

# Image Stitching Baseado em Bifurcações Vasculares Aplicado a Retinografias de Baixa Resolução

Guilherme G. S. Nunes<sup>1</sup>, João D. S. Almeida<sup>1</sup>, Darlan B. P. Quitanilha<sup>1</sup>, António Cunha<sup>1</sup>

<sup>1</sup>Núcleo de Computação Aplicada – Universidade Federal do Maranhão (UFMA)  
CEP 65085-580 – São Luís – MA – Brasil

guilherme.gabriel@discente.ufma.br, jdallyson@nca.ufma.br,  
darlan.quintanilha@ufma.br

<sup>2</sup>Universidade de Trás-os-Montes e Alto Douro - Quinta de Prados  
5000-801 - Vila Real - Portugal

acunha@utad.pt

**Abstract.** *Image stitching is a technique that allows combining multiple images, forming a single image with a wide field of view. In the context of retinography, this technique is crucial to capture a detailed view of the retina, allowing broader examinations to be performed. In this work, we present a method for stitching low-resolution images, using bifurcation points as image features. The proposed method presents an increase in the correspondences obtained in relation to the detectors in the literature, obtaining the RMSE result with a reduction of approximately 14% compared to SIFT (24.29) and ORB (24.32). Furthermore, the proposed method obtained a higher average PSNR, reaching 26.27 for glaucoma images, 26.72 for normal images and 26.71 for images of suspected patients, while the methods based on SIFT and ORB presented lower values, confirming the effectiveness of the proposed approach.*

**Resumo.** *O image stitching é uma técnica que permite combinar múltiplas imagens, formando uma imagem única com um campo de visão amplo. No contexto da retinografia, essa técnica é crucial para capturar uma visão detalhada da retina, permitindo que exames mais amplos sejam realizados. Neste trabalho, é apresentado um método de costura de imagens de baixa resolução, utilizando pontos de bifurcação como características da imagem. O método proposto apresenta um aumento das correspondências obtidas em relação aos detectores da literatura, obtendo o resultado de RMSE com uma redução de aproximadamente 14% em comparação ao SIFT (24,29) e ORB (24,32). Além disso, o método proposto obteve um PSNR médio superior, atingindo 26,27 para imagens de glaucoma, 26,72 para imagens normais e 26,71 para imagens de pacientes suspeitos, enquanto os métodos baseados em SIFT e ORB apresentaram valores inferiores, confirmando a eficácia da abordagem proposta.*

## 1. Introdução

Com o aumento nos números de doenças que podem afetar a visão humana, o olho humano vem sendo estudado para a criação de técnicas voltadas ao aprendizado de máquina visando auxiliar especialistas que atuam na detecção de tais problemas. Como foco para

essas enfermidades, temos o glaucoma, que representa uma grande quantidade de casos que levam pacientes à perda da visão, sendo considerado como causa principal da perda de visão daqui a algumas décadas, devido a tal doença ser silenciosa, podendo chegar a 118,8 milhões de pessoas afetadas [Wu et al. 2022].

Visando combater a cegueira que é causada pela patologia, exames podem ser realizados com a ajuda de um especialista para verificar e detectar os problemas previamente, analisando a estrutura da retina. Graças ao trabalho humano posto nesta ação, algoritmos que utilizam Deep Learning para a segmentação das estruturas do olho têm sido desenvolvidos e utilizados por pesquisadores atuantes no campo oftalmológico devido à sua praticidade e precisão [Sarkar 2021].

O diagnóstico precoce para identificação de patologias na visão é fundamental para a cura e tratamento das mesmas em um paciente. Com o uso de técnicas de mosaico em imagens de baixa resolução, facilitaria o procedimento de classificação, auxiliando o oftalmologista, visto que o custo para obter as imagens seria reduzido, possibilitando o uso da câmera de um dispositivo baseado em smartphone. Levando em consideração a alta quantidade de exames e cirurgias relacionadas ao olho humano [Costa et al. 2019], a integração de técnicas utilizadas neste artigo com as retinografias de alta e baixa resolução pode gerar impacto positivo para pessoas que apresentam problemas na visão.

Para isso, existem técnicas para a realização do *image stitching*, que consistem na deformação de imagens para alinhar suas perspectivas, criando uma única imagem panorâmica ou de maior resolução. Portanto, é necessário implementar o *stitching* em imagens retinianas de forma a obter um resultado da área retiniana ampla e precisa. Tal como a segmentação de imagens retinianas, um campo de pesquisa amplamente explorado, sendo fundamental para auxiliar diagnósticos oftalmológicos. Trabalhos como o de Fonseca et al. (2022) demonstram a eficácia de métodos de segmentação baseados em aprendizado profundo para a análise de estruturas da retina.

