizinhos. O custo da verificação é razoavelmente baixo devido ao fato de que a maioria dos pontos de amostragem é eliminado após as primeiras verificações.

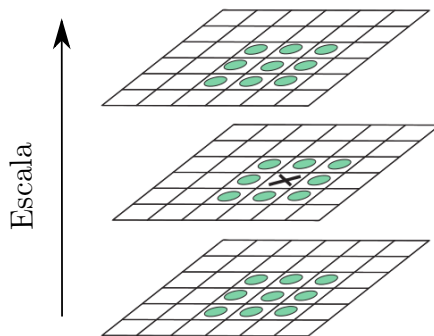


Figura 2.3: 26 pixels vizinhos analisados para detectar os máximos e mínimos da diferença-Gaussianas. Adaptada de Lowe (2004).

Uma vez que o candidato a ponto de interesse é encontrado, o próximo passo é realizar um ajuste detalhado da localização, escala e proporção de curvaturas principais de sua vizinhança. Esta informação permite que pontos que tenham baixo contraste (e, portanto, sensíveis ao ruído) ou pontos que estejam mal localizados ao longo de uma borda, sejam rejeitados (Lowe (2004)).

2.2.2 Localização exata do ponto de interesse

Depois que um ponto de interesse candidato foi encontrado, comparando um pixel com seus vizinhos, o próximo passo é realizar um ajuste detalhado da vizinhança para a localização, escala e proporção de curvaturas principais. Esta informação permite rejeitar pontos com baixo contraste (e, portanto, sensíveis ao ruído) ou que estão localizados ao longo de uma borda (Lowe (2004)).

Para a eliminação destes pontos a implementação inicial é abordada em Brown and Lowe (2002) usando a expansão de Taylor (até os termos quadrático) da função de espaço-escala, $D(x, y, \sigma)$, de modo que a origem está no ponto de amostragem, por conseguinte

$$D(\mathbf{x}) = D + \frac{\partial D^T}{\partial \mathbf{x}} \mathbf{x} + \frac{1}{2} \mathbf{x}^T \frac{\partial^2 D^T}{\partial \mathbf{x}^2} \mathbf{x}, \quad (2.5)$$

D e as derivadas são avaliadas sempre no ponto amostra e $\mathbf{x} = (x, y, \sigma)^T$ é a posição relativa do ponto. Derivando a aproximação da Equação (2.5) e igualando a zero, obtemos

$$\hat{\mathbf{x}} = -\frac{\partial D^{-1}}{\partial \mathbf{x}^2} \frac{\partial D}{\partial \mathbf{x}}. \quad (2.6)$$

Substituindo a Equação (2.6) na Equação (2.5) pode-se obter o valor do máximo local (Equação 2.7).

$$D(\hat{\mathbf{x}}) = D + \frac{1}{2} \frac{\partial D^T}{\partial \mathbf{x}} \hat{\mathbf{x}} \quad (2.7)$$

Lowe (2004) propõe que caso o ponto de interesse encontrado seja $D(x) < 0,03$, este é eliminado da lista dos pontos de interesse, e assume que D toma valores entre 0 e 1. Para a estabilidade, não é suficiente rejeitar pontos de interesse com baixo contraste. A função de diferença-de-gaussiana terá uma resposta forte ao longo das bordas, mesmo se a localização ao longo da borda é mal determinada e portanto instável a pequenas quantidades de ruído. Um ponto de interesse mal definido terá grandes curvaturas principais através da borda, mas uma pequena na direção perpendicular. As curvaturas principais podem ser calculadas a partir de uma matriz de Hessiana 2 x 2, \mathbf{H} (Equação (2.8)), calculados à localização e dimensão do ponto de interesse. As derivadas serão avaliadas tomando diferenças de pontos de amostra vizinhos.

$$\mathbf{H} = \begin{bmatrix} D_{xx} & D_{xy} \\ D_{xy} & D_{yy} \end{bmatrix} \quad (2.8)$$

Os autovalores de \mathbf{H} são proporcionais às curvaturas principais de D , entretanto, segundo Harris and Stephens (1988) pode-se evitar explicitamente pode-se evitar calcular explicitamente os autovalores, através do calculo de apenas suas relações. Denomina-se γ o autovalor com menor magnitude e β com a maior magnitude. Em seguida, pode-se calcular o traço de \mathbf{H} com a soma dos autovalores (Equação (2.9) e a determinante com o produto (Equação (2.10)).

$$\text{Traza}(H) = \frac{\partial^2 D^T}{\partial \mathbf{x}^2} + \frac{\partial^2 D^T}{\partial \mathbf{y}^2} = \gamma + \beta, \quad (2.9)$$

$$\text{Det}(H) = \frac{\partial^2 D^T}{\partial \mathbf{x}^2} x \frac{\partial^2 D^T}{\partial \mathbf{y}^2} = \gamma\beta. \quad (2.10)$$

Se $\gamma = r\beta$, a condição é reduzida à Equação (2.11), a qual depende só da relação dos autovalores.

$$\frac{\text{Traza}(H)}{\text{Det}(H)} = \frac{(\gamma + \beta)^2}{\gamma\beta} = \frac{(r\beta + \beta)^2}{r\beta^2} = \frac{r + 1^2}{r}. \quad (2.11)$$

Lowe (2004) propõe um limiar de $r = 10$. Pode-se observar que com o aumento da relação entre ambos os valores próprios acontece o mesmo com relação entre o quadrado do traço da matriz Hessiana e seu determinante. Isso elimina os pontos de interesse que têm curvaturas principais superiores a 10.

2.2.3 Alocação de Orientação

Ao atribuir uma orientação consistente para cada ponto de interesse com base nas propriedades locais da imagem, o descritor de ponto de interesse pode ser representado em relação a esta orientação e, por conseguinte, alcançar invariância de rotação da imagem. A escala do ponto de interesse é usada para selecionar a Gaussiana da imagem analisada (Equação (2.2)), com a escala mais próxima, de modo que todos os cálculos são efetuados de maneira escala invariante. Para cada imagem amostra $G(x, y)$ a esta escala são calculados, a magnitude do gradiente $m(x, y)$, dada por

$$m(x, y) = \sqrt{\Delta L_x^2 + \Delta L_y^2},$$

$$m(x, y) = \sqrt{(L(x+1, y) - L(x-1, y))^2 + (L(x, y+1) - L(x, y-1))^2}, \quad (2.12)$$

e a orientação $\theta(x, y)$, dada por

$$\theta(x, y) = \tan^{-1}\left(\frac{\Delta L_y}{\Delta L_x}\right),$$

$$\theta(x, y) = \tan^{-1}\left(\frac{L(x, y+1) - L(x, y-1)}{L(x+1, y) - L(x-1, y)}\right). \quad (2.13)$$

Um histograma de orientações é formado a partir dos gradientes dentro de uma região em torno do ponto de interesse. O histograma está distribuído em 36 conjuntos de ângulos cobrindo a faixa de 360 graus de orientações. Cada amostra adicionada ao histograma é ponderada pela sua magnitude de gradiente e por uma janela circular Gaussiano ponderada com um σ que é 1,5 vezes maior do que a escala do ponto de interesse.

Os picos no histograma correspondem a orientação das direções dominantes dos gradientes locais. O pico mais alto do histograma é detectado, e em seguida, qualquer outro pico local que está dentro de 80% do pico mais alto é usado para também criar um ponto de interesse com a orientação. Portanto, para locais com picos múltiplos de magnitude semelhantes, haverá múltiplos pontos de interesse criados no mesmo local e escala, mas diferentes orientações. Apenas cerca de 15% dos pontos são atribuídas várias orientações, mas elas contribuem de forma significativa para a estabilidade do alinhamento (Lowe (2004)).

2.2.4 Descritor dos pontos de interesse

As operações anteriores atribuem uma localização na imagem, escala e orientação a cada ponto de interesse. Estes parâmetros impõem um sistema de coordenadas locais 2D repetível no qual descreve a região local da imagem, e, portanto, fornece invariância a esses parâmetros. O próximo passo descrito por Lowe (2004) é calcular um descritor para a região local do ponto de interesse na imagem, que seja altamente distintivo. O descritor é ainda tão invariável quanto possível às variações restantes, como a mudança na iluminação ou ponto de vista 3D.

A Figura 2.4 ilustra o cálculo do descritor do ponto de interesse. Primeiro as magnitudes de gradiente de imagem e orientações são recolhidos em torno da localização do ponto de interesse, usando a escala do ponto de interesse para selecionar o nível de suavização Gaussiano para a imagem. A fim de atingir a invariância orientação, as coordenadas do descritor e as orientações de gradiente são rodados em relação à orientação do ponto de interesse. Para a eficiência, os gradientes são pré-calculados para todos os níveis da pirâmide, como descrito na detecção de extremos no espaço/escala. Estes são ilustrados com pequenas setas em cada local de amostra no lado esquerdo da Figura 2.4.

A função de ponderação Gaussiana com σ igual a metade da largura da janela do descritor é usada para atribuir um peso para a magnitude de cada ponto de amostragem. Isto é ilustrado com uma janela circular no lado esquerdo da Figura 2.4, e o peso vai diminuindo ao se afastar do ponto de interesse. O objetivo desta janela de Gauss é para evitar alterações súbitas no descritor com pequenas mudanças na posição

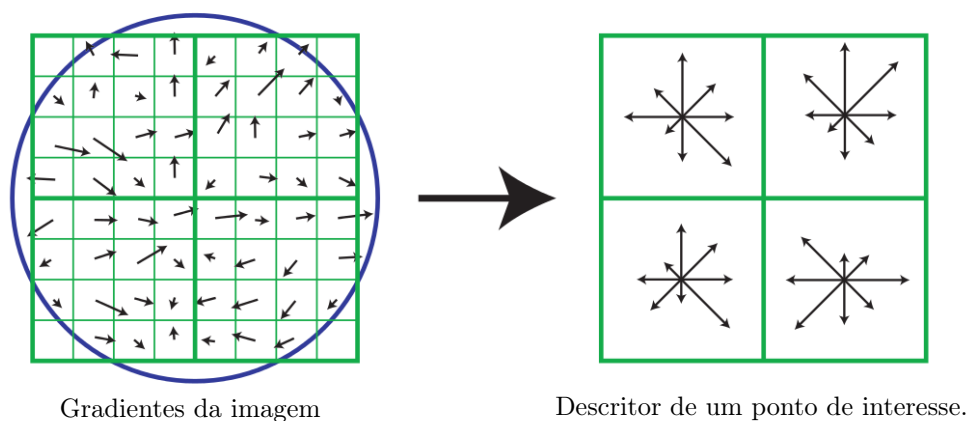


Figura 2.4: Descritor de ponto de interesse do método SIFT. Adaptada de Lowe (2004).

da janela, e para dar menos ênfase aos gradientes que estão muito longe do centro do descritor, já que estes são os mais afetados pelos erros de registo incorreto.

O descritor do ponto de interesse é mostrado no lado direito da Figura 2.4. O descritor permite a mudança significativa em posições dos gradientes através da criação de histogramas de orientação sobre regiões da amostra 4x4. A figura mostra oito direções para cada histograma orientação, com o comprimento de cada seta correspondente à magnitude da entrada do histograma. Uma amostra de gradiente na esquerda pode deslocar-se para 4 posições de amostra ao mesmo tempo que contribui para o mesmo histograma no lado direito, conseguindo assim o objetivo de permitir maiores mudanças de posicionamento locais.

Lowe (2004) expõe a importância de evitar todos os efeitos de fronteira limite em que o descritor muda abruptamente de um histograma para outro ou de uma orientação para o outra. Portanto, a interpolação trilinear é usada para distribuir o valor de cada amostra de gradiente intervalos adjacentes do histograma. Em outras palavras, cada entrada em um compartimento é multiplicada por um peso de $(1 - d)$ para cada dimensão, em que (d) é a distância da amostra a partir do valor central da caixa que é medido em unidades de deslocamento entre os intervalos do histograma.

O descritor é formado a partir de um vetor que contém os valores de todas as entradas de orientação do histograma, correspondente aos comprimentos das setas, apresentados no lado direito da Figura 2.4. Essa figura mostra uma matriz de 2x2 de histogramas de orientação. As experiências em Lowe (2004) mostram que os melhores resultados são obtidos com a matriz 4x4 de histogramas com 8 caixas de orientação em cada. Portanto, um vetor de características $4 \times 4 \times 8 = 128$ elementos para cada ponto de interesse.

Finalmente, o vetor característico é modificado para reduzir os efeitos das alterações de iluminação. Em primeiro lugar, o vetor é normalizado para a unidade de comprimento. Uma alteração no contraste da imagem em que cada valor do pixel é multiplicado por uma constante que irá multiplicar pelos gradientes da mesma constante de modo que, essa alteração de contraste será cancelada pelo vetor normalizado. Uma alteração no brilho, em que uma constante é adicionado a cada pixel da imagem não irá afetar os valores do gradiente como eles são calculados, a partir das diferenças de pixel. Portanto, o descritor é invariante às mudanças na iluminação.

2.2.5 Casamento SIFT

Em Lowe (2004) o casamento de dois pontos de interesse de diferentes imagens é encontrado através da identificação do ponto de interesse mais próximo da segunda imagem referente a cada ponto de interesse da primeira. O algoritmo para comparar a distância entre dois pontos é a distância euclidiana dos vetores descritores. O método SIFT tem seu próprio critério de filtragem de casamentos *outliers*. O método implementa três filtros, as filtragens são:

- O primeiro filtro elimina os casamentos onde os pontos tiverem o mesmo vetor representativo na mesma escala e na mesma posição nas duas imagens.
- O segundo filtro elimina os pontos de interesse casados que estiverem em diferentes posições na primeira imagem e na mesma posição e escala na segunda imagem.
- E o terceiro filtro elimina os pontos de interesse casados que estiverem em diferentes posições na segunda imagem e na mesma posição e escala na primeira imagem.

2.3 SURF (Speeded Up Robust Features)

O método SURF desenvolvido por Bay et al. (2006), implementa o detetor de características *Fast-Hessian* baseado na matriz Hessiana e o descritor descreve a distribuição da transformada *Haar-wavelet* na vizinhança do ponto de interesse. O método utiliza imagens integrais para reduzir o tempo computacional em comparação com o método tradicional SIFT. Adicionalmente, o descritor é de 64 dimensões reduzindo o cálculo do casamento e simultaneamente, aumentando a robustez.

2.3.1 Detector Fast-Hessian

O detector de característica é baseado na matriz Hessiana pelo bom desempenho no tempo computacional e na acurácia. A matriz Hessiana é utilizada para calcular a localização e a escala. A Equação (2.14) representa a matriz Hessiana $\mathcal{H}(x, y, \sigma)$ no ponto (x, y) e na escala (σ) , sendo

$$\mathcal{H}(x, y, \sigma) = \begin{bmatrix} L_{x,x}(x, y, \sigma) & L_{x,y}(x, y, \sigma) \\ L_{x,y}(x, y, \sigma) & L_{y,y}(x, y, \sigma) \end{bmatrix} \quad (2.14)$$

A função $L_{x,x}(x, y, \sigma)$ é a derivada de segunda ordem ($\frac{\partial^2}{\partial x^2}$) da função gaussiana de $L(x, y, \sigma)$ descrita na Equação (2.1), de forma similar para $L_{x,y}(x, y, \sigma)$ e $L_{y,y}(x, y, \sigma)$.

Bay et al. (2006) realizam a aproximação da diferença-de-gaussianas, implementada por Lowe (2004), com filtro *Box* mostrado na Figura 2.5. Este aproxima a derivada de segunda ordem da função gaussiana, e pode ser avaliada rapidamente com imagens integrais independentemente do tamanho.

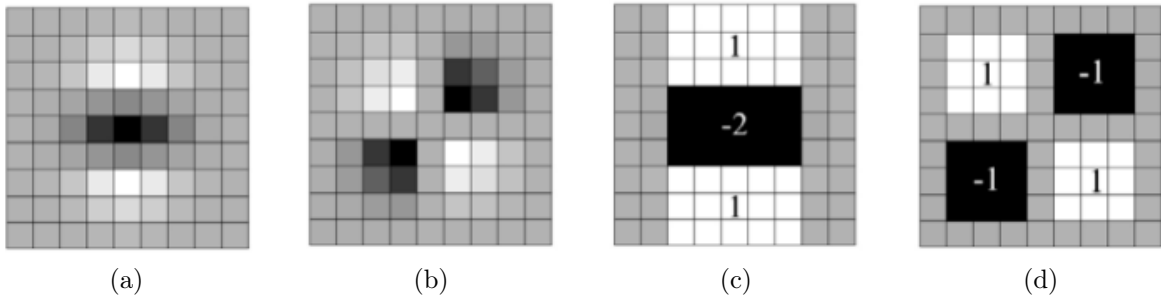


Figura 2.5: Filtro *Box*. Derivada parcial de segunda ordem da função gaussiana em imagens discretas na direção y (Figura 2.5a) e na direção xy (Figura 2.5b). A Figura 2.5c e 2.5d representam a aproximação do filtro *Box* da Figura 2.5a e 2.5b, respectivamente. Adaptada de Bay et al. (2006).

A máscara filtro *Box* de tamanho 9x9 da Figura 2.5 são aproximações da derivada parcial de segunda ordem da função gaussiana com $\sigma=1,2$. Para uma execução computacionalmente eficiente os pesos são armazenados em uma máscara, porém precisam ser balanceados pela determinante aproximada da matriz Hessiana. Bay propôs que $\det(\mathcal{H}_{aproximada}) = D_{xx}D_{yy} - (0.9D_{xy})^2$ e as respostas do filtro devem ser normalizadas com respeito ao tamanho da máscara.

Com respeito à implementação de espaço-escala aplica-se filtros *Box* de diferentes tamanhos sobre a imagem original, podendo ser paralelizados. Portanto, o espaço-escala é analisado pelo sub-escalamento do tamanho do filtro do mesmo modo, que o método SIFT implementa a função gaussiana. A saída do filtro com a máscara de

tamanho 9×9 corresponde à escala inicial com $\sigma = 1,2$. As imagens na pilha de cada oitava são obtidas pela filtragem com máscaras de tamanho gradualmente maiores. Em escalas maiores, o passo entre os tamanhos dos filtros consecutivos também deve ser dimensionado adequadamente. Assim, para cada nova oitava, o aumento de tamanho do filtro é dobrado (seguindo de 6 a 12 a 24). Simultaneamente, os intervalos de amostragem para a extração dos pontos de interesse pode ser duplicada também.

A localização dos pontos de interesse na imagem e nas escalas é aplicada em uma área de no máximo $3 \times 3 \times 3$. Os máximos do determinante da matriz Hessiana são então interpolados em escala e espaço da imagem com o método proposto por Brown and Lowe (2002). Interpolação espaço-escala é importante, porque a diferença de escala entre a primeira pilha de cada oitava é relativamente grande.

2.3.2 Descritor SURF

O descritor SURF é baseado em propriedades semelhantes ao descrito SIFT. Este é desenvolvido em duas etapas, a etapa de alocação da orientação e a etapa dos componentes do descritor. A primeira etapa consiste em fixar uma orientação reproduzível com base em informação a partir de uma região circular em torno do ponto de interesse. Entretanto, na segunda etapa construímos uma região quadrada alinhada na orientação selecionada, e extrai-se o descritor SURF a partir dele.

2.3.2.1 Alocação da Orientação

A fim de ser invariante em rotação, identifica-se uma orientação reproduzível para os pontos de interesse. Para esse efeito, calcula-se as respostas da transformada *Haar* – *wavelet* da direção x e na direção y , mostrado na Figura 2.6, em uma região circular em todo do ponto de interesse com raio de 6 vezes q_{Escala} que é a escala onde foi detectado o ponto de interesse.

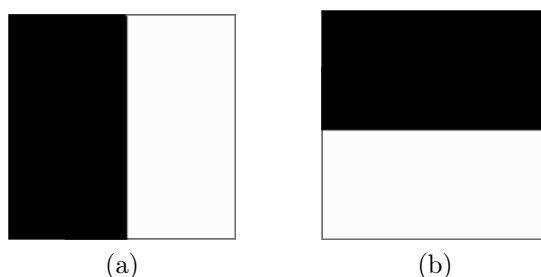


Figura 2.6: Transformada *Haar* – *wavelet* em direção x (Figura 2.6a) e direção y (Figura 2.6b).

Uma vez que as respostas da transformada *Haar – wavelet* são calculadas e ponderadas com uma gaussiana ($\sigma = 2.5(q_{Escala})$) centrada no ponto de interesse, as respostas são representadas como vetores em um espaço com a força da resposta horizontal ao longo da abscissa e a força da resposta vertical ao longo da ordenada. A orientação dominante é estimada pelo cálculo da soma de todas as respostas dentro de uma janela deslizante na orientação cobrindo um ângulo de $\frac{\pi}{3}$. As respostas horizontais e verticais dentro da janela são somadas. As duas respostas somadas, produzem um novo vetor. O mais longo destes vetores confere a orientação para o ponto de interesse. Bay escolheu o tamanho da janela deslizante de forma experimental.

2.3.2.2 Componentes do Descritor

Inicialmente, para a extração do descritor, constrói-se uma região quadrada centrada em torno do ponto de interesse com a orientação selecionada na seção anterior. A região é dividida regularmente em 4x4 sub-regiões quadradas. Isso mantém a informação espacial importante. Para cada sub-região, calcula-se algumas características simples em 5x5 pontos amostrais regularmente espaçados. Bay define d_x como a resposta da transformada *Haar – wavelet* no sentido horizontal e d_y a resposta transformada *Haar – wavelet* na direção vertical. Elas são definidas em relação à orientação do ponto de interesse selecionado. Para aumentar a robustez em relação às deformações geométricas e erros de localização, as respostas d_x e d_y são ponderadas com uma gaussiana ($\sigma = 3.3(s_{Escala})$) centrada no ponto de interesse.

