

Dissertação apresentada à Pró-Reitoria de Pós-Graduação e Pesquisa do Instituto Tecnológico de Aeronáutica, como parte dos requisitos para obtenção do título de Mestre em Ciências no Curso de Engenharia Eletrônica e Computação, Área de Informática.

**Bruno Duarte Corrêa**

**AVALIAÇÃO DE TÉCNICAS DE  
RECONHECIMENTO DE PADRÕES EM  
AMBIENTES AERONAUTICOS**

Dissertação aprovada em sua versão final pelos abaixo assinados:

Prof. Dr. Luís Gonzaga Trabasso  
Orientador

Nelson José Issa de Macedo  
Coorientador

Prof. Dr. Luiz Carlos Sandoval Góes  
Pró-Reitor de Pós-Graduação e Pesquisa

Campo Montenegro  
São José dos Campos, SP - Brasil  
2015

**Dados Internacionais de Catalogação-na-Publicação (CIP)**  
**Divisão Biblioteca Central do ITA/CTA**

Duarte Corrêa, Bruno

Avaliação de técnicas de reconhecimento de padrões em ambientes aeronauticos / Bruno Duarte Corrêa.

São José dos Campos, 2015.

65f.

Dissertação de Mestrado – Curso de Engenharia Eletrônica e Computação. Área de Informática – Instituto Tecnológico de Aeronáutica, 2015. Orientador: Prof. Dr. Luís Gonzaga Trabasso. Coorientador: Nelson José Issa de Macedo.

1. Augmented reality. 2. Opencv. 3. Computer vision. 4. Maintenance. 5. Feature matching. 6. Descriptors. 7. Markerless. I. Centro Técnico Aeroespacial. Instituto Tecnológico de Aeronáutica. Divisão de Ciência da Computação. II. Título.

## **REFERÊNCIA BIBLIOGRÁFICA**

DUARTE CORRÊA, Bruno. **Avaliação de técnicas de reconhecimento de padrões em ambientes aeronauticos**. 2015. 65f. Dissertação de Mestrado – Instituto Tecnológico de Aeronáutica, São José dos Campos.

## **CESSÃO DE DIREITOS**

NOME DO AUTOR: Bruno Duarte Corrêa

TÍTULO DO TRABALHO: Avaliação de técnicas de reconhecimento de padrões em ambientes aeronauticos.

TIPO DO TRABALHO/ANO: Dissertação / 2015

É concedida ao Instituto Tecnológico de Aeronáutica permissão para reproduzir cópias desta Dissertação e para emprestar ou vender cópias somente para propósitos acadêmicos e científicos. O autor reserva outros direitos de publicação e nenhuma parte desta Dissertação pode ser reproduzida sem a autorização do autor.

---

Bruno Duarte Corrêa

Av. Juscelino Kubistcheck, 6701 ap 14 bl 21

CEP 12.220-000 – São José dos Campos–SP

# **AVALIAÇÃO DE TÉCNICAS DE RECONHECIMENTO DE PADRÕES EM AMBIENTES AERONAUTICOS**

**Bruno Duarte Corrêa**

Composição da Banca Examinadora:

Prof. Dr.	Luís Gonzaga Trabasso	Presidente	-	ITA
Prof. Dr.	Luís Gonzaga Trabasso	Orientador	-	ITA
	Nelson José Issa de Macedo	Coorientador	-	EMBRAER
Prof. Dr.	Ricardo Bedin França	Membro Externo	-	EMBRAER
Prof. Dr.	Emília	Membro	-	ITA

**ITA**

Dedico esse trabalho primeiramente a Deus e a todas as pessoas que me apoiaram, acreditaram e incentivaram, sem os quais com toda certeza não teria obtido os resultados que esperava.

*"Persistence is the shortest path to success."*

— CHARLES CHAPLIN

# Resumo

O reconhecimento de objetos em uma cena para posterior uso em realidade aumentada depende de diversas variáveis, causando a necessidade do uso de diversas técnicas específicas para cada cenário, sendo portanto, um estudo de fronteiras para a melhor escolha do algoritmo de registro de acordo com a aplicação em questão de grande valia para o meio acadêmico. Esta tese se propõe a pesquisar, categorizar e traçar fronteiras das técnicas conhecidas tendo como caso de uso manutenção de aeronaves feita dentro de centros fechados, utilizando as técnicas BRISK,FAST,FREAK,GFTT,MSER, ORB,STAR,SURF,SIFT em uma análise aplicada com imagens reais de janelas de inspeção do Embraer ERJ-190 para reconhecimento de objetos em posteriores aplicações de manutenção. Comparando todas as técnicas quanto à cadência e a precisão de reconhecimento de características é possível selecionar GFTT e ORB como técnicas mais apropriadas ao contexto, por terem seus resultados de variação de rotação, escala, briho e blur dentro de uma faixa esperada.

# Abstract

Recognition of objects in a scene for later use in augmented reality depends on several variables, causing the need to use several specific techniques for each scenario, and therefore a border analysis to the best choice of the registration algorithm according to application in question, of great value to academia. This thesis aims to examine, categorize and draw boundaries of the known techniques having as use case maintenance of aircraft made in closed centers.

# Sumário

LISTA DE FIGURAS . . . . .	xi
LISTA DE TABELAS . . . . .	xiii
LISTA DE ABREVIATURAS E SIGLAS . . . . .	xiv
<b>1 INTRODUÇÃO . . . . .</b>	<b>15</b>
1.1 Motivação . . . . .	15
1.2 Objetivos . . . . .	16
1.3 Justificativa . . . . .	17
1.4 Contextualização . . . . .	18
1.4.1 Variáveis de contorno . . . . .	18
1.4.2 Cenário . . . . .	18
1.5 Recursos e Métodos . . . . .	19
1.6 Estrutura . . . . .	22
<b>2 FUNDAMENTAÇÃO TEÓRICA . . . . .</b>	<b>23</b>
2.1 Realidade Virtual . . . . .	23
2.2 Realidade Aumentada . . . . .	24
2.3 Dispositivos . . . . .	25
2.3.1 <i>Head-Mounted Displays</i> . . . . .	26



---

2.3.2	Projetores . . . . .	27
2.3.3	Monitores . . . . .	27
<b>2.4</b>	<b>Camera . . . . .</b>	<b>28</b>
2.4.1	Distorção Radial . . . . .	28
2.4.2	Distorção Tangencial . . . . .	29
<b>2.5</b>	<b>Modelo de Câmera . . . . .</b>	<b>30</b>
<b>2.6</b>	<b>Reconhecimento . . . . .</b>	<b>31</b>
<b>2.7</b>	<b>Características Locais . . . . .</b>	<b>32</b>
2.7.1	Propriedades da Característica Local ideal . . . . .	32
2.7.2	Características . . . . .	33
2.7.3	Deteção de características . . . . .	34
2.7.4	Correspondência de características . . . . .	35
<b>2.8</b>	<b>Algoritmos de Reconhecimento . . . . .</b>	<b>35</b>
2.8.1	FAST - Features from Accelerated Segment Test . . . . .	36
2.8.2	BRISK - Binary Robust Invariant Scalable Keypoints . . . . .	36
2.8.3	FREAK - Fast Retina Keypoint . . . . .	38
2.8.4	GFTT - Good Features To Track . . . . .	40
2.8.5	MSER - Maximal Stable Extremal Regions . . . . .	40
2.8.6	ORB - Oriented Fast and Rotated Brief . . . . .	41
2.8.7	SIFT - Scale-Invariant Feature Transform . . . . .	41
2.8.8	SURF - Speeded Up Robust Feature . . . . .	44
2.8.9	STAR . . . . .	45
<b>2.9</b>	<b>Cadência . . . . .</b>	<b>45</b>
<b>3</b>	<b>APLICAÇÃO E SELEÇÃO DE ALGORITMOS . . . . .</b>	<b>47</b>
<b>3.1</b>	<b>Metodologia . . . . .</b>	<b>47</b>

---

3.1.1	Definição de parâmetros . . . . .	47
<b>3.2</b>	<b>Análise de dados . . . . .</b>	<b>47</b>
3.2.1	Protótipo . . . . .	48
<b>4</b>	<b>ANÁLISE DE RESULTADOS . . . . .</b>	<b>52</b>
4.1	Taxa de Acertos . . . . .	52
4.2	Análise de Tempo . . . . .	56
4.2.1	Seleção da Técnica . . . . .	60
<b>5</b>	<b>CONCLUSÃO . . . . .</b>	<b>62</b>
5.1	Atendimento dos objetivos . . . . .	62
5.2	Proposta de Trabalhos Futuros . . . . .	63
	REFERÊNCIAS BIBLIOGRÁFICAS . . . . .	64
	GLOSSÁRIO . . . . .	66

# Lista de Figuras

FIGURA 1.1 – Posicionamento da LRU . . . . .	19
FIGURA 1.2 – Categorias de pesquisas científicas . . . . .	20
FIGURA 2.1 – Aplicação de realidade virtual . . . . .	24
FIGURA 2.2 – Pipeline Canônico de Realidade Aumentada . . . . .	25
FIGURA 2.3 – Arquitetura do <i>Feed-Trough</i> . . . . .	26
FIGURA 2.4 – Arquitetura do <i>See-Trough</i> . . . . .	27
FIGURA 2.5 – Realidade Aumentada com projetores. . . . .	28
FIGURA 2.6 – Distorção radial. . . . .	29
FIGURA 2.7 – Componentes de Distorção radial. . . . .	29
FIGURA 2.8 – Distorção tangencial. . . . .	30
FIGURA 2.9 – Componentes de Distorção tangencial. . . . .	30
FIGURA 2.10 – Camera no espaço vetorial. . . . .	31
FIGURA 2.11 – Ilustração do procedimento de reconhecimento com features locais. . . . .	31
FIGURA 2.12 – Reconhecimento de padrões.(UNDERSTANDING..., 2014) . . . . .	34
FIGURA 2.13 – Regiões de reconhecimento de padrões.(UNDERSTANDING..., 2014) . . . . .	34
FIGURA 2.14 – FAST . . . . .	36
FIGURA 2.15 – Padrão de amostras no BRISK . . . . .	37
FIGURA 2.16 – Padrão de amostragem do descritor FREAK . . . . .	38
FIGURA 2.17 – Distribuição de receptores na retina . . . . .	39

---

FIGURA 2.18 – Pares selecionados para calcular a orientação . . . . .	39
FIGURA 2.19 – Regiões MSER . . . . .	41
FIGURA 2.20 – SIFT . . . . .	42
FIGURA 2.21 – Difference of Gaussian . . . . .	42
FIGURA 2.22 – Formação do descritor SIFT . . . . .	44
FIGURA 2.23 – Box Filtering . . . . .	45
FIGURA 3.1 – Imagem utilizada no protótipo . . . . .	49
FIGURA 4.1 – Resultado de performance de variação de escala . . . . .	53
FIGURA 4.2 – Resultado de performance de variação de rotação . . . . .	54
FIGURA 4.3 – Resultado de performance de variação de blur . . . . .	55
FIGURA 4.4 – Resultado de performance de variação de blur . . . . .	56
FIGURA 4.5 – Análise de tempo de variação de escala . . . . .	57
FIGURA 4.6 – Análise de tempo de variação de rotação . . . . .	58
FIGURA 4.7 – Análise de tempo de variação de blur . . . . .	59
FIGURA 4.8 – Análise de tempo de variação de iluminação . . . . .	60
FIGURA 4.9 – Spiderweb Diagram (8 Dimensions, 2-Notch Scale, 3 Samples) . . .	61

# Lista de Tabelas

TABELA 1 – Lista de Acrônimos . . . . .	xiv
TABELA 1.1 – Classificação da pesquisa . . . . .	20
TABELA 1.2 – Resumo de Recursos e Métodos . . . . .	22
TABELA 4.1 – Decisão de técnica à utilizar . . . . .	60

# Lista de Abreviaturas e Siglas

TABELA 1 – Lista de Acrônimos

HMD	Head-Mounted Display
AR	Augmented Reality
VR	Virtual Reality
DoG	Difference of Gaucians
LoG	Laplacian of Gaucians
BRIEF	Binary Robust Independent Elementary Features
BRISK	Binary Robust Invariant Scalable Keypoints
FAST	Features from Accelerated Segment Test
FREAK	Fast Retina Keypoint
GFTT	Good Features To Track
MSER	Maximally Stable Extremal Regions
ORB	Oriented Fast and Rotated Brief
SIFT	Scale-Invariant Feature Transform
SURF	Speed Up Robust Feature
STAR	Is not an Acronim
GPU	Graphics Processing Unit

# 1 Introdução

## 1.1 Motivação

O reconhecimento de estruturas e sistemas de forma automática no campo da manutenção auxilia além de garantir maior confiabilidade no diagnóstico de problemas pode propiciar uma ferramenta de capacitação entre várias possibilidades. Um dos mais básicos problemas atualmente limitando o ramo da Realidade Aumentada é a etapa de registro. A Realidade Aumentada prevê imersão entre o mundo virtual e o mundo real e por isso para que a experiência de imersão seja coerente é necessário que os dois mundos estejam bem sincronizados e propriamente alinhados. Para algumas aplicações tal sincronia aumenta a experiência, entretanto existem aplicações que tal alinhamento é primordial, por exemplo em aplicações médicas por exemplo em uma aplicação de biopsia. Se o objeto não estiver no espaço e tempo da realidade, a informação fornecida ao cirurgião poderá por em risco a vida do paciente. Na maioria das aplicações de tempo real, problemas de registro podem invalidar o uso da Realidade Aumentada. Um outro problema que pode ocorrer com falhas de registro é acentuado por um fenômeno conhecido como visual capture ([WELCH, 1978](#)) que é a tendência do cérebro acreditar mais no que vê do que no que sente, ouve, etc. Nesses casos o sentido da visão tende a sobrepor os outros sentidos. Assim como um ventríloco consegue enganar quem assiste um show acreditando que o som sai da boca do boneco o usuário de uma aplicação de realidade aumentada tenderá a acreditar no que vê, mesmo que esteja defasado no espaço/tempo. No caso do erro se tornar sistemático o usuário tende a se acostumar inconscientemente e se adaptar ao erro, corrigindo o efeito. Erros de registro são difíceis de controlar adequadamente devido à grande precisão requerida das diversas fontes de erro. As fontes de erro podem ser divididas em estáticas e dinâmicas sendo as estáticas contornadas com calibração prévia de sensores entretanto

os erros dinâmicos são mais difíceis porque são susceptíveis a tempo diferença de tempo entre o real e o apresentado na tela e com o acúmulo de erro. O reconhecimento de objetos na cena permeia também:

- O contexto da cena, sendo que com conhecimento prévio do cenário se torna bem mais fácil;
- O material do qual o objeto é feito, porque caso seja feito de materiais reflexivos, os algoritmos podem confundir o reflexo de outros objetos com informações a reconhecer;
- O tamanho do objeto, pois de acordo com a escala do objeto, muitas informações que poderiam ser boas para o reconhecimento podem estar próximas demais dificultando o posterior casamento de informações. . . .

Portanto para que as diversas fontes de erros dinâmicos não sejam um impeditivo para o reconhecimento, de acordo com a cena, algoritmos diferentes devem ser selecionados por terem peculiaridades e características que garantam um registro direcionados ao tipo de desafio que encontrarão, além de já ter informações prévias, o que facilita na seleção de características.

## 1.2 Objetivos

O presente trabalho tem como objetivo geral avaliar o ambiente de manutenção aeronáutico, no contexto de janelas de inspeção, traçando estratégias de reconhecimento de itens de manutenção. Para a consecução do objetivo geral, foram definidos os seguintes objetivos específicos:

- Avaliar as algoritmos clássicos de reconhecimento;
- Aplicar os algoritmos clássicos à situações reais.
- Selecionar algoritmo mais adequado para o contexto.