A principal contribuição deste trabalho é a proposta de um método para o *stitching* de imagens de baixa resolução utilizando pontos de bifurcação como características. Essa abordagem mostra um aumento na correspondência de características em comparação com detectores da literatura, reduzindo o erro médio quadrático (RMSE) em aproximadamente 14% em relação ao SIFT e ORB.

Este artigo está organizado da seguinte forma: na Seção 2, são apresentados os trabalhos relacionados a *Image stitching*. Na Seção 3, são descritos os materiais e o método. Na Seção 4, são apresentados os experimentos e resultados para validar o método proposto, além de um estudo comparativo com detectores de características. Finalmente, a conclusão e trabalhos futuros são apresentados na Seção 5.

## 2. Trabalhos Relacionados

A pesquisa de Feng et al. (2020) apresenta uma técnica para a extração de características a partir de pontos de bifurcação, utilizando processos de segmentação em imagens de alta resolução. O método faz uso de operações morfológicas para identificar e localizar os pontos característicos, resultando na criação de mosaicos de imagens de retina.

Os autores Jalili et al. (2020) apresentam um método para a criação de mosaicos de imagens retinianas, visando ampliar o campo de visão, tornando a retina mais ampla. A

técnica emprega o algoritmo Scale-Invariant Feature Transformation (SIFT) para identificar características nas imagens retinianas. Em seguida, utiliza-se o diagrama de Voronoi para dividir as imagens em regiões, permitindo uma correspondência eficiente entre pares de regiões adjacentes.

Hu et al. (2022) propõem um método para aprimorar a capacidade diagnóstica de câmeras de fundo baseadas em smartphones e oftalmoscópios, unindo múltiplas imagens retinianas em um mosaico coeso. A técnica utiliza uma rede neural convolucional para estimar a homografia entre as imagens, facilitando o alinhamento e a combinação de retinografias.

Agarwal et al. (2022) apresentam técnicas de detecção e combinação de regiões de imagens retinianas de baixa qualidade, utilizando dispositivos portáteis. O estudo emprega redes neurais convolucionais para localizar a região da retina e técnicas de stitching para criar imagens panorâmicas com maior campo de visão.

Em Zhang et al. (2022) é apresentada uma estrutura para a reconstrução de imagens retinianas de alta resolução a partir de múltiplas imagens de baixa resolução capturadas com câmeras móveis de baixo custo. O estudo utiliza o rastreamento do disco óptico para selecionar as melhores visualizações e combina as imagens utilizando técnicas de stitching para gerar uma imagem de alta qualidade.

Em Santos et al. (2024) os autores propõem um método que utiliza a estimativa de homografia para registrar e unir essas imagens em um mosaico coeso. Utilizando uma rede convolucional para estimar a matriz de homografia das imagens. Os autores utilizam imagens de baixa resolução, demonstrando os desafios do stitching em retinografia, especialmente quando se trata de imagens adquiridas com dispositivos portáteis, como a câmera fundoscópica D-EYE acoplada a smartphones.

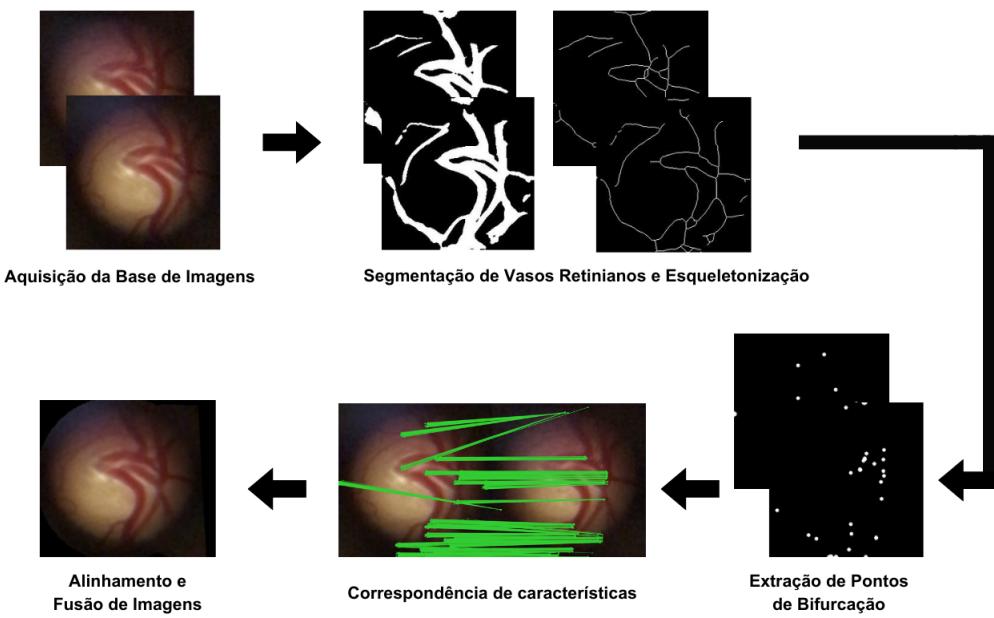
Os trabalhos supracitados se baseiam em métodos como SIFT e aprendizado profundo para a extração de características, ao passo que este estudo investiga uma opção que se baseia em segmentação semântica e operações morfológicas para detectar pontos de bifurcação. Ao contrário de estudos que utilizam segmentação em imagens de alta resolução, este estudo se concentra na extração de características em retinografias de baixa resolução obtidas por um dispositivo móvel, onde a qualidade inferior apresenta desafios extras no processo de stitching. Portanto, o objetivo é propor uma nova abordagem de *image stitching* que produza um mosaico a partir de retinografias de baixa resolução, possibilitando um melhor alinhamento em situações onde métodos convencionais podem não funcionar devido à baixa qualidade das imagens.

