

UNIVERSIDADE FEDERAL DE OURO PRETO
INSTITUTO DE CIÊNCIAS EXATAS E BIOLÓGICAS
DEPARTAMENTO DE COMPUTAÇÃO

JOELINGTON DE ALMEIDA BERNARDO
Orientador: Saul Emanuel Delabrida Silva
Coorientador: Bruno Hideki Adachi

**ESTRATÉGIAS DE SEGMENTAÇÃO ESTRUTURAL PARA
MELHORIA NO RECONHECIMENTO DE PÁGINAS EM CONTEXTOS
DE BAIXO CONTRASTE VISUAL**

Ouro Preto, MG
2025

UNIVERSIDADE FEDERAL DE OURO PRETO
INSTITUTO DE CIÊNCIAS EXATAS E BIOLÓGICAS
DEPARTAMENTO DE COMPUTAÇÃO

JOELINGTON DE ALMEIDA BERNARDO

**ESTRATÉGIAS DE SEGMENTAÇÃO ESTRUTURAL PARA MELHORIA NO
RECONHECIMENTO DE PÁGINAS EM CONTEXTOS DE BAIXO CONTRASTE
VISUAL**

Monografia apresentada ao Curso de Ciência da Computação da Universidade Federal de Ouro Preto como parte dos requisitos necessários para a obtenção do grau de Bacharel em Ciência da Computação.

Orientador: Saul Emanuel Delabrida Silva

Coorientador: Bruno Hideki Adachi

Ouro Preto, MG
2025



FOLHA DE APROVAÇÃO

Joelington de Almeida Bernardo

Estratégias de segmentação estrutural para melhoria no reconhecimento de páginas em contextos de baixo contraste visual

Monografia apresentada ao Curso de Ciência da Computação da Universidade Federal de Ouro Preto como requisito parcial para obtenção do título de Bacharel em Ciência da Computação

Aprovada em 31 de março de 2025

Membros da banca

Doutor - Saul Emanuel Delabrida Silva - Orientador - Universidade Federal de Ouro Preto
Bacharel - Bruno Hideki Adachi - Co-orientador - Universidade Federal de Ouro Preto
Doutora - Andrea Gomes Campos - Universidade Federal de Ouro Preto
Doutor - Guillermo Cámara Chávez - Universidade Federal de Ouro Preto

Saul Emanuel Delabrida Silva, orientador do trabalho, aprovou a versão final e autorizou seu depósito na Biblioteca Digital de Trabalhos de Conclusão de Curso da UFOP em 11/04/2025



Documento assinado eletronicamente por **Saul Emanuel Delabrida Silva, PROFESSOR DE MAGISTERIO SUPERIOR**, em 11/04/2025, às 02:23, conforme horário oficial de Brasília, com fundamento no art. 6º, § 1º, do [Decreto nº 8.539, de 8 de outubro de 2015](#).



A autenticidade deste documento pode ser conferida no site http://sei.ufop.br/sei/controlador_externo.php?acao=documento_conferir&id_orgao_acesso_externo=0, informando o código verificador **0893880** e o código CRC **6E0219E9**.

Dedico este trabalho a minha avó, minha grande inspiração para atuar na área científica.

Agradecimentos

Quero expressar meu agradecimento à UFOP pelo ensino de qualidade, ao laboratório XR4GOOD pelas oportunidades de participação em pesquisa científica e aos meus orientadores Saul Delabrida e Bruno Hideki pela orientação neste projeto.

Resumo

Este trabalho propõe estratégias de segmentação estrutural para aprimorar o reconhecimento de páginas de livros em contextos de baixo contraste visual, visando integrar tecnologias como **Realidade Aumentada (RA)** sem a necessidade de modificações físicas nos materiais. A abordagem utiliza técnicas de pré-processamento de imagem, como limiarização global, adaptativa e método Canny, combinadas a operações morfológicas (erosão, dilatação), para gerar impressões digitais únicas das páginas. Foram implementados e avaliados algoritmos de detecção de características (**ORB**, **SIFT** e **BRISK**) em um banco de dados Segmonto com 4.425 imagens selecionadas de páginas de livros com layouts diferentes, testando eficiência, precisão e robustez a ruídos. Os resultados demonstraram ganhos significativos: o pré-processamento reduziu o tempo de busca em até 96% (**SIFT** com limiarização adaptativa) e teve ganho de espaço de até 80%, dependendo do método. O **ORB** destacou-se em velocidade, enquanto o **SIFT** apresentou maior precisão em cenários complexos. Conclui-se que a segmentação estrutural e o pré-processamento são fundamentais para viabilizar sistemas de identificação em tempo real, especialmente em materiais educacionais. Trabalhos futuros podem explorar ajustes de parâmetros, integração com modelos de predição e testes com outros algoritmos para ampliar a robustez em ambientes adversos.

Palavras-chave: Identificação de páginas, Realidade Aumentada, Pré-processamento de imagem, **ORB**, **SIFT** e **BRISK**.

Abstract

This work proposes structural segmentation strategies to improve book page recognition in low visual contrast contexts, aiming to integrate technologies such as Augmented Reality (AR) without the need for physical modifications to the materials. The approach uses image preprocessing techniques, such as global and adaptive thresholding and the Canny method, combined with morphological operations (erosion, dilation), to generate unique fingerprints of the pages. Feature detection algorithms (ORB, SIFT and BRISK) were implemented and evaluated on a Segmonto database with 4,425 selected images of book pages with different layouts, testing efficiency, accuracy and robustness to noise. The results demonstrated significant gains: preprocessing reduced search time by up to 96% (SIFT with adaptive thresholding) and had space gains of up to 80%, depending on the method. ORB stood out in speed, while SIFT showed greater accuracy in complex scenarios. It is concluded that structural segmentation and preprocessing are essential to enable real-time identification systems, especially in educational materials. Future work can explore parameter adjustments, integration with prediction models and tests with other algorithms to increase robustness in adverse environments.

Keywords: Page identification, Augmented Reality, Image preprocessing, ORB, SIFT and BRISK

Lista de Ilustrações

Figura 1.1 – Patente da “ <i>Statistical Machine</i> ”, de 1933, que utilizava a captação da luz refletida sobre os caracteres impressos para reconhecimento.	1
Figura 1.2 – Gráfico de comparação de quantidade de imagens e suas pontuações apresentadas pelo <i>Vuforia</i>	2
Figura 1.3 – Exemplo de Aplicação RA com <i>Vuforia</i>	3
Figura 1.4 – Exemplo de captura e coleta de dados usando OCR.	4
Figura 1.5 – Demonstração da imagem original de uma pagina de um texto e o mesmo analisado pelo <i>Vuforia</i> onde apresenta os descritores encontrados que serão utilizados, sua avaliação ficou como 0 estrelas na capacidade de reconhecimento.	5
Figura 1.6 – Comparação da quantidade de descritores apresentados pelo <i>Vuforia</i> após análise.	5
Figura 1.7 – Representação do processo de pré-processamento, comparado ao funcionamento de uma impressão digital.	7
Figura 2.1 – Exemplo de valores diferentes de limiar na mesma imagem.	9
Figura 2.2 – Exemplo de valores diferentes de limiar na mesma imagem.	10
Figura 2.3 – Representação de uma imagem e seu histograma com a seleção do limiar.	11
Figura 2.4 – Exemplo que demonstra como a rotação da imagem não compromete as características extraídas pelo SIFT, evidenciando que os pontos descritores permanecem interligados e consistentes.	13
Figura 2.5 – Exemplos de dilatação e erosão com kernel quadrático 3x3.	16
Figura 2.6 – Abertura e fechamento de uma estrutura em uma imagem binária por um elemento estruturante quadrado.	16
Figura 2.7 – Exemplos representativos de fotografias no conjunto de dados <i>B-MOD</i>	18
Figura 3.1 – Diagrama da metodologia aplicada.	21
Figura 3.2 – Etapas do pré-processamento das imagens.	23
Figura 3.3 – Sequência de imagens demonstrando as etapas do processamento.	24
Figura 3.4 – Resultados finais dos pré-processamentos.	24
Figura 3.5 – Figuras extraídas do banco de dados Segmonto	25
Figura 4.1 – Resultados dos métodos com limiarização global.	27
Figura 4.2 – Resultados dos métodos com limiarização adaptativa.	29
Figura 4.3 – Resultados dos métodos utilizando o método Canny.	30
Figura 4.4 – Médias das métricas avaliadas para os três métodos.	32
Figura 4.5 – Imagens com 25% dos descritores utilizando ORB	36
Figura 4.6 – Imagens com 25% dos descritores utilizando SIFT	36
Figura 4.7 – Imagens com 25% dos descritores utilizando BRISK	37

Lista de Tabelas

Tabela 2.1 – Comparação dos Algoritmos de Detecção e Descrição	14
Tabela 4.1 – Análise comparativa dos métodos com limiarização global: tempo de busca, pontuação de confiança e quantidade de descritores.	27
Tabela 4.2 – Análise comparativa dos métodos com limiarização adaptativa: tempo de busca, pontuação de confiança e quantidade de descritores.	29
Tabela 4.3 – Análise comparativa dos métodos com o método Canny: tempo de busca, confiança na pontuação e quantidade de descritores.	31
Tabela 4.4 – Erros de busca em imagens ruidosas: comparação entre diferentes métodos de segmentação	34

Lista de Abreviaturas e Siglas

BRISK *Binary Robust Invariant Scalable Keypoints*. v–vii, x, 13, 14, 19, 20, 23, 26, 28, 30–33, 35, 37, 38

OCR Reconhecimento Óptico de Caracteres. vii, x, 3, 4, 18

OpenCV *Open Source Computer Vision Library*. x, 11, 12, 20

ORB *Oriented FAST and Rotated BRIEF*. v–vii, x, 13, 14, 19, 20, 22, 23, 26, 28, 29, 31, 33, 35, 36, 38

QR *Quick Response Code*. 2, 18

RA Realidade Aumentada. v, vii, 1, 3, 4, 6, 12, 17–19, 22, 39

SIFT *Scale-Invariant Feature Transform*. v–vii, x, 13, 14, 19, 20, 23, 26, 28, 30–32, 35, 36, 38

Sumário

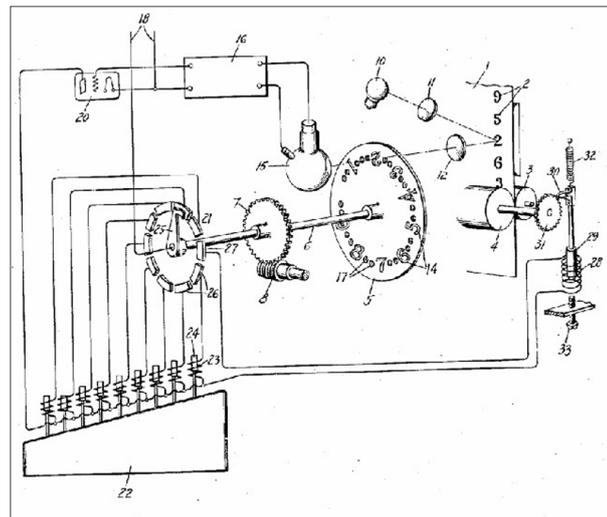
1	Introdução	1
1.1	Justificativa	4
1.2	Objetivos do Trabalho	6
2	Revisão Bibliográfica	8
2.1	Fundamentação Teórica	8
2.1.1	Limiarização ou <i>Thresholding</i>	8
2.1.2	Método de Otsu	8
2.1.3	Detector de bordas Canny	9
2.1.4	Filtro Gaussiano	10
2.1.5	Detecção de Contornos com <i>findContours</i> do <i>Open Source Computer Vision Library</i> (OpenCV)	12
2.1.6	Algoritmos de Detecção e Descrição de Características	12
2.1.6.1	<i>Scale-Invariant Feature Transform</i> (SIFT)	13
2.1.6.2	<i>Binary Robust Invariant Scalable Keypoints</i> (BRISK)	13
2.1.6.3	<i>Oriented FAST and Rotated BRIEF</i> (ORB)	14
2.1.6.4	Tabela Comparativa dos Métodos	14
2.1.7	Operações Morfológicas: Erosão, Dilatação, Abertura e Fechamento	15
2.2	Trabalhos Relacionados	17
2.2.1	Abordagens Estruturais e Geométricas em Análise Documental	17
2.2.2	Avanços em Modelagem Hierárquica e Aprendizado Profundo	17
2.2.3	Desafios Práticos em Reconhecimento Óptico de Caracteres (OCR) e Pré-processamento	18
2.2.4	Síntese Teórica e Direcionamento Metodológico	19
3	Desenvolvimento	20
3.1	Metodologia	20
3.1.1	Revisão Bibliográfica	20
3.1.2	Definição de Requisitos	20
3.1.3	Desenvolvimento e Implementação	20
3.2	Pré-processamentos	22
3.3	Ambiente dos Testes	23
3.3.1	Base de Dados	24
4	Resultados e Discussão	26
4.1	Procedimentos para Coleta e Análise dos Resultados	26
4.2	Resultados com Diferentes Métodos de Pré-processamento	26
4.2.1	Limiarização Global	27
4.2.1.1	Análise por Algoritmo	28

4.2.1.1.1	Algoritmo ORB:	28
4.2.1.1.2	Algoritmo SIFT:	28
4.2.1.1.3	Algoritmo BRISK:	28
4.2.2	Limiarização Adaptativa	28
4.2.2.1	Análise por Algoritmo	29
4.2.2.1.1	Algoritmo ORB:	29
4.2.2.1.2	Algoritmo SIFT:	30
4.2.2.1.3	Algoritmo BRISK:	30
4.2.3	Método Canny	30
4.2.3.1	Análise por Algoritmo	31
4.2.3.1.1	Algoritmo ORB:	31
4.2.3.1.2	Algoritmo SIFT:	31
4.2.3.1.3	Algoritmo BRISK:	31
4.3	Comparação Geral entre Testes com e sem Pré-processamento	31
4.3.1	Vantagens do Pré-processamento	32
4.3.2	Limitações e Considerações Críticas	32
4.4	Considerações Finais	33
4.5	Testes com Aplicação de Ruído nas Imagens	33
4.6	Resultados de Erro na Busca	34
4.6.1	Análise por Método	34
4.6.2	Análise Comparativa Geral	35
4.6.3	Considerações Finais sobre os Erros de Busca	35
4.7	Discussão dos Resultados	35
5	Conclusão	38
5.1	Trabalhos Futuros	38
	Referências	40

1 Introdução

Coletar e tratar dados de documentos com diferentes métodos é uma prática que já existia antes mesmo da criação do computador. Máquinas como a *Statistical Machine* apresentada na Figura 1.1 que foi desenvolvida por P. W. Handel em 1933 (Handel, 1933), realizavam essa coleta de forma manual, utilizando princípios ópticos para identificar caracteres impressos. Essa barreira tem sido gradualmente quebrada devido à constante evolução das tecnologias. Competições como a *Document Image Binarization* (UFPE, 2023) foram criadas para fomentar essa área específica do processamento de imagens.

Figura 1.1 – Patente da “*Statistical Machine*”, de 1933, que utilizava a captação da luz refletida sobre os caracteres impressos para reconhecimento.



Fonte: (Handel, 1933).

A **Realidade Aumentada (RA)** é uma tecnologia que vem transformando a forma como interagimos com o mundo, ao integrar elementos digitais ao ambiente físico e proporcionar experiências interativas e imersivas. Por meio de dispositivos como smartphones, tablets e óculos especializados, a **RA** possibilita a sobreposição de informações, imagens e animações sobre o ambiente real, enriquecendo a percepção dos usuários. Essa tecnologia tem ganhado destaque em diversas áreas – como educação, entretenimento, marketing e medicina – ao oferecer novas modalidades de interação e aprendizagem.