A proposta dessa tese é a partir do cenário de manutenção de aeronaves e da proposta de utilização de realidade aumentada, determinar a melhor estratégia de reconhecimento



de peças para que posteriormente seja utilizado em ferramentas de auxílio na manutenção por meio da realidade aumentada.

### 1.3 Justificativa

Muitas são as abordagens de comparação entre detectores de características ultimamente. Uma análise de desempenho entre descritores locais feita em ([MIKOLAJCZYK, 2005](#)) aponta a metodologia, bem como a métrica utilizada nesse trabalho. Uma análise comparativa feita com três descritores binários (ORB, BRIEF e BRISK) ([E.; J.-M., 2012](#)) utilizando os mais conhecidos detectores (ORB, MSER, SIFT, SURF, FAST e BRISK) é realizada comparando-se o efeito de transformações geométricas e fotométricas.

Também um estudo comparativo de desempenho demonstrado em ([M.; H., 2013](#)) apresenta resultados para FAST-SIFT comparando variações de *blur*, iluminação, escala e rotação.

A análise do desempenho de vários descritores foi feita no contexto de navegação robótica em ([SCHMIDT, 2012](#)) decidindo pelo melhor detector e descritor para navegação de robôs.

O desempenho dos detectores (FREAK vs. SURF vs. BRISK) é examinado no contexto de detecção de pedestres em ([SCHAEFFER, 2013](#)). Estudos sobre a descritores locais também são feitos em ([TUYTELAARS, 2007](#)), descrevendo e comparando os principais descritores quanto à suas peculiaridades.

Uma análise de sete combinações de conhecidos detectores e descritores, a saber, SIFT com SIFT, SURF com SURF, MSER com SIFT, BRISK com FREAK, BRISK com BRISK, ORB com ORB e FAST com BRIEF é apresentado em ([OZKAN, 2014](#))

O presente trabalho apresenta a comparação de nove combinações de detectores e descritores, a saber, BRISK com BRISK, STAR com BRIEF, MSER com SIFT, ORB com ORB, FAST com BRIEF, FREAK com FREAK, GFTT com BRIEF, SURF com SURF e SIFT com SIFT e propõe um estudo de caso do ambiente de aeronáutico com um método pré estabelecido de seleção de limites de restrições, com janelas de decisão pré de tal forma a eliminar as combinações que não estejam dentro da região desejada.

## 1.4 Contextualização

O contexto dessa tese prevê o cenário de manutenção com o uso de realidade aumentada como uma ferramenta para auxiliar nas tarefas rotineiras portanto algumas variáveis devem ser consideradas para garantir a viabilidade de implantação da abordagem:

- Velocidade de reconhecimento
- Qualidade do reconhecimento
- Invariância quanto à parâmetros ambientais

### 1.4.1 Variáveis de contorno

O cenário de reconhecimento de objetos dentro da aeronave traz alguns desafios que devem ser contornados

- Pouca iluminação em ambientes internos;
- Objetos muito parecidos entre si;
- Alguns objetos com textura;
- Objetos brilhantes.

### 1.4.2 Cenário

O uso da realidade aumentada em manutenção de aeronaves pode trazer ganho no que tange fornecer informações de procedimentos ao mecânico ou mesmo previsão de falhas ou reconhecimento de regiões com falha.

Como caso de uso será adotado a janela de inspeção frontal como mostrado na imagem [1.1](#) que mostra onde fica localizado na aeronave Embraer 190.



FIGURA 1.1 – Posicionamento da LRU

## 1.5 Recursos e Métodos

A metodologia é a aplicação de procedimentos e técnicas que devem ser observadas para construção do conhecimento de comprovar sua validade e utilidade nos diversos âmbitos da sociedade. E em um nível aplicado, examina, descreve e avalia métodos e técnicas de pesquisa que possibilitam a coleta e o processamento de informações, visando ao encaminhamento e à resolução de temas de investigação (PRODANOV CLEBER CRISTIANO; FREITAS, 2013). Pesquisa científica é a realização de um estudo planejado, sendo o método de abordagem do problema o que caracteriza o aspecto científico da investigação. Sua finalidade é descobrir respostas para questões mediante a aplicação do método científico. A pesquisa sempre parte de um problema, de uma interrogação, uma situação para a qual o repertório de conhecimento disponível não gera resposta adequada. Para solucionar esse problema, são levantadas hipóteses que podem ser confirmadas ou refutadas pela pesquisa. Portanto, toda pesquisa se baseia em uma teoria que serve como ponto de partida para a investigação. Utilizando-se as formas clássicas de classificação das pesquisas descritas por (PRODANOV CLEBER CRISTIANO; FREITAS, 2013) há a compilação dos tipos de pesquisa como mostra na figura 1.2.

A tabela 1.1 sumariza como o presente trabalho se categoriza segundo o método clássico de classificação dos critérios

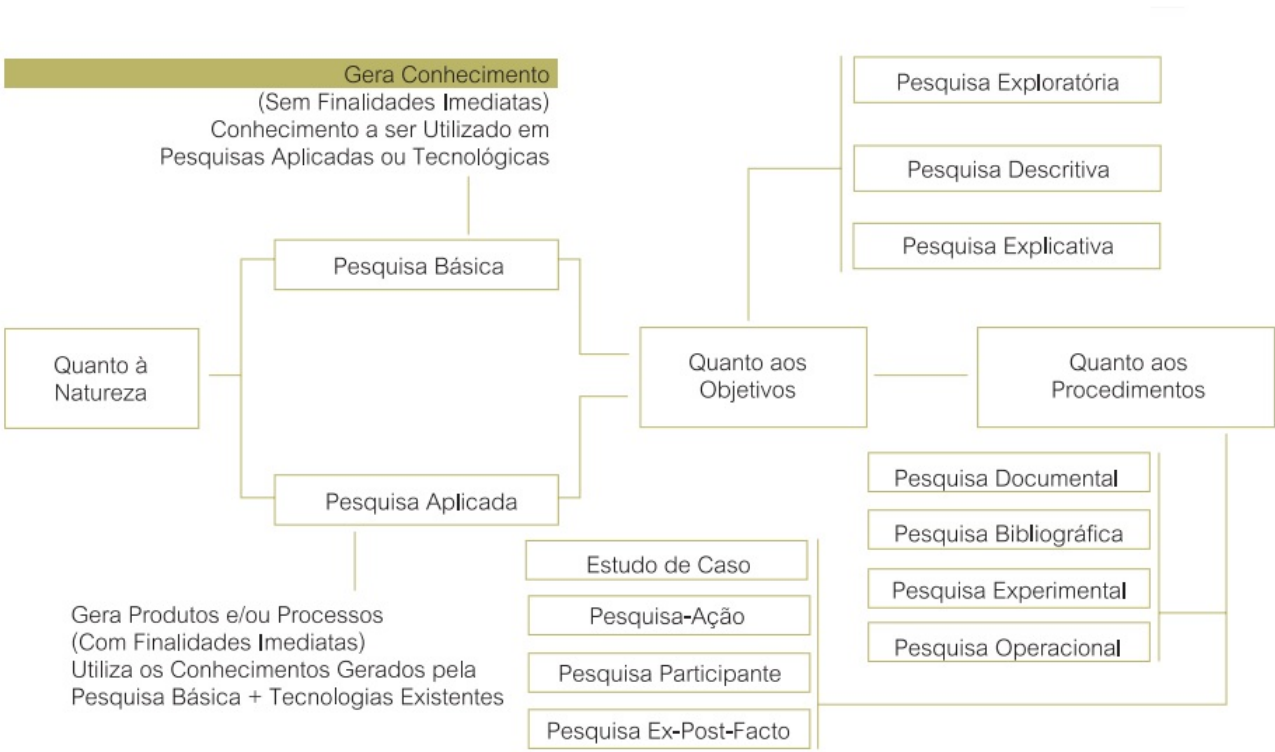


FIGURA 1.2 – Categorias de pesquisas científicas

TABELA 1.1 – Classificação da pesquisa

Ponto de Vista	Tipo de pesquisa utilizada
Natureza	Pesquisa Aplicada
Abordagem do Problema	Pesquisa Quantitativa
Objetivos	Pesquisa Explicativa
Procedimentos Técnicos	Pesquisa Bibliográfica, Pesquisa Experimental

1. Natureza

A pesquisa aplicada objetiva gerar conhecimentos para aplicação prática dirigidos à solução de problemas específicos (PRODANOV CLEBER CRISTIANO; FREITAS, 2013). Assim a pesquisa sobre reconhecimentos de padrão aplicados à realidade aeronáutica tem como objetivo selecionar as técnicas mais adequadas para situações reais em futuras aplicações de RA para manutenção, portanto se categoriza como uma aplicação prática a um problema específico.

2. Abordagem do Problema

A pesquisa quantitativa objetiva garantir a precisão dos resultados, evitando con-

tradições no processo de análise e interpretação, para tanto é feita uma hipótese prévia e traçada uma estratégia para provar tal hipótese. O estudo da melhor técnica de reconhecimento de padrões é realizada com experimentos empíricos e seu ambiente é artificial, emulando situações que caso feitos de forma natural demandaria muito tempo e custo, pois deveria ser realizado em diversas aeronaves e em diversos ambientes, desde cenários noturnos até cenários de neve.

### 3. Objetivos

A pesquisa explicativa tem por objetivo explicar os porquês das coisas e suas causas, por meio do registro, da análise, da classificação e da interpretação dos fenômenos observados. Visa a identificar os fatores que determinam ou contribuem para a ocorrência dos fenômenos; aprofunda o conhecimento da realidade porque explica a razão, o porquê das coisas.([GIL, 2010](#)).

### 4. Procedimentos Técnicos

Este trabalho utilizou as seguintes formas de procedimentos: Pesquisa Bibliográfica e Pesquisa Experimental. A pesquisa bibliográfica é elaborada a partir de material já publicado, constituído principalmente de: livros, revistas, publicações em periódicos e artigos científicos, jornais, boletins, monografias, dissertações, teses, material cartográfico, internet, sempre com o objetivo de proporcionar ao pesquisador um contato direto com todo material já existente sobre o assunto da pesquisa. Para elaborar o presente trabalho, foi feita uma intensa pesquisa em técnicas de reconhecimento de padrão com o objetivo em posterior uso em aplicações de RA, pesquisa sobre o uso de RA na manutenção e comparações de técnicas em outros contextos que não o aeronáutico. Na Pesquisa Experimental determinamos um objeto de estudo, selecionamos as variáveis que seriam capazes de influenciá-lo, definimos as formas de controle e de observação dos efeitos que a variável produz no objeto. No presente trabalho, foram gerados experimentos refazendo situações reais, avaliando o impacto de variações de variáveis ambientais por meio de aproximações matemáticas aos eventos reais, como por exemplo, variação de brilho, escala ou rotação do ponto de vista do usuário em relação ao objeto de estudo.

Com uma abordagem um pouco mais pragmática, temos a tabela [1.2](#)

TABELA 1.2 – Resumo de Recursos e Métodos

Objetivo	Recurso	Método
Avaliar as algoritmos clássicos de reconhecimento	Artigos, Livros e Sites	Pesquisa Bibliográfica
Aplicar os algoritmos clássicos à situações reais	Recursos fornecidos pelo OpenCV	Simulações
Selecionar algoritmo mais adequado para o contexto	Dados coletados de simulação	Análise de cadência e qualidade de reconhecimento

## 1.6 Estrutura

Este trabalho fundamenta-se em 5 capítulos, conforme descritos abaixo:

- Capítulo 1 apresenta a motivação do presente trabalho, levantando as necessidades e limitações impostas pelo ambiente e o fato de haverem diversas técnicas de reconhecimento de características e a necessidade de selecionar a adequada para o contexto, bem como descreve o escopo do trabalho e o contexto em que os testes são feitos;
- Capítulo 2 tem informações suficientes para o entendimento das análises, descrevendo conceitos básicos e as técnicas de reconhecimento que foram comparadas;
- Capítulo 3 descreve a metodologia de análise adotada bem como o protótipo desenvolvido para a análise;
- Capítulo 4 descreve os resultados obtidos comparando-se as técnicas, selecionando qual a técnica mais adequada para reconhecimento de características para o caso de uso descrito;
- Capítulo 5 conclui o trabalho apresentando como os objetivos foram atingidos, limitações da pesquisa e possíveis trabalhos futuros.

## 2 Fundamentação Teórica

Para compreender o presente trabalho se faz necessário elucidar fundamentos básicos, fornecendo assim subsídios necessários para o leitor avaliar com mais propriedade os resultados apresentados. Os conceitos apresentados seguem uma ordem crescente de conhecimento para que seja construído um arcabolso de conhecimento suficiente. O capítulo inicia com uma explanação sobre realidade virtual e realidade aumentada, elucida o uso de dispositivos para imersão. Para a realização de aplicações de realidade aumentada o uso de sensores para abstrair informações do ambiente é fundamental, sendo as câmeras de baixo custo os mais comuns, portanto, é descrito no capítulo os problemas de distorção inerentes a tais dispositivos e a modelagem matemática adotada para as câmeras utilizada em todas as abordagens presentes nesse trabalho. A abordagem utilizada nesse trabalho é de reconhecimento de características locais, diferente do conceito comum de reconhecer padrões como retângulos, círculos, ou contornos, portanto conceituar o reconhecimento de características se faz tão importante. É também apresentada conceituação básica de cada um dos algoritmos utilizados e de parâmetros utilizados para a análise dos resultados

### 2.1 Realidade Virtual

A Realidade Virtual (RV) é uma interface avançada do usuário para acessar aplicações executadas no computador, propiciando a visualização, movimentação e interação do usuário, em tempo real, em ambientes tridimensionais gerados por computador, como mostra a figura 2.1. O sentido da visão costuma ser preponderante em aplicações de realidade virtual, mas os outros sentidos, como tato, audição, etc. também podem ser usados para enriquecer a experiência do usuário. (KIRNER C.; SISCOOTTO, 2007)