### **3. Materiais e Método**

Nesta seção, são descritos os procedimentos adotados para o desenvolvimento dos mosaicos retinianos. O método proposto está organizado nas seguintes etapas: aquisição da base de imagens, segmentação de vasos retinianos e esqueletização, extração de pontos de bifurcação, *feature matching*, alinhamento e fusão de imagens. A Figura 1 ilustra as etapas do método proposto.

#### **3.1. Aquisição da Base de Imagens**

Nos experimentos, foi selecionado o dataset privado D-Eye [Russo et al. 2015], no qual exames de retinografia foram feitos com um dispositivo conectado à câmera de um



**Figura 1. Etapas do método proposto.**

*smartphone*. O método proposto, aplicado neste dataset de baixa resolução, visa melhorar a área da retina visível para a análise do especialista.

O dataset é composto por vídeos de olhos humanos capturados para a detecção de glaucoma, sendo dividido em três conjuntos de vídeos: glaucoma, suspeito e normal. Para obter as imagens, utiliza-se a Yolov8 [Jocher 2023], o modelo foi treinado em 150 épocas utilizando *frames* extraídos manualmente dos vídeos presentes no dataset para treinamento.

Cada paciente teve seis imagens extraídas por olho, totalizando 2060, 228 e 708 imagens para os conjuntos normal, glaucoma e suspeito, respectivamente, possuindo variações nas dimensões entre 224x470 até 570x600.

### 3.2. Segmentação de Vasos Retinianos e Esqueletização

Na etapa de segmentação dos vasos, as imagens foram dispostas a um modelo já treinado para segmentação de vasos retinianos, sendo utilizada a U-Net padrão [Ronneberger et al. 2015], no qual os vasos sanguíneos são extraídos. A rede foi treinada por 150 épocas, utilizando como base a implementação feita por [Sarkar 2021]. A realização dessa etapa se mostra necessária para a extração de características, visto que o processo realizado na identificação das bifurcações é obtido por meio de uma imagem binária, onde os vasos sanguíneos são identificados como pontos brancos.

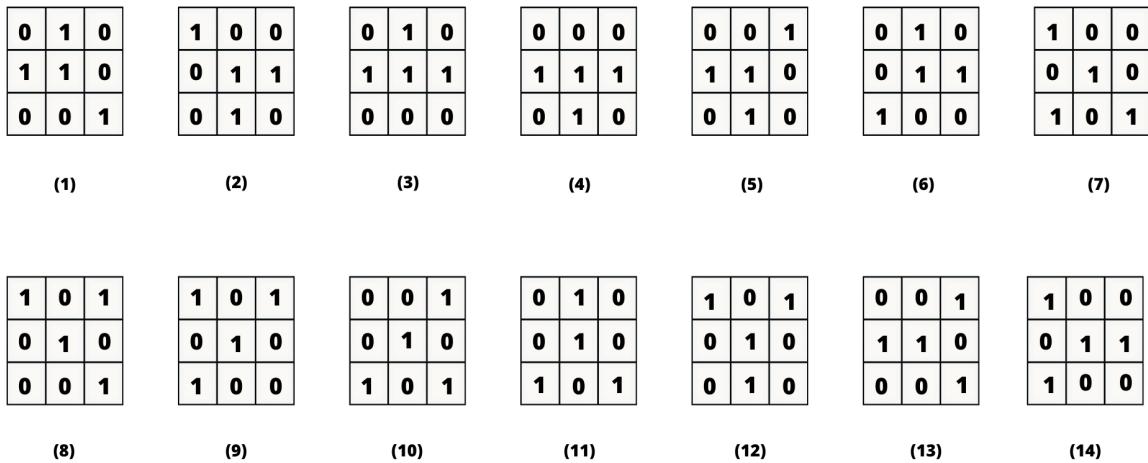
A segmentação dos vasos sanguíneos se mostrou essencial para a extração de características, visto que a quantidade de pontos obtidos em uma imagem não segmentada se mostrou menor em relação a uma com os vasos sanguíneos já detectados. O modelo utilizado para a segmentação se baseia na implementação feita por [Sarkar 2021], com alterações no tamanho do *patch* e no conjunto de dados.