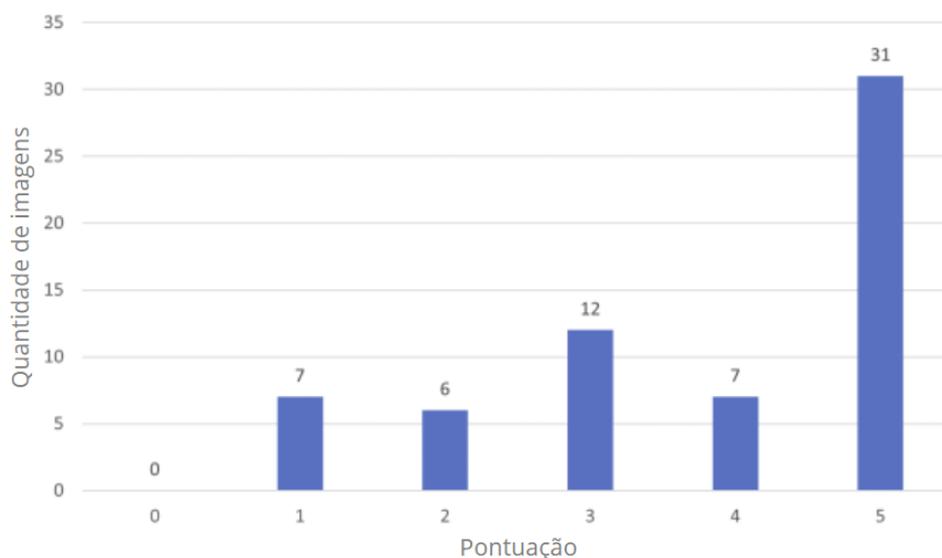
Entretanto, um dos maiores desafios na aplicação de tecnologias de **RA** em livros refere-se à precisão no reconhecimento de páginas que apresentam características visuais escassas ou variações limitadas de design e cor, o que dificulta a identificação correta de uma página específica. Conforme verificado no estudo “*Augmented Reality in Books: An Evaluation of Alan Turing Book*” (Adachi *et al.*, 2023), que avaliou 63 páginas do livro “Alan Turing: suas máquinas

e seus segredos” utilizando o *Vuforia*, constatou-se que páginas sem ilustrações apresentam desempenho significativamente inferior na detecção e rastreamento.

A Figura 1.2 ilustra, a distribuição das pontuações atribuídos pelo *Vuforia*, numa escala de 0 a 5 pontos. O gráfico evidencia que aproximadamente 40% das páginas – majoritariamente aquelas sem elementos ilustrativos – foram avaliadas com pontuações inferiores a 3 estrelas, sinalizando que estes elementos visuais são determinantes para a eficácia do reconhecimento. Por outro lado, as páginas com escores de 4 e 5 estrelas correspondem a 60% do total, demonstrando uma capacidade satisfatória de detecção e rastreamento para a maioria das imagens. Ademais, o estudo ressalta que a inserção de *Quick Response Code (QR)* nas páginas pode funcionar como uma estratégia eficaz para maximizar a identificação, aumentando o número de pontos detectados e, conseqüentemente, o contraste dos elementos, o que aprimora a performance do algoritmo do *Vuforia*.

Dessa forma, a análise gráfica apresentada na Figura 1.2 não só confirma as limitações enfrentadas na detecção de páginas com poucas características visuais, mas também aponta para possíveis soluções tecnológicas que podem ser exploradas em futuras adaptações e pesquisas na área de realidade aumentada aplicada à literatura.

Figura 1.2 – Gráfico de comparação de quantidade de imagens e suas pontuações apresentadas pelo *Vuforia*



Fonte: Adaptado de (Adachi *et al.*, 2023) pelo autor.

Os algoritmos, por sua vez, representam a espinha dorsal dos sistemas computacionais, consistindo em sequências lógicas de passos bem definidos que orientam a resolução de problemas específicos. Tais algoritmos podem ser expressos por meio de fluxogramas, pseudocódigo ou linguagens de programação, e são essenciais para o desenvolvimento de software, otimização de processos e análise de dados. Na área de reconhecimento e classificação de páginas de livros, os algoritmos fundamentam-se, majoritariamente, em técnicas de reconhecimento de padrões

– que buscam identificar características distintivas para agrupar ou categorizar informações – e em processos de segmentação de texto por meio de sistemas de **Reconhecimento Óptico de Caracteres (OCR)**.

A aplicação de tais algoritmos em documentos com escassez de imagens ou elementos de design apresenta desafios que devem ser superados para garantir um reconhecimento eficiente, sem a necessidade de reimpressão ou modificação física dos materiais. Dessa forma, este estudo evidencia a necessidade de aprimorar as técnicas empregadas, considerando as limitações impostas pela ausência de marcadores visuais claros, que comprometem o desempenho das tecnologias de **RA** em ambientes reais.

A abordagem proposta neste trabalho consiste na transformação das páginas dos livros em representações digitais que evidenciam suas estruturas geométricas. Essa estratégia gera uma “impressão digital” que atua como âncora para sistemas de **RA**. Ao identificar e extrair os contornos e limites das páginas, os algoritmos de reconhecimento capturam descritores capazes de estimar parâmetros como rotação, escala e orientação da página no mundo físico. Tais parâmetros são fundamentais para a integração precisa de elementos virtuais – por exemplo, modelos 3D, conforme ilustrado na Figura 1.3 – ao ambiente físico, garantindo a ancoragem consistente dos objetos virtuais em conformidade com o posicionamento da página no mundo real. Assim, a proposta não apenas aprimora o desempenho dos algoritmos em cenários de baixo contraste visual, mas também viabiliza a aplicação da **RA** sem a necessidade de modificar fisicamente o material original.

Figura 1.3 – Exemplo de Aplicação **RA** com *Vuforia*



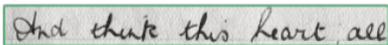
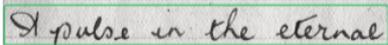
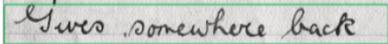
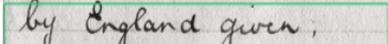
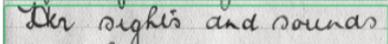
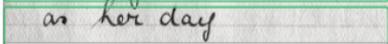
Fonte: (Colégio Santa Joana D’Arc, 2019)

1.1 Justificativa

A partir da criação do computador e da evolução da área de visão computacional, pesquisadores se empenham em aplicar características de percepção que são naturais e intuitivas em seres humanos. Aspectos como detecção de bordas, determinação de profundidade, percepção de cores e identificação de objetos são extremamente úteis quando aplicados no âmbito tecnológico. A reflexão sobre como o computador poderia processar informações do mundo real e entregar dados ou resultados de forma arbitrária, ou automatizada é uma constante.

Quando se trata de detecção e classificação em livros, a análise fundamenta-se na interpretação de textos por meio do OCR. Essa abordagem é considerada a mais lógica, uma vez que reflete a atuação humana, na qual o leitor, ao examinar uma página, determina qual página está visualizando com base em seu conteúdo. Conforme descrito por (Karthick *et al.*, 2019), o OCR é uma tecnologia que possibilita a conversão de texto presente em imagens ou documentos impressos em formato digital e editável, desempenhando papel crucial na extração e identificação de caracteres contidos em mídias digitalizadas, o que permite a manipulação eletrônica do texto. Como ilustrado na Figura 1.4, o processo envolve a análise da imagem, a identificação de padrões e o reconhecimento dos caracteres, resultando em um texto que pode ser pesquisado e armazenado de maneira eficiente. A aplicação do OCR é ampla, abrangendo desde a digitalização de documentos até a automação de processos envolvendo a manipulação de texto em formatos não digitais (Nguyen *et al.*, 2021). Entretanto, essa abordagem privilegia a interpretação do conteúdo textual, sem explorar a análise e segmentação de textos com base em sua estrutura geométrica, a qual pode ser mais adequada em determinados contextos, conforme apresentado inicialmente por (Cattoni *et al.*, 1998).

Figura 1.4 – Exemplo de captura e coleta de dados usando OCR.

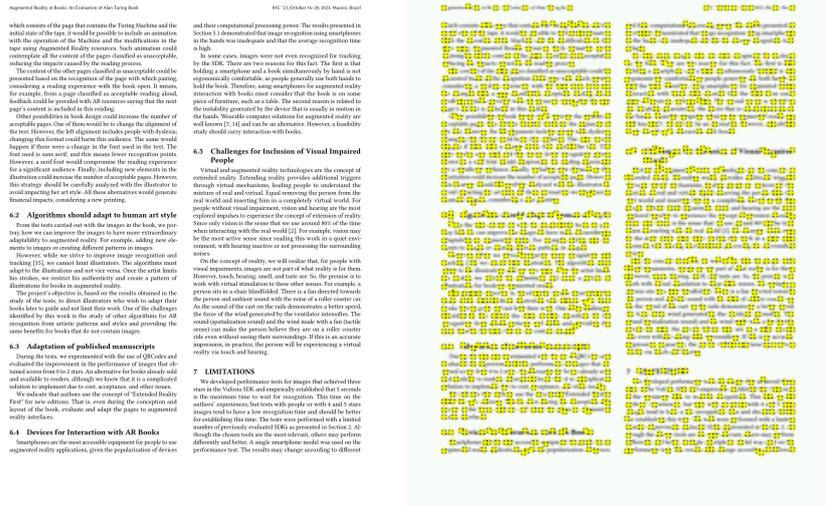
	And thik this heart all	And think this heart all
	A pulse in the elernal	A pulse in the eternal
	Yives somewhere back	Gives somewhere back
	by England given,	by England given,
	Her sighis and sounds	Her sights and sounds
	on her day	on her day

Fonte: (Klippa, 2024)

Os algoritmos de reconhecimento e ancoragem empregados em aplicações de RA encontram-se consolidados na literatura, tendo como fundamento a extração de características visuais – tais como formato, cor e texturas – para a construção de descritores robustos. Entretanto, tais algoritmos demonstram limitações em cenários com baixo contraste de cores, como em páginas predominantemente brancas (conforme apresentado em Figura 1.5), ou na ausência de imagens bem definidas, o que resulta em uma redução significativa no desempenho da detecção e do

rastreamento dos elementos . A Figura 1.6 ilustra, por meio de análise comparativa, a expressiva disparidade na quantidade de descritores gerados pelo sistema *Vuforia* em duas imagens distintas.

Figura 1.5 – Demonstração da imagem original de uma pagina de um texto e o mesmo analisado pelo *Vuforia* onde apresenta os descritores encontrados que serão utilizados, sua avaliação ficou como 0 estrelas na capacidade de reconhecimento.

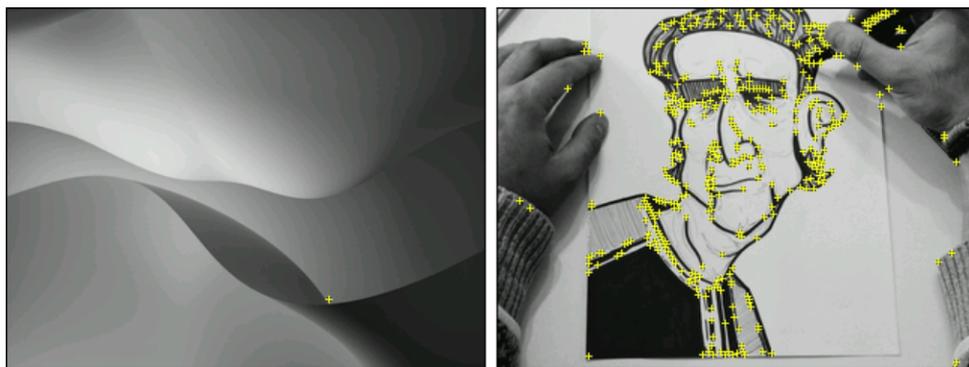


(a) Página original.

(b) Página com seus descritores.

Fonte: A Página original foi retirada do artigo (Adachi *et al.*, 2023) a Página com os descritores foi adaptada pelo autor.

Figura 1.6 – Comparação da quantidade de descritores apresentados pelo *Vuforia* após análise.



Fonte: Adaptado de (Adachi *et al.*, 2023) pelo autor.

Tecnologias amplamente disseminadas, como *Vuforia*, *ARCore* e *ARKit*, recomendam a utilização de imagens com elevada riqueza visual, conforme explicitado em suas documentações oficiais, para viabilizar uma detecção mais precisa e eficaz dos pontos de interesse. Nesse contexto, o processo de ancoragem – que realiza a indexação do conteúdo tridimensional com o ambiente real – assume papel fundamental, pois associa, de forma dinâmica e precisa, os elementos virtuais ao espaço físico, permitindo a interação em tempo real com a realidade aumentada.

Ademais, as aplicações da RA transcendem o entretenimento e encontram significativa relevância em contextos educacionais, onde a sobreposição de informações digitais sobre o meio físico promove uma experiência de aprendizagem interativa e imersiva como ilustrado na Figura 1.3. A proposta deste trabalho, ao adotar uma nova abordagem para o reconhecimento de páginas sem a necessidade de reimpressão ou modificações físicas, visa oferecer uma solução escalável e de baixo custo, contribuindo para a transição gradual para a era digital em ambientes com infraestrutura limitada. Neste sentido, a criação de “impressões digitais” únicas para cada página, possibilitada por uma segmentação estrutural acurada, evidencia o potencial transformador da integração entre tecnologias de processamento de imagem e realidade aumentada fortalecendo as bases para o avanço de sistemas interativos em diversas áreas do conhecimento.

A pertinência deste estudo reside, portanto, na oportunidade de explorar e aprimorar técnicas que permitam a aplicação eficiente da RA em diferentes contextos, com foco na robustez da ancoragem e na otimização da indexação tridimensional. Ao conciliar métodos clássicos de processamento de imagem com algoritmos de detecção modernos, como os empregados por *Vuforia*, este trabalho propõe avanços que não somente potencializam o reconhecimento de páginas, mas também abrem caminho para a implementação de soluções inovadoras em ambientes educacionais e demais setores.

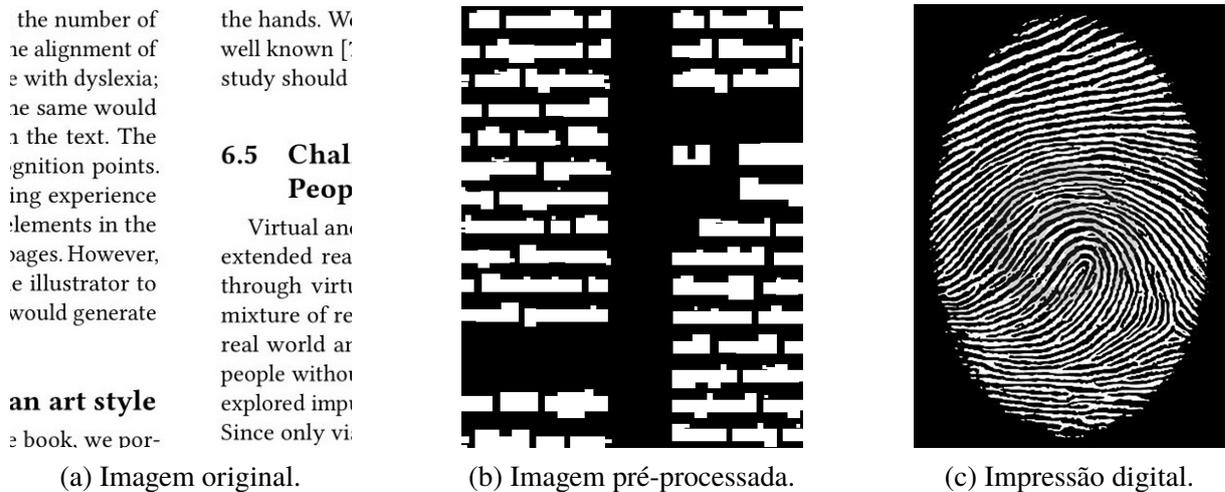
É vislumbrada uma abordagem fundamentada na segmentação estrutural das páginas, para superar os desafios associados à ausência de elementos visuais, o principal objetivo é desenvolver um método que permita a identificação eficiente e precisa das páginas, estabelecendo, após um pre-processamento, identificadores únicos para cada uma delas se assemelhando a digitais como pode ser observado na Figura 1.7. Espera-se que essa inovação contribua de maneira substancial para a superação das barreiras iniciais à adoção de novas tecnologias no contexto da literatura, representando, assim, um avanço significativo nesse cenário acadêmico e tecnológico.

1.2 Objetivos do Trabalho

Como objetivo principal, propõe-se o desenvolvimento de um algoritmo capaz de identificar páginas de livros existentes por meio de técnicas avançadas de processamento de imagens, com ênfase na geração de “impressões digitais” exclusivas para cada página. A proposta busca aumentar a precisão do reconhecimento sem a necessidade de modificar os materiais originais. Espera-se que os resultados obtidos contribuam para a aplicação da RA em contextos educacionais, bem como em outras áreas que envolvam a digitalização e o reconhecimento de documentos.