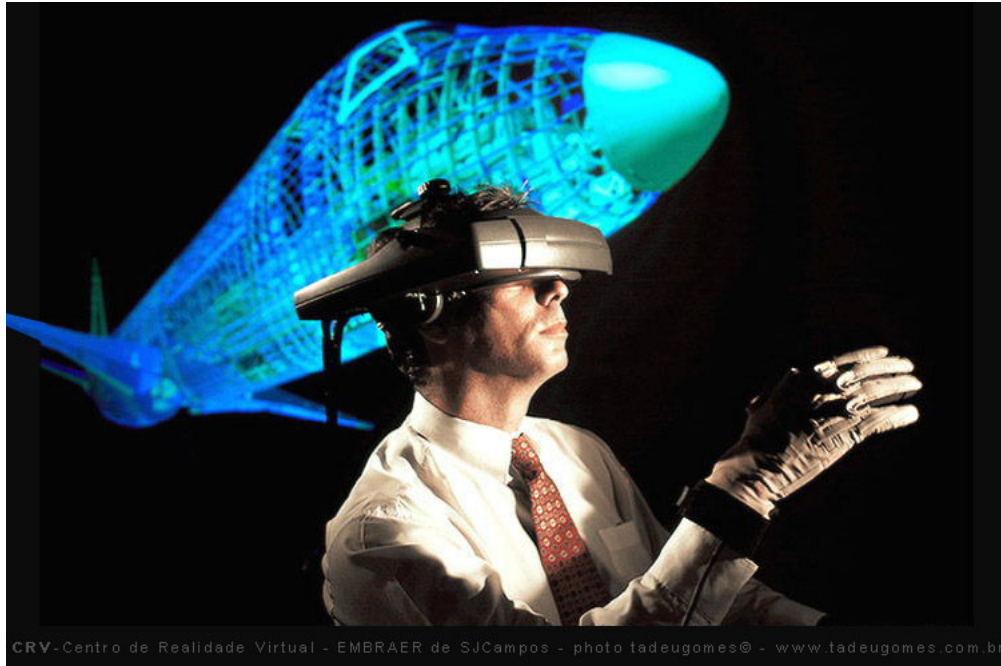


FIGURA 2.1 – Aplicação de realidade virtual

## 2.2 Realidade Aumentada

A realidade aumentada como citado em (AZUMA, 1997) é uma técnica de visão computacional em que valendo-se de artefatos do mundo real tem por objetivo causar sensação de imersão do usuário em um ambiente aumentado por artefatos virtuais, ao contrário de ambientes puramente virtuais como é comum em aplicações de realidade virtual. Idealmente o mundo virtual se torna imersivo o suficiente para que o usuário não consiga distinguir o real do virtual. Alguns autores definem AR como tendo a necessidade de utilizar-se interfaces visuais portáteis para que a usabilidade tenha mais coerência com a proposta inicial de garantir uma experiência imersiva. As imagens são obtidas por câmeras e o resultado apresentado em dispositivos como projetores ou displays como monitores, tablets ou *head-mounted display* (HMD). Realidade aumentada pode ser realizada com ou sem marcadores para facilitar o reconhecimento e posicionamento de entidades. No presente trabalho é utilizada a abordagem sem marcadores para que aplicações no cenários de manutenção se torne mais flexível no que tange à aplicabilidade e configuração inicial, não sendo necessário modificar o ambiente. A figura 2.2 apresenta um pipeline básico de realidade aumentada de forma canônica. O presente trabalho, trata de aspectos até a etapa de reconhecimento. As etapas são representadas por:



- **Captura:** Etapa de obtenção de imagens, feita por sensores como câmeras;
- **Preparação:** Etapa de preparação da imagem, aplicando filtros para a etapa de detecção;
- **Detecção:** Etapa de detecção de padrões, em que são removido informações das imagens;
- **Reconhecimento:** Etapa de reconhecimento de padrões e comparação das características reconhecidas na etapa anterior;
- **Rastreio:** Etapa que garante-se que a imagem reconhecida continua no contexto, reconhecendo apesar de movimentações ou outras variações;
- **Apresentação:** Etapa em que são desenhado na tela representações dos objetos reconhecidos de acordo com a aplicação.

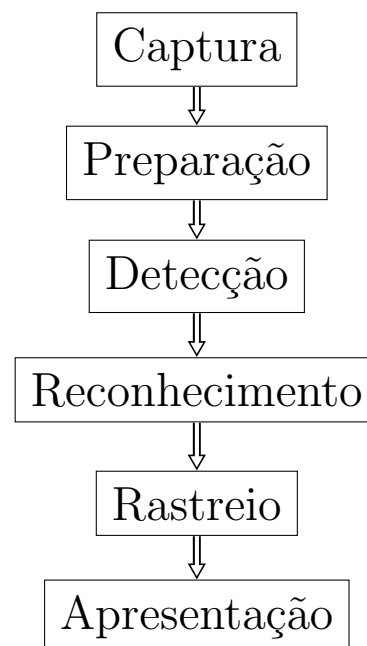


FIGURA 2.2 – Pipeline Canônico de Realidade Aumentada

## 2.3 Dispositivos

Como mostrado em (SILVA *et al.*, 2003), para que a experiência de imersão seja completa, podem ser usados alguns dispositivos como óculos especiais, ou ainda realizar com dispositivos comuns como monitores ou projetores

### 2.3.1 *Head-Mounted Displays*

É um equipamento utilizado na cabeça de forma que as duas mãos do usuário fiquem livres e tem por objetivo exibir imagens e áudio, sendo uma interface muito utilizada tanto em RV quanto em RA. Os HMD basicamente são dispositivos constituídos de duas telas posicionadas frente ao olho do usuário. Com duas telas, a tecnologia pode ser empregada para exibir imagens estereoscópicas apresentando os respectivos pontos de vista de cada olho para cada tela, o que contribui em muito na experiência de imersão. Os HMDs funcionam também como dispositivos de entrada de dados, porque contém sensores de rastreamento que medem a posição e orientação da cabeça, transmitindo esses dados ao computador. Existem dois tipos de HMDs: *Feed-Through* e *See-Through*

#### 2.3.1.1 *Feed-Through*

São dispositivos que representam um sistema fechado de visualização de imagens, como mostrado na figura 2.3 , em que o usuário consegue enxergar somente o que é mostrado no *display*, sendo assim o resultado apresentado é sempre a soma da imagem real com informações superpostas.

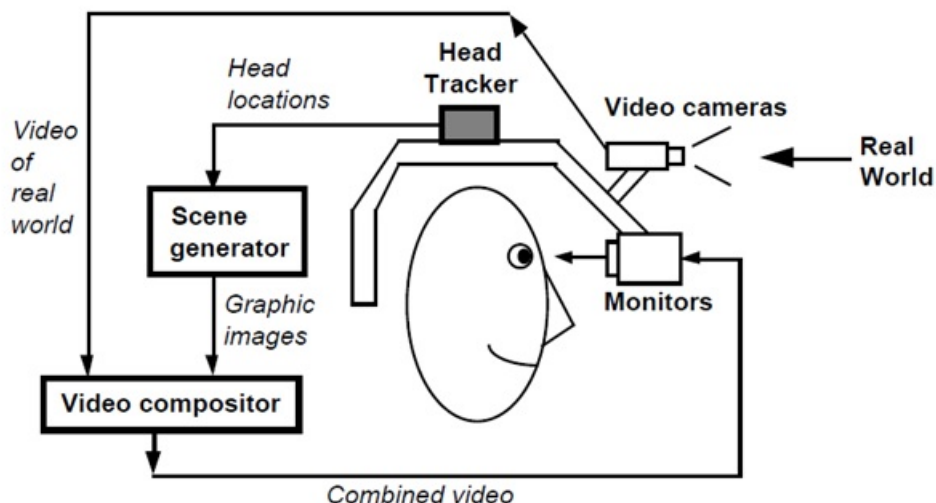


FIGURA 2.3 – Arquitetura do *Feed-Trough*.

### 2.3.1.2 *See-Through*

São dispositivos construídos com lentes translúcidas em que o usuário enxerga o mundo real e com algum tipo de sistema que sobreponha na lente as informações adicionais. Como mostrado na figura 2.4

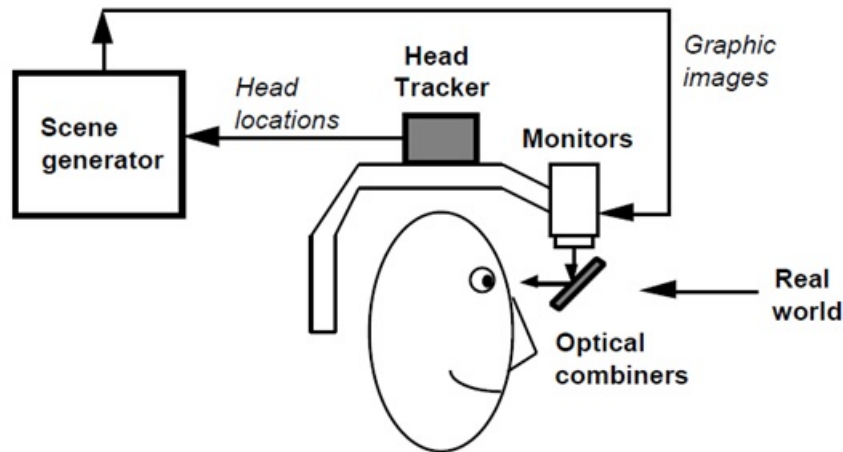


FIGURA 2.4 – Arquitetura do *See-Through*.

### 2.3.2 Projetores

O uso de projetores possibilita uma abordagem de realidade aumentada diferente porque pode ser utilizada para cobrir superfícies largas, projetando sobre objetos como carros, pessoas, prédios, etc. Um problema dessa abordagem é que a calibração se faz necessária em várias situações.

### 2.3.3 Monitores

O uso de monitores reduz bastante o custo da aplicação apesar de ter perda de imersão por ser um método de visualização indireta, o que implica o usuário ficar olhando na direção do monitor, entretanto existe a possibilidade de compartilhar os resultados da RA com mais de uma pessoa ao mesmo tempo. Como mostrado na figura 2.5



Fonte: (IMMERSION, 2014)

FIGURA 2.5 – Realidade Aumentada com projetores.

## 2.4 Camera

Como descrito em (BRADSKI, 2008), as câmeras de baixo custo vendidas atualmente introduzem diversos erros de distorção. Duas dos maiores são as distorções radiais e as distorções tangenciais. Devido a distorções radiais, linhas retas vão parecer curvadas tendendo a aumentar quanto mais o ponto se distancia do centro da imagem. Na prática essas distorções são pequenas e podem ser caracterizadas pelos primeiros termos de uma expansão da série de Taylor.

### 2.4.1 Distorção Radial

Distorções radiais como mostradas na figura 2.6, são mais associadas a lentes de grandes ângulos, em particular zoom *wide-angles*. Esse efeito faz com que as arestas da imagem apareçam curvadas a olho humano. A distorção é resultado da fabricação por ser bem mais barato e prático produzir lentes esféricas do que lentes mais matematicamente ideais com perfis parabólicos.

A imagem 2.7 mostra como o deslocamento de um retângulo devido à distorção radial, mais acentuado quanto mais próximo das bordas da imagem.

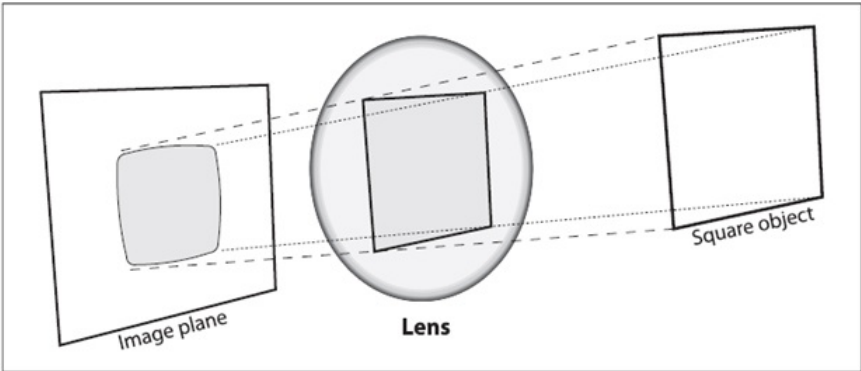


FIGURA 2.6 – Distorção radial.

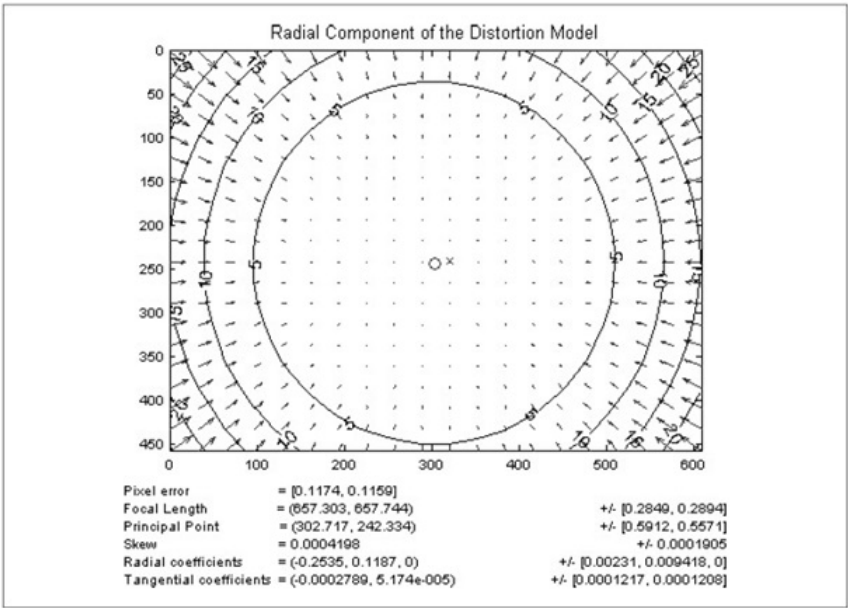


FIGURA 2.7 – Componentes de Distorção radial.

### 2.4.2 Distorção Tangencial

Como mostradas na figura 2.8 Essa distorção se deve a defeitos de manufatura causadas pelo fato de o plano da lente não ser totalmente paralelo ao plano da imagem.

A imagem 2.9 mostra o efeito da distorção tangencial em um grid de pontos retangular. Os pontos se deslocam deformando elipticamente quanto mais próximos da borda.



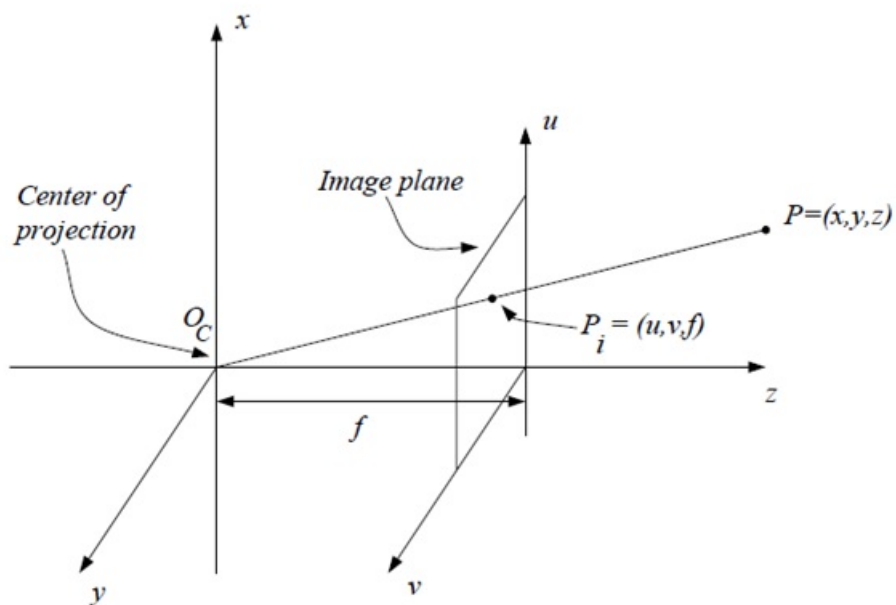


FIGURA 2.10 – Camera no espaço vetorial.

## 2.6 Reconhecimento

É a etapa em que padrões são identificados e comparados para posteriormente identificar objetos

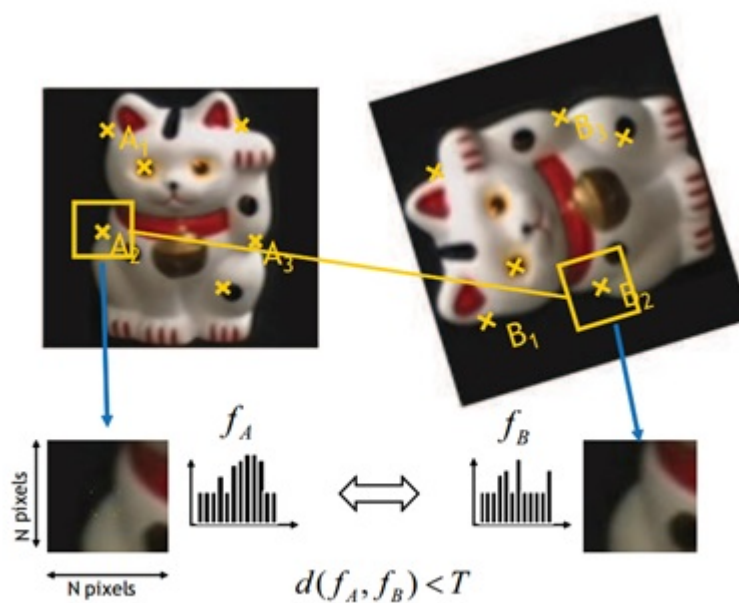


FIGURA 2.11 – Ilustração do procedimento de reconhecimento com features locais.