A esqueletização [Abu-Ain et al. 2013] é aplicada na imagem binária, resultante da segmentação dos vasos sanguíneos, sendo um processo que visa suavizar ele-

mentos de uma imagem. Pela extração de características por pontos de bifurcação, os vasos sanguíneos se mostravam espessos para a realização da etapa; com isso, a esqueletização foi realizada para afinar os vasos retinianos.

### 3.3. Extração de Características

Os pontos de bifurcação foram separados em 14 elementos estruturantes, sendo cada um representado por uma matriz de 3x3 pixels, conforme proposto por Gonzalez (2009). Foram definidos formatos em T e Y, distribuídos nas estruturas, como mostra a Figura 2.



**Figura 2. Estruturas**

As imagens resultantes da etapa de esqueletização são transformadas em binárias e submetidas a uma operação morfológica de erosão, resultando em 14 imagens, cada uma com seu respectivo elemento estruturante. Por fim, as imagens são somadas, gerando uma imagem final com apenas os pontos de bifurcação, sendo esses pontos característicos que serão utilizados nas etapas seguintes.

Foram utilizadas como base para o processo de matching o cálculo da distância euclidiana, como descrito na seguinte equação:

$$d = \sqrt{(x[0] - x[1])^2 + (y[0] - y[1])^2}$$

onde X e Y representam o conjunto de pontos de características extraídos de uma imagem referência e alvo, respectivamente. O cálculo é realizado utilizando as coordenadas dos pontos, sendo X[0] e Y[0] os valores correspondentes ao eixo das abscissas e X[1] e Y[1] o eixo das ordenadas.

O cálculo é feito para cada ponto da imagem referência em relação a todos os pontos da imagem alvo, sendo condicionado ao resultado da distância euclidiana mínima ser menor ou igual a 0,3 baseado nos experimentos de Feng et al. (2020), para ser aproveitado para a realização do *stitching*.

### 3.4. Alinhamento e Fusão de Imagens

A etapa de *Warping* foi iniciada com a transformação homográfica, sendo esta necessária para permitir a deformação em perspectiva. A construção da matriz de homografia per-

mitiu o mapeamento entre duas imagens distintas, possibilitando a fusão de diferentes perspectivas em uma única representação coerente.

Dada uma correspondência entre dois planos projetivos, podemos definir a homografia como uma matriz  $H$  de  $3 \times 3$ , tal que para cada ponto  $x$  no plano original e seu correspondente  $x'$  no plano transformado, tem-se:

$$\begin{bmatrix} x' \\ y' \\ w' \end{bmatrix} = \begin{bmatrix} h_{11} & h_{12} & h_{13} \\ h_{21} & h_{22} & h_{23} \\ h_{31} & h_{32} & h_{33} \end{bmatrix} \begin{bmatrix} x \\ y \\ w \end{bmatrix} \quad (1)$$

onde  $w$  e  $w'$  são fatores de escala homogêneos.

O algoritmo RANSAC proposto por Fischler and Bolles (1981) é utilizado para estimar a matriz homográfica de maneira robusta, diminuindo a influência de correspondências errôneas. Inicialmente, um pequeno subconjunto de pontos correspondentes é selecionado para estimar uma matriz de transformação. O algoritmo RANSAC foi aplicado para avaliar a matriz de homografia em busca de encontrar pontos de correspondência válidos no método proposto, SIFT e ORB.

Em seguida, a matriz estimada é avaliada, projetando os demais pontos e verificando quais deles se ajustam ao modelo em um limite de erro aceitável. Sendo repetido o mesmo processo diversas vezes, buscando maximizar a quantidade de pontos corretamente mapeados, garantindo assim que a matriz homográfica final seja determinada de forma confiável, mesmo na presença de pontos que não possuem correspondências entre imagens.

Para alcançar transição suave entre imagens mescladas, adotou-se a técnica de *Multiband Blending* proposta por Burt and Adelson (1983), que utiliza pirâmides gaussianas e laplacianas para estruturar as imagens em diferentes frequências espaciais. Portanto, a fusão é realizada ao combinar gradualmente as frequências, de forma que garanta a preservação dos detalhes finos e das transições suaves nas áreas de intersecção. Essa abordagem visa suavizar a sobreposição das bordas onde as imagens se encontram, evitando discrepâncias.