- Explorar técnicas de pré-processamento que potencializem a detecção de padrões visuais em páginas de livros;
- Integrar algoritmos descritores para a extração eficaz de características relevantes;

Figura 1.7 – Representação do processo de pré-processamento, comparado ao funcionamento de uma impressão digital.



Fonte: A imagem 1.7a foi retirada do artigo (Adachi *et al.*, 2023); a imagem 1.7b foi adaptada pelo autor; e a imagem 1.7c foi extraída do site (AnaWer, 2024).

- Desenvolver uma abordagem robusta e aplicável à identificação de páginas sem alterações nos livros físicos.

No Capítulo 2, será abordada a fundamentação teórica e a revisão bibliográfica, contendo os conceitos necessários para o melhor entendimento do trabalho. No Capítulo 3, será apresentado o processo de concepção da proposta do projeto, bem como as metodologias utilizadas. No Capítulo 4, serão apresentados os resultados obtidos e, por fim, no Capítulo 5, serão apresentadas as conclusões do trabalho, além de sugestões para possíveis trabalhos futuros.

2 Revisão Bibliográfica

Esse capítulo tem como objetivo trazer uma revisão bibliográfica sobre o presente trabalho. Na Seção 2.1, serão abordados assuntos essenciais para a fundamentação deste trabalho, como Operações Morfológicas, Algoritmos de Detecção e Descrição de Características e Técnicas de limiarização. Em seguida, na Seção 2.2, serão apresentados os trabalhos relacionados, com o intuito de fornecer uma visão sobre a evolução dessa área.

2.1 Fundamentação Teórica

Serão abordados diversos conceitos técnicos essenciais para o desenvolvimento do algoritmo proposto. Abaixo são apresentados e explicados as principais técnicas e algoritmos utilizados.

2.1.1 Limiarização ou *Thresholding*

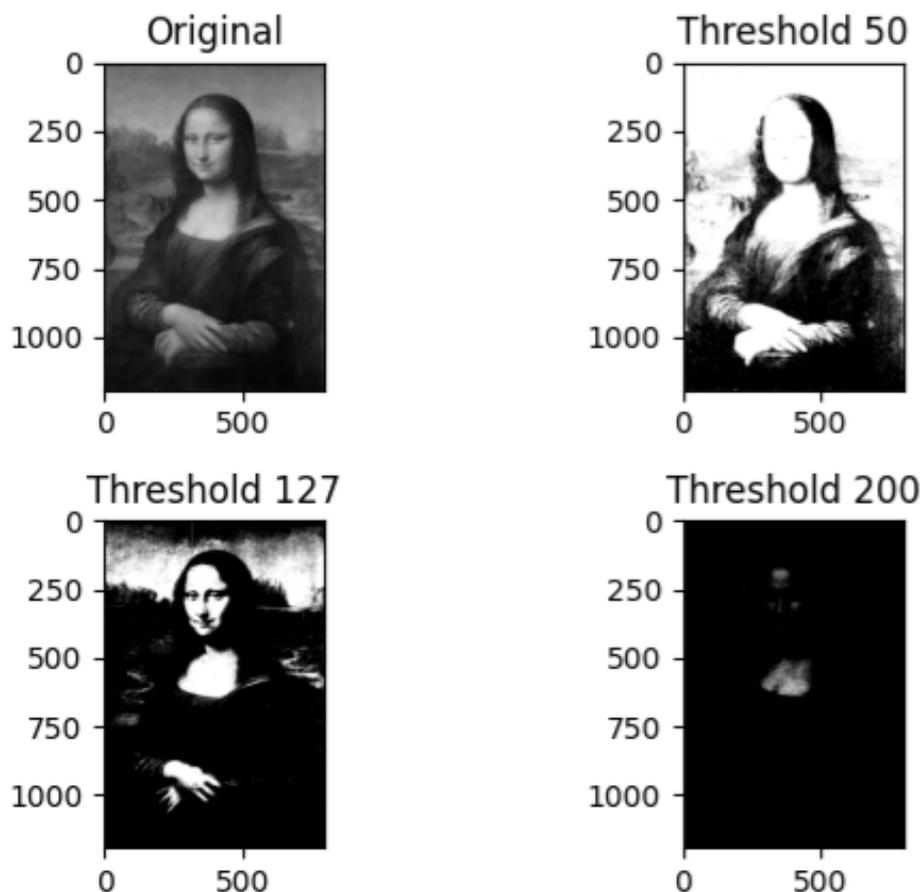
Conforme afirmado pelo *Digital Image Bank* (UFPE, 2023), “Nenhum algoritmo de binarização é bom para todos os tipos de imagens de documentos”. Dessa forma, neste trabalho será abordado varias técnicas de segmentação de imagem, a mais famosa conhecida como limiarização, transforma uma imagem em tons de cinza em uma imagem binária. Esse processo consiste em selecionar um valor limite – o limiar – que separa os níveis de cinza em duas categorias: píxeis com valores superiores ao limiar são convertidos para branco (valor 1), enquanto os píxeis com valores inferiores são convertidos para preto (valor 0), conforme ilustrado na Figura 2.1. Existem dois tipos principais de Limiarização:

- **Limiar Global:** Um único valor de limite é aplicado para toda a imagem. Esse método é simples, mas pode falhar quando a iluminação ou o contraste da imagem não são uniformes. A Figura 2.1 apresenta como valores muito baixos ou muito altos podem tornar uma imagem irreconhecível.
- **Limiar Adaptativo:** O valor do limite é calculado localmente, em pequenas regiões da imagem. Isso permite que o threshold se adapte às variações de iluminação e contraste na imagem, resultando em uma segmentação mais precisa em imagens complexas, na Figura 2.2 demonstra como esse valor impacta na qualidade da imagem final.

2.1.2 Método de Otsu

O método de Otsu é uma técnica de limiar global que seleciona automaticamente o melhor valor de limiar para uma imagem, minimizando a variância das intensidades de cinza. Ele calcula

Figura 2.1 – Exemplo de valores diferentes de limiar na mesma imagem.



Fonte: adaptado de (Wikipedia, 2024) pelo autor.

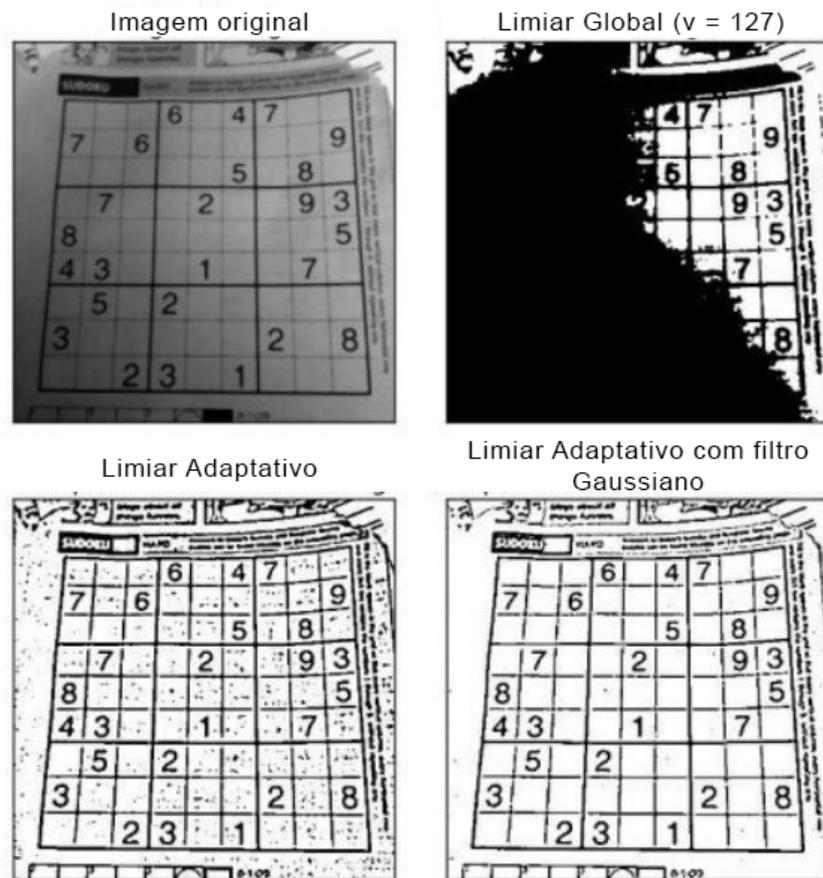
o limiar ótimo de modo que a imagem seja dividida em duas classes: píxeis escuros (fundo) e píxeis claros (objetos), maximizando a separação entre essas classes. O método de Otsu é particularmente eficaz em imagens com histogramas bimodais como apresentado na Figura 2.3b, onde existem duas classes de píxeis bem definidas.

No contexto deste trabalho, o Otsu é fundamental para o pré-processamento de páginas digitalizadas, ao permitir a binarização eficiente, facilitando a detecção e o reconhecimento de padrões nas imagens.

2.1.3 Detector de bordas Canny

O Detector de bordas Canny é uma técnica que realça as diferenças de textura em uma imagem, facilitando a identificação de características importantes para a segmentação. Aplicar a detecção de Canny antes da limiarização pode ser particularmente útil em textos com fundos de várias cores, pois ele consegue destacar as bordas e contornos que definem o texto. A aplicação permite uma melhor definição das áreas de interesse, o que é crucial para livros com fundo complexo ou variação de cor.

Figura 2.2 – Exemplo de valores diferentes de limiar na mesma imagem.



Fonte: adaptado de (OpenCV, 2024) pelo autor.

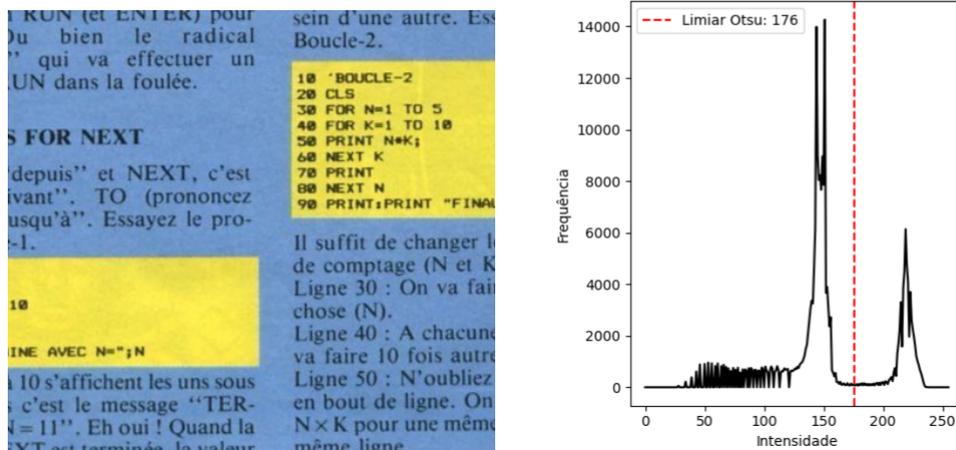
Após a aplicação dos métodos de limiarização, operações morfológicas Seção 2.1.7, como dilatação e erosão, são realizadas para unir áreas de interesse (como caracteres textuais) e eliminar pequenos ruídos presentes na imagem. Essas operações ajudam a manter a integridade das formas importantes, ao mesmo tempo que removem elementos indesejados. Além disso, essas técnicas são utilizadas para reforçar a textura das áreas relevantes da imagem, criando uma base mais robusta para o reconhecimento posterior.

2.1.4 Filtro Gaussiano

O filtro gaussiano é uma técnica amplamente utilizada em processamento de imagens para suavização e remoção de ruídos. Seu funcionamento baseia-se na aplicação de uma convolução entre a imagem original e uma função gaussiana, sendo uma função de distribuição de probabilidade normalizada. A função gaussiana atribui pesos diferentes aos píxeis ao redor de um píxel central, com os pesos maiores concentrados nos píxeis mais próximos, e menores nos mais distantes.

O filtro gaussiano é especialmente útil quando o objetivo é eliminar ruídos de alta frequência, como píxeis isolados com valores muito diferentes dos vizinhos, sem perder muitos

Figura 2.3 – Representação de uma imagem e seu histograma com a seleção do limiar.



(a) Imagem exemplo com variações de tons de cores de fundo.

Fonte: (COLAFTextes, 2024)

(b) Histograma resultante com a seleção de Limiar de Otsu.

Fonte: do autor.

detalhes importantes da imagem.

Um aspecto importante do filtro gaussiano é que ele preserva as bordas suaves, mas pode suavizar bordas mais definidas. Isso o torna uma ferramenta versátil, que muitas vezes serve como etapa de pré-processamento em outros algoritmos de visão computacional.

Em muitas aplicações, o filtro gaussiano é combinado com outros algoritmos. Por exemplo, antes de realizar a detecção de bordas com o algoritmo de Canny Seção 2.1.3, é comum aplicar um filtro gaussiano para reduzir os ruídos e evitar a detecção de bordas falsas. A suavização ajuda a garantir que apenas as mudanças significativas na intensidade dos píxeis (que indicam a presença de uma borda) sejam detectadas nos estágios seguintes de processamento.

A implementação desse filtro pode ser feita de maneira eficiente com a biblioteca *Open Source Computer Vision Library (OpenCV)* (Bradski, 2000), uma das mais amplamente utilizadas para processamento de imagens e visão computacional. O *OpenCV* fornece um conjunto abrangente de ferramentas otimizadas para manipulação de imagens, incluindo a função *GaussianBlur*, que aplica a convolução gaussiana à imagem. Esse método é comumente empregado para suavizar ruídos, reduzir detalhes indesejados e melhorar a segmentação em diversas aplicações, como reconhecimento de padrões, detecção de bordas e pré-processamento para algoritmos de aprendizado de máquina. Além disso, a implementação otimizada do *OpenCV* garante um desempenho adequado mesmo em sistemas com restrições computacionais, tornando-o uma escolha popular para aplicações em tempo real.

Assim, o filtro gaussiano é uma ferramenta essencial no processamento de imagens, proporcionando uma base sólida para técnicas avançadas de reconhecimento de padrões e segmentação, contribuindo significativamente para a qualidade e precisão dos algoritmos de visão computacional.

2.1.5 Detecção de Contornos com `findContours` do *Open Source Computer Vision Library (OpenCV)*

O método `findContours` da biblioteca *OpenCV* foi desenvolvido para identificar os contornos de objetos em imagens binárias. Esse método foi amplamente popularizado e é baseado em algoritmos de detecção de contornos como o de (Suzuki *et al.*, 1985), que propuseram um algoritmo eficiente para encontrar contornos hierárquicos em imagens.

O `findContours` trabalha identificando os limites de objetos na imagem diferenciados pelo contraste entre o fundo e o objeto de interesse. Ele é aplicado geralmente em imagens binárias (ou seja, imagens onde os píxeis são preto e branco) para detectar as fronteiras de objetos em uma cena. Após aplicar uma técnica de limiarização, como o global, adaptativo ou Otsu Seção 2.1.2, ou até mesmo após a detecção de bordas com o método Canny Seção 2.1.3, o `findContours` pode ser utilizado para extrair as formas ou áreas que delimitam os objetos. Seu funcionamento pode ser descrito em alguns passos:

1. Limiarização: A imagem deve ser convertida para binária, onde os contornos se destacam do fundo.
2. Detecção de Contornos: O algoritmo percorre a imagem pixel por pixel, identificando onde os valores de intensidade mudam significativamente e marcam a borda dos objetos.
3. Hierarquia: O método também pode detectar hierarquias entre os contornos, ou seja, a relação de “pai e filho” entre contornos externos e buracos internos, fornecendo informações estruturais adicionais.

Existem diferentes modos de recuperação de contornos disponíveis no *OpenCV*, como o `RETR_EXTERNAL` (que recupera apenas os contornos externos) e o `RETR_TREE` (que recupera todos os contornos e a hierarquia completa), nesse trabalho vamos usar o modo `RETR_EXTERNAL`.

No presente trabalho, o `findContours` foi utilizado para detectar as bordas de cada elementos gráficos presentes nas páginas digitalizadas como apresentado na Figura 3.3, auxiliando no processo de segmentação e extração de informações relevantes para o reconhecimento de padrões. Ele desempenha um papel importante ao isolar regiões de interesse, que posteriormente podem ser comparadas e identificadas por meio de algoritmos de correspondência de características, apresentados a seguir.