Seguidamente, d_x e d_y são adicionadas em cada sub-região e forma um primeiro conjunto de entradas para o vetor de característica. Com o propósito de trazer as informações sobre a polaridade das alterações de intensidade, extrai-se a soma dos valores absolutos das respostas, $|d_x|$ e $|d_y|$. Assim, cada sub-região tem um vetor descritor \mathbf{v} de quatro dimensões para a estrutura de intensidade subjacente $\mathbf{v} = (\sum d_x, \sum d_y, \sum |d_x|, \sum |d_y|)$. \mathbf{v} é o vetor descritor para todas as sub-regiões 4x4 de 64 dimensões. As respostas da transformada *Haar – wavelet* são invariantes para um viés na iluminação. A invariância ao contraste é conseguida normalizando o vetor descritor.

A Figura 2.7 apresenta as propriedades do descritor SURF para três imagens com diferentes padrões de intensidade dentro de uma sub-região.

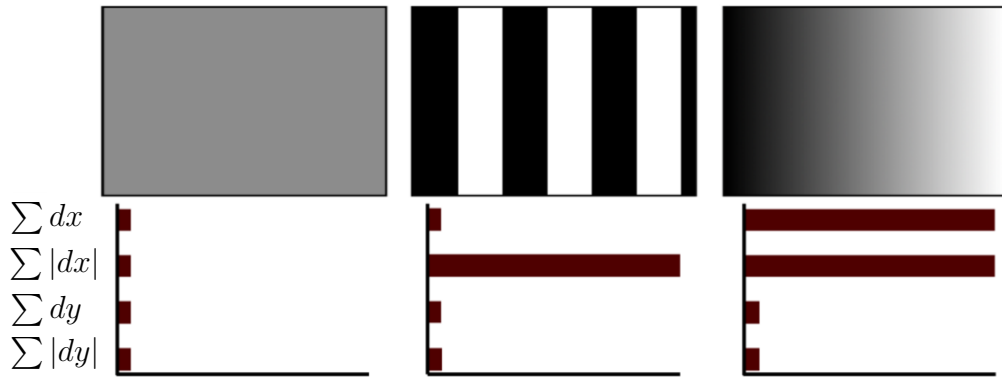


Figura 2.7: Propriedades do descritor em sub-regiões para três imagens com diferentes padrões de intensidade. Adaptada de Bay et al. (2006).

2.4 ASIFT (Affine-SIFT)

O método ASIFT foi desenvolvido por Morel and Yu (2009), este método aborda a detecção de características para a comparação de imagens totalmente invariante à transformada afim. ASIFT complementa SIFT através da simulação de dois parâmetros (θ e ϕ) do modelo da câmera óptica (Figura 2.8) que definem a orientação do eixo da câmera. Portanto, ASIFT simula três parâmetros da transformada afim (ângulo de longitude (ϕ), ângulo de latitude (θ) e escala (λ)) e normaliza os outros três parâmetros (rotação (ψ) e translação (x_1, y_1)).

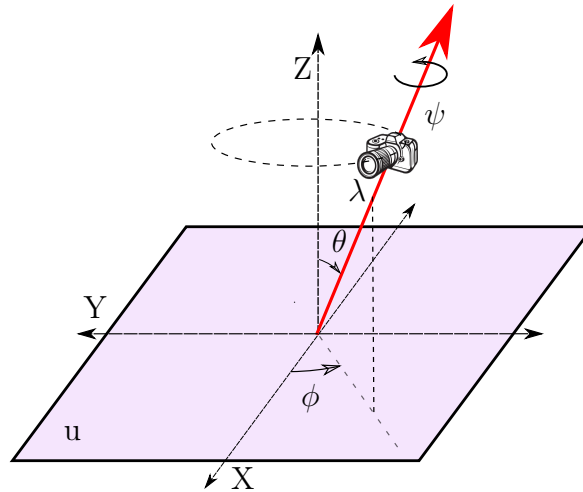


Figura 2.8: Interpretação geométrica dos parâmetros da transformada afim.

Uma interpretação geométrica da transformação afim da imagem capturada pela câmera mostrada na Figura 2.8 é representada por

$$A = \mathbf{H}_\lambda \mathbf{R}_1(\psi) \mathbf{T}_1(\theta) \mathbf{R}_2(\phi) + \mathbf{T}_2(x_1, y_1), \quad (2.15)$$

onde A é a projeção da imagem frontal u_0 devido ao movimento da câmera, dado que:

- λ : *escala*, medição da dimensão com respeito à imagem frontal. $\lambda > 0$.
- ψ : *rotação*, ângulo da câmera em torno do eixo óptico. $\psi \in [0^\circ, 360^\circ]$.
- θ : *latitude*, ângulo entre o eixo óptico e a normal ao plano de imagem. $\theta \in [0^\circ, 90^\circ]$.
- ϕ : *longitude*, ângulo entre o eixo óptico e o plano vertical fixo. $\phi \in [0^\circ, 180^\circ]$
- x_1, y_1 : *Translação*, movimento inicial da câmera (não presente na Figura 2.8).

\mathbf{H}_λ é a componente de escala da imagem capturada pela câmera representada pelo valor do parâmetro λ . \mathbf{R}_1 e \mathbf{R}_2 descrevem as rotações de ψ e ϕ respectivamente, cada rotação é representada por uma matriz de dimensão 2×2 . \mathbf{T}_1 é uma inclinação, a qual é uma matriz diagonal com dois autovalores, no primeiro $t > 1$ ($\theta = \arccos(1/t)$) e o segundo é 1. \mathbf{T}_2 é o movimento inicial da câmera, onde o plano da imagem inicial u_0 e o plano da imagem capturada pela câmera estão em paralelo, representado por uma matriz de dimensão 2×1 , o primeiro termo é x_1 e o segundo é y_1 os quais representam o deslocamento no eixo x e o eixo y respectivamente. Por conseguinte, a Equação (2.15) pode ser reescrita como:

$$A = \lambda \begin{bmatrix} \cos\psi & -\sin\psi \\ \sin\psi & \cos\psi \end{bmatrix} \begin{bmatrix} t & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \cos\phi & -\sin\phi \\ \sin\phi & \cos\phi \end{bmatrix} + \begin{bmatrix} x_1 \\ y_1 \end{bmatrix}. \quad (2.16)$$

Os parâmetros da posição inicial da câmera para capturar a vista frontal u_0 na Equação (2.16) são $\lambda = 1$, $t = 1$, $\phi = \psi = 0$ e $x_1, y_1 = 0$.

2.4.1 Algoritmo ASIFT

ASIFT procede da seguinte forma:

1. Cada imagem é transformada pela simulação de distorções causadas pela mudança do eixo óptico da câmera a partir de uma posição frontal, utilizando a Equação (2.16). Esta distorção depende de dois parâmetros: o ângulo de longitude e o ângulo de latitude. A imagem é submetida a rotações do ângulo longitude seguido pelas inclinações através do parâmetro $t = 1/|\cos\theta|$. Morel and Yu (2009) determinaram 6 amostras do ângulo θ (n_θ) para cada uma das imagens casadas, a mudança de θ é associada com a função aritmética $1, a, a^2, \dots, a^{n_\theta-1}$, com $a = \sqrt{2}$.

O número de amostras do ângulo ϕ é diferente para cada valor da série do ângulo θ , $\Delta\phi = \frac{\pi}{n_\phi}$, no qual $n_\phi = \frac{\text{Round}(10\theta)}{2}$ (a função *Round* retorna um valor inteiro arredondado para o número inteiro mais próximo do valor passado como parâmetro). Os pontos apresentados na Figura 2.9 representam todas as simulações dos ângulos θ e ϕ . Para cada uma das imagens obtidas implementa-se o algoritmo SIFT com a finalidade de conseguir um banco de dados dos possíveis pontos de interesse.

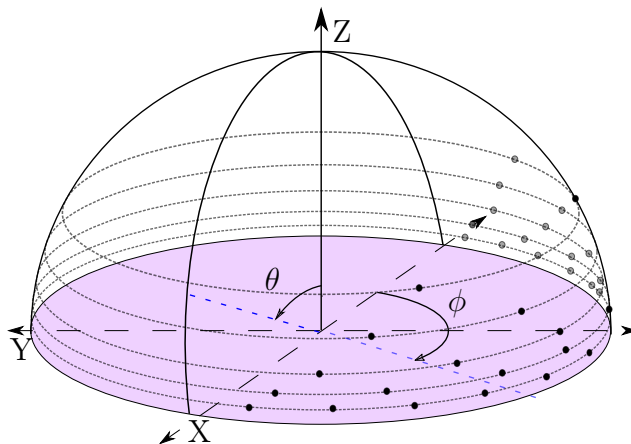


Figura 2.9: As amostras dos parâmetros $\theta = \arcsin(1/t)$ e ϕ realizadas pelo algoritmo ASIFT. As amostragens são representadas pelos pontos pretos da imagem.

2. O algoritmo de casamento invariante à similaridade exposto em Lowe (2004) fundamentada na medida da distância *City Block* descrita em Everitt et al. (2011) é aplicado ao banco de dados dos possíveis candidatos a pontos de interesse de cada uma das duas imagens. O casamento de dois pontos de interesse de diferentes imagens é encontrado através da identificação do ponto de interesse mais próximo da segunda imagem referente a cada ponto de interesse da primeira.
3. O algoritmo de casamento invariante à similaridades geralmente deixam falsos casamentos. É importante filtrar estes casamentos devido ao grande número de *outliers* obtidos. O critério usado é que os casamentos devem ser compatíveis com uma geometria epipolar.

2.5 RANSAC (Random Sample Consensus)

RANSAC é um algoritmo iterativo proposto por Fischler and Bolles (1981). Este algoritmo é um estimador de parâmetros de um modelo matemático desenhado para lidar com uma grande proporção de *outliers* nos dados de entrada. A ideia do algoritmo é utilizar a menor quantidade de pontos possíveis para estimar o modelo matemático,

em seguida ver quantos dados ajustam-se ao modelo estimado e encontrar o melhor modelo representativo dos dados, eliminando os *outliers*.

A implementação do RANSAC para casamento de imagens através de pontos característicos, estima transformada homográfica. Hartley and Zisserman (2003) define uma transformada homográfica (denominada H) como uma matriz não-singular de dimensão 3×3 que pode mapear o mesmo ponto em dois planos diferentes, isto é, que o ponto $\mathbf{x} = (x, y)$ do plano da Imagem₁ e o ponto $\mathbf{x}' = (x', y')$ do plano da Imagem₂ da Figura 2.10 podem ser representados como $\mathbf{x}' = H\mathbf{x}$ (Equação (2.17)), o que resulta uma transformação com oito graus de liberdade.

$$\begin{pmatrix} x'_1 \\ x'_2 \\ x'_3 \end{pmatrix} = \begin{bmatrix} h_{11} & h_{12} & h_{13} \\ h_{21} & h_{22} & h_{23} \\ h_{31} & h_{32} & h_{33} \end{bmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix}. \quad (2.17)$$

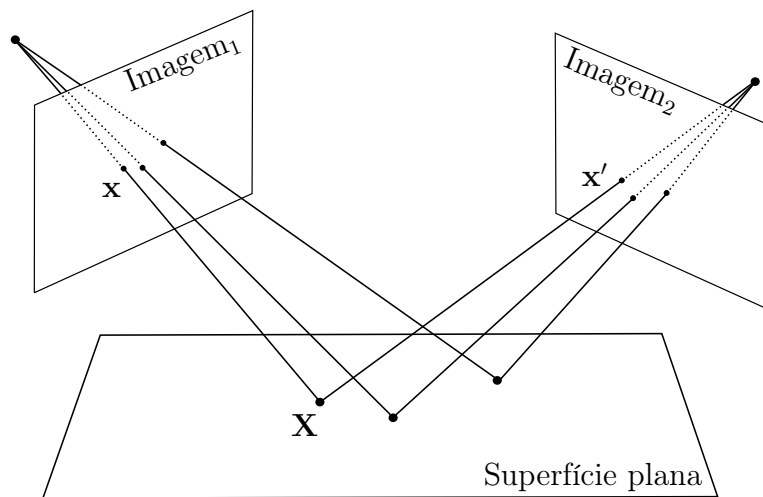


Figura 2.10: Representação do ponto \mathbf{X} que encontra-se dentro de uma superfície planar, no plano da Imagem₁ (\mathbf{x}) e da Imagem₂ (\mathbf{x}').

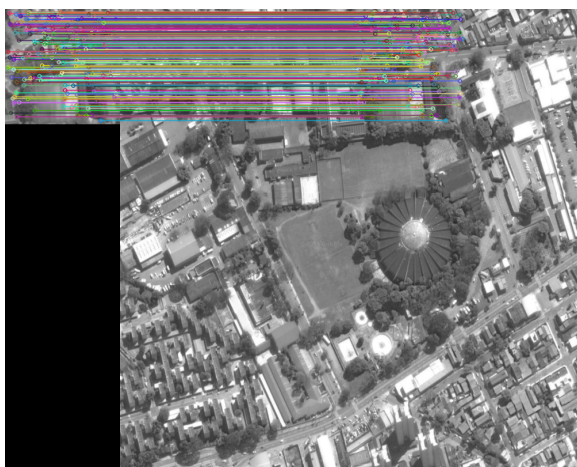
Utilizando a Equação (2.17), o ponto \mathbf{x}' pode ser dado pela valor de x' e y' da Equação (2.18). Cada ponto casado gera duas equações lineares com os elementos de H (Equação (2.19)). Por conseguinte, quatro pontos correspondentes produzem oito equações lineares que são suficientes para calcular H , a única restrição é que os pontos não podem ser colineares. Desta forma, o RANSAC calcula a transformada homográfica aproximada \hat{H} a partir de 4 pontos casados aleatórios e procura quantos pontos dentro de todos os casamentos são compatíveis à transformada \hat{H} . Se a quantidade de pontos é suficiente, utiliza-se \hat{H} como transformada homografia entre as duas imagens. Desta forma, o RANSAC estima a transformada H entre duas imagens com presença de

outliers.

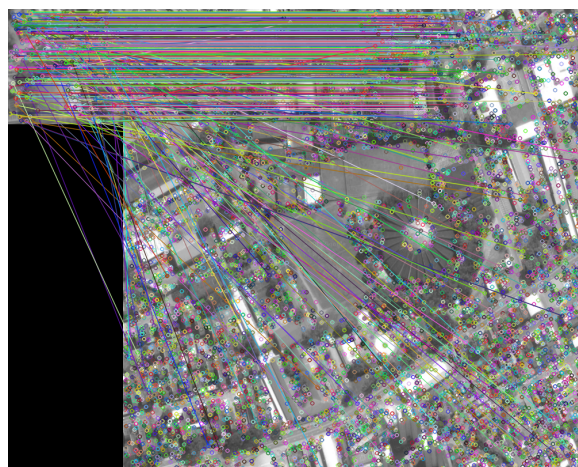
$$x' = \frac{x'_1}{x'_3} = \frac{h_{11}x + h_{12}y + h_{13}}{h_{31}x + h_{32}y + h_{33}} \quad y' = \frac{x'_2}{x'_3} = \frac{h_{21}x + h_{22}y + h_{23}}{h_{31}x + h_{32}y + h_{33}} \quad (2.18)$$

$$\begin{aligned} x'(h_{31}x + h_{32}y + h_{33}) &= h_{11}x + h_{12}y + h_{13} \\ y'(h_{31}x + h_{32}y + h_{33}) &= h_{21}x + h_{22}y + h_{23} \end{aligned} \quad (2.19)$$

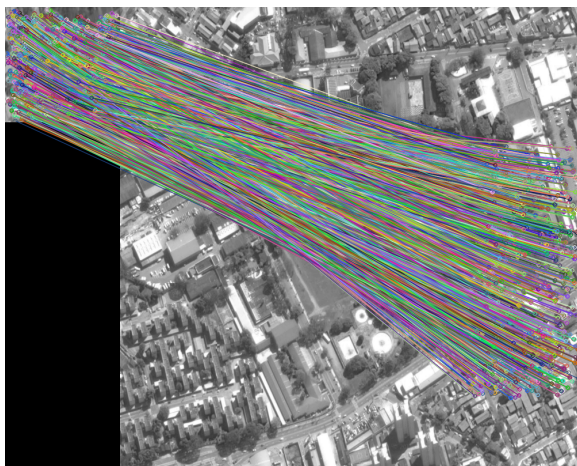
A Figura 2.11 representa a implementação da filtragem RANSAC nos métodos SIFT e ASIFT. Nas imagens da esquerda observe-se que os pontos de interesse entre as duas imagens correspondem pontos casados com a mesma transformada H .



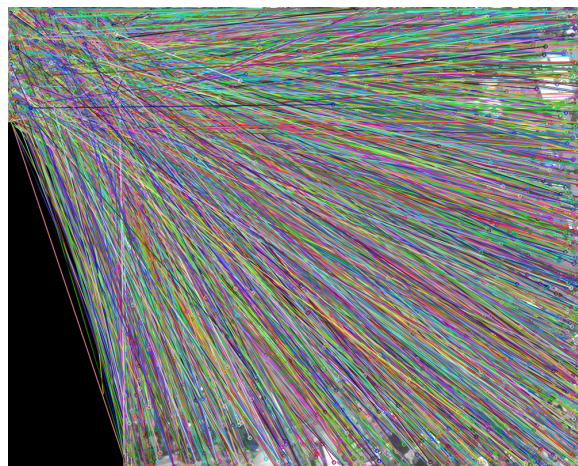
(a) SIFT com RANSAC



(b) SIFT sem RANSAC



(c) ASIFT com RANSAC



(d) ASIFT sem RANSAC (1)

Figura 2.11: Casamento de imagens, as linhas a cor representam os pontos casados entre as duas imagens. A Figura 2.11a e Figura 2.11b são a implementação do método SIFT com, e sem filtragem de *outliers* RANSAC, respectivamente. A Figura 2.11c e Figura 2.11d são a implementação do método ASIFT com, e sem filtragem de *outliers* RANSAC, respectivamente.

Hartley and Zisserman (2003) descrevem o algoritmo RANSAC como um ajuste robusto de um modelo para um conjunto de dados S que contém *outliers*, e resume RANSAC em 5 passos:

1. Escolher aleatoriamente uma amostragem s (quatro pontos dentro de todos os casamentos) de um conjunto de dados S (todos os pontos casados entre as duas imagens) e imediatamente estimar o modelo deste subconjunto de dados.
2. Determinar o subconjunto de pontos S_i que estão dentro da distância limite t_H do modelo. O conjunto de S_i é o consenso definido da amostragem e define as *inliers* de S .
3. Se o tamanho de S_i é maior que ao conjunto consenso S_t que é o número mínimo de pontos que podem definir um modelo, deve-se estimar de novo o modelo usando todos os pontos em S_i e estabelecer S_t como S_i .
4. Se o tamanho de S_i é inferior a S_t , deve-se selecionar um novo subconjunto e repetir o procedimento de acima.
5. Depois de executado N testes o maior consenso definido S_i é selecionado, e o modelo é estimado utilizando todos os pontos no subconjunto S_i .

A distância limite (t_H), o tamanho do conjunto consenso para ser aceitável (S_t) e o número de iterações (N), são parâmetros do algoritmo que devem ser definidos antes de sua execução. Nas seguintes sub-seções explica-se cada um destes parâmetros.

2.5.1 Distância limite (threshold)

Deve-se escolher uma distância limite, t_H , tal que a probabilidade $(1 - \alpha)$ de pontos sejam *inliers*. Na prática a distância limite t_H é usualmente escolhida de forma empírica. De modo que, cada ponto é classificado em *outlier* ou *inlier* dependendo da distância t_H , isto é:

$$x_{ij} = \begin{cases} \textit{inlier} & d_{\perp}^2 < t_H^2 \\ \textit{outlier} & d_{\perp}^2 \geq t_H^2 \end{cases} \quad (2.20)$$

Para medir a distância d_{\perp}^2 , Hartley and Zisserman (2003) define que a forma mais factível é avaliar o erro de um casamento a partir de uma homografia \hat{H} , e usar seu erro de transferência simétrica, $d_{transferencia}^2 = d(\mathbf{x}, \hat{H}^{-1}\mathbf{x}')^2 + d'(\mathbf{x}', \hat{H}\mathbf{x})^2$, onde $x \longleftrightarrow x'$ são os pontos casados (Figura 2.12) e a função d é a distância euclidiana entre dois pontos no mesmo plano. A medida da distância é a projeção do erro é representada na

Equação (2.21). Esta medida é custosa computacionalmente por que $\hat{\mathbf{x}}$ e $\hat{\mathbf{x}'}$ tem que ser calculados.

$$d_{\perp}^2 = d(\mathbf{x}, \hat{\mathbf{x}})^2 + d'(\mathbf{x}', \hat{\mathbf{x}'})^2. \quad (2.21)$$

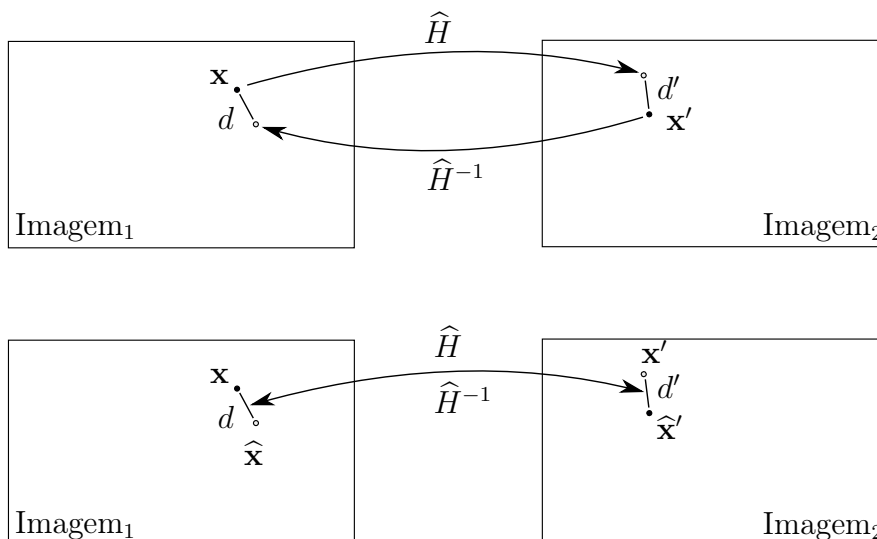


Figura 2.12: Transformação projetiva, estimação 2D. Adaptada de Hartley and Zisserman (2003).