A efetividade do método de *Multiband Blending* é conduzida pelo número de bandas utilizadas durante a decomposição da imagem. Para imagens de baixa resolução, são utilizadas apenas três bandas para assegurar uma fusão precisa e a preservação dos detalhes em várias escalas. Um número menor de camadas se mostra adequado para garantir uma transição suave e reduzir a complexidade computacional envolvida no processo, além de diminuir o desfoque final.

O processo do blending, baseado no método de [Feng et al. 2020], pode ser descrito nos seguintes passos:

1. As imagens de entrada  $I_A$  e  $I_B$  são ajustadas para possuírem as mesmas dimensões. Máscaras associadas  $M_A$  e  $M_B$  também são redimensionadas;
2. Cada imagem passa por sucessivas reduções de escala, aplicando um filtro *low-pass* seguido de subamostragem;

$$P_G^l(I) = \text{pyrDown}(P_G^{l-1}(I)) \quad (2)$$

Esse processo resulta em uma representação *multi-scale* da imagem, útil para suavização e remoção de detalhes de alta frequência.

3. A pirâmide *Laplacian* é obtida pela subtração entre cada nível da pirâmide *Gaussian* e sua versão ampliada:

$$P_L^l(I) = P_G^l(I) - \text{pyrUp}(P_G^{l+1}(I)) \quad (3)$$

Essa pirâmide armazena os detalhes de alta frequência, permitindo reconstruir a imagem original após o *blending*.

4. Os coeficientes de cada pirâmide de Laplace são combinados usando máscaras suavizadas para garantir transições naturais entre imagens:

$$C^l = P_L^l(I_A) \cdot M_A^l + P_L^l(I_B) \cdot M_B^l \quad (4)$$

5. A pirâmide fundida é reconstruída somando iterativamente os detalhes em diferentes escalas:

$$I_{blend}^l = \text{pyrUp}(I_{blend}^{l+1}) + C^l \quad (5)$$

Esse processo garante uma fusão suave das imagens, preservando detalhes importantes.

### 3.5. Métricas de Avaliação

As métricas foram utilizadas nas imagens resultantes de *stitching*, totalizando 521, 207 e 247 para os conjuntos normal, glaucoma e suspeito respectivamente.

A etapa de avaliação das técnicas de *stitching* apresentadas foi realizada com base no estudo de Feng et al. (2020), usando as métricas RMSE. A equação do RMSE utilizada pode ser denotada por:

$$RMSE = \sqrt{\frac{1}{n} \sum_{k=1}^n e^2 k}$$

onde n simboliza o número de amostras para calcular o erro dos pontos de correspondências e k representa o índice da correspondência, k = 1, 2 ... para cada ponto identificado como correspondência ( $X_k, Y_k$ ). O símbolo “e” representa a matriz de homografia dos pontos de correspondências. Portanto, o RMSE seria calculado entre os pontos projetados a partir de uma homografia e os pontos reais correspondentes. A métrica RMSE avalia as correspondências de forma que, quanto menor o resultado obtido utilizando a métrica, mais desejável é a correspondência.

Outra métrica utilizada foi o *Peak Signal-to-Noise Ratio* (PSNR), uma métrica empregada na análise de qualidade de imagens e vídeos. O PSNR mede a relação entre a informação da imagem e o ruído introduzido durante o processamento, sendo expresso em decibéis. Quanto maior o valor do PSNR, melhor a qualidade da imagem reconstruída, indicando que ela está mais próxima da original.

O cálculo do PSNR parte do *Mean Squared Error* (MSE), que quantifica a diferença média ao quadrado entre os pixels das duas imagens. Essa métrica é definida como:

$$MSE = \frac{1}{m \times n} \sum_{i=0}^{m-1} \sum_{j=0}^{n-1} [I(i, j) - K(i, j)]^2 \quad (6)$$

onde  $I(i, j)$  e  $K(i, j)$  representam os valores de pixel das imagens original e processada, respectivamente, e  $m$  e  $n$  correspondem às dimensões da imagem.