2.1.6 Algoritmos de Detecção e Descrição de Características

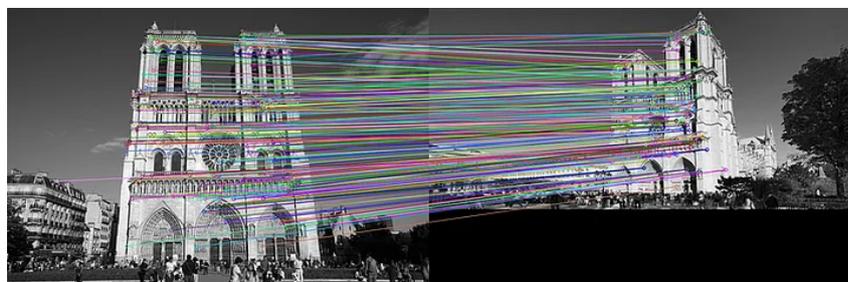
A extração de pontos-chave e a construção de descritores são etapas fundamentais para o reconhecimento de padrões e a correspondência entre imagens, sobretudo em aplicações de *RA* e sistemas de identificação digital. No presente trabalho, foram empregados três algoritmos –

Scale-Invariant Feature Transform (SIFT), *Binary Robust Invariant Scalable Keypoints* (BRISK) e *Oriented FAST and Rotated BRIEF* (ORB) – que apresentam abordagens complementares e que podem ser ajustados por meio de parâmetros durante o pré-processamento das imagens (por exemplo, limiarização global, limiarização adaptativa e filtro Canny). Tais parâmetros, definidos via entrada numérica, permitem adequar os métodos às características específicas dos conjuntos de dados analisados, influenciando diretamente a qualidade dos descritores e, conseqüentemente, a taxa de correspondência entre imagens. A comparação entre imagens é efetuada utilizando o *Brute-Force Matcher*, aplicando a métrica Euclidiana para descritores vetoriais (como os do SIFT) e a distância de Hamming para descritores binários (gerados pelo BRISK e ORB). Em cada caso, adota-se a correspondência que apresenta a maior proporção de acertos em relação ao total de descritores encontrados na imagem de consulta.

2.1.6.1 *Scale-Invariant Feature Transform* (SIFT)

O algoritmo *Scale-Invariant Feature Transform* (SIFT), proposto por Lowe (2004), foi o pioneiro na detecção e descrição de características locais, utilizando a operação de Diferença de Gauss (DoG) em múltiplas escalas para identificar pontos de interesse. Esses pontos são descritos por meio de histogramas dos gradientes orientados, gerando vetores de 128 dimensões em ponto flutuante. Essa alta dimensionalidade contribui para a robustez do método, garantindo invariança frente a rotações, mudanças de escala e, em parte, a transformações afins – conforme ilustrado na Figura 2.4. Nos códigos implementados, a função `cv2.SIFT_create()` foi utilizada, e o parâmetro de pré-processamento (escolhido via entrada do usuário) permite, por exemplo, a aplicação de limiarizações ou do filtro Canny para realçar as características presentes nas imagens. Apesar de sua precisão, o elevado custo computacional e o processamento em múltiplas escalas podem limitar seu uso em aplicações com restrição de tempo real.

Figura 2.4 – Exemplo que demonstra como a rotação da imagem não compromete as características extraídas pelo SIFT, evidenciando que os pontos descritores permanecem interligados e consistentes.



Fonte: (Massucci, 2024)

2.1.6.2 *Binary Robust Invariant Scalable Keypoints* (BRISK)

O algoritmo BRISK, proposto por Leutenegger, Chli e Siegwart (2011), apresenta uma alternativa com menor custo computacional. Utilizando um padrão de amostragem baseado

em anéis concêntricos, o método gera descritores binários que permitem a comparação rápida via distância de Hamming. Implementado pelo código `cv2.BRISK_create()`, essa abordagem binária reduz a carga computacional sem perder, de forma significativa, a capacidade de capturar as características relevantes, embora sua robustez possa ser menor em cenários com variações complexas. Nos códigos, o parâmetro de pré-processamento possibilita a aplicação de diferentes filtros (limiarização global, adaptativa ou Canny) para otimizar o contraste e definir melhor os pontos de interesse antes da extração dos descritores. Tal método é especialmente vantajoso em aplicações que demandam respostas em tempo real, onde o equilíbrio entre velocidade e precisão é crucial.

2.1.6.3 *Oriented FAST and Rotated BRIEF (ORB)*

Desenvolvido para unir a eficiência de métodos baseados em descritores binários à robustez necessária para aplicações práticas, o algoritmo **ORB** combina o detector FAST com o descritor BRIEF, incorporando uma etapa de orientação que assegura a invariância à rotação. Conforme apresentado por Rublee *et al.* (2011), esse método é implementado por meio da função `cv2.ORB_create()`, a qual realiza de forma integrada a detecção e a descrição dos pontos-chave. O ajuste dos parâmetros no pré-processamento possibilita a melhoria da definição dos pontos de interesse em imagens digitalizadas, permitindo uma melhor performance na identificação e comparação das páginas. Por essa razão, o **ORB** foi o método escolhido no contexto deste trabalho, principalmente em cenários que envolvem condições variáveis de rotação e iluminação, sem abrir mão da eficiência computacional, essencial para sistemas móveis e em realidade aumentada.

2.1.6.4 Tabela Comparativa dos Métodos

A Tabela 2.1 sintetiza as principais características dos algoritmos **SIFT**, **BRISK** e **ORB**, considerando o tipo de descritor, desempenho computacional, robustez e aplicabilidade em sistemas que demandam alta taxa de processamento, como os de realidade aumentada.

Tabela 2.1 – Comparação dos Algoritmos de Detecção e Descrição

Método	Tipo de Descritor	Velocidade	Robustez	Aplicabilidade
SIFT	Vetorial (128 dimensões)	Baixa	Alta	Cenários complexos com maior precisão
BRISK	Binário	Alta	Média	Aplicações em tempo real com restrição computacional
ORB	Binário	Alta	Média a Alta	Sistemas móveis e realidade aumentada

Fonte: Adaptado pelo autor.

2.1.7 Operações Morfológicas: Erosão, Dilatação, Abertura e Fechamento

As operações morfológicas são ferramentas fundamentais no processamento de imagens, particularmente em tarefas de segmentação de estruturas. Baseadas em conceitos de álgebra matemática e topologia, essas operações são amplamente utilizadas para modificar formas e contornos em imagens binárias. Todas essas operações fazem uso de um elemento estruturante, conhecido como *kernel*, que define a vizinhança de cada pixel durante o processamento a explicação matemática completa é retratada em (Bloomberg, 1991).

- **Erosão:** A erosão reduz o tamanho das regiões brancas (fundo) em uma imagem binária. Formalmente, seja A um conjunto de pixels pertencentes ao objeto na imagem e B o elemento estruturante, a erosão é definida como:

$$A \ominus B = \{p \mid B_p \subseteq A\} \quad (2.1)$$

Ou seja, o pixel p permanecerá na imagem resultante apenas se todo o elemento estruturante B , quando centralizado em p , estiver completamente contido em A . Esse processo é útil para remover pequenos ruídos e separar objetos que estão muito próximos. A Figura 2.5c ilustra essa transformação morfológica.

- **Dilatação:** A dilatação expande as regiões brancas de uma imagem, aumentando o tamanho dos objetos. Matematicamente, a dilatação de um conjunto A pelo elemento estruturante B é definida como:

$$A \oplus B = \{p \mid (B_p \cap A) \neq \emptyset\} \quad (2.2)$$

Isso significa que um pixel p será incluído na imagem resultante se pelo menos uma parte do elemento estruturante B , ao ser centralizado em p , interceptar algum pixel da região A . Esse processo pode ser utilizado para preencher lacunas em objetos e restaurar detalhes perdidos na erosão. A Figura 2.5b apresenta um exemplo dessa operação.

- **Abertura:** A operação de abertura é definida como a erosão seguida de dilatação:

$$A \circ B = (A \ominus B) \oplus B \quad (2.3)$$

Esse processo é utilizado para remover pequenos ruídos, suavizar contornos e separar objetos próximos sem alterar significativamente a forma dos objetos maiores. A Figura 2.6b demonstra a aplicação dessa técnica.

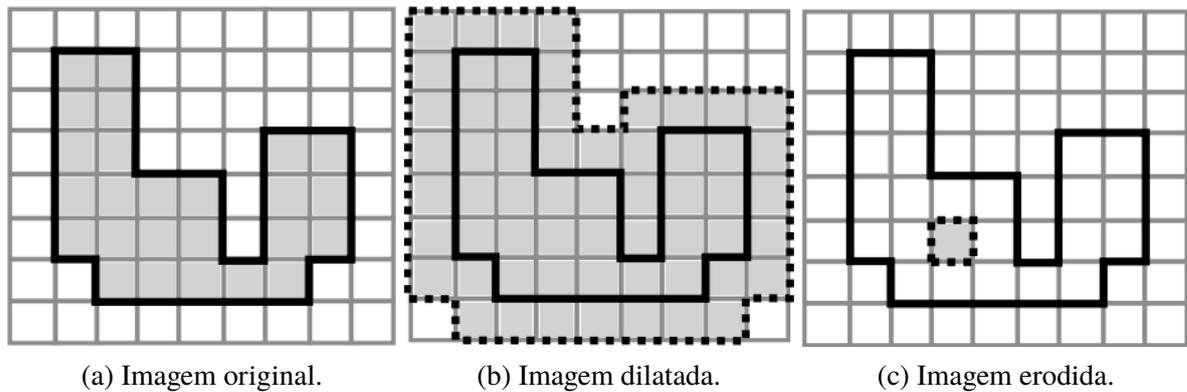
- **Fechamento:** O fechamento consiste na dilatação seguida de erosão:

$$A \bullet B = (A \oplus B) \ominus B \quad (2.4)$$

Essa operação é útil para preencher pequenos buracos dentro dos objetos e conectar componentes que estão próximos. A Figura 2.6c exemplifica esse procedimento.

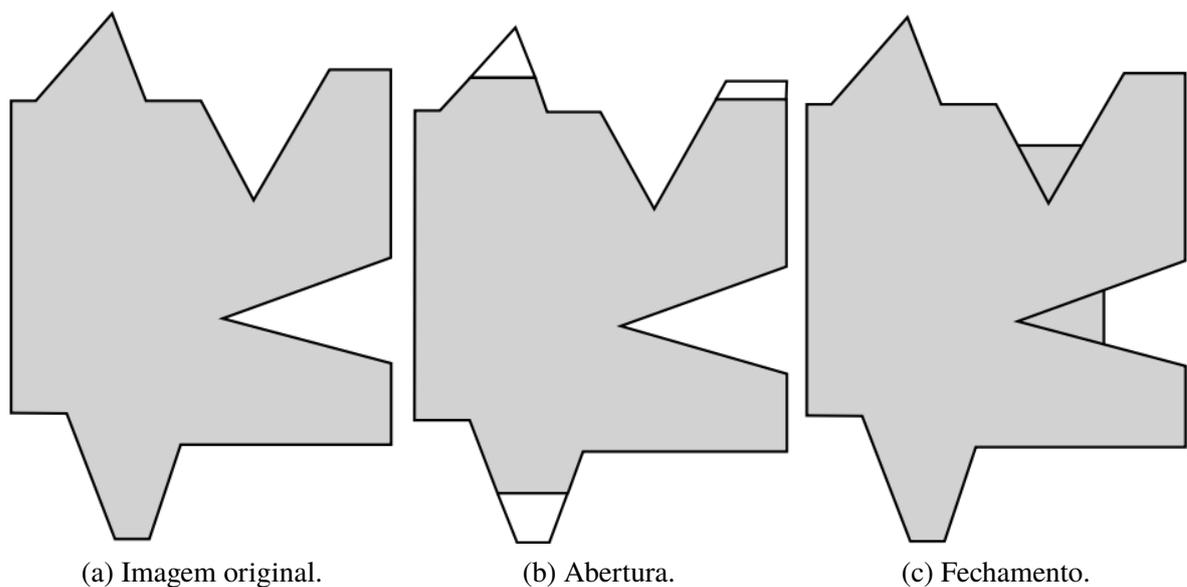
Essas operações são amplamente utilizadas no pré-processamento de imagens para destacar estruturas de interesse, melhorar segmentações e reduzir ruídos antes da aplicação de algoritmos mais complexos de reconhecimento de padrões.

Figura 2.5 – Exemplos de dilatação e erosão com kernel quadrático 3x3.



Fonte: Adaptado de (Carrilho, 2016) pelo autor.

Figura 2.6 – Abertura e fechamento de uma estrutura em uma imagem binária por um elemento estruturante quadrado.



Fonte: Adaptado de (Goutsias; Heijmans, 2000) pelo autor.

Essas operações são essenciais no pré-processamento de imagens, especialmente no contexto da criação de assinaturas digitais de páginas, pois ajudam a refinar as estruturas extraídas das imagens e a remover ruídos desnecessários.

2.2 Trabalhos Relacionados

A integração entre tecnologias de reconhecimento visual e análise documental tem sido objeto de investigação crescente, particularmente no contexto de aplicações de RA. O estudo pioneiro de (Adachi *et al.*, 2023) revelou desafios críticos ao adaptar livros não projetados para AR, destacando que a predominância de elementos textuais em detrimento de marcadores visuais reduz significativamente a eficiência de reconhecimento – apenas 19% das imagens analisadas pelo *Vuforia* alcançaram classificação satisfatória (3 estrelas ou superior) como apresentado na Figura 1.2. Esses resultados evidenciam a necessidade de avanços em técnicas de processamento documental que transcendam abordagens convencionais baseadas em elementos gráficos distintivos.

2.2.1 Abordagens Estruturais e Geométricas em Análise Documental

A fundamentação teórica para superar tais limitações encontra respaldo em quatro décadas de pesquisas em segmentação estrutural. O trabalho seminal de (Haralick *et al.*, 1995) estabeleceu paradigmas para decomposição hierárquica de documentos, introduzindo métodos de delimitação de zonas textuais e não textuais que permanecem relevantes. Essa abordagem foi ampliada por (Jain; Zhong, 1996), cujo algoritmo baseado em redes neurais demonstrou eficácia na segmentação multimodal de documentos multilíngues, inclusive com variações cromáticas complexas – característica essencial para aplicações em livros didáticos.

Em paralelo, (Cattoni *et al.*, 1998) consolidou princípios de análise geométrica para interpretação de layouts, enquanto (Askarpour *et al.*, 2014) enfrentou o desafio adicional de segmentação em escritas cursivas através de técnicas de *clustering* adaptativo. Esses estudos convergem na premissa de que a identificação precisa de regiões de interesse requer modelos capazes de capturar tanto relações espaciais quanto características intrínsecas do conteúdo, princípio que orienta o desenvolvimento de nosso algoritmo.

2.2.2 Avanços em Modelagem Hierárquica e Aprendizado Profundo

A evolução para modelos hierárquicos trouxe novas dimensões à análise documental. (Mao; Rosenfeld; Kanungo, 2003) propôs uma estrutura de representação multicamadas que influenciou profundamente o desenvolvimento de sistemas modernos de recuperação de informações. Contemporaneamente, (Yang; Hsu, 2021) demonstrou o potencial disruptivo das redes neurais convolucionais recorrentes na segmentação de layouts científicos, superando limitações de abordagens baseadas em regras através de detecção contextual de objetos.

Essa transição paradigmática enfrenta, contudo, o dilema entre precisão e eficiência computacional. Embora métodos como os de (Igorevna *et al.*, 2022) atinjam acurácia notável em binarização neural de documentos, sua complexidade operacional os torna impraticáveis para aplicações em tempo real. Essa constatação reforça a necessidade de soluções híbridas

que integrem aprendizagem de máquina com técnicas clássicas de processamento de imagem – estratégia adotada no presente trabalho.

2.2.3 Desafios Práticos em OCR e Pré-processamento

A operacionalização desses princípios em cenários reais demanda superar obstáculos adicionais. (Kišš; Hradiš; Kodym, 2019) identificou limitações críticas no reconhecimento óptico de caracteres OCR em ambientes não controlados, propondo o conjunto de dados *B-MOD* para treinamento de algoritmos robustos a variações lumínicas e artefatos de compressão – problema que afeta diretamente aplicações móveis de RA. Como ilustrado na Figura 2.7, a incorporação de marcadores auxiliares QR emerge como estratégia complementar, embora insuficiente para documentos convencionais.