*Pipeline* de reconhecimento como ilustrado na imagem 2.11 :

- Encontrar um grupo de *keypoints* distintos
- Definir uma região em torno de cada *keypoint*
- Extrair e normalizar o conteúdo da região
- Calcular um descritor para a região normalizada
- Encontrar correspondências de descritores.

## 2.7 Características Locais

Características locais são padrões em imagens que diferenciam de seu vizinho imediato, geralmente caracterizadas por mudanças em propriedades das imagens ou diversas propriedades simultaneamente. As propriedades mais comuns são intensidade, cor e texturas.

### 2.7.1 Propriedades da Característica Local ideal

Algoritmos de reconhecimento baseam-se em comparação de características recuperadas da cena. A recuperação e comparação de pontos tem um custo computacional relevante perto do tempo de execução da aplicação, portanto selecionar o menor número possível de características aumenta o desempenho e diminui o tempo de resposta da aplicação. Garantir que estamos selecionando boas características pode ser crucial na eficácia do reconhecimento. Segundo (TUYTELAARS, 2007), boas características devem ter as seguintes propriedades:

- **Repetibilidade:** Dadas duas imagens do mesmo objeto ou cena, tomadas em condições ou pontos de vista diferentes, uma porcentagem alta de características deve ser reconhecidas se estiverem visíveis.
- **Distinção:** Os padrões reconhecidos têm de ser possíveis de serem distinguidos entre si para facilitar o casamento.
- **Localidade:** As características devem ser locais para reduzir a probabilidade de oclusão.



- **Quantidade:** O número de pontos detectados tem que ser o suficiente para que mesmo objetos pequenos tenham minimamente características que possam ser localizados e para que o objeto possa sofrer oclusão e ainda assim ser reconhecido.
- **Exatidão:** As características detectadas tem que ser localizadas com o máximo de exatidão possível com respeito tanto referente à posição quanto à escala.
- **Eficiência:** De preferência a detecção deve ser o mais rápido possível.

### 2.7.2 Características

Antes de compreender como é feito o reconhecimento e registro de imagens é importante nos perguntar como nós conseguimos reconhecer objetos em uma cena, como conseguimos comparar facilmente objetos em duas imagens distintas. Somos treinados desde cedo a diferenciar formas geométricas, perceber escalas diferentes ou mesmo reconhecer o mesmo objeto independente de como está posicionado na cena buscando padrões que categorizem e diferenciem o objeto. Instintivamente conseguimos reconhecer boas características e localizar objetos. Na imagem 2.12 temos uma imagem de um prédio e seis recortes, dos quais conseguimos facilmente reconhecer com precisão a de letra E e F, as de letra A,B,C,D podemos identificar possíveis localizações mas não podemos dizer com certeza onde estão na imagem.

As características E e F são o que chamamos de good features pois o nível de certeza é bem alto.

A Imagem 2.13 ilustra tipos de características. A região azul não possibilita diferenciar onde está na imagem, a região preta pode ser confundida com qualquer uma das regiões ao deslocarmos horizontalmente, a imagem vermelha nos possibilita diferenciar e reconhecer o canto da imagem verde com precisão milimétrica. Podemos então concluir que uma característica é boa para ser utilizada como parâmetro de entrada para algoritmos de reconhecimento, quanto maior foi o nível de certeza da sua localização, o que facilita o **Feature Detection**. Para localizar o mesmo objeto em outra imagem é necessário identificarmos a região onde se encontra, caso contrário no exemplo da imagem 2.12 seria impossível localizar uma janela específica. Tal descrição de contexto é chamada de **Feature Description**. Uma vez de posse da característica e do seu contexto é possível



FIGURA 2.12 – Reconhecimento de padrões.(UNDERSTANDING..., 2014)

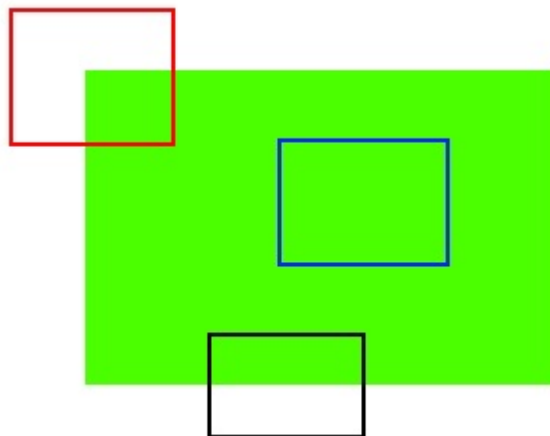


FIGURA 2.13 – Regiões de reconhecimento de padrões.(UNDERSTANDING..., 2014)

reconhecer o objeto de fato.

### 2.7.3 Detecção de características

O primeiro passo para o reconhecimento de objetos é a detecção de características. Para que possa ser feita a comparação da imagem referência com a imagem sentida. A abstração

de informações a ser reconhecida tem que ser suficiente para lidar com escalas diferentes, rotações entre as imagens, e pequenas distorções. Para que o reconhecimento seja eficiente e eficaz são necessários que uma massa de pontos mínima seja selecionada e feita a correspondência entre as imagens sentidas e referência. As características utilizadas são em geral cantos, linhas, curvas, padrões ou regiões. O tipo de característica selecionada é dependente do tipo de imagem provida. Imagens de cenas feitas a mão geralmente são compostas de segmentos de retas enquanto imagens de satélite são geralmente compostas de contornos e regiões. Quanto mais invariantes forem as características encontradas, mais robusto e preciso é o processo de comparação.

#### 2.7.4 Correspondência de características

A fase de correspondência de características é feita tanto selecionando características na imagem referência e procurando a correspondente na imagem sentida ou mesmo selecionando características nas duas imagens independentemente e procurando a correspondência entre elas. Quando a característica selecionada não for do tipo ponto, é importante para cada par de correspondências, pelo menos um ponto ser determinado para que seja utilizado para determinar posteriormente os parâmetros de transformação. Por exemplo, se forem selecionados padrões como tipo de característica, o centro do padrão é considerado o ponto, se for selecionado uma região, o centro de massa da região é o ponto de apoio, se linhas forem tomadas como tipo de característica, devem ser tomados intersecções como ponto de apoio e finalmente se forem selecionadas curvas, os máximos locais são considerados os pontos correspondentes.

### 2.8 Algoritmos de Reconhecimento

Existem diversos algoritmos de reconhecimento de características, entretanto nesse artigo os testes serão restritos aos algoritmos BRISK, FAST, FREAK, GFTT, MSER, ORB, STAR, SURF, SIFT. Este capítulo se propõe a dar uma visão geral dos algoritmos, pois o foco do presente trabalho está na análise comparativa entre os mesmos e não em cada um dos algoritmos, visto existir implementações conceituadas no OpenCV, framework utilizado.

### 2.8.1 FAST - Features from Accelerated Segment Test

Segundo (ROSTEN; DRUMMOND, 2006), é um método de reconhecimento baseado em detecção de arestas originalmente desenvolvido por Edward Rosten e Tom Drummond. A maior promessa do método é a eficiência computacional. O método considera um círculo de dezesseis pixels ao redor da aresta considerada  $p$ . O detector original (ROSTEN; DRUMMOND, 2005), (ROSTEN; DRUMMOND, 2005) classifica  $p$  como uma aresta se existirem  $n$  pixels contíguos em um círculo que são mais brilhantes do que o pixel candidato de intensidade  $I_p$  mais um threshold  $t$  ou mais escuros do que  $I_p - t$

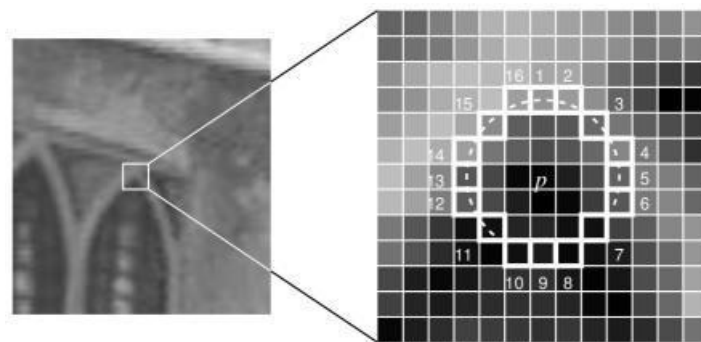


FIGURA 2.14 – FAST

Na imagem 2.14 foi escolhido  $n=12$ . Um teste rápido é

1. Selecionar o ponto é testar primeiro as extremidades. No caso da imagem, escolhido o ponto  $p$
2. Comparar os pontos 1 e 9, e verificar se o ponto  $p$  tem intensidade com diferença de módulo  $t$ , ou seja os pontos 1 e 9 são mais claros ou mais escuros do que o ponto  $p$  pelo fator de  $t$
3. Avaliar se o ponto  $p$  continue sendo um candidato considerando os pontos 5 e 13
4. Analisar se  $p$  é uma aresta, sendo pelo menos 3 desses pontos devem ser mais brilhantes ou mais escuros do que  $p$  então o teste pode ser feito nos demais pontos.

### 2.8.2 BRISK - Binary Robust Invariant Scalable Keypoints

Como descrito em (LEUTENEGGER *et al.*, 2011), ao contrário dos descritores vetoriais, como SURF e SIFT, BRISK utiliza descritores no espaço binário o que para dispositi-

vos com restrição de recursos e poder computacional como dispositivos mobiles pode ser interessante. Baseado no detector da técnica FAST, o processo consiste de três partes.

### Amostragem de padrões

Retirando um padrão de amostras ao redor do *keypoint* referentes a pontos espalhados em círculos concêntricos, que são usados para determinar se o ponto deve ou não ser selecionado como em um detector FAST. As amostras são separadas em pares de dois subsets, curta distância e longa distância.

### Compensação de Orientação

Para atingir invariância a rotação, a direção de cada *keypoint* é determinada tomando a soma dos gradientes locais calculados entre pares de longa distância e rotacionando os pares de curta distância baseados na orientação obtida.

### Comparação de pares de amostragem

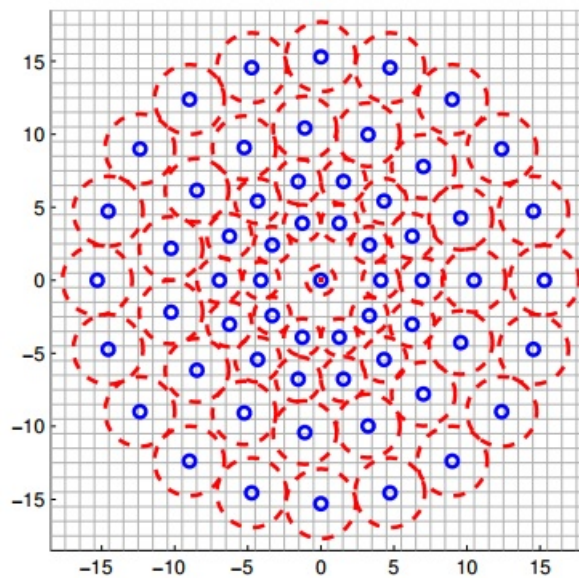


FIGURA 2.15 – Padrão de amostras no BRISK

Como mostrado na figura 2.15, BRISK possui um descritor binário de 512 bits que calcula a média ponderada utilizando uma Gaussiana sobre um padrão de pontos mais próximos do ponto selecionado como mostrado na imagem 2.15 no ponto (0,0). Ao longo do padrão descrito, são aplicadas suavizações Gaussianas sendo que os círculos vermelhos representam o tamanho do desvio padrão do filtro Gaussiano aplicado a cada ponto. É baseado em comparações entre pares de janelas de Gaussianas, resultando em 1 ou 0,

dependendo de qual janela no par for maior. Os pares então são pré selecionado no BRISK, criando descritores binários que são posteriormente utilizados para casamento de padrão, utilizando distâncias Hamming ao invés de Euclidianas.

### 2.8.3 FREAK - Fast Retina Keypoint

FREAK é um descritor binário, composto por três etapas:

#### Amostragem de padrão

FREAK propõe uma abordagem biológica para o reconhecimento de características, emulando o funcionamento da retina para amostragem de padrão, como demonstrado na imagem 2.16 que é um padrão circular com maior densidade de pontos próximo do centro, decrescendo exponencialmente.

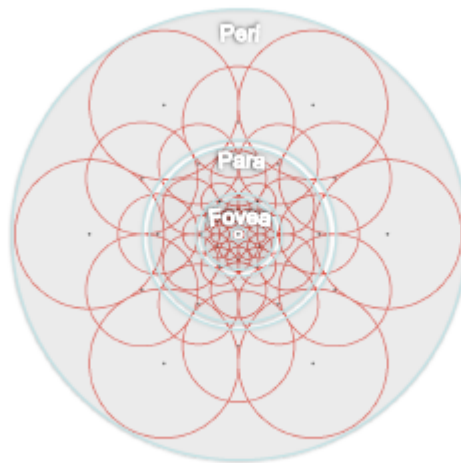


FIGURA 2.16 – Padrão de amostragem do descritor FREAK

Cada amostra é suavizada com um kernel Gaussiano em que o raio do círculo ilustra o tamanho do desvio padrão do kernel. Como pode ser observado na figura 2.17 o padrão de amostragem corresponde com a distribuição de receptores na retina.