Com o MSE calculado, o PSNR é obtido pela seguinte equação:

$$PSNR = 20 \cdot \log_{10} \left( \frac{MAX}{\sqrt{MSE}} \right) \quad (7)$$

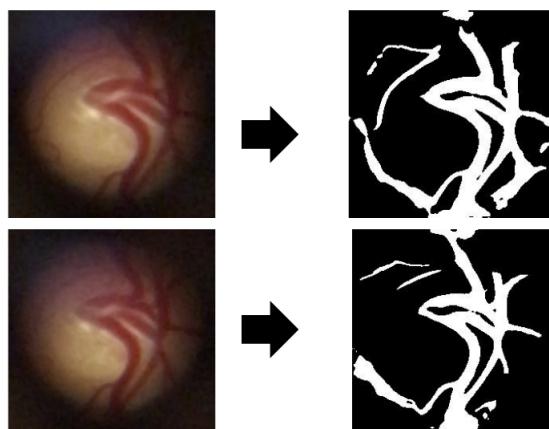
onde MAX representa o valor máximo possível de um pixel na imagem.

#### 4. Experimentos e Resultados

Nesta seção, são apresentados os resultados obtidos pelo método proposto e verifica-se sua efetividade em relação a outros métodos já apresentados na literatura. Destaca-se que o propósito dos experimentos é ampliar a área da retina de modo a aprimorar a qualidade dos relatórios médicos em retinografias de baixa resolução.

Os experimentos foram realizados no ambiente google colaboratory utilizando somente CPU. O tempo de processamento em média (em segundos) para o método, sift e orb foram respectivamente 3,66 , 0,76 e 0,35.

A segmentação dos vasos sanguíneos se mostrou efetiva para a realização do trabalho, a Figura 3 mostra que mesmo em retinografias de baixa resolução foi possível segmentar os vasos mais robustos. O modelo proposto por Sarkar (2021) que foi treinado para os *datasets* DRIVE, STARE e HRF, se mostrou bem eficaz para a segmentação das imagens de *datasets* diferentes, como o D-EYE, sendo segmentados somente os vasos sanguíneos, excluindo o disco óptico, conforme ilustrado na Figura 3. Devido à baixa resolução, a segmentação das imagens extraídas do dataset D-Eye se mostrou eficiente para os vasos sanguíneos mais robustos.

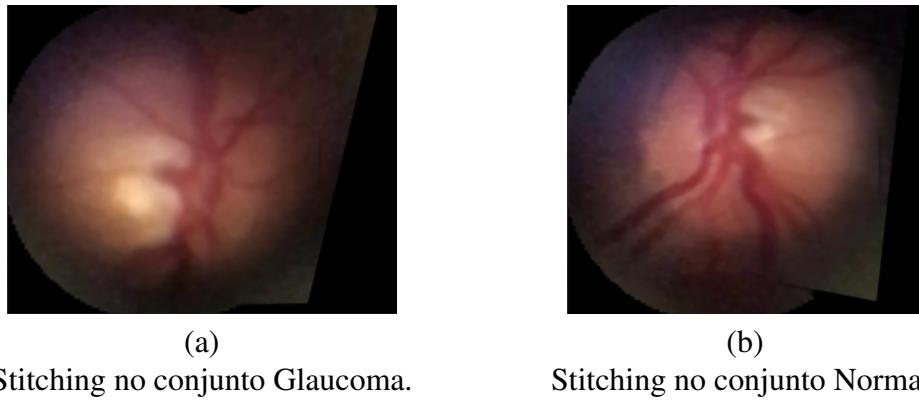


**Figura 3. Segmentação**

Após a etapa de stitching das retinografias de baixa resolução, o mosaico resultante apresentou imagens com uma área ampla de retina, preservando a continuidade es-

trutural das imagens e permitindo uma melhor visualização. Mesmo com a limitação na qualidade das imagens originais, a fusão dos quadros manteve a coerência estrutural e permitiu uma análise geral da retina.

No entanto, observou-se que imagens mais borradadas impactaram negativamente a precisão do alinhamento, dificultando a sobreposição, tal como a extração de correspondências. Ainda assim, em áreas com contraste suficiente, o método demonstrou robustez na reconstrução do mosaico final, como mostra a Figura 4.



**Figura 4. Exemplos de mosaicos geradas a partir do método proposto.**

A extração de pontos se mostrou bem eficaz em relação aos detectores da literatura, conforme mostrado na Tabela 1. O método com base na extração baseada em pontos de bifurcação retorna mais *features*, logo torna mais acessível a busca por perspectiva na criação da matriz de homografia, visto que aumenta significativamente o número de correspondências obtidas, facilitando a aquisição dos quatro pontos obrigatórios para a criação da matriz .

**Tabela 1. Número de correspondências em média no dataset D-Eye**

Conjunto	SIFT	ORB	Método proposto
Conjunto Glaucoma	2,96	0,15	<b>192,66</b>
Conjunto Normal	1,94	0,09	<b>162,78</b>
Conjunto Suspeito	2,01	0,23	<b>154,90</b>

A Tabela 2 apresenta a média do RMSE para os três conjuntos de imagens no dataset D-eye, comparando o método utilizado com detectores automáticos SIFT e ORB. O RMSE obtido entre os pontos projetados a partir da matriz de homografia e pontos reais correspondentes indica a precisão do *stitching* entre as imagens. A Tabela também mostra o número de instâncias, que é a quantidade de imagens utilizadas no cálculo da média aritmética, onde os detectores apresentam valores inferiores devido à impossibilidade de criar uma matriz de homografia com poucos pontos de correspondência extraídos.