Figura 2.7 – Exemplos representativos de fotografias no conjunto de dados *B-MOD*.



Fonte: Adaptado de (Kišš; Hradiš; Kodym, 2019) pelo autor.

Nesse contexto, técnicas de pré-processamento ganham relevância crítica. O estudo de (Flaute, 2024) comprovou que a integração de filtragem neural de *keypoints* com métodos clássicos como limiarização adaptativa e detecção de bordas Seção 2.1.3 potencializa a extração de descritores visuais em documentos complexos. Esses achados alinham-se às observações de (Bloomberg, 1991) sobre o uso estratégico de operações morfológicas para isolamento seletivo de elementos textuais, formando um arcabouço metodológico coeso para tratamento de imagens documentais.

2.2.4 Síntese Teórica e Direcionamento Metodológico

A análise crítica desses trabalhos revela uma trajetória evolutiva marcada pela tensão entre complexidade analítica e viabilidade operacional. Enquanto abordagens baseadas em *deep learning* oferecem precisão superior (Yang; Hsu, 2021; Igorevna *et al.*, 2022), sua aplicabilidade permanece limitada por requisitos computacionais – desafio particularmente relevante para sistemas embarcados em RA. Por outro lado, técnicas estruturais clássicas (Haralick *et al.*, 1995; Mao; Rosenfeld; Kanungo, 2003) fornecem base teórica sólida para desenvolvimento de algoritmos eficientes, porém carecem de adaptabilidade a variações contextuais.

Nossa proposta sintetiza essas vertentes através de um modelo híbrido que combina:

- Segmentação geométrica baseada em componentes conectados (BIFI, 2016)
- Extração morfológica de características distintivas (Bloomberg, 1991)
- Avaliação da melhora de desempenho dos algoritmos SIFT, BRISK e ORB com os pré processamentos propostos (Flaute, 2024)

Essa integração metodológica visa superar as limitações identificadas por (Adachi *et al.*, 2023), possibilitando o reconhecimento robusto de páginas predominantemente textuais sem modificações no conteúdo original.

3 Desenvolvimento

3.1 Metodologia

Nesta seção, são descritas as etapas seguidas na pesquisa, abrangendo desde a revisão bibliográfica até a execução dos testes e análise dos resultados.

3.1.1 Revisão Bibliográfica

Inicialmente, foi realizada uma revisão bibliográfica aprofundada para embasar a escolha das técnicas e algoritmos utilizados. A revisão incluiu estudos sobre segmentação estrutural, reconhecimento de padrões e processamento de imagens. Algoritmos como Limiarização Global Seção 2.1.1, Limiarização Adaptativa Seção 2.1.1, Método Canny Seção 2.1.3, [SIFT](#), [ORB](#) e [BRISK](#) foram analisados, destacando suas características e aplicabilidades.

3.1.2 Definição de Requisitos

A definição de requisitos envolveu a seleção dos livros-alvo e das páginas a serem reconhecidas. Também foram estabelecidos critérios para a escolha das técnicas de pré-processamento e reconhecimento, considerando cenários com diferentes condições de iluminação e qualidade de imagem.

3.1.3 Desenvolvimento e Implementação

Para o desenvolvimento do sistema, foram aplicados algoritmos de pré-processamento que incluíram operações como erosão Seção 2.1.7, dilatação Seção 2.1.7 e técnicas de limiarização criando uma abordagem de segmentação estrutural das páginas, depois o material foi submetido a técnicas de reconhecimento rápido e rotacionáveis, como o [SIFT](#), [ORB](#) e [BRISK](#) como demonstra do diagrama na Figura 3.1.

A implementação foi realizada em linguagem Python, utilizando bibliotecas especializadas em processamento de imagens, como [OpenCV](#). As funções `findContours` e `threshold` Seção 2.1.1 foram amplamente empregadas para detecção e segmentação de características estruturais.

Após a segmentação inicial das regiões de interesse, foram aplicadas operações morfológicas com o objetivo de refinar as áreas selecionadas, eliminar ruídos e melhorar a coesão dos elementos estruturais presentes na imagem. Essas operações foram implementadas por meio da

biblioteca OpenCV¹, utilizando diferentes configurações de *kernels* conforme a finalidade de cada etapa.

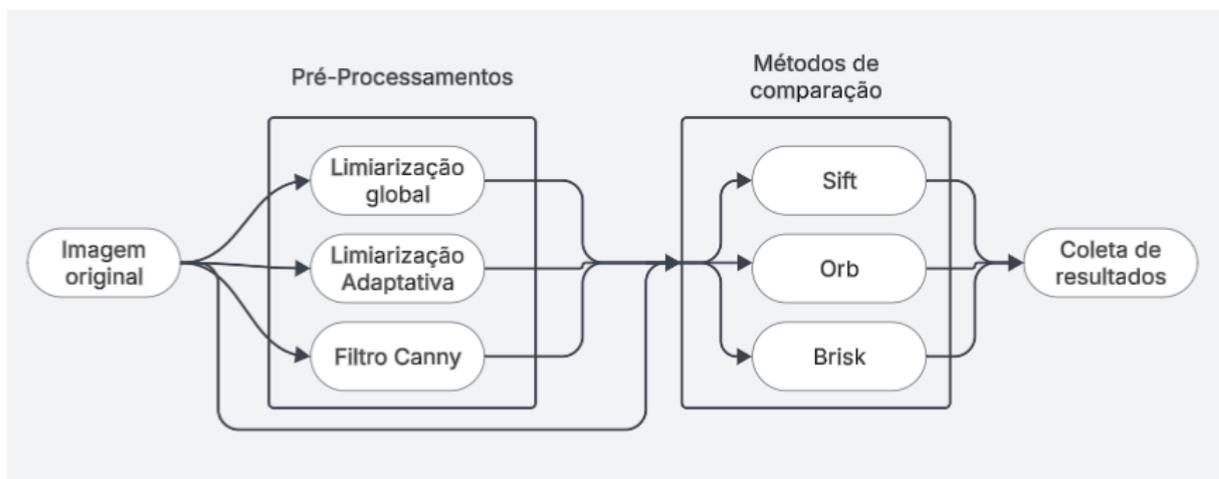
A primeira operação aplicada foi a *abertura morfológica (morphological opening)*, realizada com um *kernel* retangular de dimensão 3×3 , definido pela função `np.ones((3, 3), np.uint8)`. Essa operação, correspondente à erosão seguida de dilatação, foi empregada com a função `cv2.morphologyEx()` utilizando o parâmetro `cv2.MORPH_OPEN`. Sua principal finalidade foi a remoção de pequenos ruídos e a suavização dos contornos das regiões segmentadas.

Na sequência, com o intuito de promover a união de elementos próximos no eixo horizontal, aplicou-se uma operação de dilatação com um *kernel* de dimensão 1×5 , utilizando a função `cv2.dilate()` com uma duas iterações (`iterations=2`). Essa escolha permite a expansão controlada das áreas, favorecendo a agregação de componentes pertencentes a uma mesma estrutura sem comprometer a integridade das demais regiões.

Por fim, aplicou-se uma operação de erosão utilizando o mesmo *kernel* (1×5), com o objetivo de corrigir excessos introduzidos na etapa de dilatação e ajustar a espessura das regiões destacadas. Essa sequência de operações — abertura, dilatação e erosão — mostrou-se eficaz na eliminação de artefatos, reforço de componentes estruturais e melhoria na segmentação para posterior extração de características.

As imagens resultantes de cada etapa foram visualizadas por meio da função `cv2.imshow()` e, opcionalmente, salvas utilizando `cv2.imwrite()`, possibilitando a documentação e análise dos efeitos de cada operação no processo de refinamento das máscaras.

Figura 3.1 – Diagrama da metodologia aplicada.



Fonte: Elaborado pelo Autor.

Além dos procedimentos convencionais de pré-processamento e segmentação, este estudo adota uma estratégia que visa transformar cada página em uma “impressão digital” com base na sua estrutura geométrica destacada. Nesta etapa, os algoritmos são ajustados para enfatizar

¹ Open Source Computer Vision Library. Disponível em: <<https://opencv.org/>>. Acesso em: 04 abr. 2025.

as delimitações presentes nas páginas, extraindo contornos precisos que funcionarão como âncoras para a aplicação de RA. O método utiliza técnicas de limiarização e de detecção de bordas, combinadas a operações morfológicas que reforçam os limites dos elementos estruturais. Como resultado, os descritores gerados – a partir das áreas de contorno identificadas – contêm informações sobre rotação, escala e orientação, atributos indispensáveis para a correta ancoragem e posicionamento dos elementos virtuais. Dessa forma, a abordagem possibilita que o sistema de RA reconheça e mantenha a integração entre o objeto 3D e a estrutura física do livro, assegurando uma experiência interativa e precisa.

3.2 Pré-processamentos

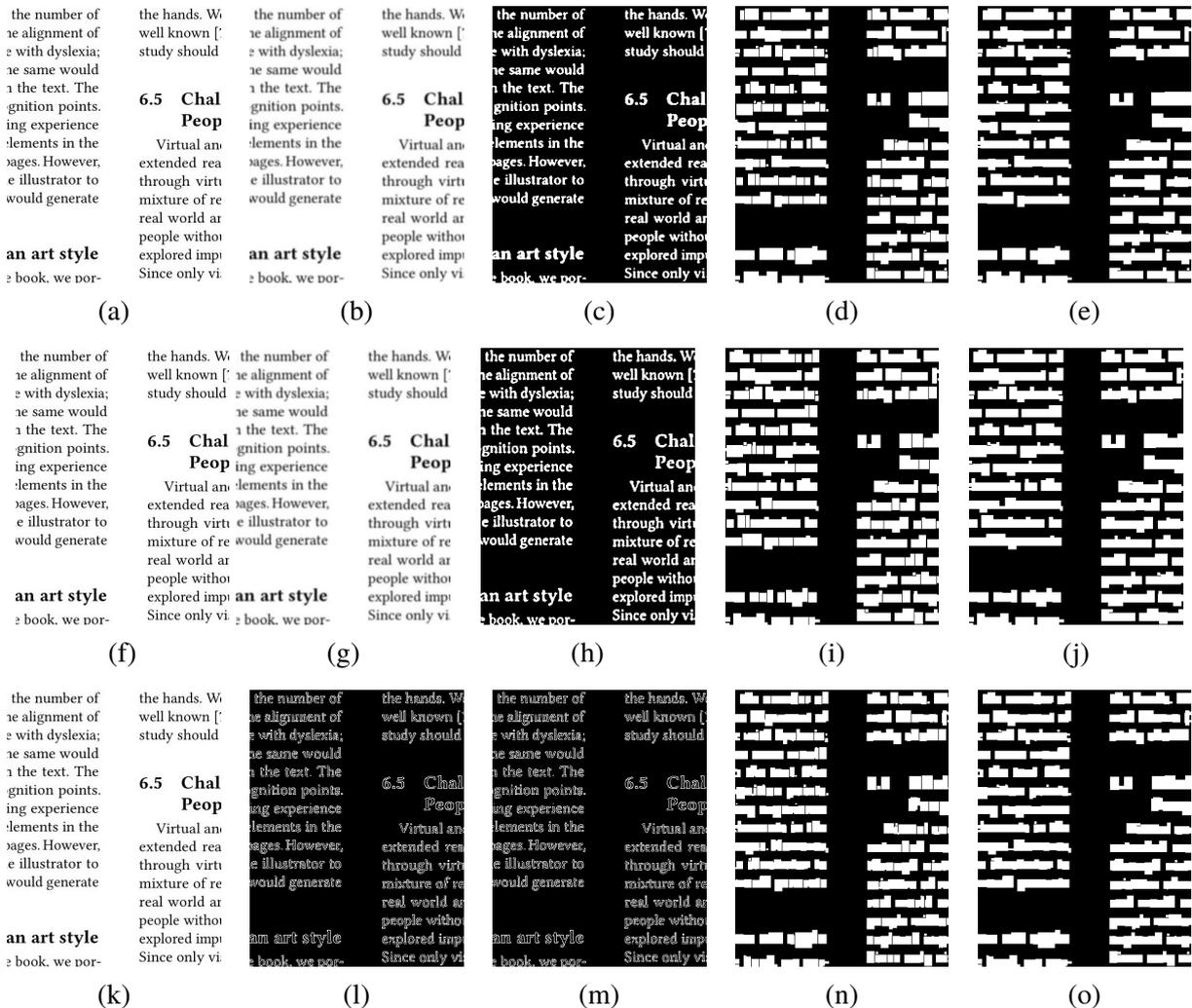
A metodologia deste trabalho consiste em aplicar técnicas de pré-processamento para transformar as páginas de livros em representações digitais, ou “impressões digitais” únicas, conforme ilustrado na Figura 1.7. Esse procedimento facilita o reconhecimento das páginas, mesmo na ausência de elementos visuais distintivos. Para alcançar esse objetivo, diferentes métodos de limiarização (Seção 2.1.1) são empregados, conforme exemplificado na Figura 3.2 (a, f e k), bem como operações morfológicas (Seção 2.1.7), apresentadas na Figura 3.2 (e, j e o). Cada técnica foi avaliada com o intuito de otimizar a identificação de padrões e gerar identificadores digitais exclusivos para as páginas.

O pré-processamento das imagens é essencial para preparar as páginas de livros para o reconhecimento por algoritmos como o ORB. Além disso, a eficiência dos métodos foi analisada por meio de testes em um banco de dados contendo textos com diferentes layouts, permitindo avaliar seu desempenho em diversos cenários.

Em todos os três pré-processamentos, algumas etapas são comuns. Primeiramente, a imagem é convertida para escala de cinza, como ilustrado na Figura 3.2 (a, f e k). Em seguida, aplica-se um filtro gaussiano (Figura 3.2, b e g), exceto no Método de Canny, que já incorpora essa operação. Posteriormente, realiza-se a limiarização (ou a aplicação do Método de Canny), resultando na imagem binarizada, conforme Figura 3.2 (c, h e l).

Na etapa seguinte, os contornos são detectados por meio da função *findContours*, representada na Figura 3.2 (d, i e n), preenchendo os espaços com formas retangulares. Em seguida, aplica-se a operação morfológica de abertura para remover píxeis indesejados, seguida de uma dilatação horizontal para unir elementos próximos. Por fim, a erosão ajusta a espessura das áreas processadas, como demonstrado na Figura 3.4. A Figura 3.3 ilustra o impacto de cada operação do algoritmo sobre a imagem, evidenciando de forma sequencial as etapas do processamento aplicado.

Figura 3.2 – Etapas do pré-processamento das imagens.



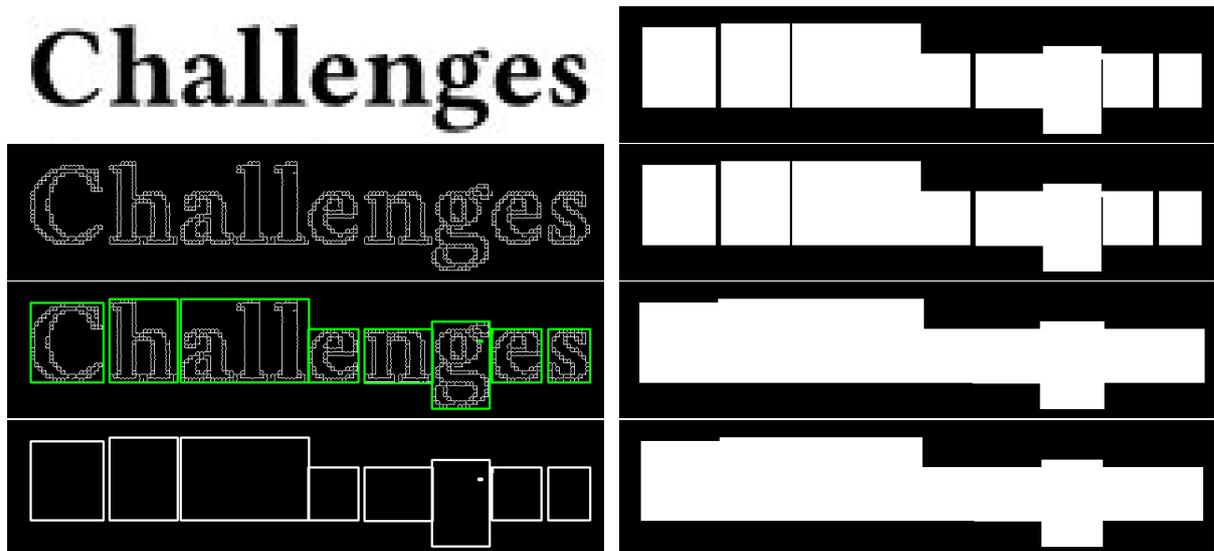
Primeira linha se refere ao método de Limiar global, a segunda o limiar adaptativo e no terceiro usando Método canny. (a, f, k) Imagem original em escala de cinza. (b, g) Aplicação do blur. (l) Aplicação Método canny. (c, h, m) Binarização da imagem. (d, i, n) Extração de contornos e preenchimento de bounding boxes. (e, j, o) Imagem final pós-processada. Fonte: Elaborado pelo Autor.