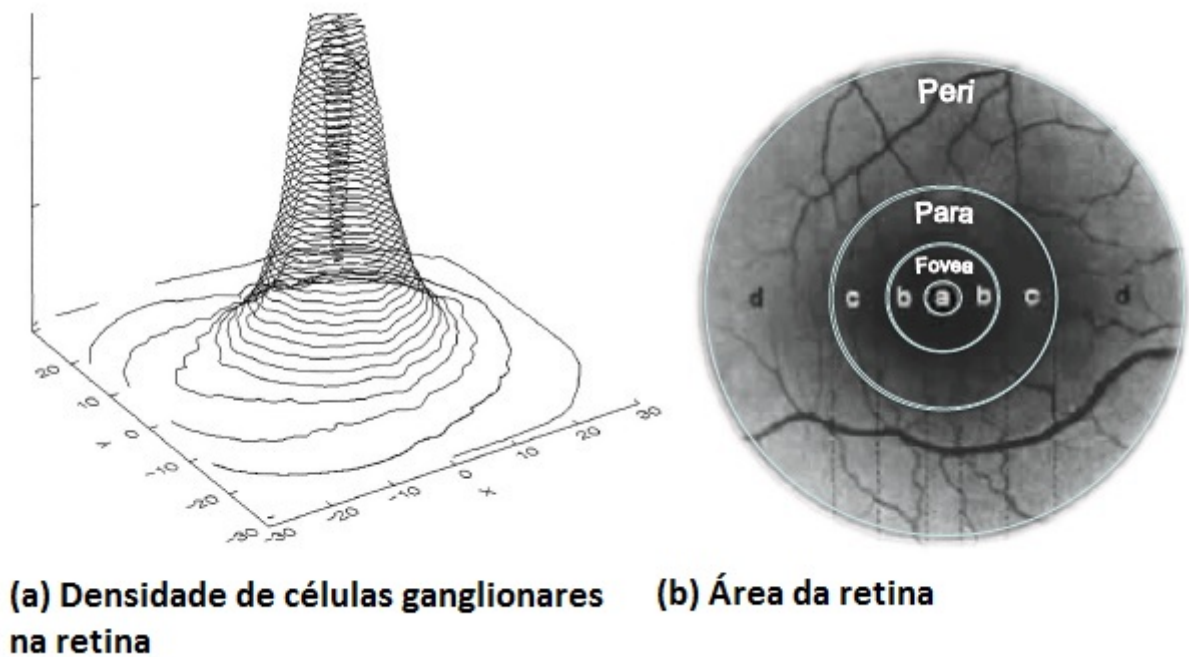


FIGURA 2.17 – Distribuição de receptores na retina

### Compensação de Orientação

Para estimar a rotação dos *keypoints*, são somados os gradientes locais assim como no BRISK, entretanto ao invés de considerar os pontos de longa distância, é considerado um padrão de 45 pontos como mostrado na figura 2.18

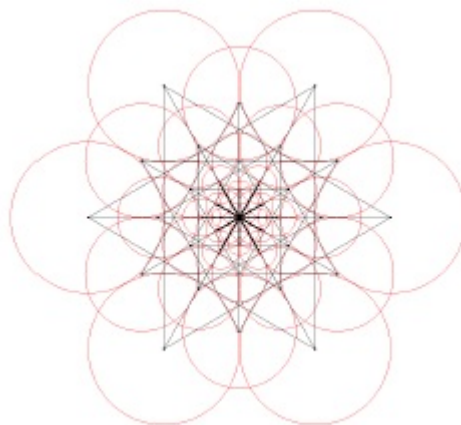


FIGURA 2.18 – Pares selecionados para calcular a orientação

Apesar de ter menos precisão para recuperar informações de rotação, como o número de pontos é bem menor do que BRISK a quantidade de memória armazenada é em geral 5 vezes menor.



### Comparação de pares de amostragem

Os pares de pontos são selecionados considerando a densidade maior no centro, como podemos observar na figura 2.17 (a). Os pares começam a ser comparados pelas extremidades e para dentro do centro, dessa forma otimizamos o reconhecimento pois com menos pontos podemos descartar casos em que a distância estiverem maiores do que um *threshold*, caso contrário prosseguimos para os outros 128 bits do descritor.

#### 2.8.4 GFTT - Good Features To Track

O método é desenvolvido para detectar padrões sem bordas. A matriz de gradientes  $G$  é calculada para cada pixel como mostrado na equação 2.1

$$G = \sum_{\Omega} \begin{bmatrix} I_x^2 & I_x I_y \\ I_x I_y & I_y^2 \end{bmatrix} \quad (2.1)$$

Sendo o valor de intensidades  $I(x, y)$  de uma imagem de tons de cinza e suas derivadas parciais  $I_x, I_y$  de uma determinada região  $\Omega$ . A matriz de gradientes é implementada por meio de uma imagem integral para  $I_x^2, I_y^2$  e  $I_x I_y$ . Devido ao uso de imagens integrais a complexidade computacional da matriz de gradientes é constante e independente do tamanho de  $\Omega$ . Uma boa característica pode ser identificada pelo menor autovalor de  $G$ . Pontos fortes aparecem em geral nas bordas, onde problemas com movimento são mais comuns, o que leva para o problema de abertura geral. Pós processamento é aplicado por uma supressão de não máximos com um *threshold* em  $q * \max(\lambda(x, y))$ , em que  $q$  é uma constante para garantir a qualidade.

#### 2.8.5 MSER - Maximal Stable Extremal Regions

O detector MSER (MATAS J., 2002) é composto por regiões de todos os pixels conectados considerando um *threshold*. Em outras palavras, as regiões selecionadas são padrões que não mudam em que binarização local é estável ao longo de uma faixa de *thresholds*. De acordo com o paper podemos fazer uma analogia com o processo de derramamento formação de poças de água para compreendermos como são formadas as regiões MSER, que são descritas em imagens em tons de cinza representados pela função  $I : \Omega \rightarrow [0 \dots 255]$



em que  $\Omega = [1]x[1]$ . O método garante localização de objetos que estão mais próximos de nossa realidade como pode ser observada na figura 2.19.



FIGURA 2.19 – Regiões MSER

Selecioneado um *threshold* de intensidades, a imagem é dividida em dois grupos, P(pretos) e B(brancos). É observada que a quantidade de regiões varia dependendo do *threshold* aplicado, de 0 à 255. A área de cada componente conectado é então armazenada como uma função. Dentre as regiões, as mais estáveis são selecionadas analisando as funções para as quais cada região em potencial mantém seu estado com a mesma função independente da variação de *threshold*. As regiões “máximamente estáveis” são chamadas regiões MSER, que mudaram apenas em tamanho variando-se pelo menos alguns níveis de *threshold*.

### 2.8.6 ORB - Oriented Fast and Rotated Brief

Como citado em (RUBLEE, 2011), é uma combinação de FAST e BRIEF. Para extrair *keypoints*, modifica o detector do FAST construindo uma pirâmide de escalas das imagens. Em cada escala, *keypoints* são detectados e a distância de Harris aplicada para selecioná-los e somente os N melhores pontos são selecionados de acordo com um *threshold*. Para obter invariância a rotação, momentos de primeira ordem são utilizados para calcular a orientação local através de um centróide que referencia a média da magnitude dos pixels em um trecho local. Posteriormente são calculados os descritores BRIEF nos trechos rotacionados e armazenadas as informações em um descritor ORB.

### 2.8.7 SIFT - Scale-Invariant Feature Transform

De acordo com (LOWE, 2004), é um método baseado em detecção de arestas que tem como proposta garantir a invariância à variação de escala. Um problema que nos

métodos de reconhecimento de arestas se não tratado pode causar diminuição na robustez do algoritmo. A figura 2.20 ilustra bem o efeito que a mudança de escala pode fazer.

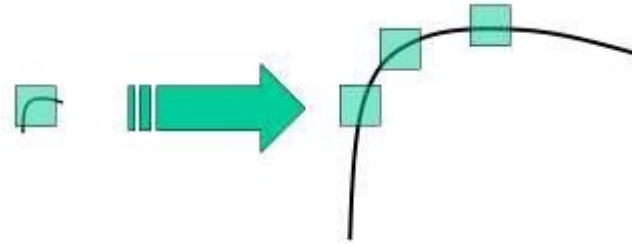


FIGURA 2.20 – SIFT

Em 2004, D.Lowe, descreveu o algoritmo Scale Invariant Feature Transform no seu paper Distinctive Image Features from Scale-Invariant Keypoints[15] como extrair keypoints e computar os descritores. 8.10.2.1. Scale-space Extrema Detection

Observando a imagem 2.20 é possível notar que não podemos utilizar a mesma janela de inspeção independente da escala do objeto, para objetos maiores temos que utilizar janelas maiores. Nesse contexto, o filtro de espaço de escalas é utilizado e são calculados Laplacianos de Gaussianos com diversos valores de  $\sigma$ . Utilizar Laplacianos de Gaussianos é uma abordagem custosa computacionalmente, como uma forma de aproximação são utilizados diferenças de Gaussianas.

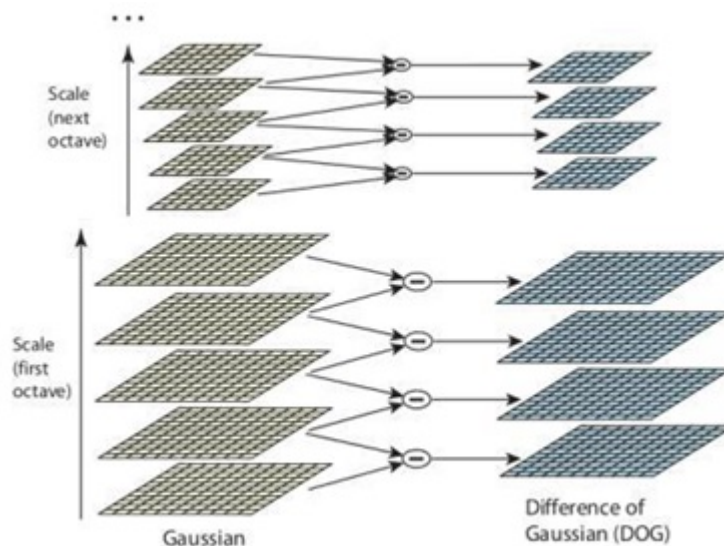


FIGURA 2.21 – Difference of Gaussian

Uma vez que as diferenças de gaussianas são calculadas, é necessário procurar por máximos entre espaços de escala diferentes. Por exemplo como mostrado na figura 2.21,

um pixel na imagem é comparado com seus 8 vizinhos e com níveis de escala próximos e anteriores. O que representa que o *keypoint* encontrado é melhor representado naquela escala.

#### 2.8.7.1 Keypoint Localization

Uma vez que *keypoints* potenciais são localizados é importante selecionar pontos de interesse com contraste alto. A localização dos pontos é refinada utilizando uma expansão de Taylor e se as intensidades dos máximos forem menores do que um threshold são rejeitados. Uma característica do DoG é a alta resposta a arestas gerando falsos positivos. Portanto é necessário eliminar algumas arestas identificadas erroneamente.

#### 2.8.7.2 Orientation Assignment

Para obter descritores que são invariantes a rotação, um histograma de orientações é gerado a partir do gradiente de orientações de cada maximo local do DoG ao redor dos *keypoints*.

#### 2.8.7.3 Keypoint Descriptor

O descritor é construído por um vetor de características considerando a direção do keypoint cujo peso no gradiente é máximo. Como mostrado na figura 2.22, tipicamente, uma região de adjacentes 16x16 é determinada partindo do keypoint no centro. Depois da região ser selecionada, SIFT divide em sub regiões de 4x4 com 8 representações de orientações em cada. Como obtemos  $4 \times 4 = 16$  histogramas, cada um com 8 items, o vetor tem 128 elementos, o que torna o descritor significativo, compacto e ainda assim robusto o suficiente para mudanças de iluminação e variações de câmera.

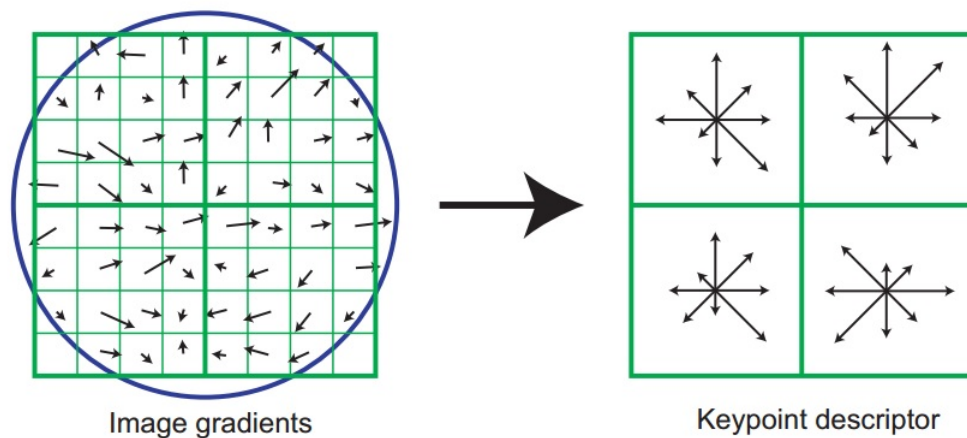


FIGURA 2.22 – Formação do descritor SIFT

#### 2.8.7.4 Casamento de Padrões

O casamento entre dois pontos é feito identificando pontos próximos. Entretanto em algumas situações existem pontos muito próximos que podem ser causados por ruídos na detecção de pontos de interesse, nesse caso é calculada uma razão de distância entre o ponto de interesse com o mais próximo, e com o segundo mais próximo, se a razão for maior do que 80 são rejeitados. Tal abordagem elimina cerca de 90 de falsos positivos e 5 de pontos corretos.

#### 2.8.8 SURF - Speeded Up Robust Feature

Segundo ([BAY TINNE TUYTELAARS, 2008](#)), em comparação com o algoritmo SIFT que aproxima LoG por Diferenças de Gaussianas (DoG), SURF vai um pouco mais além e aproxima LoG de *Box-type Filter* como mostrado na figura 2.23 e não é utilizado nenhum tipo de suavização entre escalas, o que garante mais agilidade nos resultados porque a convolução com *box filters* são muito mais rápidos com o uso de *integral images* e pode ser paralelizado

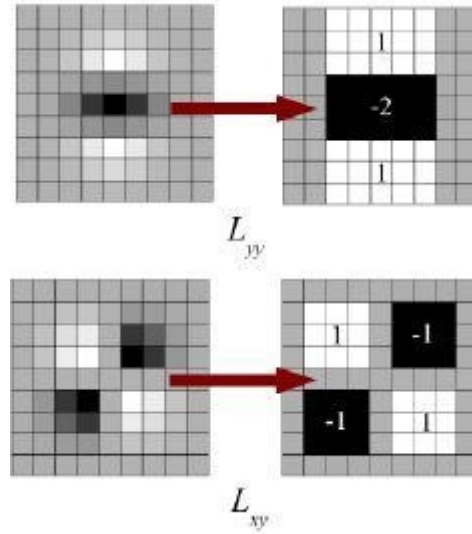


FIGURA 2.23 – Box Filtering

Em geral SURF se apresenta cinco vezes mais rápido do que DoG. Assim como SIFT, SURF cria um grade ao redor do *keypoint* e divide a grade em sub-grades.

### 2.8.9 STAR

## 2.9 Cadência

É a medida do número de quadros individuais que um determinado dispositivo óptico ou eletrônico processa e exibe por unidade de tempo. Em geral a cadência é medida em fps. Em cinema, a cadência de projeção padrão desde 1929 foi fixada em 24fps, mas no período do cinema mudo a maioria dos filmes eram rodados com cadência entre 16 e 20fps. Em vídeo, os principais sistemas lidam com cadência entre 25fps(PAL) e 30fps(NTSC). As aplicações devem ter cadência tolerável dependendo de seu uso, segundo (TANG *et al.*, 1993) para aplicações interativas o mínimo tolerável é de 5fps enquanto para aplicações de animações fluidas de 30fps. Sendo a cadência, a frequência entre frames, deve ser contabilizado o tempo de geração de informações e o tempo de dispor a informação no dispositivo óptico. O tempo de cada frame é calculado como mostrado na equação 2.2

$$t_{frame} = \frac{1}{fps} \quad (2.2)$$

---

No caso de cadência mínima de 5fps, temos quadros com tempo menor que 200ms, portanto as análises devem ser balisadas a tempos menores.