2.5.2 Número de iterações

É computacionalmente inviável e desnecessário iterar com todas as amostras. O número de iterações (N) escolhida tem que ser o suficientemente alto para garantir com uma probabilidade p , que pelo menos uma das amostras aleatórias de s pontos seja livre de *outliers*.

Portanto o número de iterações N é

$$N = \frac{\log(1 - p)}{\log((1 - (1 - \alpha)^s))}. \quad (2.22)$$

Na Tabela 2.1 apresenta-se exemplos do números de iterações requeridas para garantir, com uma probabilidade $p = 0,99$. Muitas vezes acontece que a proporção pontos *outliers* é desconhecida, nesses casos o algoritmo é inicializado utilizando um caso de pior estimativa de α , e esta estimativa pode ser atualizada com os conjuntos mais consistentes encontrados. O algoritmo(1) mostra RANSAC com o α adaptativo.

Tabela 2.1: Número de N iterações requeridas para garantir, com uma probabilidade $p = 0,99$, uma boa estimativa do modelo matemático. No caso de casamento de imagens $s = 4$. Adaptada de Hartley and Zisserman (2003).

Tamanho de Amostragem	Proporção de outliers							
	s	5	10	20	25	30	40	50
2	2	3	5	6	7	11	17	
3	3	4	7	9	11	19	35	
4	3	5	9	13	17	34	72	
5	4	6	12	17	26	57	146	
6	4	7	16	24	37	97	293	

Algoritmo 1: RANSAC

início

$N = \infty, \text{conteo}_{amostragem} = 0;$

repita

Escolha uma amostragem e conte o número de *inliers*;

O conjunto $\alpha = 1 - (\text{número de } inliers) / (\text{número total de pontos});$

Encontre N de α e $p = 0,99$ com (2.22).;

Acrescente $\text{conteo}_{amostragem}$ em 1.;

até $N > \text{conteo}_{amostragem};$

fim

2.5.3 Tamanho do conjunto consenso

Hartley and Zisserman (2003) determina que o tamanho do conjunto consenso tem que ser semelhante ao número de *inliers* acreditado para estar no conjunto de dados. Portanto, para n pontos de dados o tamanho do conjunto consenso é $S_t = (1 - \alpha)n$.

2.6 Resumo do capítulo

Neste capítulo, foram apresentados os métodos SIFT, SURF e ASIFT. Estes métodos extraem pontos característicos em imagens, cada ponto característico é descrito com valores reais da região em torno do ponto de interesse. Os métodos SIFT e SURF são invariantes a rotação, translação e escala. Entretanto o método ASIFT é invariante a rotação, translação, escala e afinidade, uma desvantagem deste método é o alto custo computacional uma vez que, para conseguir sua invariância à afinidade realiza 41 transformações de em cada imagem. Estes métodos são muito eficientes para o casamento de imagens, mas precisa-se que as imagens casadas sejam capturadas pelo mesmo sensor ou sensores muito próximos. Desta forma, a melhor técnica de navegação autônoma para sua implementação é a odometria, na qual se estima o deslocamento do VANT através do casamento de imagens capturadas de forma sequencial.

Capítulo 3

Propostas

Este trabalho apresenta duas propostas de extratores de características invariantes a afinidade, focados no posicionamento de VANT. Os extratores são baseados no método ASIFT para a invariância à transformada afim, entretanto limita as simulações com a implementação dos dados fornecidos pelo INS, além do método ser paralelizado para reduzir o tempo computacional. A primeira proposta, denominada ASIFT-V, e a segunda proposta, ASURF-V, são apresentadas nas seções 3.1 e 3.2, respectivamente.

3.1 ASIFT-V (Affine-SIFT-VANT)

O método de extração de características ASIFT-V é baseado no método ASIFT, e utiliza os dados fornecidos pelo INS para diminuir a amostragem dos ângulos θ e ϕ , e reduzir o custo computacional do algoritmo original. O INS descreve a orientação de um corpo rígido, como o VANT, girando em um espaço euclidiano tridimensional mediante o uso dos ângulos Euler. Os ângulos são definidos no sentido positivo dos eixos do sistema do corpo da aeronave. *Roll* (ϕ) em torno do eixo X, *Pitch* (θ) em torno do eixo Y e *Yaw* (ψ) em torno do eixo Z, conforme representado na Figura 3.1. Portanto, o conhecimento a priori dos ângulos *Pitch* e *Roll* é utilizado para diminuir a amostragem dos ângulos de longitude e latitude do método ASIFT.

O método ASIFT-V procede da seguinte forma:

1. A imagem territorial é transformada pela simulação de distorções causadas pela mudança do eixo óptico da câmera a partir de uma posição frontal, utilizando a Equação (2.16). Do mesmo modo que o ASIFT a distorção depende dos parâmetros θ e ϕ . A amostragem dos θ e ϕ é baseada nas simulações dadas por Morel and Yu (2009) no método ASIFT apresentados na Tabela 3.1. Porém, a imagem só simula os ângulos θ e ϕ mais próximos da orientação real do VANT no momento da captura da imagem, dados fornecidos pelo INS (*Pitch* e *Roll*). En-

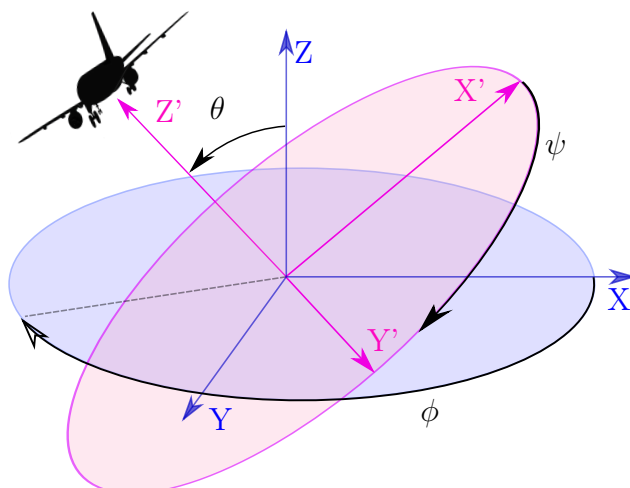


Figura 3.1: Conversão dos ângulos Euler.

tretanto, se o ângulo θ é menor que 32° não é realizada nenhuma simulação. Por exemplo, se *Pitch* é 65° e *Roll* é 100° a imagem seria transformada para os ângulos θ e ϕ de: $(60^\circ, 72^\circ)$, $(60^\circ, 108^\circ)$, $(69^\circ, 77^\circ)$ e $(69^\circ, 102^\circ)$. Para cada uma das imagens obtidas implementa-se o algoritmo SIFT com a finalidade de conseguir um banco de dados dos possíveis pontos de interesse. Na Figura 3.2 a localização da aeronave é interpretada pelo vetor vermelho, e a otimização proposta diminui a amostragem de ASIFT à simulação dos ângulos mais próximos representados pelos pontos azuis.

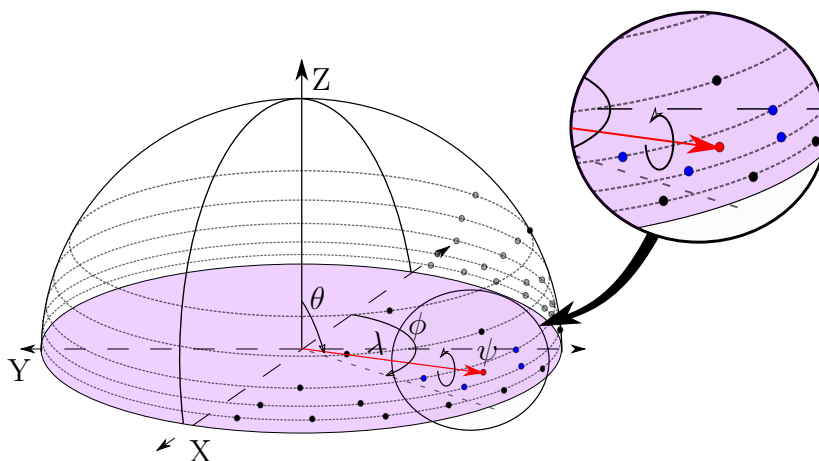


Figura 3.2: Representação das simulações realizadas pelo ASIFT-V.

2. A imagem capturada pelo VANT não realiza nenhuma simulação dos ângulos θ e ϕ para garantir que o casamento seja executado em imagens com os mesmos ângulos da transformada homográfica. Implementa-se o algoritmo SIFT com a finalidade de conseguir um banco de dados dos possíveis pontos de interesse.

Tabela 3.1: Valores dos ângulos θ e ϕ simulados pelo método ASIFT.

θ	ϕ														
0	0														
45	0	45	90	135											
60	0	36	72	108	144										
69	0	26	51	77	102	128	154								
75	0	18	36	54	72	90	108	126	144	162					
80	0	13	26	39	51	64	77	90	103	116	129	141	154	167	

3. O algoritmo de casamento invariante á similaridade exposto em Lowe (2004) fundamentado na medida da distância *City Block* descrita em Everitt et al. (2011) e o método de filtragem de *outliers* RANSAC são implementados para obter os casamentos compatíveis com uma mesma geometria epipolar.

Com o objetivo de diminuir o custo computacional, no primeiro passo foi paralelizada a implementação do algoritmo SIFT em cada uma das transformações da imagem territorial. Deste modo, o primeiro passo do algoritmo é implementado em 4 linhas de execução (em inglês *threads*), uma para cada combinação de θ e ϕ . Cada *thread* recebe o a imagem original (I_1) e os ângulos θ e ϕ , e executa SIFT à imagem transformada para obter os possíveis pontos de interesse da primeira imagens. O algoritmo 2 mostra a paralelização do primeiro passo de ASIFT-V.

Algoritmo 2: Paralelização do primeiro passo de ASIFT-V

Entrada: $I_1 = Imagem_{Territorial}$; $\theta_x = Pitch$; $\phi_x = Roll$;**Saída:** \mathbf{X}_{I_1} =Pontos $Imagem_{Territorial}$ **início** $h_{(\theta_x, \phi_x)}$. Calcula-se a transformada homográfica para θ_x e ϕ_x ; $I'_1 = h_{(\theta_x, \phi_x)} I_1$. Aplica-se a $h_{(\theta_x, \phi_x)}$ para a imagem original; $\mathbf{X}'_i = SIFT(I'_1)$, $i = 1, \dots, k$. Aplica-se o algoritmo SIFT na imagem transformada e obtêm-se os pontos de interesse ; $h'_{(\theta_x, \phi_x)}$. Calcula-se a transformada homográfica inversa para θ_x , e ϕ_x ; $\mathbf{X}_i = h'_{(\theta_x, \phi_x)} \mathbf{X}'_i$, $i = 1, \dots, k$. Aplica-se para cada ponto de interesse a transformada homográfica inversa para levar os pontos à imagem original;**fim** $\mathbf{X}_{I_1} = X_1 \cup X_2 \cup X_3 \cup X_4$

A Figura 3.3 apresenta o diagrama de fluxo da paralelização do primeiro passo para *Pitch* de 65° e *Roll* de 100° para o primeiro passo do algoritmo ASIFT-V.

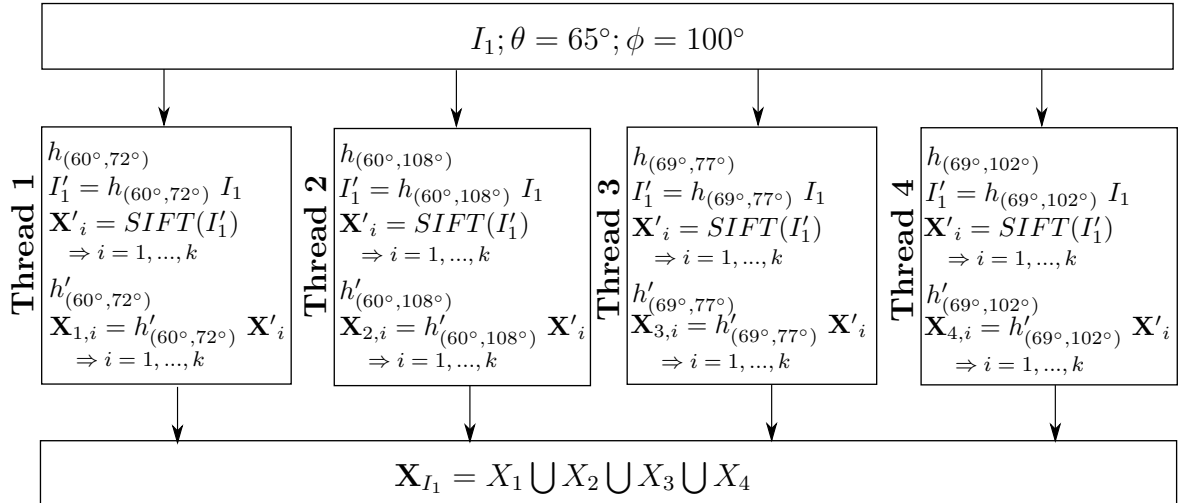


Figura 3.3: Diagrama de Fluxo do primeiro passo de ASIFT-V paralelizado

O diagrama de fluxo da Figura 3.4 e o Algoritmo 3 representa a implementação referente a ASIFT-V. Ptos é a localização e Dtor o descritor de cada um dos pontos quando é implementado o método SIFT. A função $SIFT()$ é a implementação do método SIFT para extrair pontos característicos. A função $Casamento()$ é algoritmo de casamento invariante á similaridade. Finalmente, a função $RANSAC()$ implementa o método RANSAC para eliminar pontos *outliers* no casamento.

3.2 ASURF-V (Affine-SURF-VANT)

O método de extração de características ASURF-V, do mesmo modo que o primeiro método proposto, é baseado em ASIFT por sua invariância da transformada afim e utiliza os dados fornecidos pelo INS para diminuir a amostragem dos ângulos θ e ϕ , e assim reduzir o custo computacional. Porém, o segundo método proposto implementa o extractor de características SURF em cada uma das imagens transformadas, em comparação com ASIFT-V que implementa o método SIFT. A escolha da implementação de SURF como extractor de características foi fundamentada em Li and Aouf (2012), o qual compara os métodos SIFT e SURF para navegação de um VANT com imagens nas bandas visíveis e infravermelho. Li and Aouf verificaram que SURF, é mais rápido que SIFT e proporciona taxas de casamento mais elevados para ambas as sequências de imagens visíveis e de infravermelhos, representando uma adequada opção para sua implementação. O diagrama de fluxo da Figura 3.5 e o Algoritmo 4 representa a implementação do método ASURF-V. A função $SURF()$ é a implementação do método SURF para extrair pontos característicos.

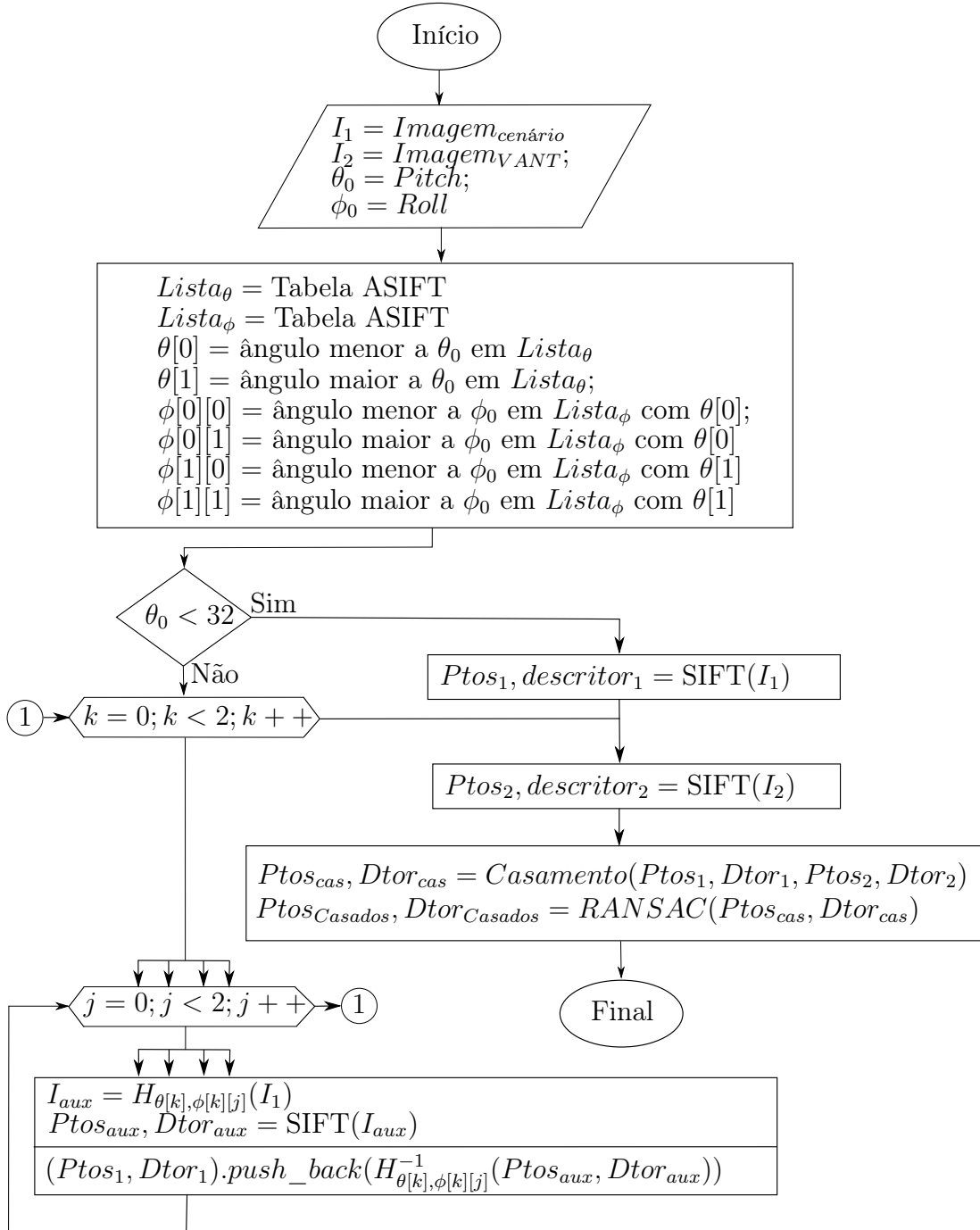


Figura 3.4: Diagrama de Fluxo do algoritmo ASIFT-V

3.3 Resumo do capítulo

Neste capítulo, foram apresentadas duas propostas (ASIFT-V e ASURF-V) de extractores de características de imagens focado à navegação autônoma de VANTs. As duas propostas são baseadas no ASIFT, desta forma são invariantes a escala, rotação, translação e afinidade. Porém, otimizam o método ASIFT reduzindo sua complexidade com a transformação das quatro (4) combinações dos ângulo θ e ϕ mais próximos ao

Algoritmo 3: Algoritmo ASIFT-V**Entrada:** $I_1 = \text{Imagem}_{\text{Terrorial}} ;$ $I_2 = \text{Imagem}_{\text{VANT}} ;$ $\theta_0 = \text{Pitch} ;$ $\phi_0 = \text{Roll} ;$ **Saída:** $Ptos_{\text{Casados}}, Dtor_{\text{Casados}}$ **início** $Lista_{\theta} = \text{Tabela 3.1} ;$ $Lista_{\phi} = \text{Tabela 3.1} ;$ $\theta[0] = \hat{\text{Ângulo menor a }} \theta_0 \text{ em } Lista_{\theta} ;$ $\theta[1] = \hat{\text{Ângulo maior a }} \theta_0 \text{ em } Lista_{\theta} ;$ $\phi[0][0] = \hat{\text{Ângulo menor a }} \phi_0 \text{ em } Lista_{\phi} \text{ com } \theta[0] ;$ $\phi[0][1] = \hat{\text{Ângulo maior a }} \phi_0 \text{ em } Lista_{\phi} \text{ com } \theta[0] ;$ $\phi[1][0] = \hat{\text{Ângulo menor a }} \phi_0 \text{ em } Lista_{\phi} \text{ com } \theta[1] ;$ $\phi[1][1] = \hat{\text{Ângulo maior a }} \phi_0 \text{ em } Lista_{\phi} \text{ com } \theta[1] ;$ **if** $\theta_0 < 32$ **then**| $Ptos_1, Dtor_1 = \text{SIFT}(I_1) ;$ **else****repita****repita**| $I_{aux} = H_{\theta[k], \phi[k][j]}(I_1) ;$ | $Ptos_{aux}, Dtor_{aux} = \text{SIFT}(I_{aux}) ;$ | $(Ptos_1, Dtor_1).push_back(H_{\theta[k], \phi[k][j]}^{-1}(Ptos_{aux}, Dtor_{aux})) ;$ | **até** $j = 0 ; j < 2 ; j ++ (\text{Paralelizado}) ;$ | **até** $k = 0 ; k < 2 ; k ++ (\text{Paralelizado}) ;$ **end** $Ptos_2, Dtor_2 = \text{SIFT}(I_2) ;$ $Ptos_{cas1}, Dtor_{cas1} = \text{Casamento}(Ptos_1, Dtor_1, Ptos_2, Dtor_2) ;$ $Ptos_{\text{Casados}}, Dtor_{\text{Casados}} = \text{RANSAC}(Ptos_{cas1}, Dtor_{cas1}) ;$ **fim**

momento da captura da imagem, a diferença do ASIFT que realiza quarenta e uma (41) transformações de cada imagem. Uma desvantagem destes métodos, assim como o SIFT, SURF e ASIFT, é que precisa-se que as imagens casadas sejam capturadas pelo mesmo sensor ou sensores muito próximos. Desta forma, a melhor técnica de navegação autônoma para sua implementação é a odometria, na qual se estima o deslocamento do VANT através do casamento de imagens capturadas de forma sequencial.