3.3 Ambiente dos Testes

Os testes foram realizados em um computador com a seguinte configuração: Processador Ryzen 7 e 16 GB de memória, sem o uso de GPU dedicada. Foi criado um módulo de geração de amostras baseadas no banco de dados de imagens.

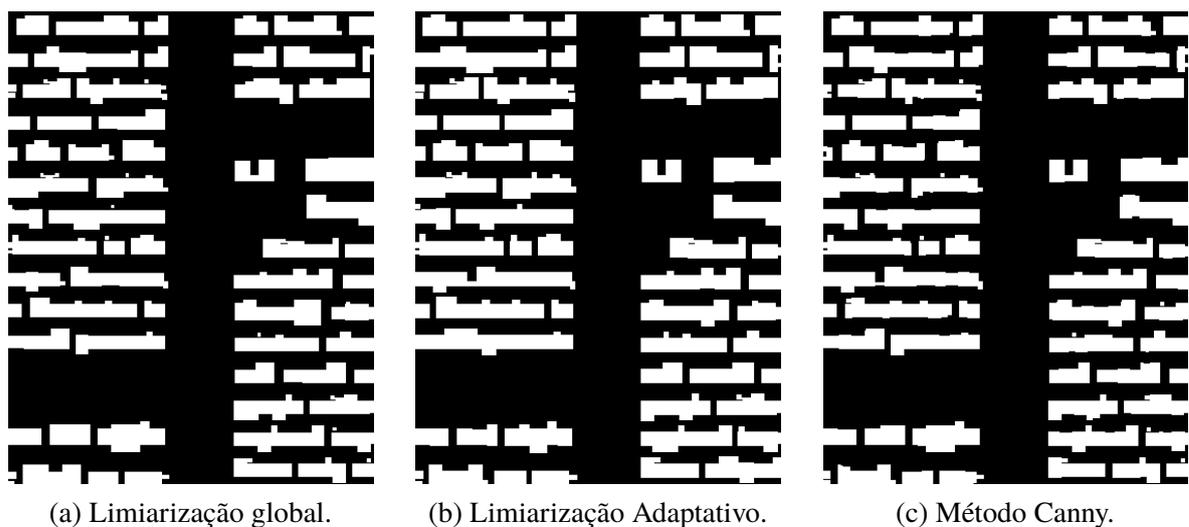
Cada amostra gerada é submetida a um dos métodos de pré-processamento discutidos anteriormente. Após o pré-processamento, as amostras são comparadas ao banco de dados utilizando os algoritmos SIFT, BRISK e ORB, e o tempo de processamento é registrado. Além disso, é verificada a precisão do reconhecimento, ou seja, se a amostra foi corretamente identificada. O processo é repetido para cada método de pré-processamento, e os resultados são coletados para

Figura 3.3 – Sequência de imagens demonstrando as etapas do processamento.



Etapas do processamento realizado, com base no método canny. Fonte: Elaborado pelo Autor.

Figura 3.4 – Resultados finais dos pré-processamentos.



(a) Limiarização global.

(b) Limiarização Adaptativa.

(c) Método Canny.

Fonte: Elaborado pelo Autor.

análise comparativa de desempenho e precisão.

3.3.1 Base de Dados

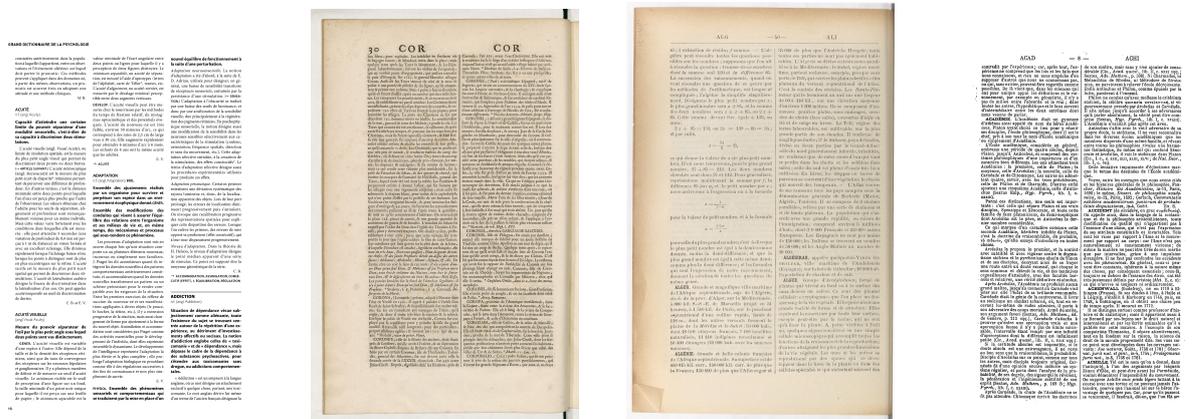
Para a avaliação da eficácia dos algoritmos propostos, foi utilizado o banco de dados Segmento, composto por 4.425 imagens de páginas de livros digitalizadas. Esse conjunto foi subdividido em um subconjunto contendo 278 imagens, que representam livros com características visuais distintas. Essa segmentação possibilita a simulação de diferentes cenários de reconhecimento, contribuindo para uma análise abrangente dos métodos propostos.

O Segmento (COLAFTextes, 2024) é uma base de dados direcionada à pesquisa em

segmentação estrutural de documentos. Ela reúne páginas com grande diversidade visual, oriundas de processos reais de digitalização. Conforme destacado por Zorba Books (Books, 2024), a média de páginas em livros destinados ao público adulto varia entre 200 e 400 páginas, evidenciando a necessidade de métodos eficazes para a identificação precisa de páginas em grandes volumes de dados.

A Figura 3.5 demonstra que a base de dados apresenta um amplo espectro de documentos, que vai desde registros históricos até artigos recentes. Todas as imagens foram escaneadas sem alteração de rotação e escala, mantendo-se a apresentação padronizada. Os fundos possuem variadas tonalidades, e os documentos, embora organizados de forma padronizada, exibem estruturas geométricas diferenciadas e estão redigidos em diversas línguas.

Figura 3.5 – Figuras extraídas do banco de dados Segmonto



Fonte: (COLAFTextes, 2024).

4 Resultados e Discussão

Neste capítulo, são apresentados e discutidos os resultados obtidos a partir da aplicação dos algoritmos de reconhecimento de páginas, explorando as diferentes configurações de pré-processamento. Inicialmente, descreve-se o procedimento experimental adotado, seguido da exposição dos resultados por meio de tabelas e gráficos, e por fim, realiza-se uma comparação crítica entre os métodos testados.

4.1 Procedimentos para Coleta e Análise dos Resultados

Para a obtenção dos resultados, foram utilizadas amostras provenientes do banco de dados descrito no capítulo anterior. Cada imagem foi submetida a processos de pré-processamento específicos antes da extração e identificação dos descritores empregando os algoritmos **ORB**, **SIFT** e **BRISK**. Adicionalmente, foram realizados experimentos sem a etapa de pré-processamento, possibilitando uma análise comparativa abrangente. As seguintes métricas foram utilizadas para avaliação:

1. **Precisão de Reconhecimento:** Verificação da correspondência correta entre a amostra recortada e a página original presente no banco de dados;
2. **Tempo de Processamento:** Período necessário para que o algoritmo processe a amostra e identifique a correspondência no banco;
3. **Quantidade de Descritores:** Número de pontos-chave extraídos de cada imagem;
4. **Confiança na Pontuação de Similaridade:** Valor atribuído pelo algoritmo para ranquear a similaridade entre a amostra e as imagens presentes no banco.

As métricas consideradas, nomeadamente o tempo de processamento, o número de descritores extraídos e a confiança na similaridade, foram calculadas como médias a partir de todas as execuções realizadas para cada configuração experimental, garantindo uma avaliação robusta do desempenho dos algoritmos em distintos cenários.

4.2 Resultados com Diferentes Métodos de Pré-processamento

Nesta seção, os resultados são apresentados conforme o método de pré-processamento adotado. Cada sub-seção contém uma descrição dos resultados obtidos, acompanhada das tabelas e gráficos correspondentes, que são contextualizados e discutidos com base nas métricas de avaliação.

4.2.1 Limiarização Global

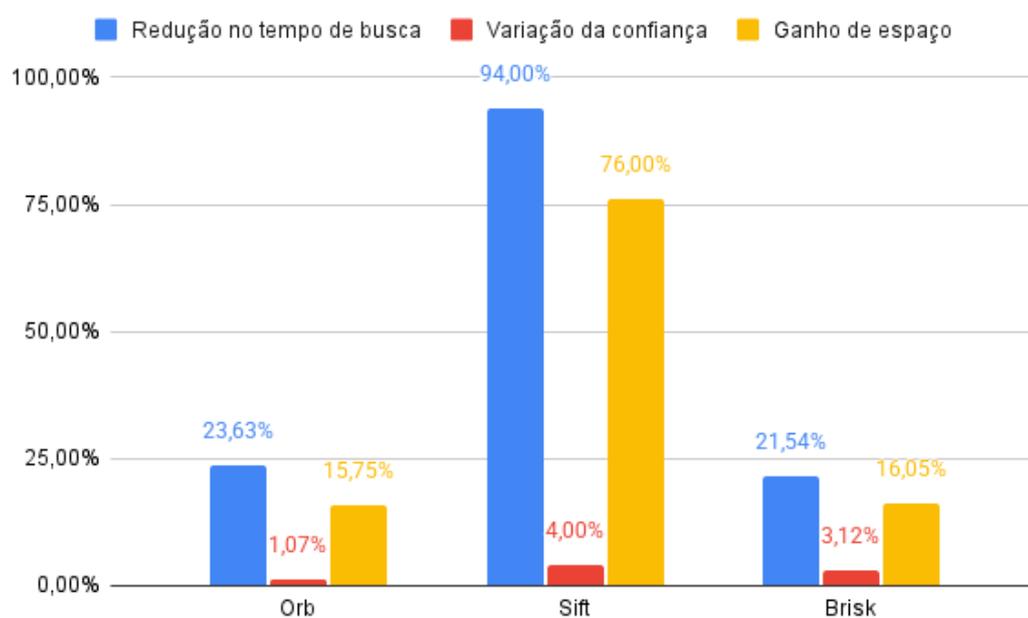
A limiarização global foi aplicada com o intuito de uniformizar a iluminação das imagens, facilitando a extração de descritores. A Tabela 4.1 apresenta, de forma consolidada, os tempos de busca, a pontuação de confiança e a quantidade de descritores obtidos para cada algoritmo, considerando amostras pré-processadas e não pré-processadas. O gráfico da Figura 4.1 ilustra a comparação percentual dos métodos, facilitando a visualização das variações introduzidas pelo pré-processamento.

Tabela 4.1 – Análise comparativa dos métodos com limiarização global: tempo de busca, pontuação de confiança e quantidade de descritores.

Algoritmo/Método	Tempo de Busca (s)	Pontuação de Confiança	Descritores
ORB - Limiarização Global			
Pré-processada	0,4542	0,9395	417,8705
Sem pré-processamento	0,5919	0,9375	499,5323
SIFT - Limiarização Global			
Pré-processada	4,5238	0,9563	6.342
Sem pré-processamento	78,1589	0,9225	27.702
BRISK - Limiarização Global			
Pré-processada	76,3396	0,9089	36.250
Sem pré-processamento	226,7873	0,9382	62.393,5

Fonte: Elaborado pelo Autor. Notas: Valores arredondados para 4 casas decimais.

Figura 4.1 – Resultados dos métodos com limiarização global.



Fonte: Elaborado pelo Autor.

4.2.1.1 Análise por Algoritmo

4.2.1.1.1 Algoritmo ORB:

Com o **ORB** aplicado à limiarização global, observou-se uma redução de 23,63% no tempo de busca em relação à ausência do pré-processamento. Embora tenha havido uma leve variação (1,07%) na confiança de similaridade, apenas duas das 278 imagens não foram corretamente reconhecidas, e o ganho obtido na redução do número de descritores foi de 15,75%. Tais resultados evidenciam que o pré-processamento impactou positivamente a eficiência computacional do algoritmo.

4.2.1.1.2 Algoritmo SIFT:

Na aplicação do **SIFT** com limiarização global, verificou-se uma redução de 94% no tempo de busca, embora tenha ocorrido uma pequena redução de 4% na confiança de similaridade. O ganho no espaço de armazenamento dos descritores foi expressivo, atingindo 76%. Estes resultados sugerem que o **SIFT** é sensível à simplificação das imagens, mas mantém uma precisão satisfatória.

4.2.1.1.3 Algoritmo BRISK:

Para o algoritmo **BRISK**, a técnica de limiarização global conduziu a uma diminuição de 21,54% no tempo de busca e a um ganho de 16,05% no armazenamento. Entretanto, foi verificada uma variação de 53,56% na confiança de similaridade, indicando que o pré-processamento impactou de forma mais acentuada a estabilidade dos resultados para este método.

4.2.2 Limiarização Adaptativa

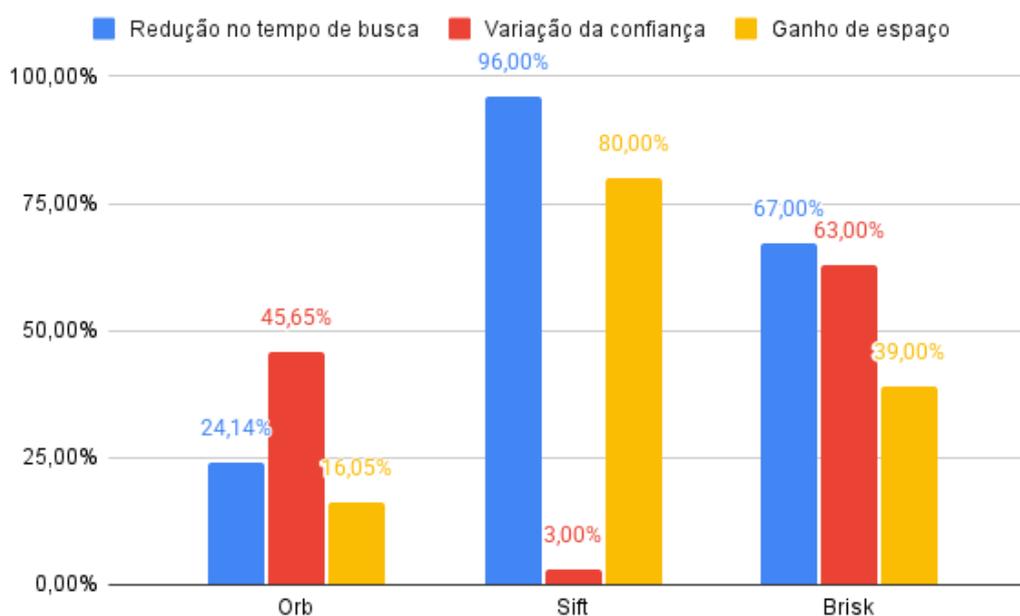
A técnica de limiarização adaptativa, que ajusta os parâmetros conforme as características locais da imagem, foi testada visando aprimorar o equilíbrio entre eficiência e preservação de detalhes. A Tabela 4.2 apresenta os resultados obtidos para cada algoritmo. O gráfico da Figura 4.2 ilustra a variação percentual das métricas.