## 3 Aplicação e Seleção de Algoritmos

### 3.1 Metodologia

#### 3.1.1 Definição de parâmetros

Os testes são realizados adicionando variáveis de forma artificial por meio de transformações afins de forma a emular comportamentos ou situações encontradas no ambiente de manutenção

- Variação de escala para simular a aproximação dentro da janela de inspeção
- Variação de rotação para simular a movimentação durante a manutenção
- Variação de iluminação para simular manutenções feitas em horários do dia e iluminação diferente como por exemplo ambientes com neve com brilho muito maior
- Adição de *Blur* para emular ambientes com muita poeira ou esfumaçados

A definição das faixas de parâmetro utilizados foi feita com inspeção em campo

### 3.2 Análise de dados

A análise de desempenho de cada técnica deve ser balizada por parâmetros que garantam que a aplicação seja viável para o contexto citado na Seção 1.4. Para tal deve-se considerar os parâmetros estimados abaixo definidos por inspeção

- Cadência  $> 5\text{fps}$  ou seja, tempo  $< 200\text{ms}$  ([TANG et al., 1993](#))

- Robustez à escala:  $0.5 < \text{escala} < 1.2$
- Robustez à rotação:  $0 < \text{rotação} < 30$
- Robustez à iluminação:  $-50 < \text{iluminação} < 50$
- Robustez à *blur*:  $\text{kernelsize} < 4$

Uma análise mais precisa dos parâmetros não faz parte do escopo desse trabalho pois envolve pesquisa de campo em situações adversas. Portanto tais parâmetros foram escolhidos para validar o método de seleção das técnicas, sendo assim selecionados parâmetros identificados por experiência como usuais para o contexto, entretanto não há perda de generalismo no método posto que pode ser executado da mesma maneira com outros valores.

### 3.2.1 Protótipo

O escopo de desenvolvimento dessa tese tem por traçar um método de escolha do algoritmo a ser utilizado dependente de aspectos descritos na Seção 1.4.1, os testes serão feitos em duas etapas: Definição de parâmetros de contorno e avaliação dos algoritmos mais conhecidos no contexto aeronáutico

#### 3.2.1.1 Testes de reconhecimento de padrões

Utilizando os parâmetros definidos na etapa anterior serão rodados todos os algoritmos para avaliar qual se encaixa melhor em cada uma das situações. O protótipo para testes será realizado utilizando o OpenCV por já disponibilizar os algoritmos mais utilizados para o reconhecimento de padrões e algoritmos de visão computacional. Nesse protótipo será avaliado a aderência dos algoritmos: BRISK, FAST, FREAK, MSER, ORB, STAR, SURF e SIFT Utilizando a figura 3.1 nos testes.





FIGURA 3.1 – Imagem utilizada no protótipo

Segundo os seguintes critérios:

**Velocidade** - para garantir que a aplicação poderá ser utilizada como uma ferramenta de tempo real é imprescindível buscar uma aplicação que rode naturalmente a 30 frames por segundo. Tal realidade é facilmente atingida em computadores com vários núcleos como os core i7 mas é importante lembrar que o cenário de realidade aumentada para ter um contexto coerente com a proposta de imersão é composto por dispositivos portáteis que via de regra não tem poder de processamento tão alto.

**Qualidade** - Características detectadas são geralmente utilizadas posteriormente em etapas de rastreamento e casamento de padrão. Para os testes de rastreamento serão adicionadas às imagens algumas transformações afins, como escala, rotação e iluminação e então estimar a qualidade das características.

**Iluminação e invariância à escala** - Detectores de características tem que ter a habilidade de reconhecer características independente do tamanho do objeto. A invariância deve ser verdadeira para variação de iluminação. Variações pequenas de iluminação e contraste não devem afetar o detector significativamente. As câmeras atuais em geral tem controle automático de ganho que ajusta automaticamente a exposição evitando sub ou super exposição.

Utilizados as seguintes transformações que tem por finalidade avaliar as variações de ambiente

- **Rotação:** Robustez à rotação também é uma característica importante pois nem sempre encontraremos situações em que o dispositivo de RA será posicionado na mesma orientação o tempo todo, garantindo uma maior autonomia a quem executa a manutenção
- **Brilho:** A variação de brilho pode emular tanto variações temporais, como dia e noite, mas também pode emular situações de muita luminosidade como neve por exemplo
- **Blur:** A robustez à *Blur* auxilia na resposta do algoritmo frente à movimentos bruscos
- **Escala:** A robustez à escala do objeto garante que o mesmo será reconhecido tanto ao longe quanto em situações bem próximas

#### 3.2.1.2 Ambiente

Os testes tem por objetivo garantir as restrições descritas na Seção 3.2 Constraints, portanto é o desempenho e o tempo de reconhecimento são uma característica relevante. Os resultados obtidos pelos testes descritos nessa tese foram executado em uma máquina com a seguinte configuração:

Processador: Core 2-duo 2.2GHz

Memória: 4GB

Placa de Vídeo: NVidia GForce 9300M GS

Biblioteca: OpenCV 2.4.10

#### 3.2.1.3 OpenCV

Distribuído sob licença BSD e portanto livre para uso acadêmico e comercial. Possui interfaces para C++, C, Python e Java suportando Windows, Linux, Mac OS, iOS e Android. OpenCV foi desenvolvido para resultados eficientes e com foco em aplicações

de tempo real, podendo tomar vantagem de processamentos paralelos, utilizando-se de OpenCL aproveitando-se da aceleração de hardware ou mesmo de plataformas heterogêneas. Adotado ao redor do mundo em várias pesquisas com uma comunidade de mais de 47 mil pessoas com mais de 9 milhões de downloads. Adotado como plataforma de desenvolvimento para aplicações de arte interativa a aplicações de robótica avançada.

## 4 Análise de Resultados

A abordagem de testes e análise de desempenho é feita como descrito em (MIKO-LAJCZYK, 2005). Como descrito na Seção 3.2, existem restrições para garantir que a técnica utilizada responde segundo as necessidades da aplicação. Nos gráficos a seguir são traçadas janelas de restrição, sendo que os valores que se encontram dentro dela são adequadas à situação. As técnicas tem suas implementações no OpenCV com a possibilidade de configurar alguns parâmetros entretanto para garantir que a comparação inicial seja independente de peculiaridades de implementação ou mesmo das diversas possibilidades de configuração foi feita a comparação utilizando os construtores padrão, com as configurações iniciais.

### 4.1 Taxa de Acertos

Foram avaliados, a razão entre a quantidade de características corretamente localizadas e o número de características na imagem inicial. Idealmente o valor dessa medida deve se aproximar de 100%. Inicialmente foram considerados os resultados que se demonstraram dentro de pelo menos mais de 50% da janela delimitada, pois a análise deve ser feita considerando tanto o parâmetro de taxa de acertos quanto o tempo gasto para realizá-lo.

Como mostrado na imagem 4.1, para os testes de Escala estão dentro da janela as técnicas FREAK, ORB, GFTT.

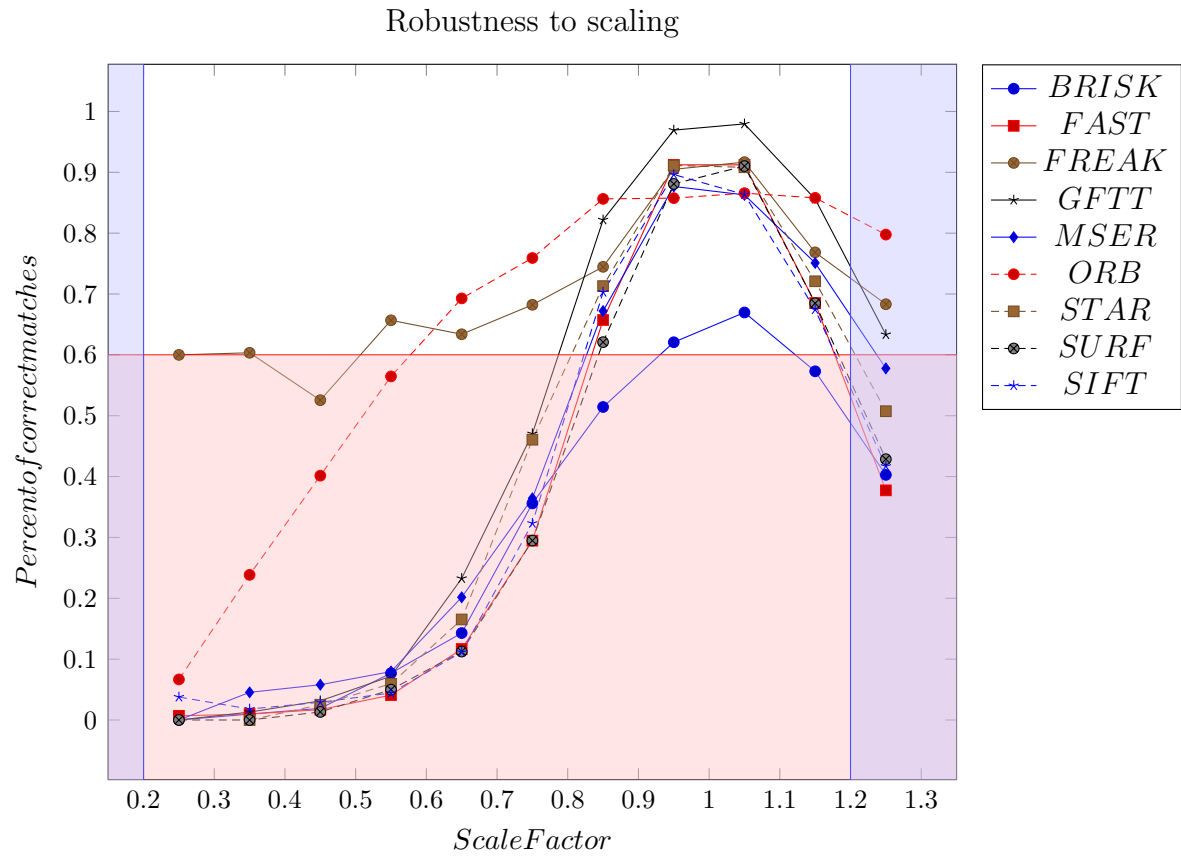


FIGURA 4.1 – Resultado de performance de variação de escala

Como mostrado na imagem 4.2, para os testes de Rotação estão dentro da janela as técnicas BRISK,FAST,FREAK,GFTT,MSER,ORB,STAR,SURF,SIFT

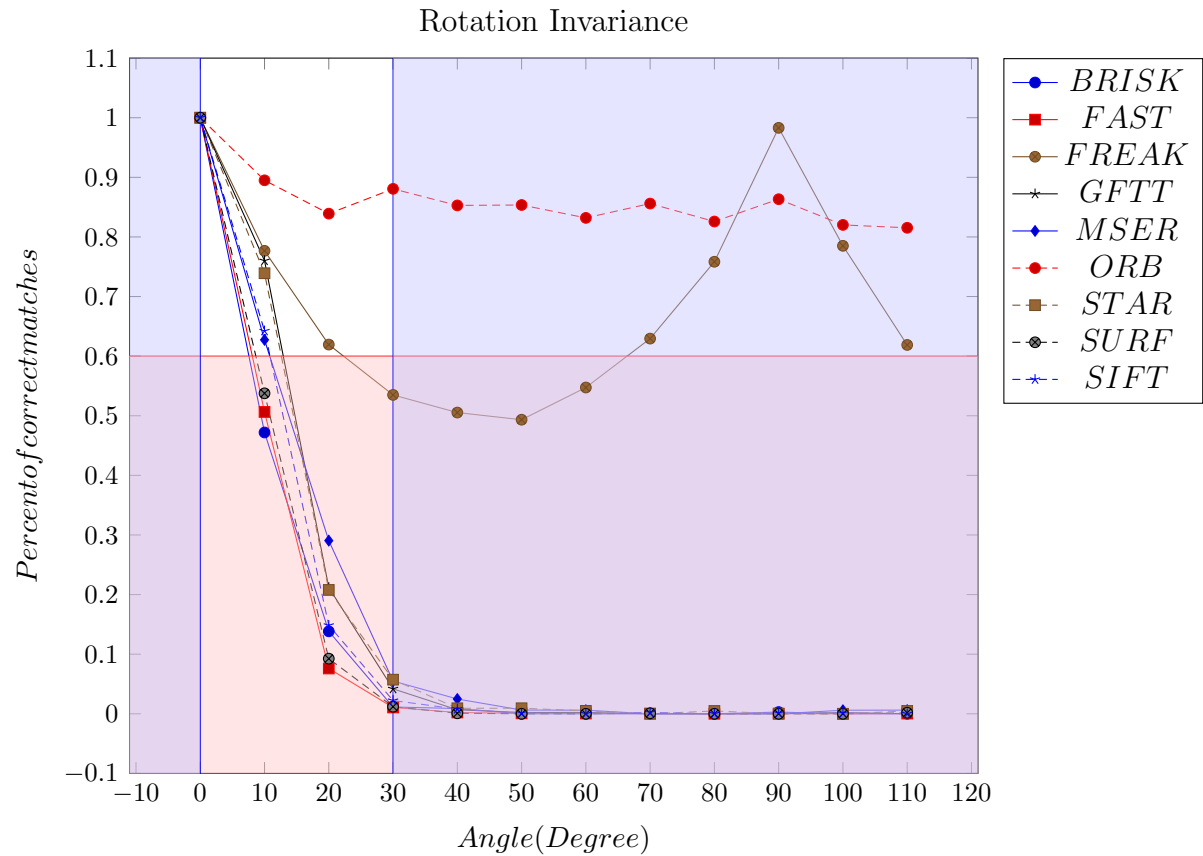


FIGURA 4.2 – Resultado de performance de variação de rotação

Como mostrado na imagem 4.3, para os testes de Blur estão dentro da janela as técnicas FAST,FREAK,GFTT,MSER,ORB,STAR,SURF,SIFT

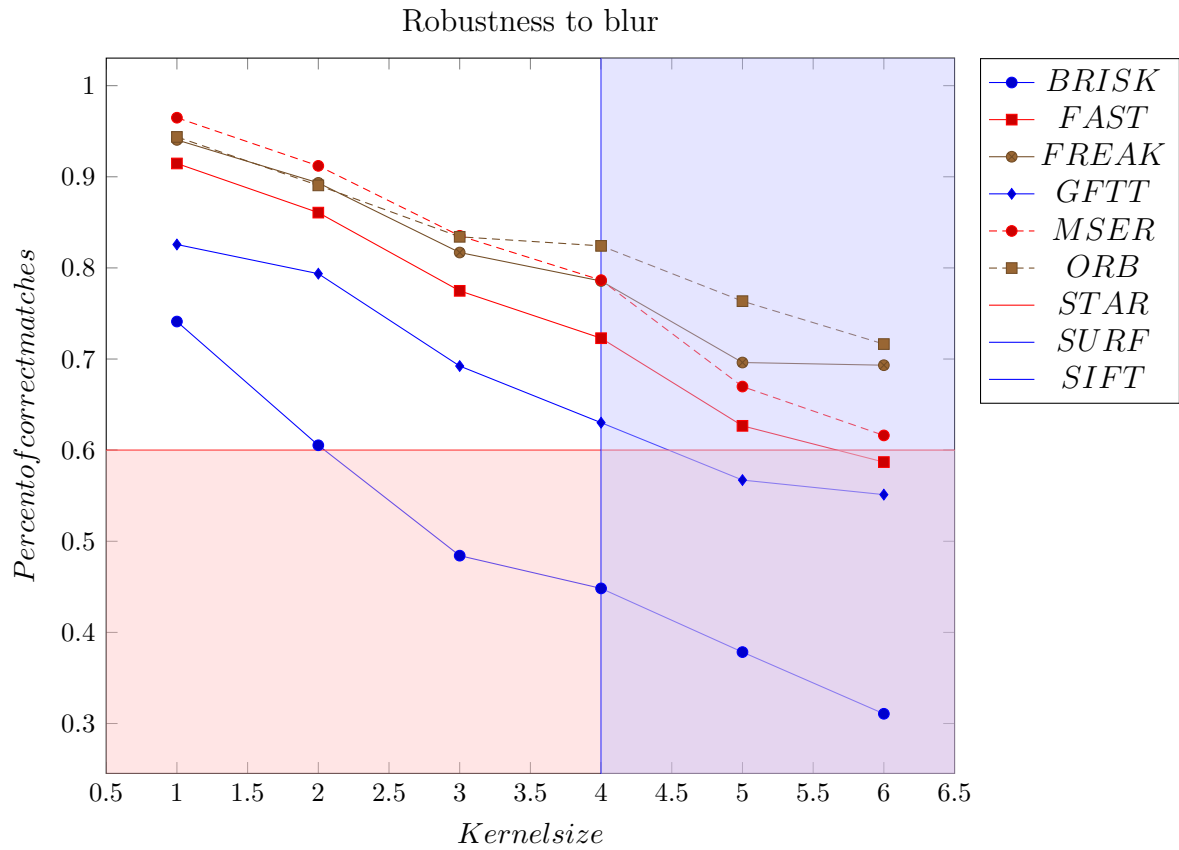


FIGURA 4.3 – Resultado de performance de variação de blur

Como mostrado na imagem 4.4, para os testes de Iluminação estão dentro da janela as técnicas BRISK,FAST,FREAK,GFTT,MSER,ORB,STAR,SURF,SIFT