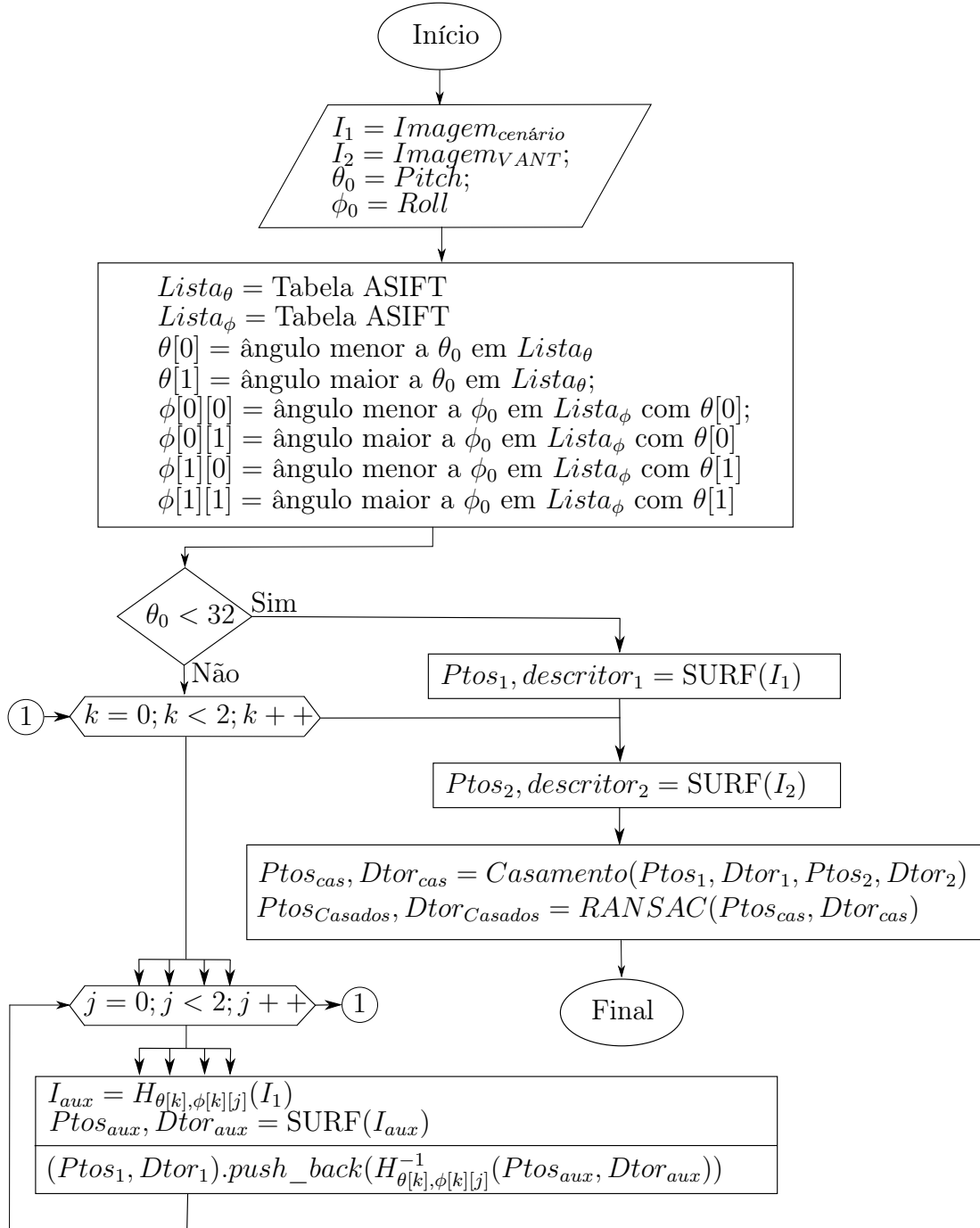


Figura 3.5: Diagrama de Fluxo do algoritmo ASURF-V

Algoritmo 4: Algoritmo ASURF-V**Entrada:** $I_1 = Imagem_{Terroriorial}$; $I_2 = Imagem_{VANT}$; $\theta_0 = Pitch$; $\phi_0 = Roll$;**Saída:** $Ptos_{Casados}$, $Dtor_{Casados}$ **início** $Lista_\theta =$ Tabela 3.1; $Lista_\phi =$ Tabela 3.1; $\theta[0] =$ Ângulo menor a θ_0 em $Lista_\theta$; $\theta[1] =$ Ângulo maior a θ_0 em $Lista_\theta$; $\phi[0][0] =$ Ângulo menor a ϕ_0 em $Lista_\phi$ com $\theta[0]$; $\phi[0][1] =$ Ângulo maior a ϕ_0 em $Lista_\phi$ com $\theta[0]$; $\phi[1][0] =$ Ângulo menor a ϕ_0 em $Lista_\phi$ com $\theta[1]$; $\phi[1][1] =$ Ângulo maior a ϕ_0 em $Lista_\phi$ com $\theta[1]$;**if** $\theta_0 < 32$ **then**| $Ptos_1, descriptor_1 = SURF(I_1)$;**else****repita****repita**| $I_{aux} = H_{\theta[k], \phi[k][j]}(I_1)$;| $Ptos_{aux1}, Dtor_{aux} = SURF(I_{aux})$;| $(Ptos_1, Dtor_1).push_back(H_{\theta[k], \phi[k][j]}^{-1}(Ptos_{aux}, Dtor_{aux}))$;| **até** $j = 0; j < 2; j ++$ (*Paralelizado*);| **até** $k = 0; k < 2; k ++$ (*Paralelizado*);**end** $Ptos_2, Dtor_2 = SURF(I_2)$; $Ptos_{cas1}, Dtor_{cas1} = Casamento(Ptos_1, Dtor_1, Ptos_2, Dtor_2)$; $Ptos_{Casados}, Dtor_{Casados} = RANSAC(Ptos_{cas1}, Dtor_{cas1})$;**fim**

Capítulo 4

Experimentos e Resultados Computacionais

Nesse capítulo são apresentados os experimentos realizados para avaliar o desempenho dos métodos SIFT (Seção 2.2), SURF (Seção 2.3), ASIFT (Seção 2.4), ASIFT-V (Seção 3.1) e ASURF-V (Seção 3.2). Os métodos são implementados para o casamento de um recorte de uma imagem satelital da cidade São José dos Campos - Brasil com imagens recortes com distorções sintéticas dos ângulos *Pitch*, *Yall* e *Roll*, o banco de imagens implementado é detalhado na Seção 4.1. O equipamento e o software onde foram realizados os experimentos são especificados na Seção 4.2. Para avaliar o desempenho dos extratores de características as imagens recortadas foram distribuídas em 6 variações do ângulo *Pitch*, onde é apresentada a maior distorção afim. Nos testes foi analisada a quantidade de pontos obtidos, a distância erro dos pontos casados e o custo computacional medido através do tempo de processamento, o anterior é explicado na Seção 4.3. A Seção 4.4 apresenta a análise estatística implementada nos testes. Finalmente, a Seção 4.5 mostra os resultados obtidos nos experimentos.

4.1 Banco de Imagens

Para a avaliação dos extratores de características foi utilizada a imagem satélite primeiramente apresentada em dos Santos et al. (2013). A imagem satelital mostrada na Figura 4.1 possui aproximadamente 212 MB com dimensões de 13724x16198 pixels, sendo dividida em 33 imagens com dimensões de 800x800 pixels com diferentes características denominadas *cenários*, mostrados na Figura 4.2. Os cenários utilizados, são imagens representativas do setor urbano e setor rural. As imagens do setor urbano contêm muitas características, incluindo quarteirões com diferentes tipos de casas delimitadas por ruas, trechos de avenidas principais e rodovias, praças esportivas, complexos de

pontes, campos de futebol, estádios, armazéns, blocos de edifícios, lotes de construção de cimento e de pasto, diferentes tipos de casas e áreas verdes incluindo árvores. As imagens representativas do setor rural contêm poucas características, incluem áreas verdes com pasto, diferentes vegetações e algumas ruas não pavimentadas.



Figura 4.1: Imagem Satélite.

Para ter controle da locação verdadeira dos pontos casados e da transformada homográfica que representa a distorção das duas imagens nos testes, foram realizadas transformações sintéticas com o intuito de simular imagens obtidas por um VANT com variações nos ângulos θ , ϕ e ψ , além disso aplicou-se um filtro gaussiano 3x3 para simular uma imagem capturada pelo VANT. As transformações foram baseadas na Equação (2.16), realizaram-se 6 variações o ângulo θ , e a amostragem implementada foi de 0° , 16° , 32° , 48° , 64° e 76° . Em cada um dos ângulos θ são realizadas 3 variações do ϕ (0° , 100° , 150°), 3 variações do ψ (0° , 90° , 150°), obtendo 54 transformações sintéticas. A Figura 4.3 mostra 5 diferentes transformações de cenário, (a) $\theta = 0^\circ$, $\phi = 0^\circ$, $\psi = 0^\circ$, (b) $\theta = 0^\circ$, $\phi = 100^\circ$, $\psi = 0^\circ$, (c) $\theta = 16^\circ$, $\phi = 100^\circ$, $\psi = 90^\circ$, (d) $\theta = 48^\circ$, $\phi = 100^\circ$, $\psi = 150^\circ$ e (e) $\theta = 64^\circ$, $\phi = 100^\circ$, $\psi = 150^\circ$.

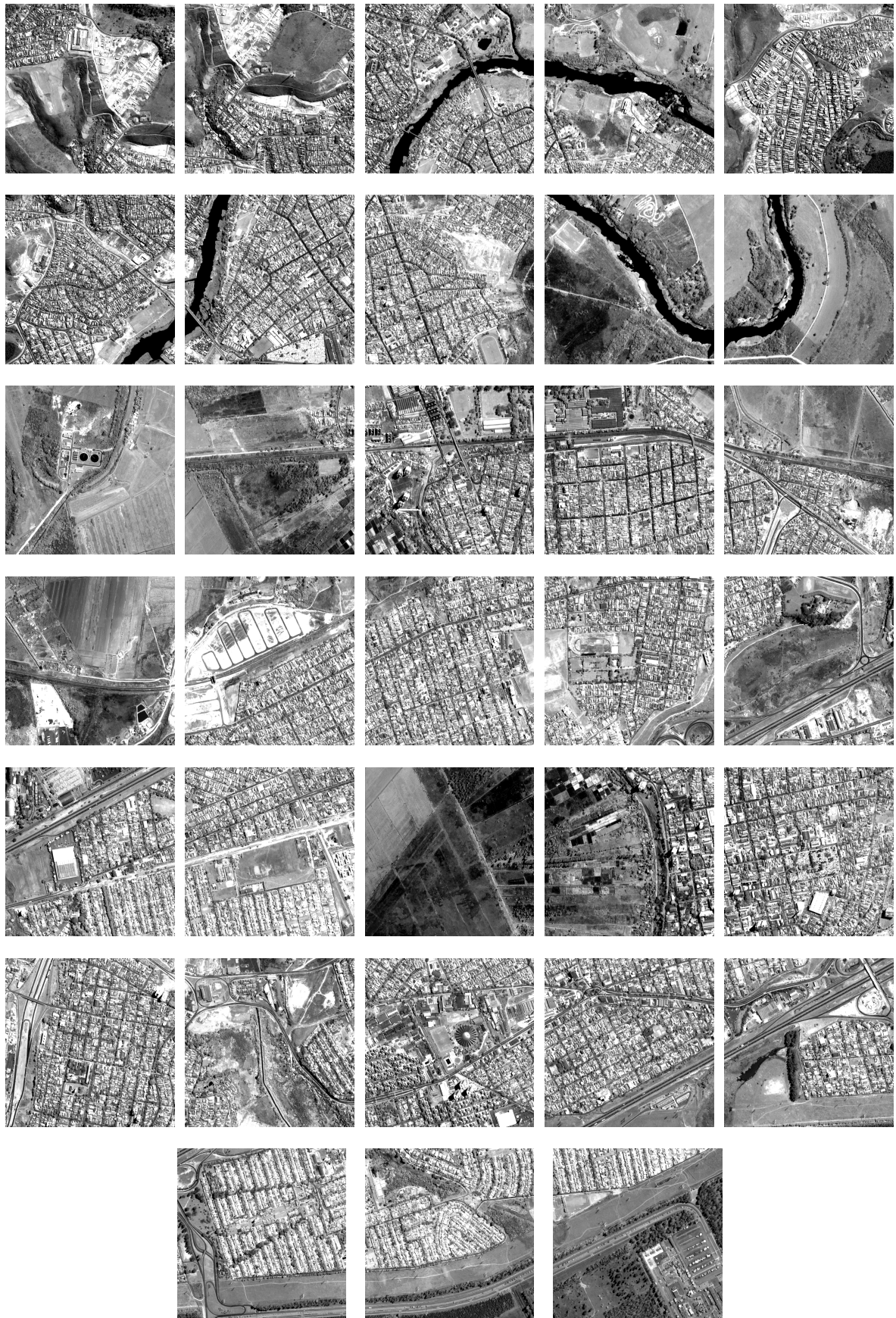


Figura 4.2: Imagens cenários.

Aplicou-se quatro recortes com dimensões de 200x200 pixels em cada transformação do cenário, alcançando um total de 216 imagens para casar com cada cenário, obtendo um total de 7128 imagens, que foram denominadas *recortes*.

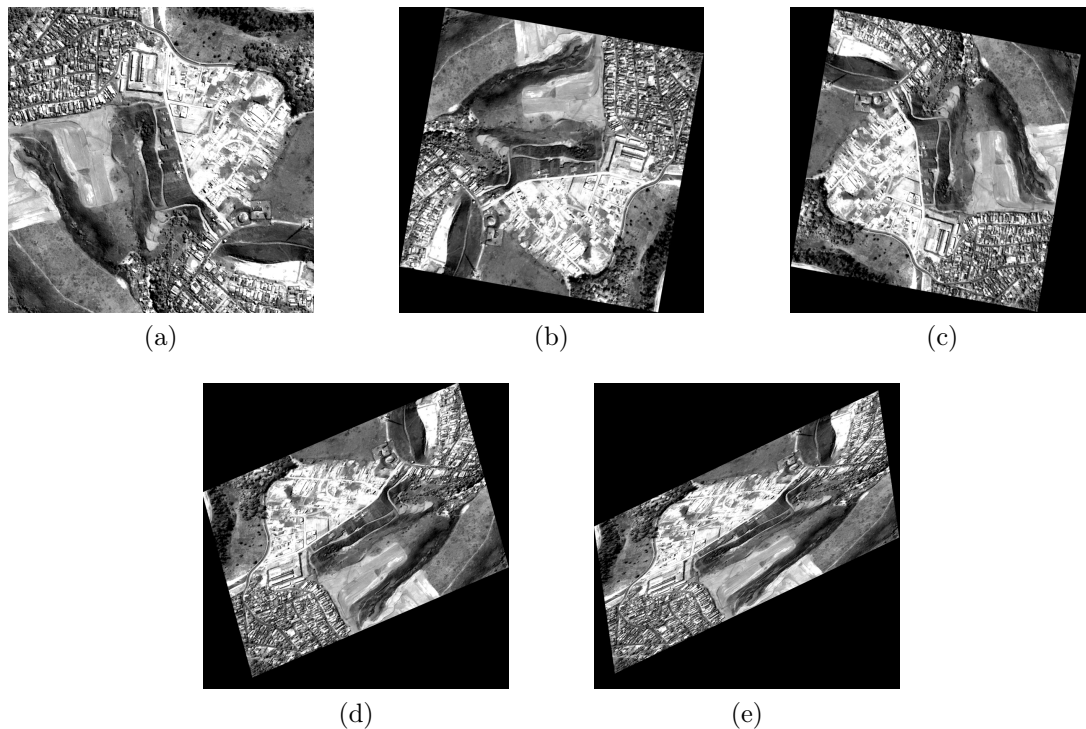


Figura 4.3: Transformações de uma imagem cenário, (a) $\theta = 0^\circ$, $\phi = 0^\circ$, $\psi = 0^\circ$, (b) $\theta = 0^\circ$, $\phi = 100^\circ$, $\psi = 0^\circ$, (c) $\theta = 16^\circ$, $\phi = 100^\circ$, $\psi = 90^\circ$, (d) $\theta = 48^\circ$, $\phi = 100^\circ$, $\psi = 150^\circ$ e (e) $\theta = 64^\circ$, $\phi = 100^\circ$, $\psi = 150^\circ$.

Desta forma, os métodos foram avaliados com casamento de 7128 imagens recortes de dimensão de 200x200 pixels com transformações sintéticas dos ângulos θ , ϕ e ψ , com 33 imagens cenários de dimensão 800x800 pixels.

4.2 Equipamento e Software

Os algoritmos avaliados neste trabalho foram escritos em *C++*, utilizando as bibliotecas: OpenCV 3,0 (*Open Source Computer Vision Library*) para visão computacional e OpenMP (*Open Multi-Processing*) para paralelizar os métodos ASIFT, ASIFT-V e ASURF-V. O sistema operacional utilizado foi Debian 8 Jessie de 64 bits. As simulações foram realizadas em um computador com dois processadores de 6 núcleos físicos intel Xenon(R)CPU X5650 @ 2,67GHzx19 (para um total de 24 threads), e 80 GB de memória RAM.

4.3 Métricas de Avaliação

A avaliação do desempenho dos métodos foi realizada por meio das seguintes métricas:

- A primeira métrica de avaliação, baseado nos trabalhos desenvolvidos por Zheng et al. (2014), Li and Aouf (2012), Romero Cortijo et al. (2009) e Mikolajczyk et al. (2005), é definida como a quantidade de pontos extraídos em cada uma das imagens, a quantidade de pontos casados após a filtragem do método RANSAC e a porcentagem de pontos casados, em cada um dos algoritmos. A porcentagem de pontos casados foi analisada com a quantidade de pontos extraídos nas imagens *recortes*, devido que são de dimensão menor e desta forma, podem estimar corretamente à porcentagem do casamento.
- A segunda métrica de avaliação, baseado em Lingua et al. (2009) y no erro do casamento dos pontos no RANSAC (Hartley and Zisserman (2003)), é definida como o erro do casamento a partir da distância da projeção dos pontos utilizando a transformada homográfica conhecida a priori. Neste trabalho, esta métrica de avaliação foi nomeada como distância-erro² e foi baseada na Equação (2.21) implementada no RANSAC, $d_{\perp}^2 = d(\mathbf{x}, \hat{\mathbf{x}})^2 + d'(\mathbf{x}', \hat{\mathbf{x}}')^2$. A distância-erro² foi dada em pixels².
- Finalmente, a última métrica é definida como o custo computacional, medido a partir do tempo de execução dos algoritmos na extração dos pontos característicos e no casamento das imagens, incluindo a filtragem feita pelo RANSAC. Para medir o tempo de execução foi utilizada a biblioteca padrão da linguagem de programação C para manipulação de datas e horários, <ctime> (time.h).

4.4 Análise Estatística

A análise dos extratores propostos foi realizada separadamente para cada um dos seis ângulos θ . A primeira métrica de avaliação foi analisar a quantidade de pontos casados obtidos. Para técnicas de navegação como reconhecimento de marcos em imagens ou "*landmarks*" e casamentos de imagens precisa-se obter um alto número de pontos casados, para uma boa localização de um objeto ou uma imagem dentro da outra. No entanto, para técnicas como a odometria é mais importante a qualidade dos pontos casados, para uma representação verdadeira do deslocamento entre duas imagens.