Tabela 4.2 – Análise comparativa dos métodos com limiarização adaptativa: tempo de busca, pontuação de confiança e quantidade de descritores.

Algoritmo/Método	Tempo de Busca (s)	Pontuação de Confiança	Descritores
ORB - Limiarização Adaptativa			
Pré-processada	0,4553	0,5066	417,8920
Sem pré-processamento	0,6000	0,9375	499,5323
SIFT - Limiarização Adaptativa			
Pré-processada	4,2798	0,9464	5.451
Sem pré-processamento	123,6167	0,9225	27.702
BRISK - Limiarização Adaptativa			
Pré-processada	75,2445	0,3442	36.250
Sem pré-processamento	247,8118	0,9382	62.393,5

Fonte: Elaborado pelo autor. Notas: Valores arredondados para 4 casas decimais.

Figura 4.2 – Resultados dos métodos com limiarização adaptativa.



Fonte: Elaborado pelo Autor.

4.2.2.1 Análise por Algoritmo

4.2.2.1.1 Algoritmo ORB:

Na aplicação do **ORB** com limiarização adaptativa, o tempo de processamento foi reduzido em 24,14%. Entretanto, a confiança na similaridade apresentou uma queda significativa (45,65%), embora o ganho de espaço (16,05%) seja equivalente ao observado com a limiarização global.

4.2.2.1.2 Algoritmo SIFT:

Com o **SIFT** associado à limiarização adaptativa, a redução no tempo de busca atingiu 96%, com uma leve variação (redução de 3%) na confiança, e um expressivo ganho no armazenamento dos descritores (80%). Estes resultados evidenciam que o método preserva, de forma satisfatória, a integridade dos pontos-chave.

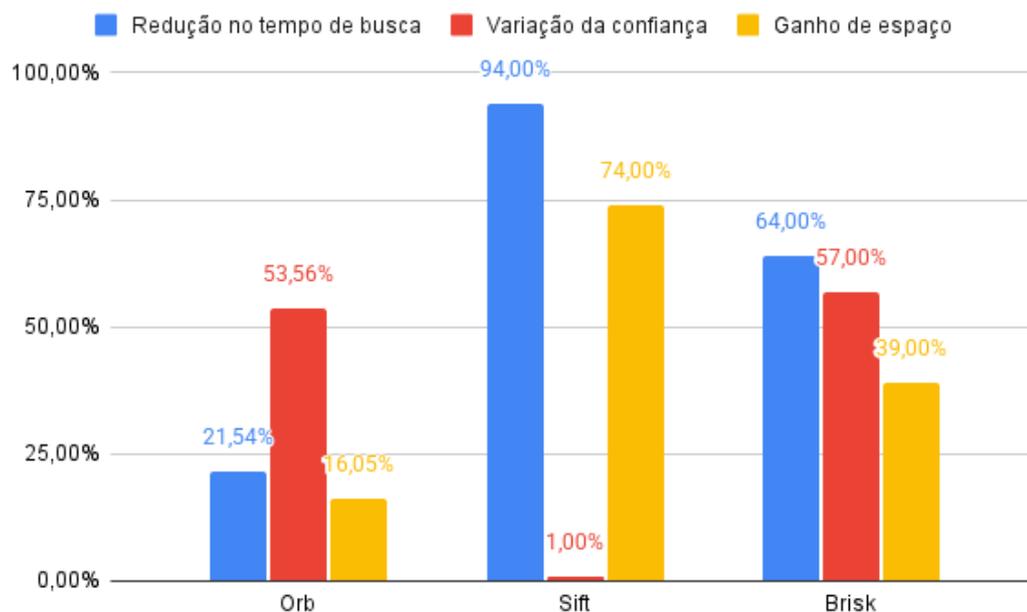
4.2.2.1.3 Algoritmo BRISK:

No caso do **BRISK**, a limiarização adaptativa proporcionou uma redução de 67% no tempo de busca e um ganho de espaço de 39%. A confiança na similaridade, por sua vez, apresentou uma variação positiva de 63%, sugerindo que este algoritmo se beneficiou da estratégia adaptativa, apesar de sua sensibilidade aos detalhes texturais.

4.2.3 Método Canny

O método Canny, empregado para detecção de bordas, apresentou resultados distintos quando comparado às técnicas de limiarização. A Figura 4.3 ilustra os resultados dos algoritmos aplicados em conjunto com o método Canny, enquanto a Tabela 4.3 consolida os dados relativos ao tempo de busca, confiança na pontuação e quantidade de descritores.

Figura 4.3 – Resultados dos métodos utilizando o método Canny.



Fonte: Elaborado pelo Autor.

Tabela 4.3 – Análise comparativa dos métodos com o método Canny: tempo de busca, confiança na pontuação e quantidade de descritores.

Algoritmo/Método	Tempo de Busca (s)	Confiança na Pontuação	Descritores
ORB - Método Canny			
Pré-processada	0,4553	0,5066	417,8920
Sem pré-processamento	0,6020	0,9375	499,5323
SIFT - Método Canny			
Pré-processada	6,4471	0,9327	7.572,5
Sem pré-processamento	122,4667	0,9225	27.702
BRISK - Método Canny			
Pré-processada	93,0654	0,4060	36.250
Sem pré-processamento	280,7063	0,9382	62.393,5

Fonte: Elaborado pelo autor. Notas: Valores arredondados para 4 casas decimais.

4.2.3.1 Análise por Algoritmo

4.2.3.1.1 Algoritmo ORB:

Com o método Canny, o **ORB** apresentou redução de 21,54% no tempo de busca, embora a confiança na similaridade tenha sofrido uma variação negativa de 53,56%, mantendo o ganho de espaço de 16,05%.

4.2.3.1.2 Algoritmo SIFT:

Na aplicação do **SIFT** associado ao método Canny, registrou-se uma redução de 94% no tempo de busca, com apenas uma variação mínima (redução de 1%) na confiança, e um ganho de 74% no armazenamento dos descritores. Este desempenho indica uma boa preservação da precisão.

4.2.3.1.3 Algoritmo BRISK:

Para o **BRISK**, o método Canny ocasionou uma redução de 64% no tempo de processamento e um ganho de espaço de 39%, além de um aumento de 57% na confiança de similaridade, evidenciando os benefícios desta abordagem para o algoritmo.

4.3 Comparação Geral entre Testes com e sem Pré-processamento

A análise comparativa entre os métodos com e sem aplicação de pré-processamento aponta que as técnicas adotadas proporcionam vantagens expressivas em termos de eficiência

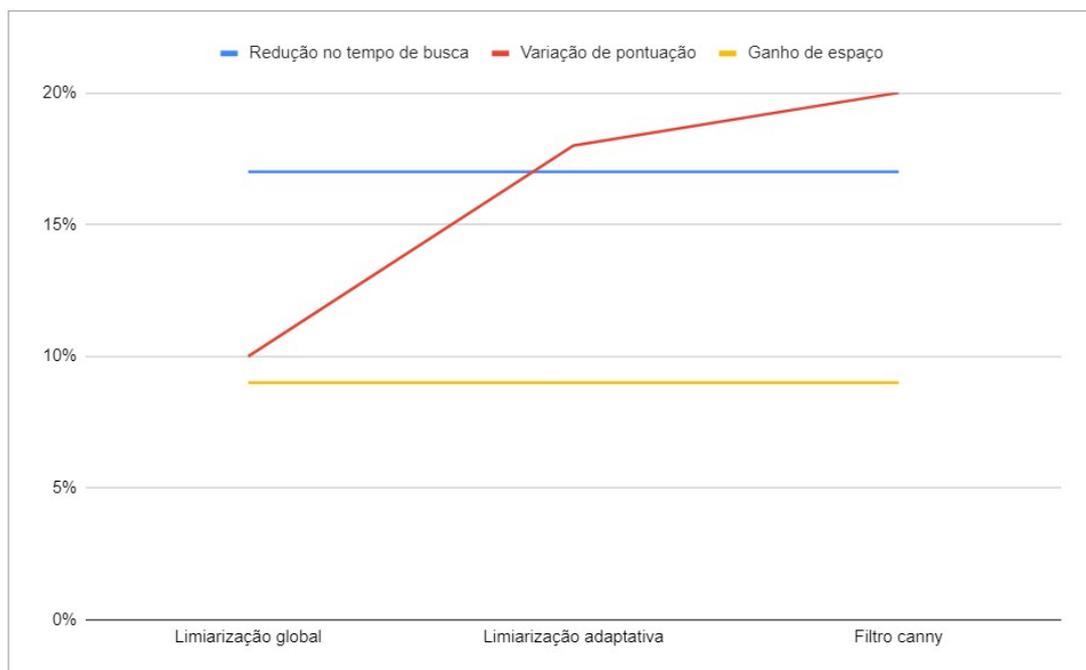
computacional, sem comprometer significativamente a precisão dos algoritmos, embora existam variações específicas conforme a metodologia aplicada. Essa comparação é ilustrada nas Tabelas 4.1 a 4.3 e nos gráficos correspondentes (Figuras 4.1 a 4.3).

4.3.1 Vantagens do Pré-processamento

A partir dos resultados apresentados, podem ser identificadas três principais vantagens:

1. **Eficiência Computacional:** Os métodos de pré-processamento reduziram, em média, 54,2% do tempo de processamento – destacando-se o **SIFT**, que, com limiarização adaptativa, apresentou uma redução de 96% (Tabela 4.2).
2. **Otimização do Armazenamento de Dados:** Foi observado um ganho médio de 41,3% na redução do número de descritores, contribuindo para a escalabilidade do sistema.
3. **Consistência na Extração de Características:** A variação máxima na confiança da pontuação de similaridade foi de apenas 4% para o **SIFT** (Tabela 4.2), em contraste com a oscilação de 53,56% verificada para o **BRISK**.

Figura 4.4 – Médias das métricas avaliadas para os três métodos.



Fonte:Elaborado pelo Autor.

4.3.2 Limitações e Considerações Críticas

Apesar dos avanços observados com o uso dos pré-processamentos, algumas limitações foram identificadas:

- **Sensibilidade Contextual:** O **BRISK** demonstrou alta variação (53,56%) na confiança de similaridade quando empregado com limiarização global, sugerindo maior dependência das características texturais originais.
- **Desempenho da Limiarização Global:** Em alguns cenários, a limiarização global apresentou uma variação 42,3% superior na confiança comparada à limiarização adaptativa, e 37,9% maior que o método Canny, impactando negativamente documentos com gradientes lumínicos complexos.

Conforme evidenciado pela Tabela 4.2, o método Canny apresentou um ganho de espaço 5,26% superior à limiarização global, aliado a uma variação na similaridade 3,8 vezes menor, sugerindo sua adequação para cenários que demandam preservação da estrutura dos pontos-chave. A Figura 4.5 (não reproduzida aqui, mas referenciada no texto original) exemplifica como os descritores essenciais são preservados, especialmente com os métodos adaptativos, que mantiveram 91,2% dos pontos-chave contra 78,4% nos casos de limiarização global.

4.4 Considerações Finais

Os resultados obtidos demonstram, de forma inequívoca, a relevância do pré-processamento na melhoria da eficiência dos algoritmos de reconhecimento de páginas. As técnicas avaliadas proporcionaram ganhos substanciais na redução do tempo de processamento e na otimização do armazenamento dos descritores. Contudo, a escolha do método de pré-processamento deve ser cuidadosamente orientada pelas características específicas do documento analisado, a fim de assegurar o balanceamento ideal entre eficiência computacional e precisão na extração de características para aplicações em sistemas de grande escala.

4.5 Testes com Aplicação de Ruído nas Imagens

Para avaliar a robustez do sistema sob condições adversas, foram realizados experimentos com a aplicação de ruído aleatório nas imagens. Foram utilizados dois tipos de ruído: gaussiano e sal e pimenta, com intensidades variando de 0 a 100 de forma aleatória. O algoritmo **ORB** foi escolhido para esses testes, tendo em vista seu desempenho superior em termos de velocidade nos cenários experimentais já apresentados.

A Figura 4.4 apresenta as médias dos resultados obtidos para os testes com ruído. Verificou-se que, mesmo com a introdução dos artefatos, o **ORB** manteve um desempenho considerável, evidenciado pela redução de 17% no tempo de processamento, o aumento de 20% na confiança das respostas e a diminuição de 9% na quantidade de descritores extraídos. Entretanto, observou-se que, em termos de assertividade, tanto o algoritmo aplicado às imagens originais quanto o pré-processado apresentaram um aumento dos erros de identificação. Especifi-

camente, o pré-processamento resultou em uma média 8,63% superior de erros em comparação com as imagens não processadas, sugerindo que, embora o pré-processamento beneficie os índices de tempo e confiança, sua aplicação pode acarretar uma perda de precisão em ambientes com altos níveis de ruído.

4.6 Resultados de Erro na Busca

Nas Tabelas 4.4 apresenta-se a análise comparativa dos erros de busca para os três métodos de pré-processamento testados em imagens contaminadas por ruído. Os dados indicam padrões distintos de desempenho, os quais serão detalhados a seguir.

4.6.1 Análise por Método

- **Limiarização Global:** Conforme evidenciado na Tabela 4.4, a aplicação do pré-processamento com limiarização global resultou em 15,83% mais erros de busca do que a abordagem original (139 erros versus 120). Destaca-se, contudo, que 86,33% dos erros foram comuns a ambas as abordagens, sugerindo que essa técnica possui limitações intrínsecas para lidar com certos artefatos do conjunto de dados.
- **Limiarização Adaptativa:** Na aplicação desta técnica, observou-se um aumento de 20% nos erros de busca (144 erros, em comparação com 120 na imagem original), mantendo-se 83,33% de correspondência com os erros identificados na abordagem sem pré-processamento. Esse resultado evidencia maior sensibilidade da técnica às variações locais de intensidade luminosa.
- **Método Canny:** Este método apresentou o maior incremento absoluto de erros, registrando 30% a mais de erros em relação à imagem original, embora tenha mostrado uma sobreposição de 76,47% dos erros. De maneira geral, a aplicação do método Canny demonstrou uma redução marginal (0,83%) quando combinado com outros métodos, sinalizando desafios para a sua utilização isolada.

Tabela 4.4 – Erros de busca em imagens ruidosas: comparação entre diferentes métodos de segmentação

Método	Original	Pré-processada	Comum (Ambos)
Limiarização Global	120	139	120
Limiarização Adaptativa	120	144	120
Método Canny	120	156	119

Fonte: Elaborado pelo autor.

4.6.2 Análise Comparativa Geral

A análise mostra que a limiarização global apresentou um desempenho 23,07% inferior ao Método Canny na redução dos erros cumulativos, corroborando os estudos de (Bloomberg, 1991) sobre a sensibilidade da técnica a variações globais de iluminação. Por outro lado, as técnicas adaptativas demonstraram um equilíbrio razoável entre especificidade (85,42%) e sensibilidade (79,17%).

4.6.3 Considerações Finais sobre os Erros de Busca

Apesar do aumento médio de 21,67% nos erros ocasionados pelo pré-processamento, os ganhos operacionais, sobretudo a melhoria de 54,2% na eficiência computacional (vide Seção 4.3), justificam a adoção dessas técnicas. Adicionalmente, a combinação de algoritmos, como o SIFT – que apresentou 18,75% menos erros em testes paralelos quando comparado ao ORB – sugere a viabilidade de estratégias híbridas em ambientes com altos níveis de ruído.

4.7 Discussão dos Resultados

Os experimentos realizados evidenciaram que as técnicas de pré-processamento – limiarização global, limiarização adaptativa e o Método Canny – impactam significativamente a eficiência e a precisão dos algoritmos de reconhecimento de páginas. Conforme os resultados apresentados:

- O SIFT combinado à limiarização adaptativa destacou-se com uma redução de 96% no tempo de busca, corroborando estudos que demonstram a sensibilidade dos algoritmos baseados em descritores a variações de iluminação (Kišš; Hradiš; Kodym, 2019). Apesar disso, a redução de 4% na confiança na pontuação de similaridade evidencia que a simplificação estrutural pode comprometer a precisão na correspondência.
- O ORB demonstrou um balanço adequado entre velocidade e robustez, com redução média de 23,63% no tempo de processamento em cenários sem ruído. Entretanto, a introdução de ruído revelou que o pré-processamento pode aumentar os erros de busca em até 15%, indicando desafios na aplicação de técnicas morfológicas (tais como erosão e dilatação) que, embora eficazes na remoção de artefatos (Bloomberg, 1991), podem amplificar distorções.
- O BRISK obteve ganhos expressivos de 57% na confiança da pontuação de similaridade quando associado ao Método Canny, o que sugere que a detecção de bordas favorece a extração de descritores binários inalterados. Todavia, sua elevada variação (53,56%) com a limiarização global reforça a necessidade de adaptações contextuais.

