Panoramas

Caio Petrucci dos Santos Rosa - RA 248245

Celso Gabriel Ferreira Marçal Prado - RA 235393

Jonatas de Sousa Santos - RA 225334

Lucas Peixoto Gonçalves - RA 233419

Randerson Araújo de Lemos - RA 103897

## **1. Introdução**

O objetivo deste trabalho é aplicar conceitos fundamentais de Visão Computacional no desenvolvimento de um sistema capaz de realizar a construção automática de panoramas a partir de múltiplas imagens com regiões parcialmente sobrepostas. Esse processo, conhecido como *image stitching*, é um dos exemplos práticos mais importantes da área, pois integra diferentes técnicas em um fluxo completo de processamento de imagens.

A tarefa envolve diversas etapas interdependentes. Inicialmente, é realizada a detecção e extração de características locais, responsáveis por identificar pontos de interesse invariantes a transformações de escala, rotação e iluminação. Em seguida, esses pontos são descritos por meio de descritores robustos, que permitem estabelecer o emparelhamento confiável entre imagens distintas.

Com os pares de pontos correspondentes, procede-se à estimação da homografia que relaciona o plano de uma imagem ao de outra, utilizando métodos robustos contra outliers, como o algoritmo RANSAC. Essa homografia garante a correta sobreposição das imagens em um mesmo sistema de coordenadas. Finalmente, a composição do panorama é realizada por meio de técnicas de blending, que atenuam descontinuidades de cor e exposição nas regiões de transição, garantindo uma junção visualmente natural.

Esse tipo de sistema simula, em menor escala, o funcionamento de ferramentas presentes em câmeras de smartphones, softwares de fotografia panorâmica e aplicações de realidade virtual e aumentada, evidenciando a relevância prática dos conceitos estudados. Além disso, o trabalho permite consolidar conhecimentos sobre processamento digital de imagens, álgebra linear e estatística aplicada, demonstrando como diferentes áreas convergem na solução de um problema real.

## **2. Coleta das Imagens**

As imagens utilizadas foram coletadas com um Poco X7 Pro, utilizando a câmera principal de 64 MP (f/1.9, 26mm, PDAF) em um escritório, ou seja, um ambiente interno. O conjunto contém cinco fotos com sobreposição parcial de cerca de 30% a 40% entre vizinhas, conforme recomendado. Durante a captura, buscou-se manter a iluminação uniforme e evitar mudanças bruscas de luz ou movimentação de objetos, de modo a reduzir inconsistências no emparelhamento. A sequência foi registrada por meio de um movimento rotacional da câmera em torno do ponto de interesse, garantindo continuidade visual entre as tomadas.

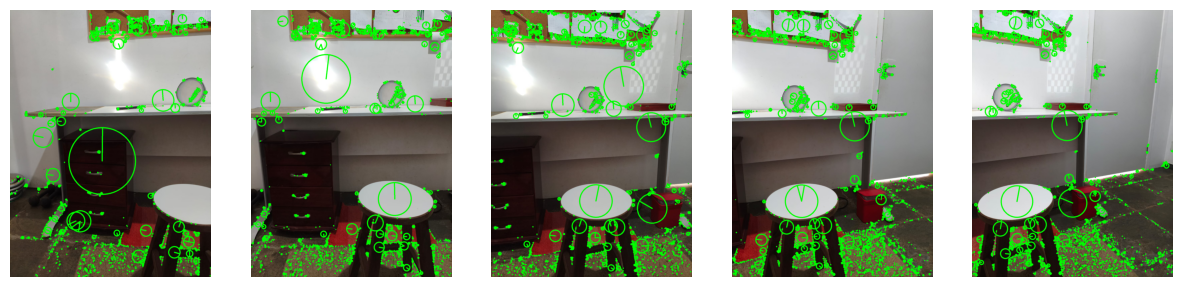
Durante a captura das imagens, percebemos que pequenas mudanças de iluminação ou a presença de objetos em movimento resultaram em keypoints inconsistentes, gerando dificuldades no emparelhamento e pequenas falhas visuais no panorama. Para mitigar, mantivemos o ambiente estático e realizamos a captura com movimento controlado da câmera.