Para avaliar os resultados da segunda métrica de avaliação e custo computacional, primeiramente foi realizada uma análise gráfica dos algoritmos de cada um dos θ , para

verificar se os dados fornecem informações conclusivas sobre a diferença entre os algoritmos. Em seguida, foi utilizado o teste estatístico Friedman (Friedman (1940)). Este teste pertence à categoria de métodos não-paramétricos para múltipla comparação de amostras pareadas. O pareamento do teste ocorrerá ao longo das diferentes instâncias em que são rodados os algoritmos, no presente caso, as instâncias são as 33 imagens do banco de imagens descritas na seção 4.1 que alimentam os algoritmos. As hipóteses para realização do teste estatístico Friedman foram formuladas da seguinte maneira:

$$\text{Teste Friedman : } \left\{ \begin{array}{l} H_0 : \text{ Não há diferença em relação ao desempenho entre todos os métodos.} \\ \quad \text{(Hipótese Nula).} \\ H_1 : \text{ Existe diferença em relação ao desempenho entre todos os métodos.} \\ \quad \text{(Hipótese Alternativa).} \end{array} \right.$$

Posteriormente, foi utilizada a análise Post-hoc de superioridade e inferioridade todos contra todos, para todos algoritmos em cada θ . A análise Post-hoc utiliza o teste estatístico Wilcoxon (*Wilcoxon signed rank test*, Wilcoxon (1945)). Este teste pertence à categoria de métodos não-paramétricos para comparação de duas amostras pareadas. Do mesmo modo que o teste Friedman, o pareamento do teste ocorrerá ao longo das as 33 imagens do banco de imagens que alimentam os algoritmos. A definição formal da hipótese do teste estatístico Wilcoxon, foi formulada da seguinte maneira:

$$\text{Teste Wilcoxon : } \left\{ \begin{array}{l} H_0 : \text{ Não há diferença em relação ao desempenho dos métodos.} \\ \quad \text{(Hipótese Nula).} \\ H_1 : \text{ Existe diferença em relação ao desempenho dos métodos.} \\ \quad \text{(Hipótese Alternativa).} \end{array} \right.$$

Os testes foram realizados com um nível de significância igual a 5%.

4.5 Resultados Experimentais

Nesta seção são apresentados os resultados estatísticos dos testes planejados na seção 4.4. A Tabela 4.1 apresenta o valor médio e o desvio padrão da quantidade de pontos característicos obtidos pelos métodos em cada uma das imagens, dos pontos casados após da filtragem do RANSAC e a porcentagem de pontos casados, como foi expressado nas métricas de avaliação (seção 4.3). Os resultados foram divididos em 6 ângulos θ como foi expressado na seção 4.1. Pontos₁ são os pontos extraídos nas imagens recortes, Pontos₂ são a quantidade de pontos extraídos nas imagens cenários, Casados_(pontos) são

a quantidade de pontos obtidos após a filtragem realizada pelo método RANSAC e $\text{Casados}_{(\%)}$ é porcentagem de pontos casados.

Tabela 4.1: Medida pontual e o desvio padrão da média da quantidade de pontos característicos obtidos pelos métodos em cada uma das imagens, da quantidade dos pontos casados após a filtragem do RANSAC e a porcentagem de casamento.

θ	Método	Pontos ₁	Pontos ₂	Casados _(pontos)	Casados _(%)
0°	SIFT	328 ± 99	4530 ± 1175	207 ± 61	63 ± 1
	SURF	266 ± 62	7026 ± 1510	141 ± 28	53 ± 3
	ASIFT	4406 ± 1223	32869 ± 8765	1579 ± 481	35 ± 2
	ASIFT-V	328 ± 99	5215 ± 1431	247 ± 74	76 ± 1
	ASURF-V	266 ± 62	5921 ± 1333	153 ± 31	58 ± 3
16°	SIFT	328 ± 101	4530 ± 1175	204 ± 61	62 ± 2
	SURF	265 ± 63	7026 ± 1510	136 ± 27	52 ± 4
	ASIFT	4391 ± 1217	32869 ± 8765	1566 ± 478	35 ± 2
	ASIFT-V	328 ± 101	5215 ± 1431	241 ± 73	73 ± 1
	ASURF-V	265 ± 63	5921 ± 1333	147 ± 29	56 ± 4
32°	SIFT	325 ± 99	4530 ± 1175	193 ± 57	59 ± 2
	SURF	268 ± 64	7026 ± 1510	118 ± 22	44 ± 4
	ASIFT	4402 ± 1220	32869 ± 8765	1527 ± 466	34 ± 2
	ASIFT-V	325 ± 99	11550 ± 13123	276 ± 85	85 ± 1
	ASURF-V	268 ± 64	13371 ± 2955	170 ± 35	64 ± 4
48°	SIFT	320 ± 97	4530 ± 1175	140 ± 39	44 ± 2
	SURF	275 ± 65	7026 ± 1510	52 ± 8	20 ± 5
	ASIFT	4398 ± 1224	32869 ± 8765	1400 ± 426	31 ± 2
	ASIFT-V	320 ± 97	9374 ± 2500	270 ± 81	84 ± 1
	ASURF-V	274 ± 65	11402 ± 2475	166 ± 34	61 ± 3
64°	SIFT	306 ± 89	4530 ± 1175	10 ± 2	4 ± 2
	SURF	282 ± 65	7026 ± 1510	7 ± 0,5	3 ± 1
	ASIFT	4302 ± 1151	32869 ± 8765	901 ± 259	21 ± 1
	ASIFT-V	306 ± 89	7046 ± 1873	245 ± 71	80 ± 1
	ASURF-V	282 ± 65	9130 ± 1944	153 ± 29	55 ± 3
76°	SIFT	249 ± 75	4530 ± 1175	7 ± 1,4	3 ± 2
	SURF	264 ± 70	7026 ± 1510	7 ± 0,6	3 ± 1
	ASIFT	3669 ± 1092	32868 ± 141	141 ± 39	4 ± 1
	ASIFT-V	242 ± 78	5438 ± 1445	112 ± 34	47 ± 2
	ASURF-V	264 ± 70	7207 ± 1489	63 ± 11	25 ± 4

Segundo os resultados apresentados na Tabela 4.1, ASIFT extrai a maior quantidade de pontos em todas as imagens e em todas as variações dos ângulos θ , com exceção ao θ de 76°, no qual ASIFT-V obteve a maior quantidade de pontos. No entanto, ASIFT apresentou uma grande quantidade de *outliers* no casamento de similaridade, nos primeiros ângulos teve aproximadamente 68% de pontos *outliers* e esta quantidade aumenta proporcionalmente ao longo do ângulo θ . Já o algoritmo SIFT foi eficiente para ângulos θ menores que 65°, com uma porcentagem em torno de 28% de *outliers* no algoritmo de similaridade.

O algoritmo SURF foi eficiente para ângulos θ menores que 48° , com uma presença de *outliers* em torno de 50% no algoritmo de similaridade. No entanto, ele apresentou uma boa quantidade de pontos casados para o ângulo θ de 48° , apesar desse valor ser inferior aos resultados obtidos para demais ângulos menores que 48° .

As propostas ASIFT-V e ASURF-V apresentaram bons resultados em todas as variações dos ângulos θ . Porém, ASIFT-V apresentou uma maior quantidade de pontos casados que ASURF-V. A percentagem de *outliers* de ASIFT-V foi ao redor de 18% e para o ASURF-V de 40%. Deste modo, os métodos ASIFT, ASIFT-V e ASURF-V mostraram-se invariantes a mudanças do ângulo θ , SIFT e SURF evidenciaram ser eficientes para ângulos menores ou iguais a 48° .

A análise anterior só foi realizada apenas a quantidade de pontos casados nas duas imagens, a seguinte análise estatística apresenta a distância-erro² dos pontos casados e o custo computacional medido com o tempo de execução dos métodos de extração de características expostos neste trabalho, para cada ângulo θ separadamente. As sub-seções 4.5.1, 4.5.2, 4.5.3, 4.5.4, 4.5.5 e 4.5.6 apresentam as análises para $\theta = 0^\circ, 16^\circ, 32^\circ, 48^\circ, 64^\circ$ e 76° respectivamente. Por conseguinte, apresenta-se uma análise completa dos métodos de extração de características para variações do ângulo *Pitch* com foco na localização em navegação autônoma de VANTs.

4.5.1 Análise estatística para θ de 0°

De acordo com a seção 4.4, primeiramente foi realizada uma análise gráfica da segunda métrica de avaliação e o custo computacional apresentados na seção 4.3. A Figura 4.4a apresenta em boxplots a distância-erro² em pixels² obtida pelo casamento dos pontos em cada um dos métodos. A Figura 4.4b apresenta em boxplots o tempo de execução em segundos que utilizou cada um dos métodos em fazer a extração, o casamento e a filtragem de pontos.

Os boxplots da distância-erro² indicaram uma maior eficiência dos métodos ASIFT-V e SIFT. Porém, todos os métodos fazem um bom casamento de ponto entre as imagens, portanto a análise estatística da distância-erro² foi realizada para todos os métodos. Os boxplots do tempo indicaram uma grande inferioridade do método ASIFT, dessa forma a análise estatística foi realizada para métodos SIFT, SURF, ASIFT-V e ASURF-V.

Aplicando-se o teste de Friedman para avaliar a distância-erro², obteve-se um p-valor igual a $2,2 \times 10^{-16}$, verificando-se que p-valor $\leq 0,05$, pode-se concluir com um nível de confiança de 95% que a hipótese nula foi rejeitada em todos os casos. Portanto, pode-se afirmar que existe uma diferença do desempenho nos métodos propostos. Aplicando-se o Post-hoc de superioridade e inferioridade todos contra todos, obteve-

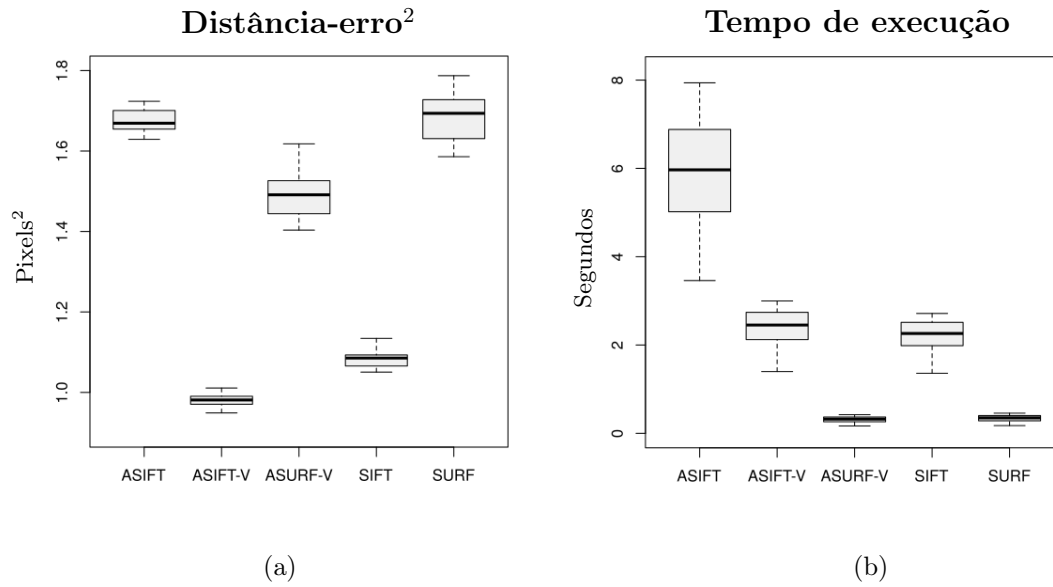


Figura 4.4: Análise gráfica de cada um dos métodos com θ de 0° , os dados são representados por boxplots. A Figura 4.4a apresenta a distância-erro² em pixels² obtida pelo casamento final dos pontos de interesse. A Figura 4.4b apresenta o tempo de execução em segundos que utilizou cada um dos métodos em fazer a extração de pontos, o casamento e a filtragem de pontos.

se os resultados de p-value para as comparações, apresentados na Tabela 4.2. Em complemento, a Figura 4.5 apresenta os resultados da análise Pos-hoc.

Tabela 4.2: Análise Post-hoc de superioridade e inferioridade todos contra todos, para avaliar a distância-erro² dos métodos ASIFT, ASIFT-V, ASURF-V, SIFT e SURF, com θ de 0°

Post-hoc :	Superioridade	Inferioridade
ASIFT-V - ASIFT	$1,0000 \times 10^0$	$0,0000 \times 10^0$
ASURF-V - ASIFT	$1,0000 \times 10^0$	$2,5432 \times 10^{-4}$
SIFT - ASIFT	$1,0000 \times 10^0$	$8,9705 \times 10^{-10}$
SURF - ASIFT	$9,9907 \times 10^{-1}$	$9,6491 \times 10^{-1}$
ASURF-V - ASIFT-V	$9,1406 \times 10^{-6}$	$1,0000 \times 10^0$
SIFT - ASIFT-V	$4,1947 \times 10^{-2}$	$1,0000 \times 10^0$
SURF - ASIFT-V	$0,0000 \times 10^0$	$1,0000 \times 10^0$
SIFT - ASURF-V	$1,0000 \times 10^0$	$9,1163 \times 10^{-2}$
SURF - ASURF-V	$6,5296 \times 10^{-4}$	$1,0000 \times 10^0$
SURF - SIFT	$2,6467 \times 10^{-9}$	$1,0000 \times 10^0$

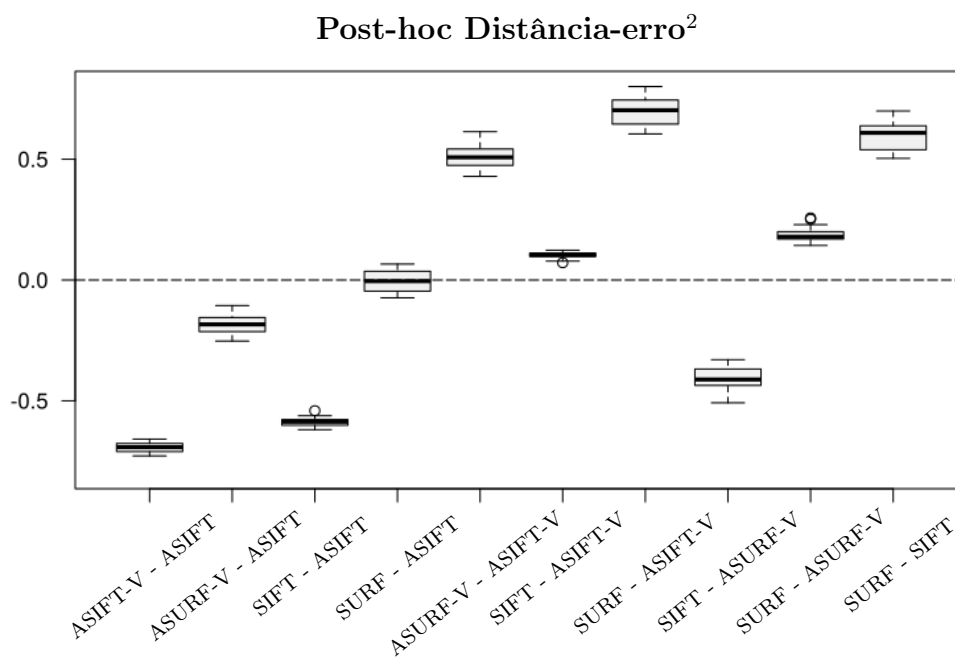


Figura 4.5: Análise Post-hoc para avaliar a distância-erro² dos métodos ASIFT, ASIFT-V, ASURF-V, SIFT e SURF, com θ de 0°

Aplicando-se o teste de Friedman para avaliar o tempo de execução, obteve-se um p-valor igual a $2,2 \times 10^{-16}$, verificando-se que p-valor $\leq 0,05$, pode-se concluir com um nível de confiança de 95% que a hipótese nula foi rejeitada em todos os casos. Portanto, pode-se afirmar que existe uma diferença do tempo de execução nos métodos propostos. Aplicando-se o Post-hoc de superioridade e inferioridade todos contra todos, obteve-se os resultados de p-value para as comparações, apresentados na Tabela 4.3. Em complemento, a Figura 4.6 apresenta os resultados da análise Pos-hoc.

Tabela 4.3: Análise Post-hoc de superioridade e inferioridade todos contra todos, para avaliar o tempo de execução dos métodos ASIFT-V, ASURF-V, SIFT e SURF, com θ de 0°

Post-hoc :	Superioridade	Inferioridade
ASURF-V - ASIFT-V	$1,1000 \times 10^0$	$0,0000 \times 10^0$
SIFT - ASIFT-V	$1,0000 \times 10^0$	$4,3838 \times 10^{-3}$
SURF - ASIFT-V	$1,0000 \times 10^0$	$6,1934 \times 10^{-10}$
SIFT - ASURF-V	$4,7341 \times 10^{-10}$	$1,0000 \times 10^0$
SURF - ASURF-V	$4,8067 \times 10^{-3}$	$1,0000 \times 10^0$
SURF - SIFT	$1,0000 \times 10^0$	$4,3838 \times 10^{-3}$

Post-hoc Custo Computacional

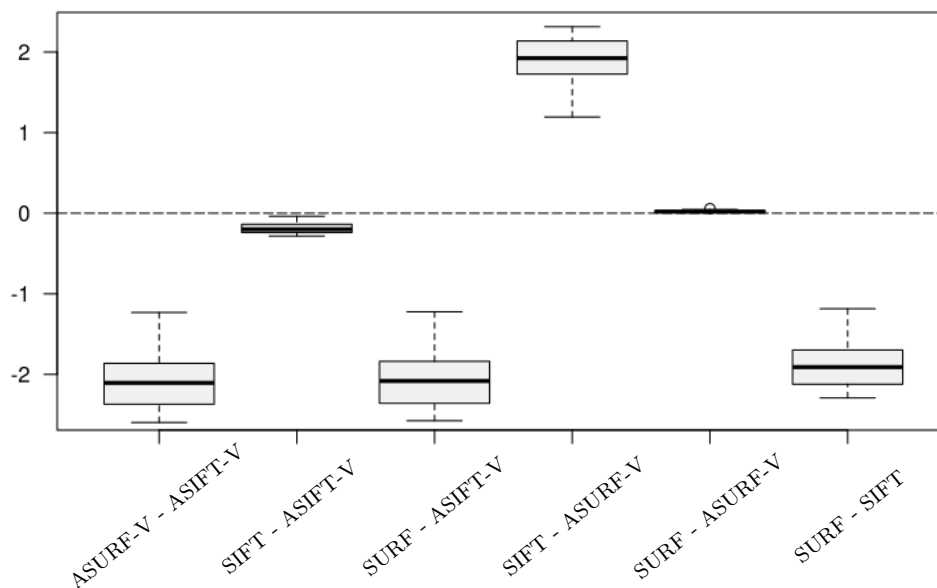


Figura 4.6: Análise Post-hoc para avaliar o tempo de execução dos métodos ASIFT-V, ASURF-V, SIFT e SURF, com θ de 0°

O testes apresentados para o θ de 0° mostraram que todos os métodos conseguiram fazer um bom casamento. Porém, uma grande desvantagem do método ASIFT é o custo computacional. Os resultados mostraram que o método não pode ser implementado em navegação autônoma por possuir um tempo de processamento muito alto. Por outro lado, os métodos ASIFT-V e ASURF-V propostos neste trabalho apresentaram um bom custo-benefício em relação à quantidade de pontos obtidos, distância-erro² do casamento e o tempo de execução. ASIFT-V foi o melhor método no teste da distância-erro² do casamento dos pontos de interesse com um nível de confiança de 95%, além de apresentar um bom tempo de execução em torno de 2,3 segundos. Com um nível de confiança de 95% ASURF-V foi o método mais rápido para fazer a extração de características e o casamento das imagens, além de apresentar uma boa distância-erro² (em torno de 1,5 pixels²).

4.5.2 Análise estatística para θ de 16°

Do mesmo modo que a análise realizada para θ de 0° , primeiramente foi realizada uma análise gráfica da segunda métrica de avaliação e o custo computacional. A Figura 4.7a apresenta em boxplots a distância-erro² em pixels² obtida pelo casamento dos pontos para cada um dos métodos. A Figura 4.7b apresenta em boxplots o tempo de execução em segundos que cada um dos métodos utilizou em fazer a extração, o casamento e a filtragem dos pontos *outliers*.

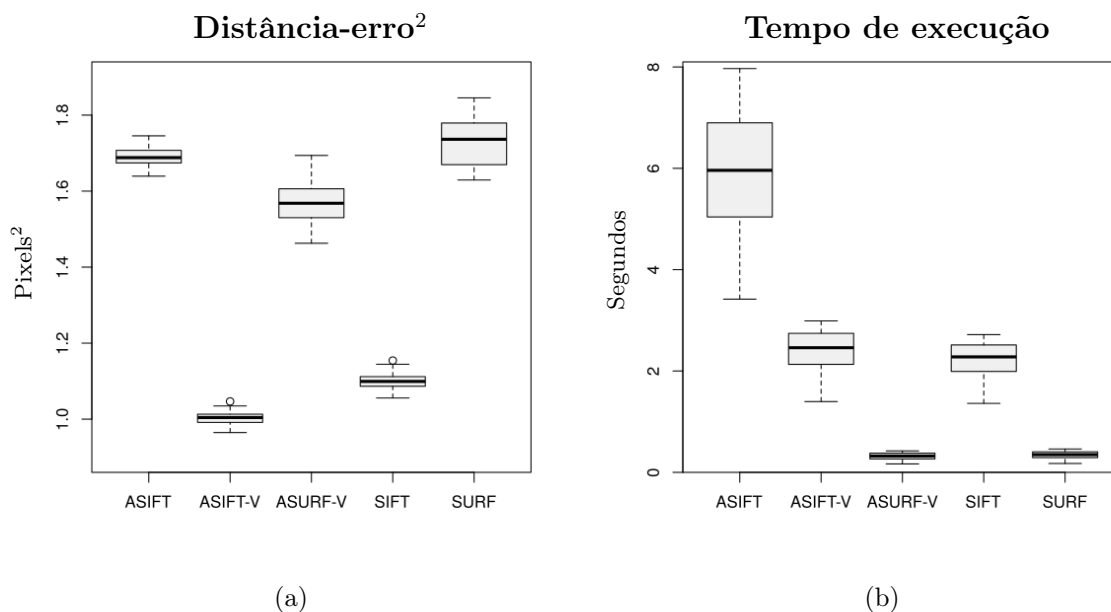


Figura 4.7: Análise gráfica de cada um dos métodos com θ de 16° , os dados são representados por boxplots. A Figura 4.7a apresenta a distância-erro² em pixels² obtida pelo casamento final dos pontos de interesse. A Figura 4.7b apresenta o tempo de execução em segundos que utilizou cada um dos métodos em fazer a extração de pontos, o casamento e a filtragem de pontos.