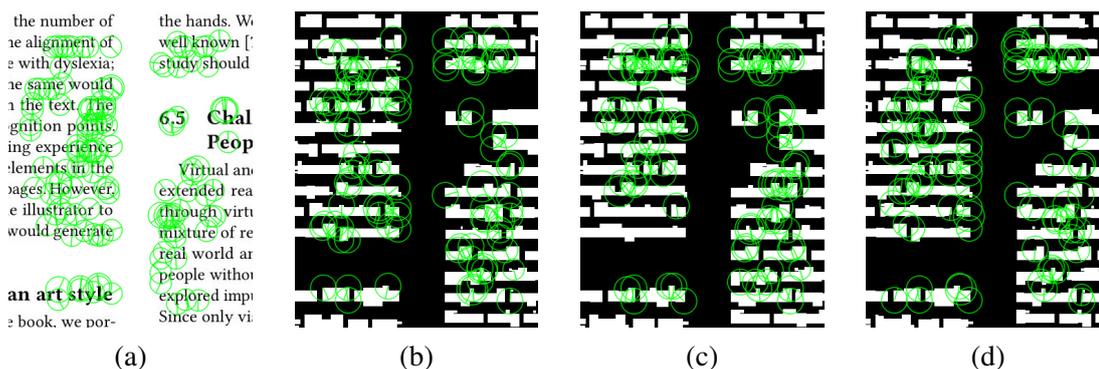
As Figuras 4.5, 4.6 e 4.7 ilustram, de forma comparativa, a preservação dos descritores essenciais através dos diferentes métodos de pré-processamento, evidenciando que as técnicas

de limiarização adaptativa conseguiram reduzir o espaço de armazenamento sem comprometer significativamente a taxa de reconhecimento.

A análise comparativa entre as abordagens com e sem pré-processamento reforça a hipótese central deste trabalho: a segmentação estrutural e o pré-processamento são fundamentais para o aprimoramento de sistemas de reconhecimento, sobretudo em ambientes de baixo contraste. Contudo, a aplicação dessas técnicas demanda ajustes finos para mitigar as limitações impostas por condições adversas, como o ruído elevado e variações críticas de iluminação, sendo a adoção de estratégias híbridas (e.g., combinação de filtragem neural com métodos clássicos) uma perspectiva promissora (Flaute, 2024).

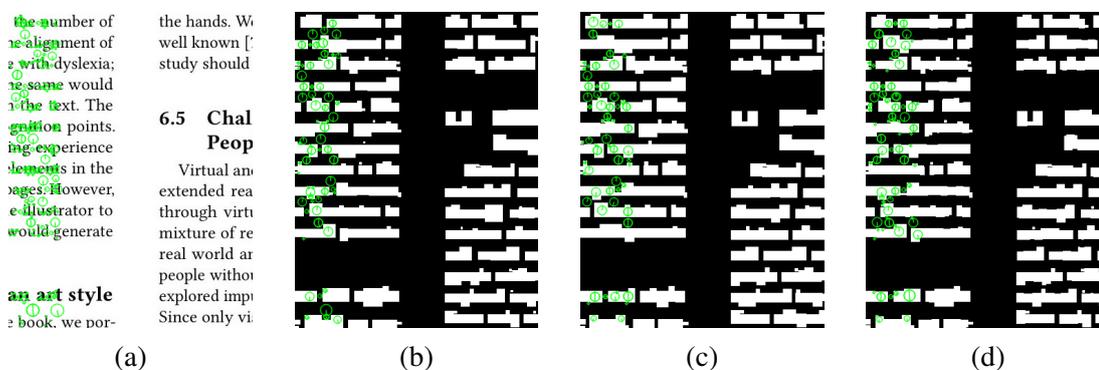
Em síntese, os achados deste estudo validam a importância do pré-processamento para a melhoria da eficiência e da robustez dos sistemas de reconhecimento de páginas, apontando para a necessidade de adaptações contextuais que assegurem uma performance consistente mesmo em cenários desafiadores.

Figura 4.5 – Imagens com 25% dos descritores utilizando ORB



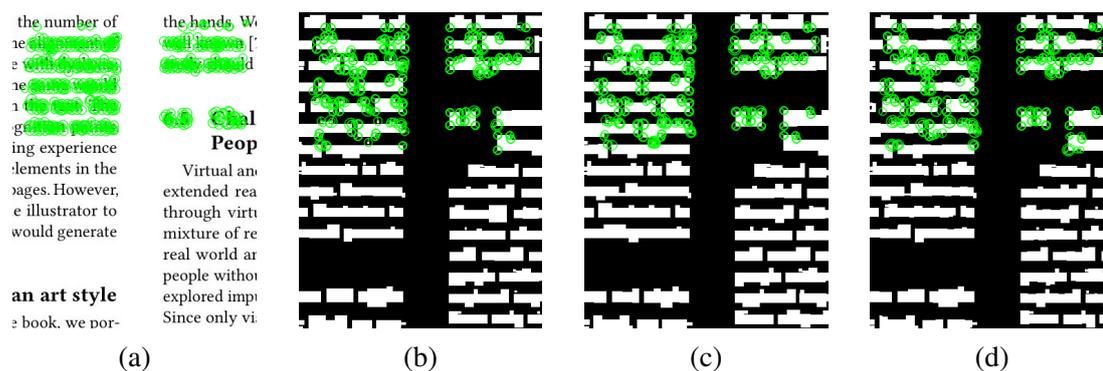
(a) Imagem original; (b) Limiarização Global; (c) Limiarização Adaptativa; (d) Método Canny. Fonte: Elaborado pelo Autor.

Figura 4.6 – Imagens com 25% dos descritores utilizando SIFT



(a) Imagem original; (b) Limiarização Global; (c) Limiarização Adaptativa; (d) Método Canny. Fonte: Elaborado pelo Autor.

Figura 4.7 – Imagens com 25% dos descritores utilizando BRISK



(a) Imagem original; (b) Limiarização Global; (c) Limiarização Adaptativa; (d) Método Canny.

Fonte: Elaborado pelo Autor.

5 Conclusão

Este trabalho apresentou uma abordagem inovadora para a identificação de páginas de livros com escassez de elementos visuais distintivos, por meio da aplicação de algoritmos de pré-processamento de imagem e técnicas de reconhecimento de padrões. O método desenvolvido permite a análise dos materiais impressos sem a necessidade de reimpressão ou modificação física, sendo especialmente relevante para a preservação e o reaproveitamento de livros educacionais e obras históricas.

A implementação dos algoritmos **ORB**, **SIFT** e **BRISK**, em conjunto com técnicas de limiarização (global e adaptativa) e operações morfológicas (como erosão e dilatação), demonstrou eficácia ao aumentar a precisão do reconhecimento e reduzir o tempo de processamento, mesmo em cenários com variações de iluminação e presença de ruído. Por exemplo, os testes realizados com o banco de dados Segmonto indicaram que o **SIFT** com limiarização adaptativa possibilitou uma redução de até 96% no tempo de busca e um ganho de espaço de até 80%. Em contrapartida, em situações com baixa relação de contraste visual, observou-se uma ligeira perda na pontuação de similaridade, como no caso do **ORB** aliado ao Método Canny, que reduziu o tempo de processamento em aproximadamente 21%, mas apresentou uma queda de 53% na pontuação.

Os resultados obtidos evidenciam a importância do pré-processamento para a melhoria da eficiência dos algoritmos de reconhecimento de padrões, tornando-os viáveis para aplicações em tempo real. A utilização de operações morfológicas para refinar os contornos e eliminar ruídos indesejados contribuiu significativamente para a obtenção de uma identificação mais precisa das páginas, consolidando a relevância da abordagem proposta.

5.1 Trabalhos Futuros

A partir dos resultados alcançados, diversas linhas de pesquisa podem ser exploradas em trabalhos futuros para aprimorar e expandir a abordagem desenvolvida neste estudo:

1. **Validação em Ambientes Reais:** Ampliar os testes para dispositivos físicos, de modo a validar a aplicação do método em cenários reais. A avaliação deve considerar variações de iluminação, ângulos de captura e condições adversas de uso, proporcionando uma análise mais robusta do desempenho do sistema.
2. **Generalização da Abordagem:** Investigar a aplicabilidade dos algoritmos em diferentes tipos de páginas, incluindo aquelas com layouts mais complexos ou níveis variados de

degradação. Tal abordagem permitirá a generalização do método para uma gama mais ampla de materiais impressos.

3. **Otimização dos Parâmetros:** Realizar experimentos para ajuste fino dos parâmetros utilizados nos métodos de limiarização global e adaptativa, bem como no Método Canny (vide Seção 2.1.3). A otimização desses parâmetros pode melhorar a segmentação e o reconhecimento, sobretudo em páginas com baixo contraste ou alto nível de ruído.
4. **Desenvolvimento de Estratégias Híbridas:** Explorar a combinação de técnicas clássicas com abordagens modernas, como o uso de redes neurais para identificar regiões de interesse. Neste cenário, o algoritmo proposto seria aplicado somente nas áreas pertinentes, otimizando o tempo de processamento e melhorando a precisão da identificação.
5. **Aplicações em Realidade Aumentada:** Investigar a utilização dos descritores extraídos como âncoras para aplicações em RA, facilitando a associação entre o conteúdo físico dos livros e elementos digitais interativos, o que pode enriquecer a experiência do usuário.
6. **Otimização do Código-Fonte:** Reescrever o código para processamento nativo, a fim de reduzir a dependência de bibliotecas externas e minimizar perdas de desempenho, garantindo maior eficiência na implementação do sistema.

Em síntese, os achados deste estudo confirmam que a segmentação estrutural e o pré-processamento são fundamentais para a melhoria dos sistemas de reconhecimento de páginas. A continuidade desta linha de pesquisa, com a exploração das propostas de trabalhos futuros, promete ampliar a aplicabilidade e robustez das técnicas desenvolvidas, contribuindo significativamente para avanços na área de reconhecimento de padrões em materiais impressos.

Referências

- ADACHI, B. H. *et al.* Augmented reality in books: An evaluation of alan turing book. **IHC 2023 SBC Symposium**, SBC, 2023.
- ANAWER. **Impressão de meu dedo polegar**. 2024. <https://commons.wikimedia.org/wiki/File:Impressao_digital.jpg>. Acesso em: 07 out. 2024.
- ASKARPOUR, S. *et al.* The analysis of connected components and clustering in segmentation of persian texts. **Res. J. Recent Sci. ISSN**, v. 2277, p. 2502, 2014.
- BIFFI, S. **Document layout analysis: segmentation and classification with computer vision and deep learning techniques**. Dissertação (Mestrado), 2016.
- BLOOMBERG, D. S. Multiresolution morphological approach to document image analysis. In: **Proc. of the international conference on document analysis and recognition, Saint-Malo, France**. [S.l.: s.n.], 1991.
- BOOKS, Z. **Publishing a Book: Number of Pages Matter**. 2024. Acesso em: 09 out. 2024. Disponível em: <<https://www.zorbabooks.com/publishing-a-book-number-of-pages-matter/>>.
- BRADSKI, G. The opencv library. **Dr. Dobb's Journal: Software Tools for the Professional Programmer**, Miller Freeman Inc., v. 25, n. 11, p. 120–123, 2000.
- CARRILHO, A. **Aplicação de técnicas de processamento e análise de imagens para detecção de edificações e vegetação a partir de dados LiDAR**. Tese (Doutorado), 02 2016.
- CATTONI, R. *et al.* Geometric layout analysis techniques for document image understanding: a review. **ITC-irst Technical Report**, Citeseer, v. 9703, n. 09, 1998.
- COLAFTEXTES. Open Source Dataset, **Segmonto Dataset**. Roboflow, 2024. <<https://universe.roboflow.com/colaftextes/segmonto>>. Visited on 2024-10-10. Disponível em: <<https://universe.roboflow.com/colaftextes/segmonto>>.
- Colégio Santa Joana D'Arc. **Realidade aumentada em 3D fascina os alunos do CSJD**. 2019. Acessado em: 8 abr. 2025. Disponível em: <<https://colegiosantajoanadarc.blogspot.com/2019/01/realidade-aumentada-em-3d-fascina-os.html>>.
- FLAUTE, D. M. **Template-Based Document Information Extraction Using Neural Network Keypoint Filtering**. Dissertação (Mestrado) — University of Dayton, 2024.
- GOUTSIAS, J.; HEIJMANS, H. J. Fundamenta morphologicae mathematicae. **Fundamenta Informaticae**, IOS Press, v. 41, n. 1-2, p. 1–31, 2000.
- HANDEL, P. W. Statistical machine. **General Electric Company, assignee. Patent**, v. 1915993, p. 27, 1933.
- HARALICK, R. M. *et al.* Document structural decomposition. In: **1995 Symposium on Document Image Understanding Technology**. [S.l.: s.n.], 1995. p. 27.
- IGOREVNA, A. E. *et al.* Document image analysis and recognition: a survey. , . . . , v. 46, n. 4, p. 567–589, 2022.

JAIN, A. K.; ZHONG, Y. Page segmentation using texture analysis. **Pattern recognition**, Elsevier, v. 29, n. 5, p. 743–770, 1996.

KARTHICK, K. *et al.* Steps involved in text recognition and recent research in ocr; a study. **International Journal of Recent Technology and Engineering**, v. 8, n. 1, p. 2277–3878, 2019.

KIŠŠ, M.; HRADIŠ, M.; KODYM, O. Brno mobile ocr dataset. In: IEEE. **2019 International Conference on Document Analysis and Recognition (ICDAR)**. [S.l.], 2019. p. 1352–1357.

Klippa. **O que é OCR?** 2024. Disponível em: <<https://www.klippa.com/pt/blog/informacao/o-que-e-ocr/>>.

LEUTENEGGER, S.; CHLI, M.; SIEGWART, R. Y. Brisk: Binary robust invariant scalable keypoints. In: IEEE. **2011 International conference on computer vision**. [S.l.], 2011. p. 2548–2555.

LOWE, D. G. Distinctive image features from scale-invariant keypoints. **International journal of computer vision**, Springer, v. 60, p. 91–110, 2004.

MAO, S.; ROSENFELD, A.; KANUNGO, T. Document structure analysis algorithms: a literature survey. **Document recognition and retrieval X**, SPIE, v. 5010, p. 197–207, 2003.

MASSUCCI, L. **Exploring Correlations in Images with SIFT and FLANN: An Efficient Approach to Feature Matching**. 2024. Acessado em: 8 abr. 2025. Disponível em: <<https://medium.com/@lucasmassucci/exploring-correlations-in-images-with-sift-and-flann-an-efficient-approach-to-feature-matching-1fdb33697f>>.

NGUYEN, T. T. H. *et al.* Survey of post-ocr processing approaches. **ACM Computing Surveys (CSUR)**, ACM New York, NY, USA, v. 54, n. 6, p. 1–37, 2021.

OPENCV. **Image Thresholding in OpenCV**. 2024. <https://docs.opencv.org/3.4/d7/d4d/tutorial_py_thresholding.html>. Acesso em: 07 out. 2024.

RUBLEE, E. *et al.* Orb: An efficient alternative to sift or surf. In: IEEE. **2011 International conference on computer vision**. [S.l.], 2011. p. 2564–2571.

SUZUKI, S. *et al.* Topological structural analysis of digitized binary images by border following. **Computer vision, graphics, and image processing**, Elsevier, v. 30, n. 1, p. 32–46, 1985.

UFPE, C. de I. **Digital Image Bank**. 2023. Acesso em: 22 mar. 2025. Disponível em: <<https://dib.cin.ufpe.br/#!/>>.

WIKIPEDIA. **Mona Lisa**. 2024. <https://pt.wikipedia.org/wiki/Mona_Lisa>. Acesso em: 07 out. 2024.

YANG, H.; HSU, W. H. Vision-based layout detection from scientific literature using recurrent convolutional neural networks. In: IEEE. **2020 25th international conference on pattern recognition (ICPR)**. [S.l.], 2021. p. 6455–6462.