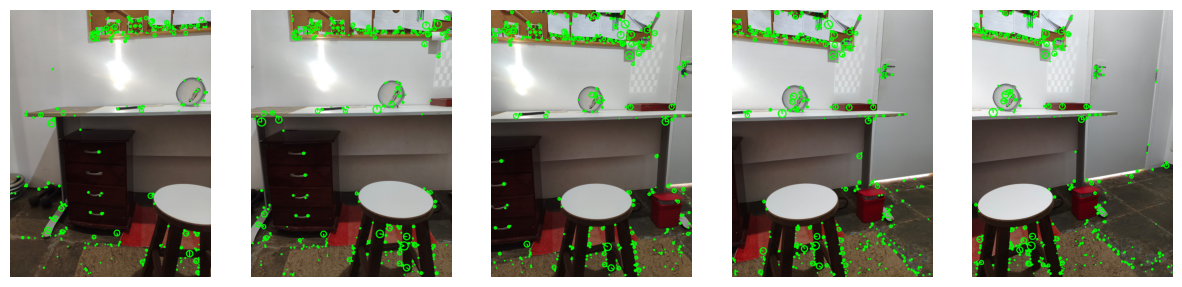


**Figura 1 – Conjunto de imagens capturadas para o panorama.**

## **3. Detecção e Extração de Características**

Para a extração de pontos de interesse, utilizou-se o detector SIFT (Scale-Invariant Feature Transform), amplamente reconhecido por sua robustez frente a variações de escala, rotação e iluminação. Esse método mostrou-se adequado ao cenário do trabalho, permitindo a obtenção de keypoints consistentes, mesmo em regiões com baixa textura ou sob pequenas mudanças nas condições de captura.



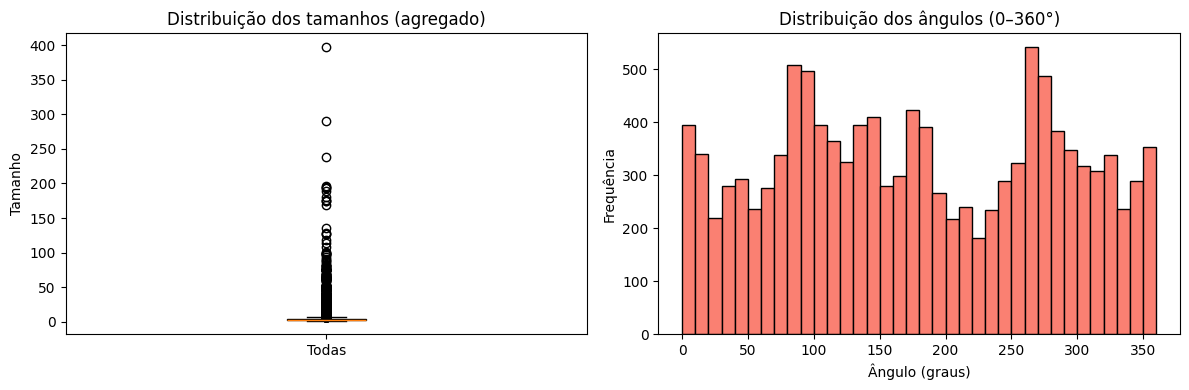


**Figura 2 – Visualização dos keypoints detectados com SIFT. As cinco imagens superiores apresentam todos os pontos detectados; as cinco inferiores mostram apenas os keypoints após filtragem (remoção dos menores e maiores).**

Na sequência, foram extraídas estatísticas descritivas dos keypoints, focando na distribuição de tamanhos e ângulos das características. A Figura 3 apresenta o boxplot dos tamanhos e o histograma das orientações. Observa-se que a maior parte dos keypoints possui tamanhos reduzidos, concentrados próximos ao limite inferior, mas com alguns outliers representando estruturas em escalas maiores. Esse comportamento é desejável, pois indica que o detector captura tanto detalhes finos quanto elementos mais amplos da cena.

No histograma de ângulos, a distribuição é aproximadamente uniforme entre 0° e 360°, evidenciando a invariância à rotação do SIFT: os pontos são detectados de forma consistente independentemente da orientação da imagem. Essa diversidade angular contribui para a robustez do emparelhamento, permitindo encontrar correspondências confiáveis mesmo com variações de perspectiva.