Os boxplots da distância-erro² indicaram uma maior eficiência dos métodos ASIFT-V e SIFT. Porém, todos os métodos fazem um bom casamento de ponto entre as imagens, portanto a análise estatística da distância-erro² foi realizada em todos os métodos. Os boxplots do tempo de execução indicaram uma grande inferioridade do método ASIFT, dessa forma a análise estatística foi realizada pelos métodos SIFT, SURF, ASIFT-V e ASURF-V.

Aplicando-se o teste de Friedman para avaliar a distância-erro², obteve-se um p-valor igual a $2,2 \times 10^{-16}$, verificando-se que p-valor $\leq 0,05$, pode-se concluir com um nível de confiança de 95% que a hipótese nula foi rejeitada para todos os casos. Portanto, pode-se afirmar que existe uma diferença do desempenho nos métodos propostos. Aplicando-se o Post-hoc de superioridade e inferioridade todos contra todos, obteve-se os resultados de p-value para as comparações, apresentados na Tabela 4.4. Em complemento, a Figura 4.8 apresenta estes resultados.

Aplicando-se o teste de Friedman para avaliar o tempo de execução, obteve-se um p-valor igual a $2,2 \times 10^{-16}$, verificando-se que p-valor $\leq 0,05$, pode-se concluir com um nível de confiança de 95% que a hipótese nula foi rejeitada em todos os casos. Portanto, pode-se afirmar que existe uma diferença do tempo de execução nos métodos propostos. Aplicando-se o Post-hoc de superioridade e inferioridade todos contra todos,

Tabela 4.4: Análise Post-hoc de superioridade e inferioridade todos contra todos, para avaliar a distância-erro² dos métodos ASIFT, ASIFT-V, ASURF-V, SIFT e SURF, com θ de 16°

Post-hoc :	Superioridade	Inferioridade
ASIFT-V - ASIFT	$1,0000 \times 10^0$	$5,5511 \times 10^{-16}$
ASURF-V - ASIFT	$1,0000 \times 10^0$	$2,1547 \times 10^{-3}$
SIFT - ASIFT	$1,0000 \times 10^0$	$3,3091 \times 10^{-8}$
SURF - ASIFT	$7,4507 \times 10^{-1}$	$1,0000 \times 10^0$
ASURF-V - ASIFT-V	$7,5425 \times 10^{-6}$	$1,0000 \times 10^0$
SIFT - ASIFT-V	$4,1924 \times 10^{-2}$	$1,0000 \times 10^0$
SURF - ASIFT-V	$0,0000 \times 10^0$	$1,0000 \times 10^0$
SIFT - ASURF-V	$1,0000 \times 10^0$	$9,1098 \times 10^{-2}$
SURF - ASURF-V	$6,2348 \times 10^{-5}$	$1,0000 \times 10^0$
SURF - SIFT	$8,6975 \times 10^{-11}$	$1,0000 \times 10^0$

Post-hoc Distância-erro²

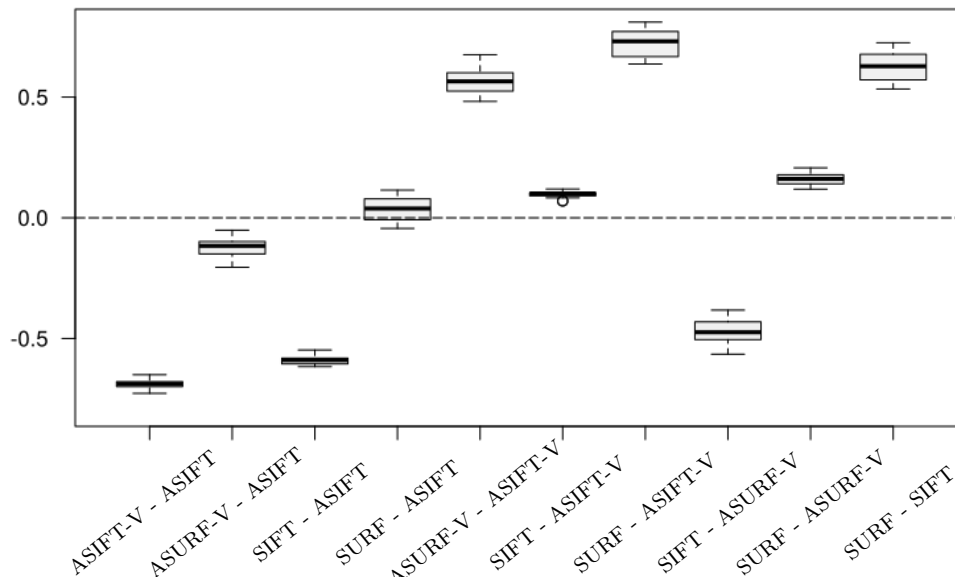


Figura 4.8: Análise Post-hoc para avaliar a distância-erro² dos métodos ASIFT, ASIFT-V, ASURF-V, SIFT e SURF, com θ de 16°

obteve-se os resultados de p-value para as comparações, apresentados na Tabela 4.5. Em complemento, a Figura 4.9 apresenta estes resultados.

Tabela 4.5: Análise Post-hoc de superioridade e inferioridade todos contra todos, para avaliar o tempo de execução dos métodos ASIFT-V, ASURF-V, SIFT e SURF, com θ de 16°

Post-hoc :	Superioridade	Inferioridade
ASURF-V - ASIFT-V	$1,0000 \times 10^0$	$0,0000 \times 10^0$
SIFT - ASIFT-V	$1,0000 \times 10^0$	$4,7356 \times 10^{-3}$
SURF - ASIFT-V	$1,0000 \times 10^0$	$1,7522 \times 10^{-8}$
SIFT - ASURF-V	$4,0132 \times 10^{-10}$	$1,0000 \times 10^0$
SURF - ASURF-V	$4,5691 \times 10^{-3}$	$1,0000 \times 10^0$
SURF - SIFT	$1,0000 \times 10^0$	$4,7356 \times 10^{-3}$

Post-hoc Custo Computacional

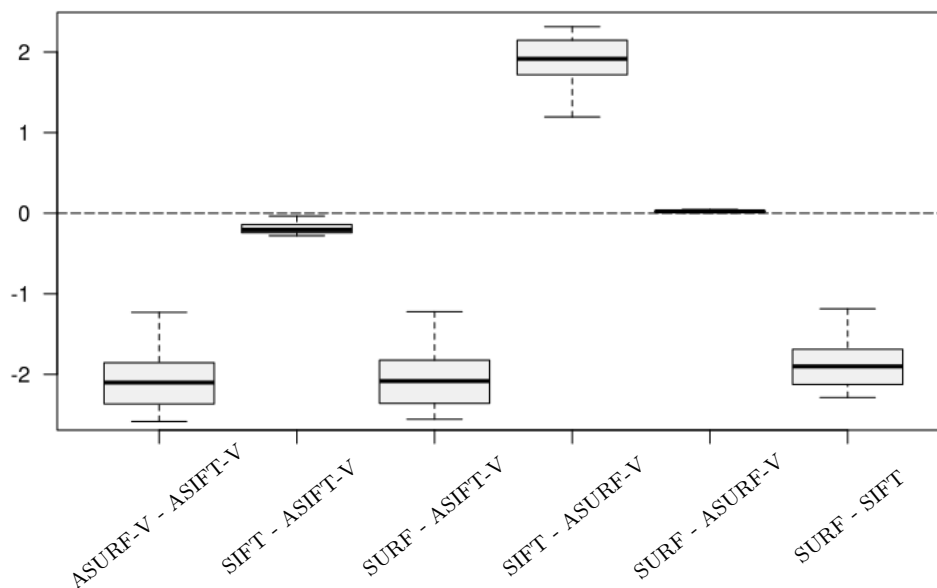


Figura 4.9: Análise Post-hoc para avaliar o tempo de execução dos métodos ASIFT-V, ASURF-V, SIFT e SURF, com θ de 16°

O testes apresentados para o θ de 16° mostraram que todos os métodos conseguiram fazer um bom casamento. Porém, uma grande desvantagem do método ASIFT é o custo computacional, os resultados mostraram que o método não pode ser implementado em navegação autônoma por possuir um tempo de processamento muito alto. Por outro lado, os métodos ASIFT-V e ASURF-V propostos neste trabalho apresentaram um bom custo-benefício em relação à quantidade de pontos obtidos, a distância-erro² do casamento e o tempo de execução. ASIFT-V foi o melhor método no teste da distância-erro² do casamento dos pontos com um nível de confiança de 95%, além de apresentar um bom tempo de execução em torno de 2,4 segundos. Com um nível de confiança de 95% ASURF-V foi o método mais rápido para fazer a extração de características e o casamento das imagens, além de apresentar uma boa distância-erro² (em torno de 1,6 pixels²).

4.5.3 Análise estatística para θ de 32°

A análise gráfica da segunda métrica de avaliação e o custo computacional para o θ de 32° foi apresentada na Figura 4.10a e na Figura 4.10b, respectivamente. Os boxplots da distância-erro² indicaram uma maior eficiência dos métodos ASIFT-V e SIFT. Da mesma forma que para o θ de 0° e 16° , todos os métodos fazem um bom casamento de ponto entre as imagens, portanto a análise estatística da distância-erro² foi realizada para todos os métodos. Os boxplots do tempo indicaram uma inferioridade do método

ASIFT, todavia não pode-se excluir da avaliação. Dessa forma a análise estatística do custo computacional foi realizada para todos os métodos.

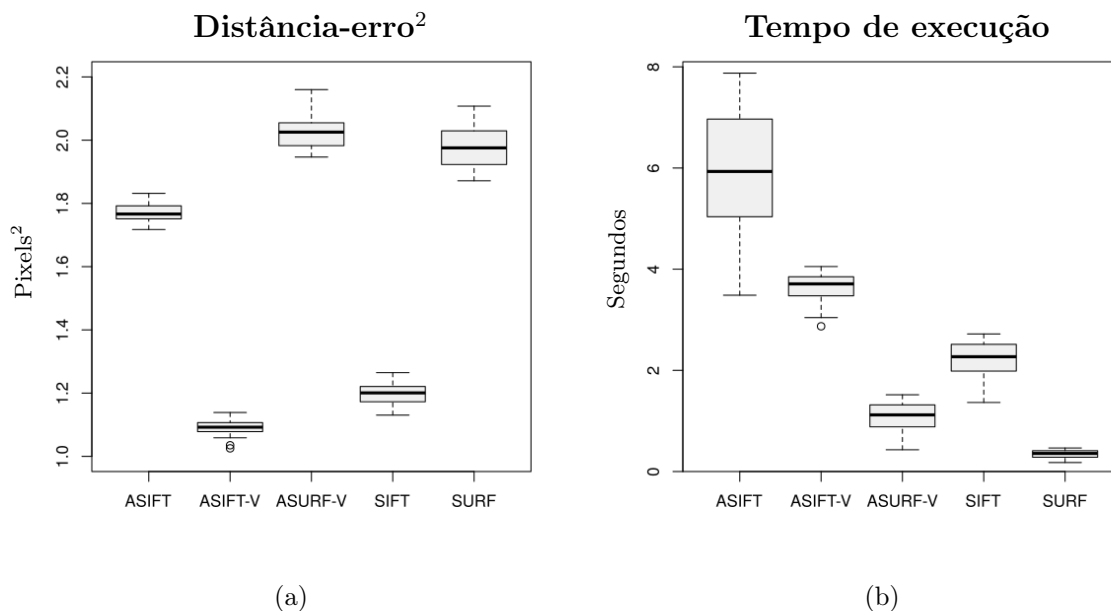


Figura 4.10: Análise gráfica de cada um dos métodos com θ de 32° , os dados são representados por boxplots. A Figura 4.10a apresenta a distância-erro² em pixels² obtida pelo casamento final dos pontos de interesse. A Figura 4.10b apresenta o tempo de execução em segundos que utilizou cada um dos métodos em fazer a extração de pontos, o casamento e a filtragem de pontos.

Aplicando-se o teste de Friedman para avaliar a distância-erro², obteve-se um p-valor igual a $2,2 \times 10^{-16}$, verificando-se que p-valor $\leq 0,05$, pode-se concluir com um nível de confiança de 95% que a hipótese nula foi rejeitada em todos os casos. Portanto, pode-se afirmar que existe uma diferença do desempenho nos métodos propostos. Aplicando-se o Post-hoc de superioridade e inferioridade todos contra todos, obteve-se os resultados de p-value para as comparações, apresentados na Tabela 4.6. Em complemento, a Figura 4.11 apresenta estes resultados.

Tabela 4.6: Análise Post-hoc de superioridade e inferioridade todos contra todos, para avaliar a distância-erro² dos métodos ASIFT, ASIFT-V, ASURF-V, SIFT e SURF, com θ de 32°

Post-hoc :	Superioridade	Inferioridade
ASIFT-V - ASIFT	$1,0000 \times 10^0$	$3,4644 \times 10^{-7}$
ASURF-V - ASIFT	$4,8084 \times 10^{-5}$	$1,0000 \times 10^0$
SIFT - ASIFT	$1,0000 \times 10^0$	$2,7241 \times 10^{-2}$
SURF - ASIFT	$1,5155 \times 10^{-1}$	$1,0000 \times 10^0$
ASURF-V - ASIFT-V	$0,0000 \times 10^0$	$1,0000 \times 10^0$
SIFT - ASIFT-V	$4,1859 \times 10^{-2}$	$1,0000 \times 10^0$
SURF - ASIFT-V	$4,1589 \times 10^{-13}$	$1,0000 \times 10^0$
SIFT - ASURF-V	$1,0000 \times 10^0$	$1,3091 \times 10^{-12}$
SURF - ASURF-V	$1,0000 \times 10^0$	$6,2668 \times 10^{-2}$
SURF - SIFT	$1,0331 \times 10^{-5}$	$1,0000 \times 10^0$

Post-hoc Distância-erro²

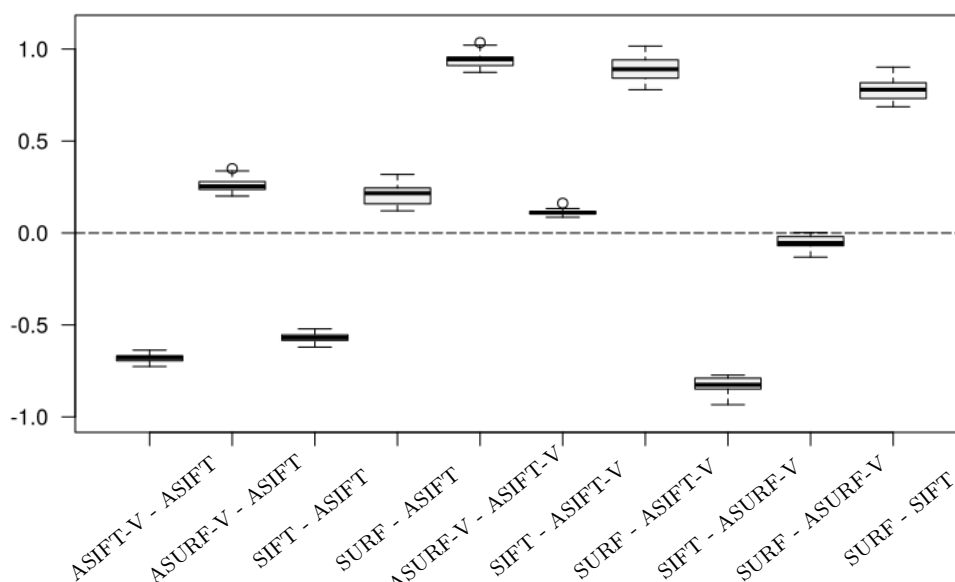


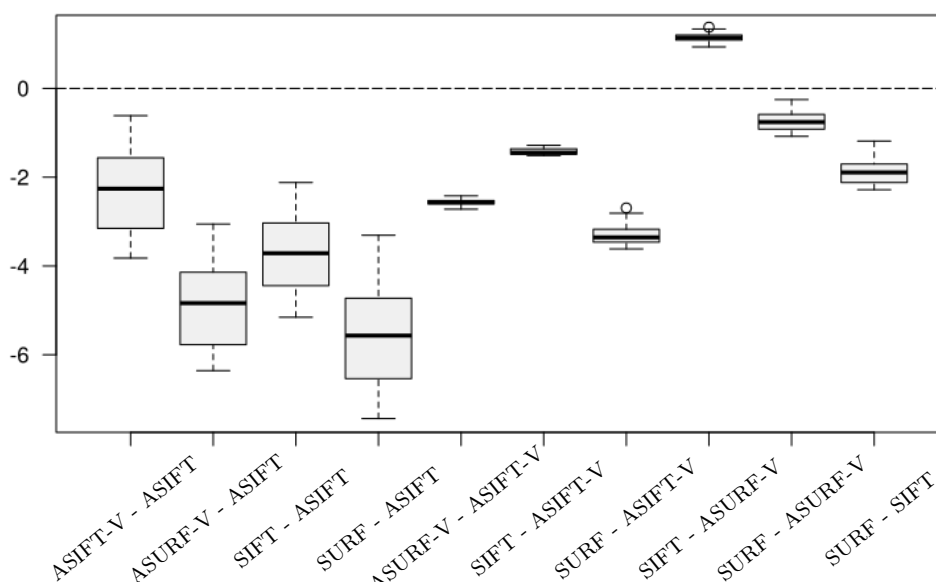
Figura 4.11: Análise Post-hoc para avaliar a distância-erro² dos métodos ASIFT, ASIFT-V, ASURF-V, SIFT e SURF, com θ de 32°

Aplicando-se o teste de Friedman para avaliar o tempo de execução, obteve-se um p-valor igual a $2,2 \times 10^{-16}$, verificando-se que p-valor $\leq 0,05$, pode-se concluir com um nível de confiança de 95% que a hipótese nula foi rejeitada em todos os casos. Portanto, pode-se afirmar que existe uma diferença do tempo de execução nos métodos propostos. Aplicando-se o Post-hoc de superioridade e inferioridade todos contra todos, obteve-se os resultados de p-value para as comparações, apresentados na Tabela 4.7. Em complemento, a Figura 4.12 apresenta estes resultados.

Tabela 4.7: Análise Post-hoc de superioridade e inferioridade todos contra todos, para avaliar o tempo de execução dos métodos ASIFT, ASIFT-V, ASURF-V, SIFT e SURF, com θ de 32°

Post-hoc :	Superioridade	Inferioridade
ASIFT-V - ASIFT	$1,0000 \times 10^0$	$4,1918 \times 10^{-2}$
ASURF-V - ASIFT	$1,0000 \times 10^0$	$5,1514 \times 10^{-14}$
SIFT - ASIFT	$1,0000 \times 10^0$	$1,5027 \times 10^{-6}$
SURF - ASIFT	$1,0000 \times 10^0$	$0,0000 \times 10^0$
ASURF-V - ASIFT-V	$1,0000 \times 10^0$	$1,5027 \times 10^{-6}$
SIFT - ASIFT-V	$1,0000 \times 10^0$	$4,1918 \times 10^{-2}$
SURF - ASIFT-V	$1,0000 \times 10^0$	$5,1514 \times 10^{-14}$
SIFT - ASURF-V	$0,0419 \times 10^0$	$1,0000 \times 10^0$
SURF - ASURF-V	$1,0000 \times 10^0$	$4,1918 \times 10^{-2}$
SURF - SIFT	$1,0000 \times 10^0$	$1,5027 \times 10^{-6}$

Post-hoc Custo Computacional

Figura 4.12: Análise Post-hoc para avaliar o tempo de execução dos métodos ASIFT, ASIFT-V, ASURF-V, SIFT e SURF, com θ de 32°

O resultado da análise Post-hoc, demonstrou que todos os métodos conseguem fazer um bom casamento para o ângulo θ de 32° . Neste análise, os métodos ASIFT-V e ASURF-V apresentaram um bom custo-benefício em relação à quantidade de pontos casados e a distancia-erro². Entretanto, nos experimentos realizados, ASIFT-V foi o algoritmo que apresento uma menor distancia-erro² do casamento dos pontos com um nível de confiança do 95%, com um bom tempo de execução em torno de 2,63 segundos. Porém, o método ASURF-V foi o método mais rápido em casar imagens, com uma distância-erro² em torno de 2 pixels².