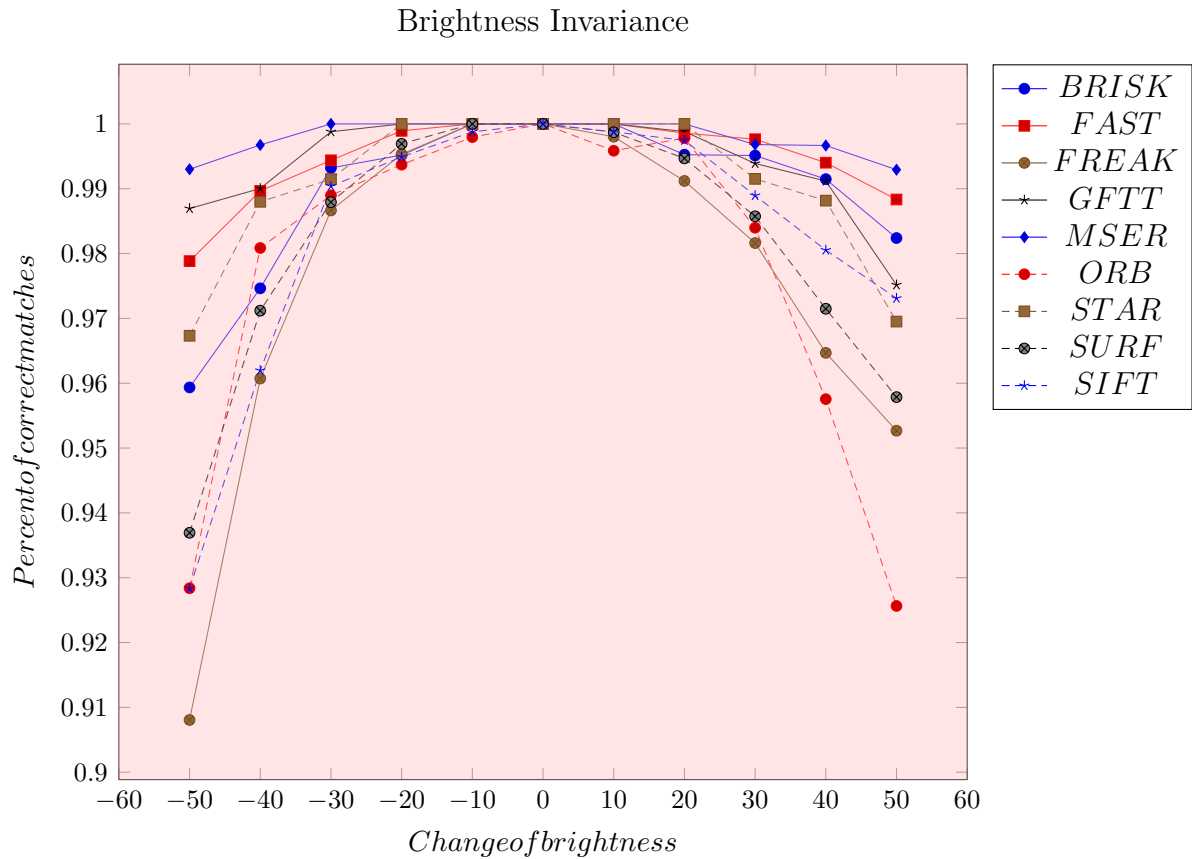


FIGURA 4.4 – Resultado de performance de variação de blur

## 4.2 Análise de Tempo

A análise de tempo, ao contrário da análise de taxa de acertos, é excludente, pois como descrito em (TANG *et al.*, 1993), os usuários não toleram aplicações com menos do que 5fps.

Como mostrado na imagem 4.5, para os testes de Escala estão dentro da janela, as técnicas BRISK,GFTT,ORB,STAR



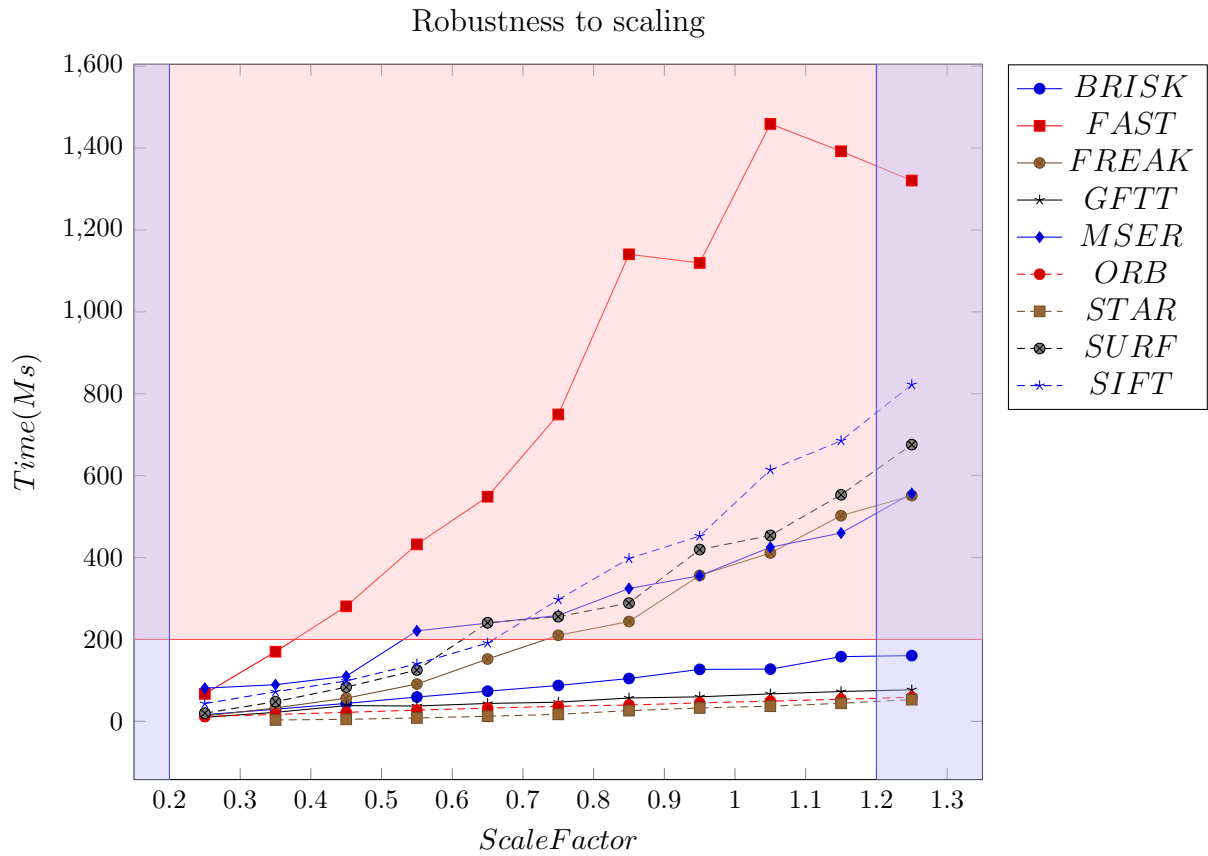


FIGURA 4.5 – Análise de tempo de variação de escala

Como mostrado na imagem 4.5, para os testes de Rotação estão dentro da janela, as técnicas BRISK,GFTT,ORB,STAR

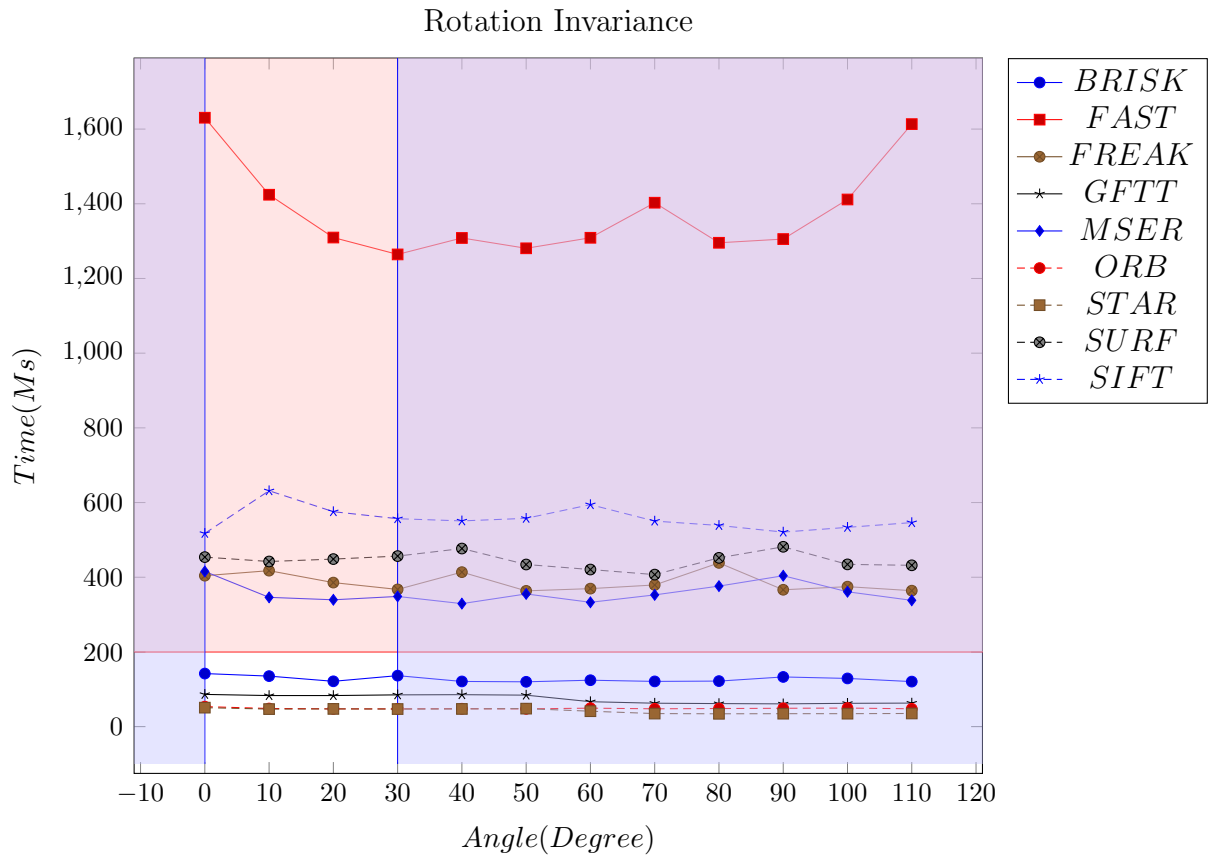


FIGURA 4.6 – Análise de tempo de variação de rotação

Como mostrado na imagem 4.7, para os testes de Blur estão dentro da janela, as técnicas BRISK,GFTT,ORB,STAR

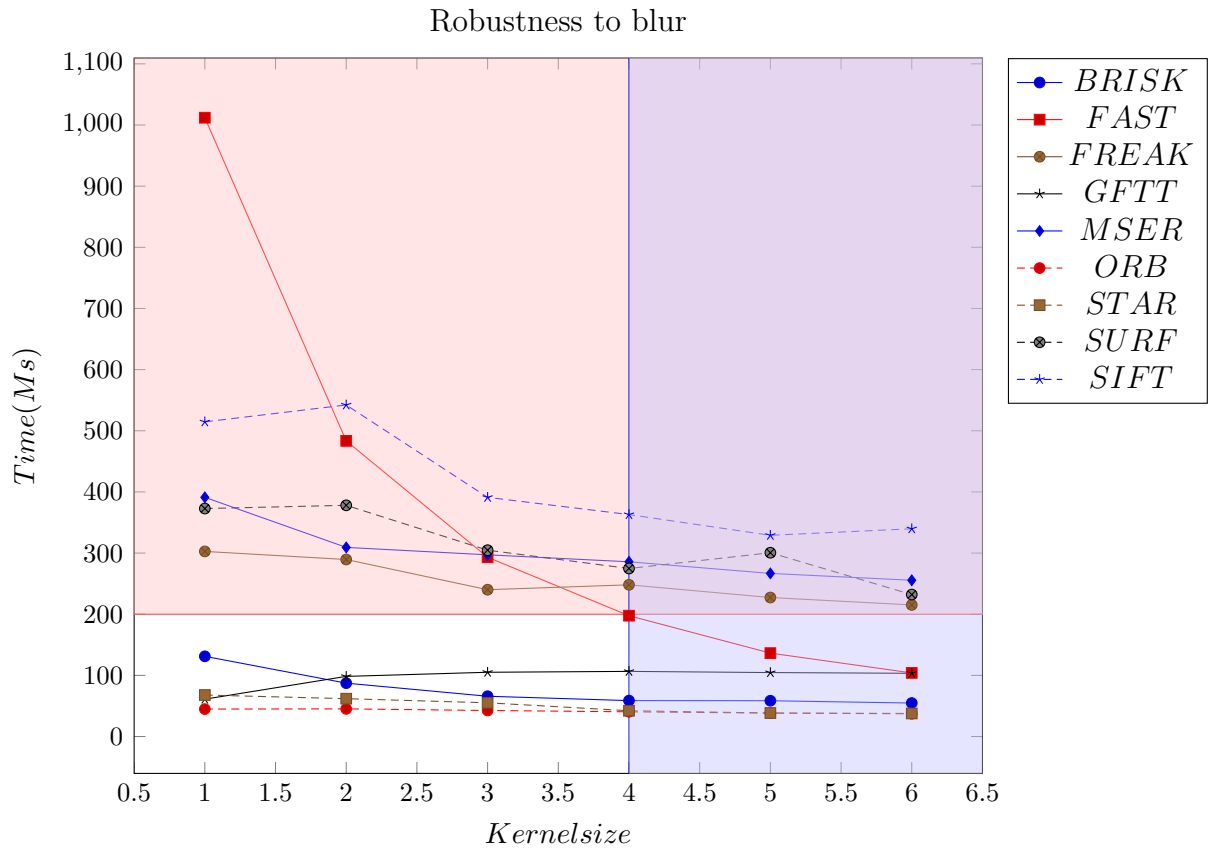


FIGURA 4.7 – Análise de tempo de variação de blur

Como mostrado na imagem 4.8, para os testes de Iluminação estão dentro da janela, as técnicas BRISK, GFTT, ORB, STAR