Em comparação com o ORB (Oriented FAST and Rotated BRIEF), usado como referência, o SIFT apresentou desempenho superior. Embora o ORB seja mais eficiente computacionalmente, ele mostrou menor dispersão em escala e maior concentração angular, resultando em menos correspondências confiáveis. Assim, a escolha pelo SIFT se justifica pela necessidade de um conjunto de keypoints mais robusto e invariável, essencial para garantir o sucesso das etapas subsequentes de emparelhamento e reconstrução.



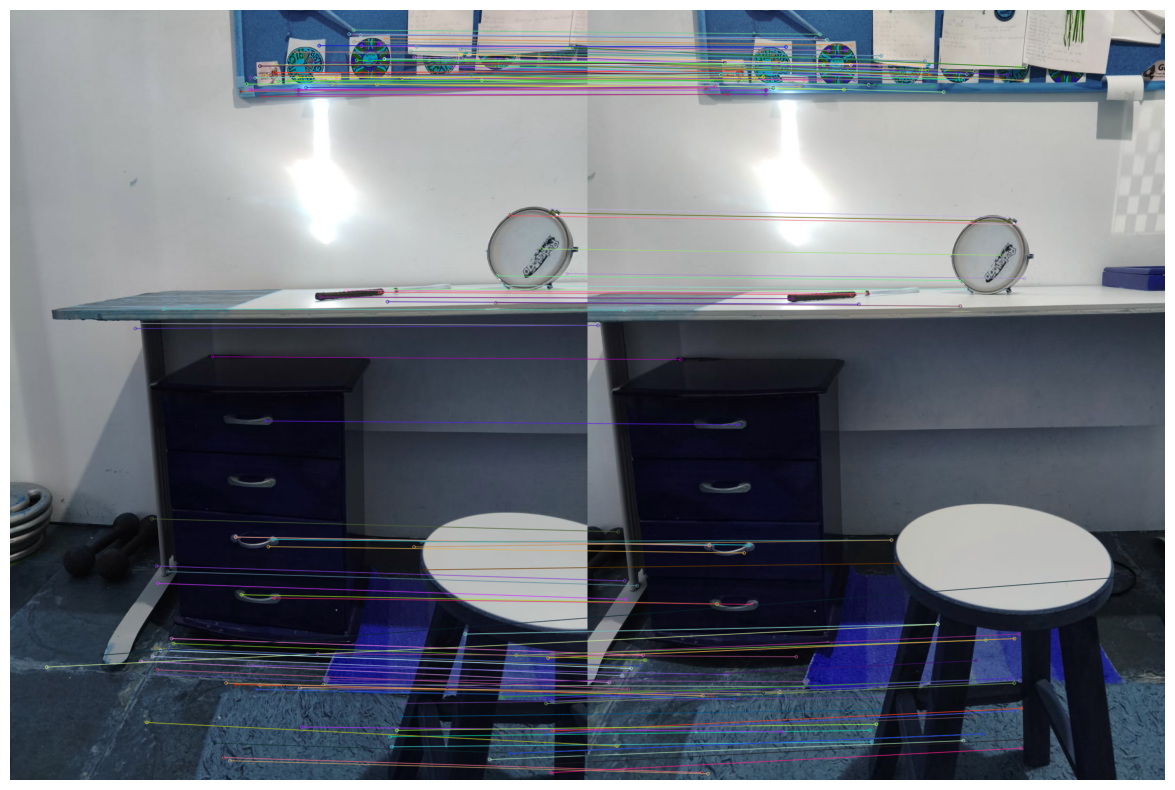
**Figura 3 – Boxplot dos tamanhos e histograma das orientações dos keypoints detectados com SIFT.**

## **4. Emparelhamento de Características**

O emparelhamento de descritores foi realizado utilizando o Brute Force Matcher, empregando a distância Euclidiana (norma L2) como métrica de similaridade. Para cada descritor detectado em uma imagem, o algoritmo buscou candidatos correspondentes na outra, calculando a distância entre os vetores de características para avaliar o grau de compatibilidade.