4.5.4 Análise estatística para θ de 48°

Seguindo a sequência da análise dos ângulos θ anteriores, primeiramente foi realizada a análise gráfica para verificar se os dados fornecem informações conclusivas sobre a diferença entre os algoritmos. A análise gráfica da segunda métrica de avaliação e o custo computacional para o θ de 48° foi apresentada na Figura 4.13a e na Figura 4.13b respectivamente. Os boxplots da distância-erro² indicaram uma maior eficiência dos métodos ASIFT-V e SIFT. Entretanto, todos os métodos fazem um bom casamento de ponto entre as imagens, portanto a análise da distância-erro² foi realizada em todos os métodos. Os boxplots do tempo indicaram uma inferioridade do método ASIFT, todavia não pode-se excluir da análise. Dessa forma a análise do custo computacional foi realizada para todos os métodos.

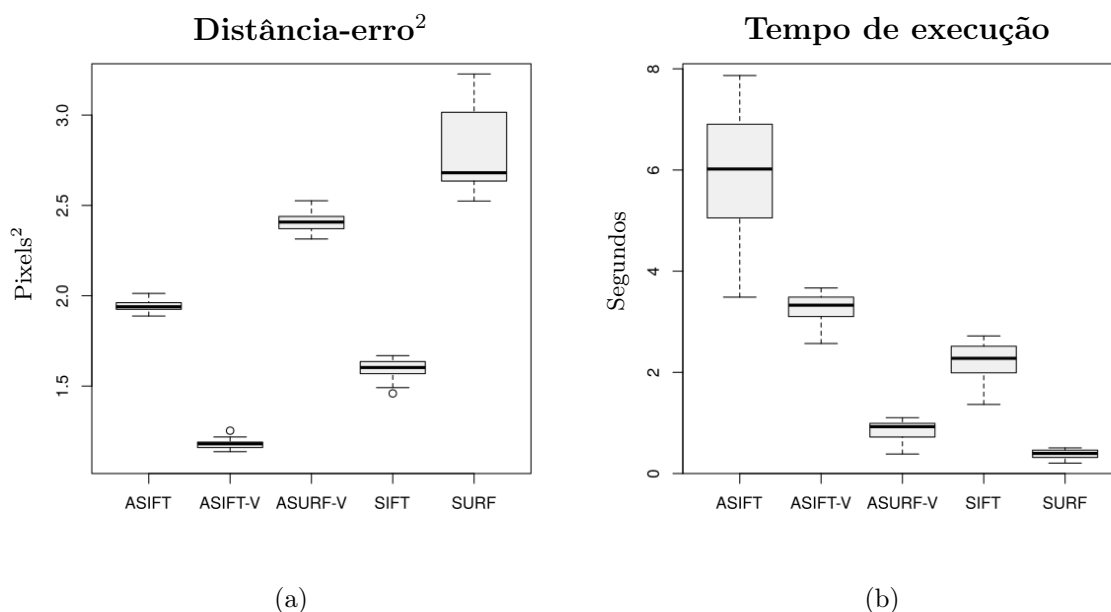


Figura 4.13: Análise gráfica de cada um dos métodos com θ de 48° , os dados são representados por boxplots. A Figura 4.13a apresenta a distância-erro² em pixels² obtida pelo casamento final dos pontos de interesse. A Figura 4.13b apresenta o tempo de execução em segundos que utilizou cada um dos métodos em fazer a extração de pontos, o casamento e a filtragem de pontos.

Aplicando-se o teste Friedman para avaliar a distância-erro², obteve-se um p-valor igual a $2,2 \times 10^{-16}$, verificando-se que p-valor $\leq 0,05$, pode-se concluir com um nível de confiança de 95% que a hipótese nula foi rejeitada em todos os casos. Sendo assim, pode-se afirmar que existe uma diferença do desempenho nos métodos propostos. Aplicando-se o Post-hoc de superioridade e inferioridade todos contra todos, obteve-se os resultados de p-value para as comparações, apresentados na Tabela 4.8. Em complemento, a Figura 4.14 apresenta estes resultados.

Tabela 4.8: Análise Post-hoc de superioridade e inferioridade todos contra todos, para avaliar a distância-erro² dos métodos ASIFT, ASIFT-V, ASURF-V, SIFT e SURF, com θ de 48°

Post-hoc :	Superioridade	Inferioridade
ASIFT-V - ASIFT	$1,0000 \times 10^0$	$6,7193 \times 10^{-8}$
ASURF-V - ASIFT	$2,3674 \times 10^{-2}$	$1,0000 \times 10^0$
SIFT - ASIFT	$1,0000 \times 10^0$	$1,7305 \times 10^{-2}$
SURF - ASIFT	$5,5703 \times 10^{-5}$	$1,0000 \times 10^0$
ASURF-V - ASIFT-V	$3,6826 \times 10^{-13}$	$1,0000 \times 10^0$
SIFT - ASIFT-V	$2,7357 \times 10^{-2}$	$1,0000 \times 10^0$
SURF - ASIFT-V	$0,0000 \times 10^0$	$1,0000 \times 10^0$
SIFT - ASURF-V	$1,0000 \times 10^0$	$1,4247 \times 10^{-5}$
SURF - ASURF-V	$4,1937 \times 10^{-2}$	$1,0000 \times 10^0$
SURF - SIFT	$7,9592 \times 10^{-13}$	$1,0000 \times 10^0$

Post-hoc Distância-erro²

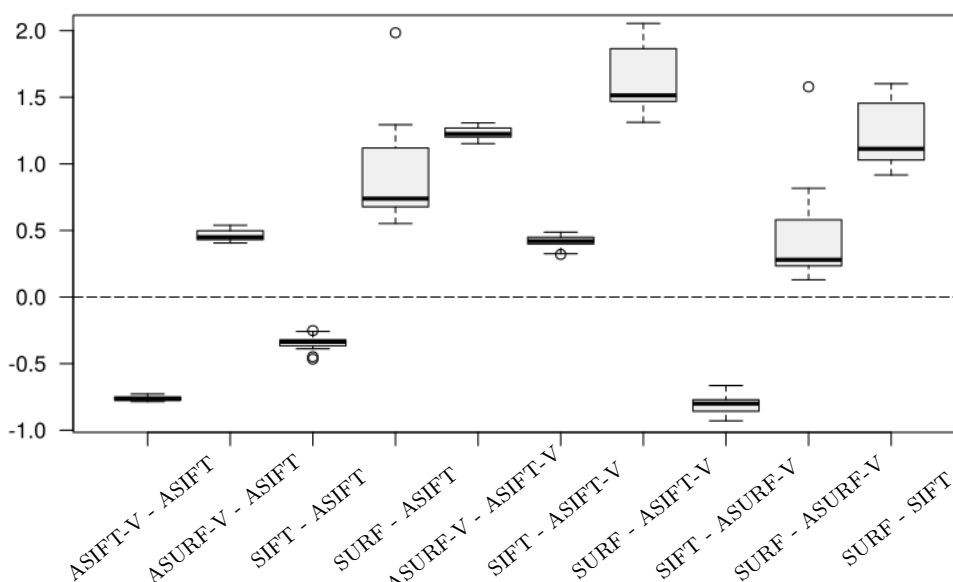


Figura 4.14: Análise Post-hoc para avaliar a distância-erro² dos métodos ASIFT, ASIFT-V, ASURF-V, SIFT e SURF, com θ de 48°

Aplicando-se o teste Friedman para avaliar o tempo de execução, obteve-se um p-valor igual a $2,2 \times 10^{-16}$, verificando-se que p-valor $\leq 0,05$, pode-se concluir com um nível de confiança de 95% que a hipótese nula foi rejeitada em todos os casos. Portanto, pode-se afirmar que existe uma diferença do tempo de execução nos métodos propostos. Aplicando-se o Post-hoc de superioridade e inferioridade todos contra todos, obteve-se os resultados de p-value para as comparações, apresentados na Tabela 4.9. Em complemento, a Figura 4.15 apresenta estes resultados.

Tabela 4.9: Análise Post-hoc de superioridade e inferioridade todos contra todos, para avaliar o tempo de execução dos métodos ASIFT, ASIFT-V, ASURF-V, SIFT e SURF, com θ de 48°

Post-hoc :	Superioridade	Inferioridade
ASIFT-V - ASIFT	$1,0000 \times 10^0$	$4,1918 \times 10^{-2}$
ASURF-V - ASIFT	$1,0000 \times 10^0$	$4,7407 \times 10^{-14}$
SIFT - ASIFT	$1,0000 \times 10^0$	$9,6587 \times 10^{-7}$
SURF - ASIFT	$1,0000 \times 10^0$	$0,0000 \times 10^0$
ASURF-V - ASIFT-V	$1,0000 \times 10^0$	$9,6587 \times 10^{-7}$
SIFT - ASIFT-V	$1,0000 \times 10^0$	$4,1901 \times 10^{-2}$
SURF - ASIFT-V	$1,0000 \times 10^0$	$4,7407 \times 10^{-14}$
SIFT - ASURF-V	$0,0420 \times 10^0$	$1,0000 \times 10^0$
SURF - ASURF-V	$1,0000 \times 10^0$	$4,1901 \times 10^{-2}$
SURF - SIFT	$1,0000 \times 10^0$	$9,6586 \times 10^{-7}$

Post-hoc Custo Computacional

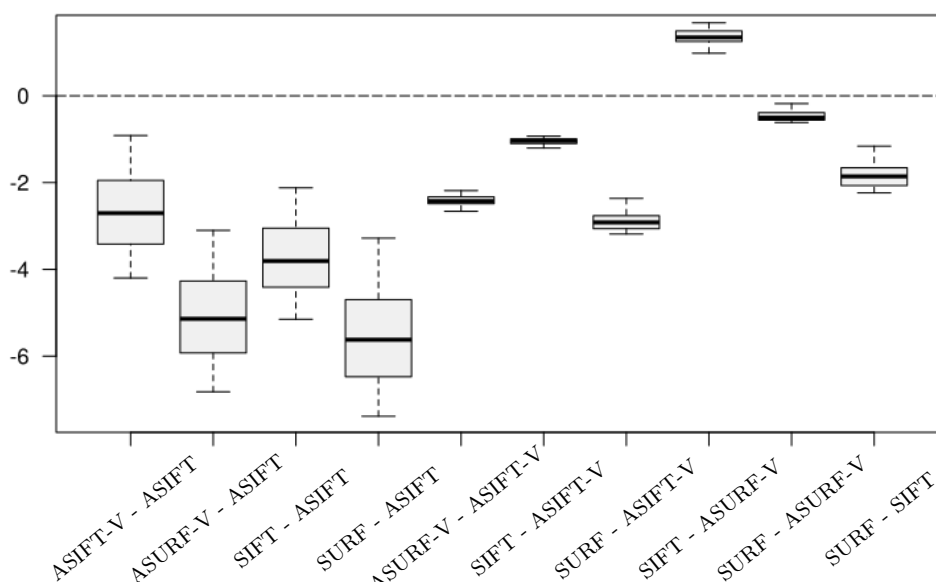


Figura 4.15: Análise Post-hoc para avaliar o tempo de execução dos métodos ASIFT, ASIFT-V, ASURF-V, SIFT e SURF, com θ de 48°

O testes apresentados para o θ de 48° mostraram que todos os métodos conseguiram fazer um bom casamento. Os métodos ASIFT-V e ASURF-V propostos apresentaram um bom custo-benefício em relação à quantidade de pontos obtidos, distância-erro² do casamento e o tempo de execução. ASIFT-V foi o melhor método com um nível de confiança de 95% no teste da distância-erro² do casamento dos pontos, além de apresentar um bom tempo de execução em torno de 3,3 segundos. SURF foi o método mais rápido com um nível de confiança de 95% para fazer a extração de características e o casamento das imagens, além de apresentar uma boa distância-erro² (em torno de 2,7 pixels²).

4.5.5 Análise estatística para θ de 64°

A análise gráfica da segunda métrica de avaliação e o custo computacional para o θ de 48° foi apresentada na Figura 4.16a e na Figura 4.16b respectivamente. Os boxplots da distância-erro² indicaram que o método SIFT e SURF não conseguiram fazer casamento de pontos característicos entre as imagens. Portanto a análise estatística da distância-erro² foi realizada para os métodos ASIFT, ASIFT-V e ASURF-V. Os boxplots do tempo indicaram uma inferioridade do método ASIFT. Portanto só os métodos propostos são eficientes para navegação autônoma em VANT. A análise estatística do custo computacional foi realizada pelos métodos ASIFT-V e ASURF-V.

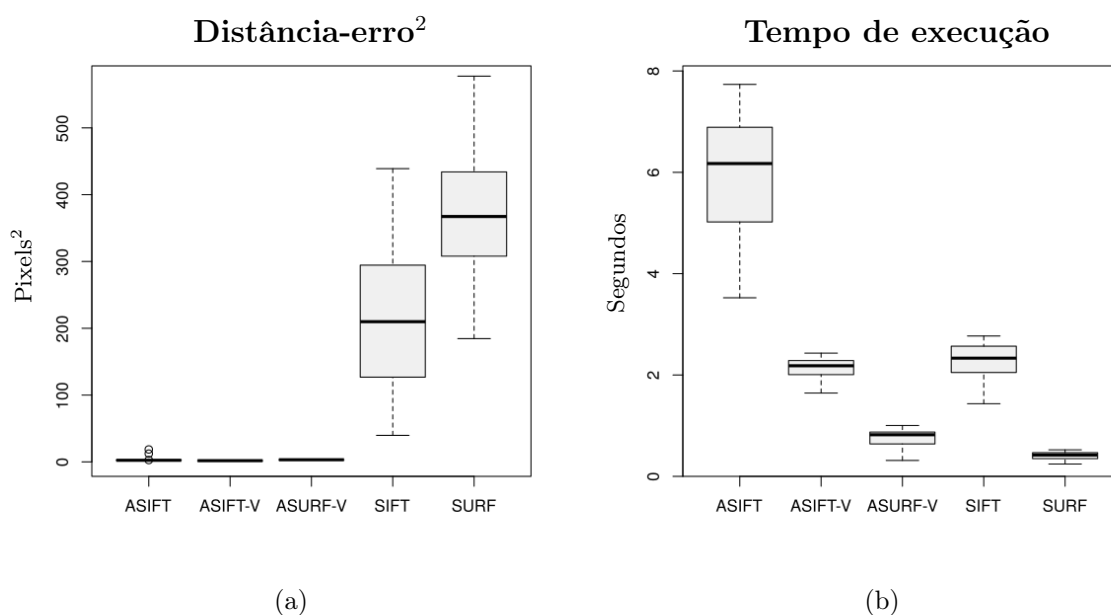


Figura 4.16: Análise gráfica de cada um dos métodos com θ de 64° , os dados são representados por boxplots. A Figura 4.16a apresenta a distância-erro² em pixels² obtida pelo casamento final dos pontos de interesse. A Figura 4.16b apresenta o tempo de execução em segundos que utilizou cada um dos métodos em fazer a extração de pontos, o casamento e a filtragem de pontos.

Aplicando-se o teste de Friedman para avaliar a distância-erro², obteve-se um p-valor igual a $5,107 \times 10^{-15}$, verificando-se que p-valor $\leq 0,05$, pode-se concluir com um nível de confiança de 95% que a hipótese nula foi rejeitada em todos os casos. Portanto, pode-se afirmar que existe uma diferença do desempenho nos métodos propostos. Aplicando-se o Post-hoc de superioridade e inferioridade todos contra todos, obteve-se os resultados de p-value para as comparações, apresentados na Tabela 4.10. Em complemento, a Figura 4.17a apresenta estes resultados.

Tabela 4.10: Análise Post-hoc de superioridade e inferioridade todos contra todos, para avaliar a distância-erro² dos métodos ASIFT, ASIFT-V e ASURF-V com θ de 64°

Post-hoc :	Superioridade	Inferioridade
ASIFT-V - ASIFT	$1,0000 \times 10^0$	$2,4212 \times 10^{-5}$
ASURF-V - ASIFT	$5,3889 \times 10^{-4}$	$1,0000 \times 10^0$
ASURF-V - ASIFT-V	$2,4425 \times 10^{-15}$	$1,0000 \times 10^0$

Aplicando-se o teste Wilcoxon pareado de inferioridade para avaliar o tempo de execução dos métodos ASIFT-V e ASURF-V, obteve-se um p-valor igual a $4,6079 \times 10^{-9}$, verificando-se que p-valor $\leq 0,05$, pode-se concluir com um nível de confiança de 95% que a hipótese nula foi rejeitada. Portanto, pode-se afirmar que o método ASURF-V tem menor custo computacional que o método ASIFT-V.

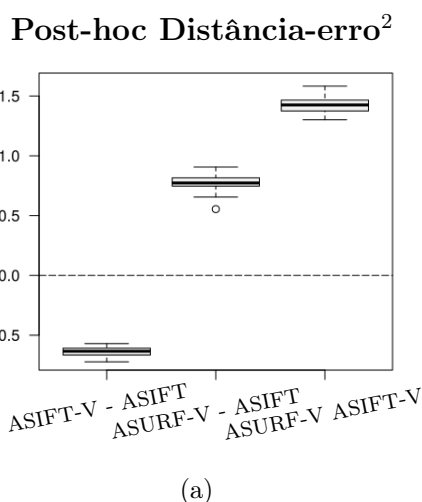


Figura 4.17: A Figura 4.17a mostra a análise Post-hoc para avaliar a distância-erro² dos métodos ASIFT, ASIFT-V e ASURF-V, com θ de 64°.

O testes apresentados para o θ de 64° mostraram que só os métodos ASIFT, ASIFT-V e ASURF-V conseguiram fazer um bom casamento. Porém, uma desvantagem do método ASIFT é o custo computacional, os resultados mostraram que o método não pode ser implementado em navegação autônoma por possuir um tempo de processamento muito alto. Por outro lado, os métodos ASIFT-V e ASURF-V propostos apresentaram um bom custo-benefício em relação à quantidade de pontos obtidos, a distância-erro² do casamento e o tempo de execução. ASIFT-V foi o melhor método com um nível de confiança de 95% no teste da distância-erro² do casamento dos pontos, além de apresentar um bom tempo de execução em torno de 2,2 segundos. Com um nível de confiança de 95%, ASURF-V foi mais rápido que o método ASIFT-V para fazer a extração de características e o casamento das imagens, além de apresentar uma boa distância-erro² (em torno de 3,23 pixels²).

4.5.6 Análise estatística para θ de 76°

A análise gráfica da segunda métrica de avaliação e o custo computacional para o θ de 76° foi apresentada na Figura 4.18a e na Figura 4.18b respectivamente. Os boxplots da distância-erro² indicaram que o método ASIFT, SIFT e SURF não conseguiram fazer casamento de pontos característicos entre as imagens. Portanto a análise estatística da distância-erro² e o tempo de execução foi realizada para os métodos ASIFT-V e ASURF-V.

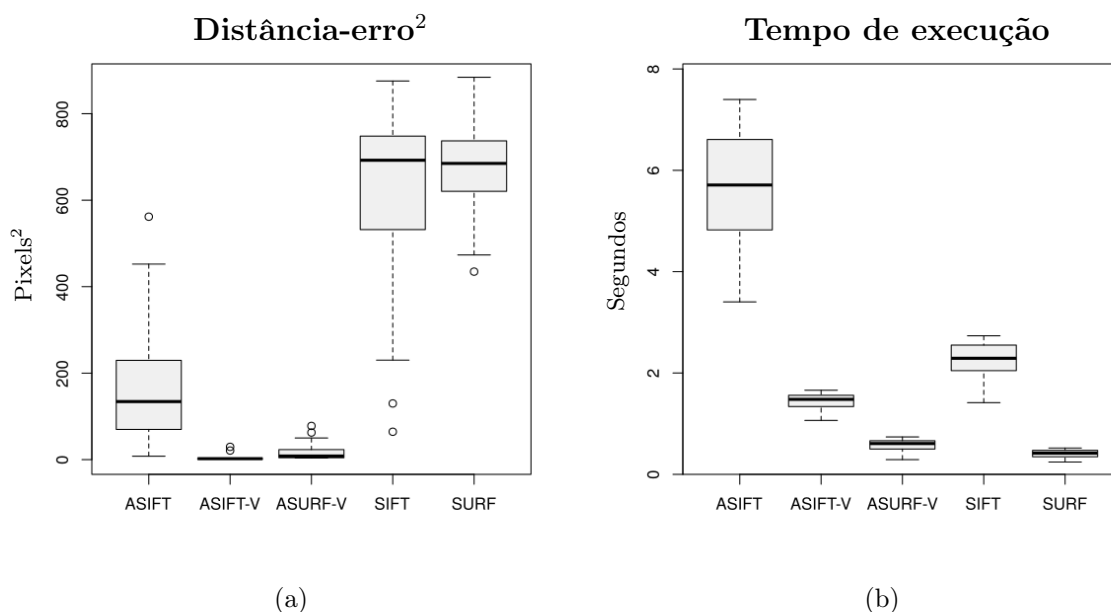


Figura 4.18: Análise gráfica de cada um dos métodos com θ de 76° , os dados são representados por boxplots. A Figura 4.18a apresenta a distância-erro² em pixels² obtida pelo casamento final dos pontos de interesse. A Figura 4.18b apresenta o tempo de execução em segundos que utilizou cada um dos métodos em fazer a extração de pontos, o casamento e a filtragem de pontos.

Aplicando-se o teste Wilcoxon pareado de inferioridade para avaliar a distância-erro² dos métodos ASIFT-V e ASURF-V, obteve-se um p-valor igual a $4,6079 \times 10^{-9}$, verificando-se que $p\text{-valor} \leq 0,05$, pode-se concluir com um nível de confiança de 95% que a hipótese nula foi rejeitada. Portanto, pode-se afirmar que o método ASIFT-V tem uma menor distância-erro² no casamento das imagens que o método ASURF-V. Aplicando-se o teste Wilcoxon pareado de superioridade para avaliar o tempo de execução dos métodos ASIFT-V e ASURF-V, obteve-se um p-valor igual a $4,6079 \times 10^{-9}$, verificando-se que $p\text{-valor} \leq 0,05$, pode-se concluir com um nível de confiança de 95% que a hipótese nula foi rejeitada. Portanto, pode-se afirmar que o método ASURF-V tem um menor custo computacional que o método ASIFT-V.