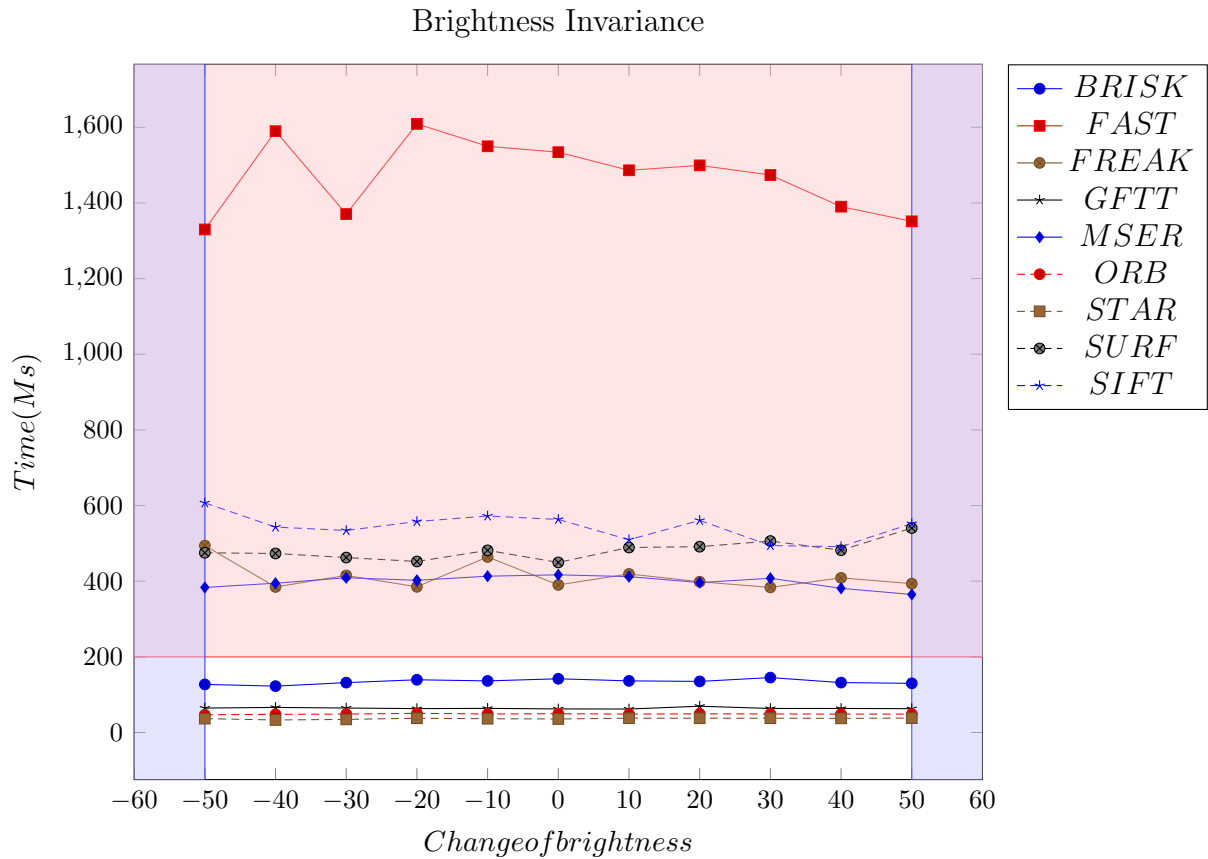


FIGURA 4.8 – Análise de tempo de variação de iluminação

### 4.2.1 Seleção da Técnica

As restrições das análises feitas como mostrado na tabela 4.1, nos levam a escolher GFTT e ORB por serem os únicos que garantem resultados dentro da janela esperada para todos os testes.

TABELA 4.1 – Decisão de técnica à utilizar

	BRISK	FAST	FREAK	GFTT	MSER	ORB	STAR	SURF	SIFT
Precisão Escala			X	X		X			
Precisão Rotação	X	X	X	X	X	X	X	X	X
Precisão Blur		X	X	X	X	X	X	X	X
Precisão Iluminação	X	X	X	X	X	X	X	X	X
Tempo Escala	X			X		X	X		
Tempo Rotação	X			X		X	X		
Tempo Blur	X			X		X	X		
Tempo Iluminação	X			X		X	X		

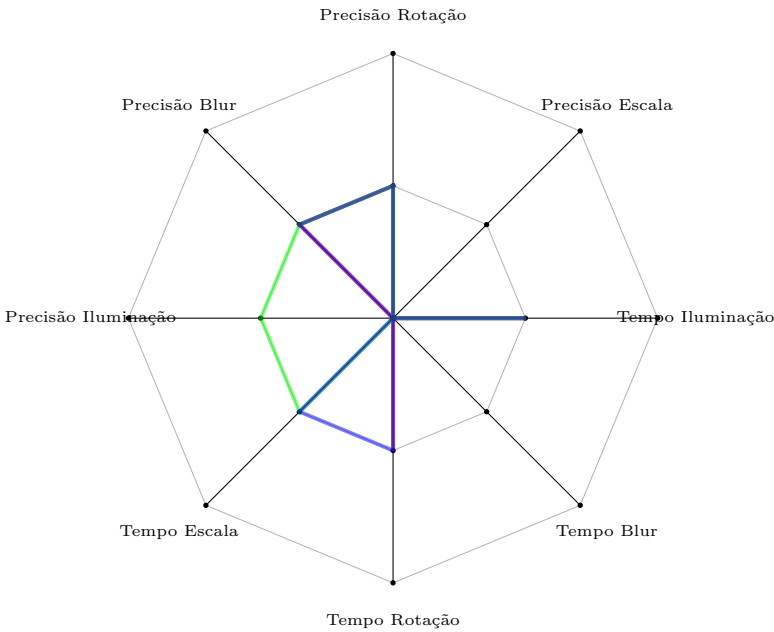


FIGURA 4.9 – Spiderweb Diagram (8 Dimensions, 2-Notch Scale, 3 Samples)

## 5 Conclusão

Este trabalho apresentou uma metodologia de comparação de algoritmos de reconhecimento de padrões, seguindo a mesma abordagem que apresentada em (COMPARISON..., 2014a), (COMPARISON..., 2014b) e (FEATURE..., 2014), tendo como caso de uso o contexto aeronáutico, mais especificamente no reconhecimento de peças do interior da aeronave, objeto de manutenção. Os conceitos básicos de Realidade Aumentada e dos algoritmos utilizados foram apresentados para fornecer meios ao conhecimento dessa tecnologia, que são fundamentais para compreender melhor a proposta do trabalho. Com o objetivo de traçar uma metodologia de decisão imparcial foram descritos algumas restrições impostas à análise que posteriormente foram aplicados nos resultados das simulações. Foram feitos testes emulando situações de variação ambiental pois fazer os testes em ambientes e situações reais demandaria um custo muito elevado, além dos resultados serem semelhantes. Dessa forma concluiu-se que GFTT Seção 2.8.4 e ORB Seção 2.8.6 são os mais favoráveis para o contexto proposto no trabalho.

### 5.1 Atendimento dos objetivos

#### **Avaliar as algoritmos clássicos de reconhecimento**

O trabalho apresentou os algoritmos clássicos na Seção 2.8, descrevendo seu funcionamento e realizando testes de desempenho com os mesmos.

#### **Aplicar os algoritmos clássicos à situações reais**

Os algoritmos foram implementados em um protótipo desenvolvido em C++ e OpenCV, configurado para rodar simulações dos algoritmos BRISK, FAST, FREAK, GFTT, MSER, ORB, STAR, SURF, SIFT, gerando ao final de cada simulação resultados de precisão de

reconhecimento e de tempo gasto.

### **Selecionar algoritmo mais adequado para o contexto**

Foi feita uma análise comparativa entre os algoritmos com as mesmas imagens de teste e configurando as técnicas com suas configurações padrão, para que a única variável no momento fosse a técnica em questão. Levantadas restrições através de parâmetros empíricos definidos por análise de engenharia, com os valores mais prováveis no contexto. Para a seleção das técnicas mais adequadas, foram utilizadas as restrições e criado nos gráficos, janelas de decisão, gerando uma matriz de decisão com os resultados de todos os testes, tanto de qualidade de reconhecimento quanto de tempo de execução.

## **5.2 Proposta de Trabalhos Futuros**

A utilização de Realidade Aumentada no campo da manutenção pode trazer muitos ganhos no que tange à usabilidade levando ao usuário uma quantidade de informações que da maneira tradicional por meio inspeção e consulta em manuais seria inviável. Este trabalho teve como foco o reconhecimento de padrões em um cenário aeronáutico específico, como próximos passos temos:

- Adequar a aplicação para dispositivos móveis como tablets, celulares ou mesmo dispositivos HMD de forma a dar mais flexibilidade ao condutor da manutenção;
- Realizar o casamento de padrões com vídeos e imagens tempo real utilizando as técnicas identificadas, otimizando a aplicação para se tornar o mais tempo real e aceitável possível;
- Adaptar a aplicação para utilizar processamento paralelo e processamento em GPU visto os algoritmos serem recursivos e localmente independentes.
- Analisar por meio de testes em campo com possíveis usuários para abstrair parâmetros de usabilidade como por exemplo determinar que tipo de informações seriam úteis ao usuário ou mesmo que tipo de configuração de dispositivo seria o mais adequado para uma aplicação desse porte.

# Referências Bibliográficas

AZUMA, R. T. A survey of augmented reality. **Hughes Research Laboratories**, aug 1997.

BAY TINNE TUYTELAARS, L. V. G. H. Surf : Speeded up robust features. **ETH Zurich**, 2008.

BRADSKI, A. K. G. **Learning OpenCv Computer vision with the Opencv Library**. [S.l.]: O'NEILL, 2008.

COMPARISON of the OpenCV feature detection algorithms. 2014. Available at <http://computer-vision-talks.com/articles/2011-01-04-comparison-of-the-opencv-feature-detection-algorithms/> (2015/19/01).

COMPARISON of the OpenCV feature detection algorithms - II. 2014. Available at <http://computer-vision-talks.com/articles/2011-07-13-comparison-of-the-opencv-feature-detection-algorithms/> (2015/19/01).

E., H. J. D.; J.-M., F. Comparative evaluation of binary features. **Computer Vision&ECCV**, Springer, 2012.

FEATURE descriptor comparison report. 2014. Available at <http://computer-vision-talks.com/articles/2011-08-19-feature-descriptor-comparison-report/> (2015/19/01).

GIL, A. C. **Como elaborar projetos de pesquisa**. [S.l.]: Atlas, 2010.

KIRNER C.; SISCOOTTO, R. **Realidade Virtual e Aumentada: Conceitos, Projeto e Aplicações**. [S.l.] : EditoraSBC, 2007.

LEUTENEGGER, S.; CHLI, M.; SIEGWART, Y. Brisk: Binary robust invariant scalable keypoints. In: **In Computer Vision (ICCV), 2011 IEEE International Conference on**. [S.l.: s.n.], 2011. p. 2548–2555.

LOWE, D. G. Distinctive image features from scale-invariant keypoints. **Computer Science Department University of British Columbia Vancouver, B.C., Canada**, 2004.

M., E. gayar; H., S. A comparative study of image low level feature extraction algorithms. **of image low level feature extraction algorithms, Egyptian Informatics Journal**, 2013.



MATAS J., e. a. Robust wide baseline stereo from maximally stable extremal regions. **British Machine Vision Conference**, 2002.

MIKOLAJCZYK, C. S. K. A performance evaluation of local descriptors. feb 2005.

OZKAN, S. I. **A Comparative Evaluation of Well-known Feature Detectors and Descriptors**. [S.l.]: International Journal of Applied Mathematics, Electronics and Computers, 2014.

PRODANOV CLEBER CRISTIANO; FREITAS, E. C. **Metodologia do Trabalho Científico: Metodos e Tecnicas da Pesquisa e do Trabalho Academico**. [S.l.]: Universidade FEEVALE, 2013.

ROSTEN, E.; DRUMMOND, T. Fusing points and lines for high performance tracking. **IEEE International Conference on Computer Vision**, vol. 2, 2005.

ROSTEN, E.; DRUMMOND, T. Machine learning for high-speed corner detection. In: **In European Conference on Computer Vision**. [S.l.: s.n.], 2006. p. 430–443.

ROSTEN, G. R. E.; DRUMMOND, T. Real-time video annotations for augmented reality. **International Symposium on Visual Computing**, 2005.

RUBLEE, E. Orb: an efficient alternative to sift or surf. 2011.

SCHAEFFER, C. **A Comparison of Keypoint Descriptors in the Context of Pedestrian Detection: FREAK vs. SURF vs. BRISK**. [S.l.]: Stanford University CS Department, 2013.

SCHMIDT, e. a. A. The comparison of point feature detectors and descriptors in the context of robot navigation. **Workshop on Perception for Mobile Robots Autonomy**, 2012.

SILVA, R.; OLIVEIRA, J. C.; GIRALDI, G. A. **Introduction to augmented reality**. [S.l.], 2003.

TANG, J. C.; ISAACS, E. A.; INC, S.; TANG, J. C.; ISAACS, E. Why do users like video. In: **Computer Supported Cooperative Work**. [S.l.]: Kluwer Academic Publisher, 1993. p. 163–196.

TUYTELAARS, K. M. T. Local invariant feature detectors: A survey. 2007.

UNDERSTANDING Features. 2014. Available at [http://docs.opencv.org/trunk/doc/py\\_tutorials/py\\_feature2d/py\\_features\\_meaning/py\\_features\\_meaning.html](http://docs.opencv.org/trunk/doc/py_tutorials/py_feature2d/py_features_meaning/py_features_meaning.html) (2015/02/04).

WELCH, R. B. Perceptual modification: Adapting to altered sensory environments. **Academic Press**, 1978.

## FOLHA DE REGISTRO DO DOCUMENTO

1. CLASSIFICAÇÃO/TIPO TM	2. DATA 24 de dezembro de 1969	3. DOCUMENTO Nº CTA/ITA - IEC/TM-002/1969	4. Nº DE PÁGINAS 65
5. TÍTULO E SUBTÍTULO: Avaliação de técnicas de reconhecimento de padrões em ambientes aeronauticos			
6. AUTOR(ES): <b>Bruno Duarte Corrêa</b>			
7. INSTITUIÇÃO(ÕES)/ÓRGÃO(S) INTERNO(S)/DIVISÃO(ÕES): Divisão de Ciência da Computação – ITA/IEC			
8. PALAVRAS-CHAVE SUGERIDAS PELO AUTOR: Teses; Estilos; Italus			
9. PALAVRAS-CHAVE RESULTANTES DE INDEXAÇÃO: Teses e Dissertações; Estilos; Usuários			
10. APRESENTAÇÃO: (X) Nacional ( ) Internacional Curso Programa Area Orientador Data..			
11. RESUMO: O reconhecimento de objetos em uma cena para posterior uso em realidade aumentada depende de diversas variáveis, causando a necessidade do uso de diversas técnicas específicas para cada cenário, sendo portanto, um estudo de fronteiras para a melhor escolha do algoritmo de registro de acordo com a aplicação em questão de grande valia para o meio acadêmico. Esta tese se propõe a pesquisar, categorizar e traçar fronteiras das técnicas conhecidas tendo como caso de uso manutenção de aeronaves feita dentro de centros fechados, utilizando as técnicas BRISK,FAST,FREAK,GFTT,MSER, ORB,STAR,SURF,SIFT em uma análise aplicada com imagens reais de janelas de inspeção do Embraer ERJ-190 para reconhecimento de objetos em posteriores aplicações de manutenção. Comparando todas as técnicas quanto à cadência e a precisão de reconhecimento de características é possível selecionar GFTT e ORB como técnicas mais apropriadas ao contexto, por terem seus resultados de variação de rotação, escala, briho e blur dentro de uma faixa esperada.			
12. GRAU DE SIGILO: ( ) OSTENSIVO ( ) RESERVADO ( ) CONFIDENCIAL (X) SECRETO			