O Método apresentou os menores valores de RMSE em todos os conjuntos de dados, com destaque para o Conjunto Glaucoma e o Conjunto Suspeito. Embora o ORB tenha obtido o melhor resultado no Conjunto Normal, utilizou um número de instâncias significativamente inferior ao Método, o que pode ter impactado sua precisão. O Método

**Tabela 2. Média do RMSE no dataset D-eye**

Conjunto	SIFT (Instâncias)	ORB (Instâncias)	Método proposto (Instâncias)
Conjunto Glaucoma	24,29 (5)	24,32 (6)	20,88 (22)
Conjunto Normal	25,81 (45)	10,24 (104)	21,21 (155)
Conjunto Suspeito	21,49 (24)	12,22 (68)	19,65 (75)

se destacou por usar um número maior de instâncias e, consequentemente, obteve um RMSE competitivo e mais robusto.

O Conjunto Normal apresentou uma diferença considerável no desempenho do ORB em relação ao método, o que pode indicar uma menor generalização desse descritor ao lidar com um volume mais significativo de dados. No Conjunto Glaucoma, onde o número de instâncias foi relativamente pequeno, o impacto do método adotado se mostrou mais relevante para a precisão dos resultados obtidos.

Essa análise reforça a importância do número de instâncias na comparação, visto que, embora o ORB tenha mostrado resultados melhores em termos de RMSE, seu desempenho pode não ser tão confiável devido ao uso limitado de dados. Mesmo apresentando o RMSE menor no conjunto normal, o número de instâncias é menor, o que demonstra a efetividade do método, que obteve um valor aproximado com mais instâncias. A média do RMSE foi calculada considerando valores menores ou iguais a 50 a termos de comparação entre os três métodos.

A Tabela 3 apresenta as métricas estatísticas do *Peak Signal-to-Noise Ratio* (PSNR) para diferentes métodos de registro de imagens em três categorias: *Glaucoma*, *Normal* e *Suspeito*. São comparados os métodos *SIFT*, *ORB* e um terceiro método, destacado em negrito, que obteve os melhores resultados em todas as categorias.

**Tabela 3. Métricas estatísticas do PSNR**

Conjunto	Média	Desvio Padrão	Mínimo	Máximo
[Santos et al. 2024]	22,90	3,29	18,04	22,69
Glaucoma SIFT	24,39	3,91	14,34	33,76
Glaucoma ORB	23,39	3,73	13,97	32,97
<b>Glaucoma Método</b>	<b>26,27</b>	4,17	14,50	33,46
Normal SIFT	25,70	3,68	11,90	34,06
Normal ORB	24,75	3,67	10,94	35,17
<b>Normal Método</b>	<b>26,72</b>	3,69	12,08	35,58
Suspeito SIFT	26,26	3,54	14,11	34,39
Suspeito ORB	24,69	3,97	11,54	34,16
<b>Suspeito Método</b>	<b>26,71</b>	3,97	13,45	36,39

Nota-se que o método proposto apresenta os maiores valores médios de PSNR em todas as categorias, indicando uma melhor qualidade da fusão das imagens. Além disso, os valores de desvio padrão são próximos entre os métodos, sugerindo uma variação

semelhante entre as amostras.

O método proposto, também, apresenta resultado médio de PSNR superior ao trabalho de [Santos et al. 2024], que utiliza um filtro de desfoque, reforçando a eficácia do método proposto. Além disso, é importante destacar que o trabalho citado foi conduzido utilizando um subconjunto de imagens, composto por recortes extraídos manualmente, enquanto o método proposto foi avaliado sobre o conjunto completo do dataset. Essa diferença metodológica pode influenciar diretamente os resultados, uma vez que a utilização do dataset completo abrange uma maior diversidade de variações presentes nas imagens, tornando a avaliação mais abrangente e representativa.

## 5. Conclusão

*Image stitching* de retina é uma tarefa desafiadora, principalmente em retinografias de baixa resolução devido à escassez de correspondências. Nesse contexto, métodos eficazes são necessários para produzir resultados precisos.

A pesquisa apresentada visou demonstrar o uso da técnica de *image stitching*, buscando ampliar o campo de visão de imagens em baixa resolução obtidas de exames de retinografia, utilizando pontos de bifurcação como características para realizar o stitching.