O testes apresentados para o θ de 76° mostraram só os métodos ASIFT-V e ASURF-V conseguiram fazer um bom casamento. Os métodos propostos ASIFT-V e ASURF-V apresentaram um bom custo-benefício em relação à quantidade de pontos obtidos, a distância-erro² do casamento e o tempo de execução. ASIFT-V foi melhor no teste da distância-erro² do casamento dos pontos, além de apresentar um bom tempo de execução em torno de 1,4 segundos. ASURF-V foi mais rápido para fazer a extração de características e o casamento das imagens, além de apresentar uma boa distância-erro² (em torno de 16 pixels²).

4.6 Resumo do capítulo

Finalmente, para fechar o capítulo dos experimentos e resultados, é apresentada uma análise dos extractores SIFT, SURF, ASIFT, ASIFT-V e ASURF-v na Tabela 4.11. Esta tabela sumariza os resultados obtidos nesta seção, além da análise dos resultados obtidos em Farinella et al. (2013); Tuytelaars and Mikolajczyk (2008); Zheng et al. (2014); Mishkin and Matas (2014); Romero Cortijo et al. (2009). A tabela apresenta uma comparação do tamanho do descritor de interesse, a repetibilidade, acurácia, robustez e eficiência da localização dos pontos casados e finalmente o tempo de execução.

Tabela 4.11: Comparação dos extractores de características SIFT, SURF, ASIFT, ASIFT-V e ASURF-V.

Método	Invariância	Tam. Descritor	Repetibilidade	Acurácia	Robustez	Eficiência	Tempo
SIFT	Escala e Rotação	128	++	+++	++	++	++
SURF	Escala e Rotação	64	+++	++	+++	++	+++
ASIFT	Escala, Rotação e Afinidade	128	++	++	++	+	+
ASIFT-V	Escala, Rotação e Afinidade	128	+++	+++	+++	+++	++
ASURF-V	Escala, Rotação e Afinidade	64	+++	++	++	+++	+++

Capítulo 5

Conclusões e Trabalhos Futuros

Neste trabalho foram apresentados os algoritmos ASIFT-V e ASURF-V, dois novos extratores de características com foco no posicionamento do VANT para navegação autônoma através da visão computacional. Foi realizada uma análise dos extratores propostos e de outros três extratores muito utilizados na literatura (SIFT, SURF e ASIFT) para casamento de imagens quando são capturadas em diferentes ângulos da transformada homográfica. Para avaliar o desempenho dos algoritmos propostos, foi utilizada a quantidade e a porcentagem de pontos casados, a distância erro da projeção do casamento dos pontos (distância-erro²) e o custo computacional medido com o tempo de execução. A análise apresentada neste trabalho verificou que a eficiência dos métodos SIFT e SURF é limitada para casamentos de imagens com ângulos *Pitch* (θ) da transformada homográfica menores que 48°, este resultado foi esperado devido os métodos serem invariantes apenas a rotação e translação. Entretanto, o método ASIFT demonstrou ser invariante ao casamento de imagens para ângulos *Pitch* menores que 76° na transformada homográfica. Porém, uma grande desvantagem do método ASIFT é o custo computacional, como foi evidenciado nos resultados do tempo de processamento. Isto se deve as 41 transformações que o algoritmo tem que fazer em cada imagem para poder conseguir a invariância a afinidade.

Os resultados mostraram a robustez dos métodos ASIFT-V e ASURF-V. Os métodos foram capazes de casar imagens capturadas pelo VANT com imagens georreferenciadas do território em toda a amostragem realizada do ângulo θ , e em todos os cenários, mesmo que estes tenham poucas características. Os métodos propostos apresentaram um bom custo-benefício em relação à quantidade de pontos obtidos, a porcentagem de *outliers*, a distância-erro² do casamento dos pontos e o tempo computacional. Como esperado, em todos os casos, o método ASURF-V foi mais rápido que o método ASIFT-V. Porém o método ASIFT-V apresentou menor distância-erro². A porcentagem de *outliers* no algoritmo de similaridade de ASIFT-V foi em torno de 18%

e para o ASURF-V de 40%. Uma desvantagem destes extratores é que para uma eficiente implementação, tem-se que garantir que as duas imagens sejam capturadas pelos mesmos sensores ou sensores muito próximos, i.e. porque os extratores são baseados em SURF e SIFT os quais apresentam a mesma problemática.

Com base na análise dos resultados obtidos neste trabalho, sugere-se a implementação do método SURF para ângulos *Pitch* menores a 48° , $\theta < 48^\circ$. Nestes ângulos, SURF apresentou a menor quantidade de pontos extraídos e casados entre as duas imagens, mas os resultados na distância-erro², a porcentagem de *outliers* e o tempo computacional demonstraram maior eficiência em comparação com os outros métodos. Seguidamente, sugere-se a implementação do método ASURF-V para ângulos menores ou iguais a 64° e maiores ou iguais a 48° , $48^\circ \leq \theta \leq 64^\circ$. Neste intervalo, o método ASURF-V obteve o melhor custo-benefício entre a quantidade de pontos casados, a porcentagem de *outliers*, a distância-erro² e o tempo computacional. Finalmente, sugere-se o método ASIFT-V para ângulos maiores a 64° , $\theta > 64^\circ$, sendo o ASIFT-V o único método que conseguiu fazer um ótimo casamento em todos os cenários utilizados.

As propostas deste trabalho proporcionam uma paralelização em um ambiente propício para ser embarcado em VANTs. Porém, a implementação utilizando núcleos dedicados poderia diminuir o tempo computacional apresentado. Desta forma, recomenda-se aplicar em técnicas de navegação autônoma com visão computacional em técnicas como a odometria, onde o casamento é feito em imagens capturadas pelo mesmo sensor, ou em técnicas como casamento de marcos ou "*landmarks*", ou casamento de imagens, mas tem-se que garantir que as imagens casadas sejam do mesmo sensor ou sensores muito próximos. Além disto, para uma eficiente implementação dos algoritmos propostos, precisa-se de um bom sincronismo entre a leitura dos dados dos sensores inerciais e a captura da imagem, desta forma, garantindo um bom casamento entre as imagens. Na atualidade, devido a implicações tecnológicas, é recomendado utilizar os extratores propostos em técnicas de navegação implementadas em uma base no solo para que possam ser executados eficientemente (i.e. devido ao custo computacional).

Como propostas para trabalhos futuros são apresentadas algumas sugestões. Em relação ao custo computacional sugere-se pesquisar extratores de características binários (como BREAK, ORB, BRIEF, KAZE e AKAZE), que podem conduzir a melhores resultados quanto ao custo computacional. Outras melhorias poderiam ocorrer com a mudança do descritor de características. Na literatura apresentam-se trabalhos como Zheng et al. (2014) que expõe um novo descritor de características SIFT invariante a rotação e afim para casamento de imagens de VANT com imagens satélite.

Referências Bibliográficas

- (2016). Insidegnss. Acessado em 28/06/2016 para url <http://www.insidegnss.com/compass>.
- (2016). Opencv. Acessado em 28/09/2016 para url <http://opencv.org/>.
- (2016). Openmp. Acessado em 28/09/2016 para url <http://www.openmp.org/>.
- Alcantarilla, P. F., Bartoli, A., and Davison, A. J. (2012). Kaze features. In *European Conference on Computer Vision*, pages 214–227. Springer.
- Alcantarilla, P. F., Nuevo, J., and Bartoli, A. (2013). Fast explicit diffusion for accelerated features in nonlinear scale spaces. In *British Machine Vision Conf. (BMVC)*.
- Anderson, B. D. and Moore, J. B. (1979). Optimal filtering. 1979.
- Andert, F., Ammann, N., Puschel, J., and Dittrich, J. (2014). On the safe navigation problem for unmanned aircraft: Visual odometry and alignment optimizations for uav positioning. In *Unmanned Aircraft Systems (ICUAS), 2014 International Conference on*, pages 734–743. IEEE.
- Bay, H., Tuytelaars, T., and Van Gool, L. (2006). Surf: Speeded up robust features. In *Computer vision–ECCV 2006*, pages 404–417. Springer.
- Benini, A., Rutherford, M., and Valavanis, K. (2016). Real-time, gpu-based pose estimation of a uav for autonomous takeoff and landing. In *2016 IEEE International Conference on Robotics and Automation (ICRA)*, pages 3463–3470. IEEE.
- Braga, J. R., Velho, H. d. C., and Shiguemori, H. (2015). Estimation of uav position using lidar images for autonomous navigation over the ocean. In *2015 9th International Conference on Sensing Technology (ICST)*, pages 811–816. IEEE.
- Brown, M. and Lowe, D. G. (2002). Invariant features from interest point groups. In *BMVC*, number s 1.

- Buyukyazi, T., Bayraktar, S., and Lazoglu, I. (2013). Real-time image stabilization and mosaicking by using ground station cpu in uav surveillance. In *Proceedings of the IEEE 6th International Conference on Recent Advances in Space Technologies (RAST)*, pages 121–126.
- Calonder, M., Lepetit, V., Strecha, C., and Fua, P. (2010). Brief: Binary robust independent elementary features. In *European conference on computer vision*, pages 778–792. Springer.
- Camargo, A., Schultz, R. R., Wang, Y., Fevig, R. A., and He, Q. (2010). Gpu-cpu implementation for super-resolution mosaicking of unmanned aircraft system (uas) surveillance video. In *Image Analysis & Interpretation (SSIAI), 2010 IEEE Southwest Symposium on*, pages 25–28. IEEE.
- Canhoto, A., Shiguemori, E. H., and Domiciano, M. A. P. (2009). Image sequence processing applied to autonomous aerial navigation. In *Signal and Image Processing Applications (ICSIPA), 2009 IEEE International Conference on*, pages 496–499. IEEE.
- Castro, A. P. A., da Silva, J. D. S., Shiguemori, E. H., Martins, M. P., and Goltz, G. (2009). Restauração de imagens e detecção automática de características aplicados à navegação aérea autônoma. *Simpósio Brasileiro de Sensoriamento Remoto (SBSR)*, 14:6813–6819.
- Cesetti, A., Frontoni, E., Mancini, A., Zingaretti, P., and Longhi, S. (2010). A vision-based guidance system for uav navigation and safe landing using natural landmarks. In *Selected papers from the 2nd International Symposium on UAVs, Reno, Nevada, USA June 8–10, 2009*, pages 233–257. Springer.
- Chaves, Á. N. (2012). *Proposta de modelo de veículos aéreos não tripulados (VANTs) cooperativos aplicados a operações de busca*. PhD thesis, Universidade de São Paulo.
- Chen, X. and Meng, Q. (2013). Vehicle detection from uavs by using sift with implicit shape model. In *Systems, Man, and Cybernetics (SMC), 2013 IEEE International Conference on*, pages 3139–3144. IEEE.
- Chunhui, Z., Rongzhi, W., Tianwu, Z., and Quan, P. (2014). Visual odometry and scene matching integrated navigation system in uav. In *Information Fusion (FUSION), 2014 17th International Conference on*, pages 1–6. IEEE.
- Conte, G. and Doherty, P. (2008). An integrated uav navigation system based on aerial image matching. In *Aerospace Conference, 2008 IEEE*, pages 1–10. IEEE.

- Del Moral, P. (1996). Non-linear filtering: interacting particle resolution. *Markov processes and related fields*, 2(4):555–581.
- dos Santos, A. L., Torres, L. C. B., de Campos Velho, H. F., Shiguemori, E. H., and de Pádua Braga, A. (2013). Controle e navegação de vants através de utilização de template matching. *Anais do Computer on the Beach*, pages 70–78.
- Everitt, B. S., Landau, S., Leese, M., and Stahl, D. (2011). Hierarchical clustering. *Cluster Analysis, 5th Edition*, pages 71–110.
- Farinella, G. M., Battiato, S., and Cipolla, R. (2013). *Advanced Topics in Computer Vision*. Springer.
- Fernández-Hernandez, J., González-Aguilera, D., Rodríguez-Gonzálvez, P., and Mancera-Taboada, J. (2015). Image-based modelling from unmanned aerial vehicle (uav) photogrammetry: An effective, low-cost tool for archaeological applications. *Archaeometry*, 57(1):128–145.
- Fischler, M. A. and Bolles, R. C. (1981). Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. *Communications of the ACM*, 24(6):381–395.
- Friedman, M. (1940). A comparison of alternative tests of significance for the problem of m rankings. *The Annals of Mathematical Statistics*, 11(1):86–92.
- Goltz, G. A. M. and Shiguemori, E. H. (2008). Aplicação do algoritmo sift em imagens de navegação autônoma. In *Workshop Anual de Pesquisa e Desenvolvimento do IEAv*, pages 35–35.
- Gonçalves, M., Souza, C., and Shiguemori, E. (2015). Uso de coordenadas geográficas para estimação automática de rumo aplicada a navegação aérea autônoma de veículos aéreos não tripulados. *Simpósio Brasileiro de Sensoriamento Remoto (SBSR)*, 17:5598–5605.
- González-Jorge, H., Puente, I., Roca, D., Martínez-Sánchez, J., Conde, B., and Arias, P. (2014). Uav photogrammetry application to the monitoring of rubble mound breakwaters. *Journal of Performance of Constructed Facilities*, 30(1):04014194.
- Guan, X. and Bai, H. (2012). A gpu accelerated real-time self-contained visual navigation system for uavs. In *Information and Automation (ICIA), 2012 International Conference on*, pages 578–581. IEEE.

- Harris, C. and Stephens, M. (1988). A combined corner and edge detector. In *Alvey vision conference*, volume 15, page 50. Citeseer.
- Hartley, R. and Zisserman, A. (2003). *Multiple view geometry in computer vision*. Cambridge university press.
- Jorge, L., Inamasu, R., and Carmo, R. (2011). Desenvolvimento de um vant totalmente configurado para aplicações em agricultura de precisão no brasil. *Anais XV Simpósio Brasileiro de Sensoriamento Remoto-SBSR*, (1979):399–406.
- Koenderink, J. J. (1984). The structure of images. *Biological cybernetics*, 50(5):363–370.
- Kwon, H., Sharma, R., Yoder, J., and Pack, D. (2012). Robust mobile ground target localization using ground image features with uav position compensation techniques. In *Control, Automation and Systems (ICCAS), 2012 12th International Conference on*, pages 454–458. IEEE.
- Li, X. and Aouf, N. (2012). Sift and surf feature analysis in visible and infrared imaging for uavs. In *Cybernetic Intelligent Systems (CIS), 2012 IEEE 11th International Conference on*, pages 46–51. IEEE.
- Li, X., Aouf, N., and Nemra, A. (2012). 3d mapping based vslam for uavs. In *Control & Automation (MED), 2012 20th Mediterranean Conference on*, pages 348–352. IEEE.
- Lin, Y. and Medioni, G. (2007). Map-enhanced uav image sequence registration and synchronization of multiple image sequences. In *Computer Vision and Pattern Recognition, 2007. CVPR'07. IEEE Conference on*, pages 1–7. IEEE.
- Lindeberg, T. (1994). Scale-space theory: A basic tool for analyzing structures at different scales. *Journal of applied statistics*, 21(1-2):225–270.
- Lingua, A. M., Marenchino, D., and Nex, F. C. (2009). A comparison between "old and new" feature extraction and matching techniques in photogrammetry. *REVISTA DE CADASTRU*, 9:43–52.
- Lowe, D. G. (1999). Object recognition from local scale-invariant features. In *Computer vision, 1999. The proceedings of the seventh IEEE international conference on*, volume 2, pages 1150–1157. Ieee.
- Lowe, D. G. (2004). Distinctive image features from scale-invariant keypoints. *International journal of computer vision*, 60(2):91–110.

- Lyu, Y., Pan, Q., Zhang, Y., Zhao, C., Zhu, H., Tang, T., and Liu, L. (2015). Simultaneously multi-uav mapping and control with visual servoing. In *Unmanned Aircraft Systems (ICUAS), 2015 International Conference on*, pages 125–131. IEEE.
- Martinez, J., Hernández, J. T., and Rodriguez, C. F. (2013). A vision based algorithm for the guidance of a glider. In *Graphics, Patterns and Images (SIBGRAPI), 2013 26th SIBGRAPI-Conference on*, pages 203–209. IEEE.
- Mikolajczyk, K., Tuytelaars, T., Schmid, C., Zisserman, A., Matas, J., Schaffalitzky, F., Kadir, T., and Van Gool, L. (2005). A comparison of affine region detectors. *International journal of computer vision*, 65(1-2):43–72.
- Mishkin, D. and Matas, J. (2014). Matching of images of non-planar objects with view synthesis. In *International Conference on Current Trends in Theory and Practice of Informatics*, pages 30–39. Springer.
- Moranduzzo, T. and Melgani, F. (2012). A sift-svm method for detecting cars in uav images. In *Geoscience and Remote Sensing Symposium (IGARSS), 2012 IEEE International*, pages 6868–6871. IEEE.
- Morel, J.-M. and Yu, G. (2009). Asift: A new framework for fully affine invariant image comparison. *SIAM Journal on Imaging Sciences*, 2(2):438–469.
- Pinage, F., Carvalho, J. R. H., and Pinheiro de Queiroz Neto, J. (2012). Natural landmark tracking method to support uav navigation over rain forest areas. In *Computing System Engineering (SBESC), 2012 Brazilian Symposium on*, pages 105–110. IEEE.
- Radecki, K., Samczyński, P., Kulpa, K., and Drozdowicz, J. (2015). A real-time unfocused sar processor based on a portable cuda gpu. In *Radar Conference (EuRAD), 2015 European*, pages 173–176. IEEE.
- Roger-Verdeguer, J. F., Mannberg, M., and Savvaris, A. (2012). Visual odometry with failure detection for the aegis uav. In *Imaging Systems and Techniques (IST), 2012 IEEE International Conference on*, pages 291–296. IEEE.
- Romero Cortijo, A. M., Cazorla Quevedo, M. Á., et al. (2009). Comparativa de detectores de características visuales y su aplicación al slam.
- Rosten, E. and Drummond, T. (2006). Machine learning for high-speed corner detection. In *European conference on computer vision*, pages 430–443. Springer.

- Rublee, E., Rabaud, V., Konolige, K., and Bradski, G. (2011). Orb: an efficient alternative to sift or surf. In *Computer Vision (ICCV), 2011 IEEE International Conference on*, pages 2564–2571. IEEE.
- Sanfourche, M., Delaune, J., Le, G., De, H., Israel, J., Cornic, P., Treil, A., Watanabe, Y., and Plyer, A. (2012). Perception for uav: Vision-based navigation and environment modeling. *AerospaceLab*, (4):p-1.
- Silva Filho, P., Rodrigues, M., Saotome, O., and Shiguemori, E. H. (2014). Fuzzy-based automatic landmark recognition in aerial images using orb for aerial auto-localization. In *International Symposium on Visual Computing*, pages 467–476. Springer.
- Sineglazov, V. (2014). Landmarks navigation system software. In *Methods and Systems of Navigation and Motion Control (MSNMC), 2014 IEEE 3rd International Conference on*, pages 62–65. IEEE.
- Sun, W., Li, D., Jia, W., Li, P., Zhao, C., and Chen, X. (2015). Small moving object tracking in dynamic video. In *2015 International Conference on Intelligent Information Hiding and Multimedia Signal Processing (IIH-MSP)*, pages 239–242. IEEE.
- Suzuki, T., Amano, Y., and Hashizume, T. (2010). Vision based localization of a small uav for generating a large mosaic image. In *SICE Annual Conference 2010, Proceedings of*, pages 2960–2964. IEEE.
- Tuytelaars, T. and Mikolajczyk, K. (2007). Local invariant feature detectors: a survey. *Foundations and trends® in computer graphics and vision*, 3(3):177–280.
- Tuytelaars, T. and Mikolajczyk, K. (2008). Local invariant feature detectors: a survey. *Foundations and trends® in computer graphics and vision*, 3(3):177–280.
- Wang, T., Zhang, Y., Wang, C., Liang, J., Gao, H., Liu, M., Guan, Q., and Sun, A. (2014). Indoor visual navigation system based on paired-landmark for small uavs. In *Robotics and Biomimetics (ROBIO), 2014 IEEE International Conference on*, pages 1703–1708. IEEE.
- Wilcoxon, F. (1945). Individual comparisons by ranking methods. *Biometrics bulletin*, 1(6):80–83.
- Witkin, A. P. (1984). Scale-space filtering: A new approach to multi-scale description. In *Acoustics, Speech, and Signal Processing, IEEE International Conference on ICASSP'84.*, volume 9, pages 150–153. IEEE.

- Ye, X., Li, P., and Zhang, J. (2015). Fully affine invariant matching algorithm based on nonlinear scale space for side scan sonar image. In *2015 IEEE International Conference on Mechatronics and Automation (ICMA)*, pages 2387–2391. IEEE.
- Zheng, M., Wu, C., Chen, D., and Meng, Z. (2014). Rotation and affine-invariant sift descriptor for matching uav images with satellite images. In *Guidance, Navigation and Control Conference (CGNCC), 2014 IEEE Chinese*, pages 2624–2628. IEEE.