Apesar de na etapa de segmentação fora utilizado o modelo U-net treinado para retinografias de alta resolução, se mostrou eficaz ao segmentar imagens de resoluções menores. Portanto, a utilização de pontos de bifurcação para a construção de mosaicos se mostrou eficiente em imagens de baixa resolução, de modo que a quantidade de características extraídas está extremamente ligada à imagem final do mosaico. Com o uso do método, foi possível realizar o *stitching* em imagens que detectores automáticos não demonstram uma boa extração de características, como SIFT e ORB.

Como trabalhos futuros, sugere-se explorar outras abordagens de pré-processamento, como a remoção de desfoque, visto que a importância da etapa de segmentação para a extração de características. Outra abordagem seria treinar um modelo de segmentação semântica para vasos sanguíneos em retinografias de baixa resolução.

## Agradecimentos

Os autores agradecem o apoio financeiro da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) - Finance Code 001, Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), e Fundação de Amparo à Pesquisa e ao Desenvolvimento Científico e Tecnológico do Maranhão (FAPEMA).

## Referências

- Abu-Ain, W., Abdullah, S. N. H. S., Bataineh, B., Abu-Ain, T., and Omar, K. (2013). Skeletonization algorithm for binary images. *Procedia Technology*, 11:704–709.
- Agarwal, S. et al. (2022). Detection and mosaicing techniques for low-quality retinal videos. *MDPI Sensors*, 22(5):2059.
- Burt, P. J. and Adelson, E. H. (1983). A multiresolution spline with application to image mosaics. *ACM Transactions on Graphics (TOG)*, 2(4):217–236.

- Costa, E. d. P. F., Gomes, T. M., Mendes, T. A., Campos, M. A. G., Bertrand, R. H. C., and Pinto, L. M. (2019). Perfil epidemiológico e prognóstico visual do trauma ocular em um centro de referência da região nordeste do brasil. *Rev Bras Oftalmol.*, 78(5):310–314.
- Feng, X., Cai, G., Gou, X., Yun, Z., Wang, W., and Yang, W. (2020). Retinal mosaicing with vascular bifurcations detected on vessel mask by a convolutional network. *healthcare engineering*.
- Fischler, M. A. and Bolles, R. C. (1981). Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. *Communications of the ACM*, 24(6):381–395.
- Fonseca, P., Araújo, A., Almeida, J., and Júnior, G. B. (2022). Treinando rede neural profunda com divisão proporcional de imagens para segmentação de estruturas da retina. In *Anais do XXII Simpósio Brasileiro de Computação Aplicada à Saúde*, pages 1–12, Porto Alegre, RS, Brasil. SBC.
- Gonzalez, R. C. (2009). *Digital image processing*. Pearson education india.
- Hu, R., Chalakkal, R. J., Linde, G., and Dhupia, J. S. (2022). Multi-image stitching for smartphone-based retinal fundus stitching. In *2022 IEEE/ASME International Conference on Advanced Intelligent Mechatronics (AIM)*, pages 179–184.
- Jalili, J., Hejazi, S. M., Riazi-Esfahani, M., Eliasi, A., Ebrahimi, M., Seydi, M., Fard, M. A., and Ahmadian, A. (2020). Retinal image mosaicking using scale-invariant feature transformation feature descriptors and Voronoi diagram. *Journal of Medical Imaging*, 7(4):044001.
- Jocher, G. (2023). Yolov8: State-of-the-art object detection. Acesso em: 17 fev. 2025.
- Köhler, T., Heinrich, A., Maier, A., Hornegger, J., and Tornow, R. P. (2016). Super-resolved retinal image mosaicing. In *2016 IEEE 13th International Symposium on Biomedical Imaging (ISBI)*, pages 1063–1067. IEEE.
- Ronneberger, O., Fischer, P., and Brox, T. (2015). U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical image computing and computer-assisted intervention*, pages 234–241. Springer.
- Russo, A., Morescalchi, F., Costagliola, C., Delcassi, L., Semeraro, F., et al. (2015). A novel device to exploit the smartphone camera for fundus photography. *Journal of ophthalmology*, 2015.
- Santos, L., Almeida, M., Almeida, J., Braz, G., Camara, J., and Cunha, A. (2024). Image stitching of low-resolution retinography using fundus blur filter and homography convolutional neural network. *Information*, 15(10).
- Sarkar, A. (2021). Retinal-vessel-segmentation-using-variants-of-unet.
- Wu, Y., Szymanska, M., Hu, Y., Fazal, M. I., Jiang, N., Yetisen, A. K., and Cordeiro, M. F. (2022). Measures of disease activity in glaucoma. *Biosensors and Bioelectronics*, 196:113700.