Para aumentar a robustez e reduzir falsos positivos, aplicou-se o teste de razão proposto por Lowe, com um limiar de 0,75. Esse critério compara a distância do melhor candidato com a do segundo melhor: apenas quando a correspondência mais próxima é significativamente superior à segunda é que o par é considerado válido. Dessa forma, correspondências ambíguas ou incorretas são descartadas, resultando em um conjunto de matches mais consistente e confiável.

Durante o processo, identificamos correspondências incorretas principalmente em regiões com padrões repetitivos. Para mitigar esse problema, ajustamos o limiar do ratio test, conseguindo equilibrar a quantidade de matches com a confiabilidade das correspondências.



**Figura 4 – Emparelhamento entre duas imagens consecutivas mostrando as correspondências de keypoints.**

## **5. Estimação de Homografia e Alinhamento**

A homografia entre os pares de imagens foi estimada utilizando o algoritmo RANSAC (Random Sample Consensus), que aumenta a robustez do processo ao descartar automaticamente correspondências inconsistentes (outliers). Dessa forma, apenas os pontos confiáveis são considerados para o cálculo da matriz de transformação.

Com a homografia definida, aplicou-se a função cv2.warpPerspective, responsável por projetar cada imagem para um sistema de coordenadas comum. Esse alinhamento progressivo permite que as imagens se sobreponham de maneira coerente, respeitando a geometria da cena.

Essa etapa é fundamental, pois estabelece a base geométrica do panorama: sem um alinhamento adequado, pequenas inconsistências entre as imagens poderiam comprometer significativamente a qualidade final. Embora o RANSAC tenha sido eficaz na remoção de outliers, foi necessário selecionar cuidadosamente os pares de imagens e ajustar parâmetros para minimizar erros de alinhamento e garantir uma composição final precisa.



**Figura 5 – Alinhamento progressivo das 3 primeiras imagens utilizando a homografia.**

## **6. Composição e Blending**

A etapa de composição e blending é crucial no processo de produção, pois é nela que os elementos obtidos ao longo do desenvolvimento são integrados, conferindo unidade e coerência visual ao panorama. Nessa fase, o objetivo não é apenas posicionar cada imagem corretamente, mas também garantir uma fusão harmoniosa entre elas. Foram testadas três estratégias de blending: soma direta, alpha blending e pixel máximo. A soma direta torna as regiões sobrepostas mais visíveis, resultando em descontinuidades perceptíveis. O alpha blending suaviza as transições aplicando uma média ponderada entre os pixels sobrepostos, produzindo um resultado mais uniforme. Já o pixel máximo preserva melhor os detalhes das bordas e se mostrou particularmente eficaz para este trabalho, pois reduziu de forma significativa os fantasmas em regiões de iluminação estável, proporcionando a melhor integração visual.

O blending foi implementado de modo a combinar as imagens sobrepostas de forma coerente: enquanto a soma direta e o alpha blending apresentaram limitações em determinadas condições, o pixel máximo garantiu maior clareza estrutural, preservando contornos e mantendo a naturalidade da cena.

Mesmo com a homografia precisa, pequenas variações de iluminação ainda geraram inconsistências em algumas áreas. Contudo, a escolha pelo pixel máximo permitiu atenuar esses efeitos, assegurando uma fusão mais nítida e consistente, equilibrando preservação de detalhes e coerência visual no panorama final.



**Figura 6 – Resultados do blending: soma direta, alpha blending e pixel máximo.**

## 

## **7. Conclusão**

O desenvolvimento deste trabalho permitiu uma compreensão completa das etapas envolvidas na construção automática de panoramas por meio de técnicas de Visão Computacional. Desde a coleta cuidadosa das imagens até a composição final, cada fase apresenta desafios específicos, como manutenção de iluminação uniforme, pontos ambíguos em áreas repetitivas e pequenas inconsistências de alinhamento, que exigiram ajustes e escolhas técnicas conscientes.

A utilização do SIFT para detecção de características mostrou-se robusta frente a variações de escala, rotação e iluminação, enquanto o emparelhamento de descritores com o Brute Force Matcher e o ratio test de Lowe garantiu correspondências confiáveis entre as imagens. A estimação de homografia com RANSAC assegurou alinhamento geométrico preciso, minimizando o impacto de outliers, e a aplicação de técnicas de blending integrou as imagens de forma natural, preservando coerência de cor, textura e iluminação.

O trabalho evidenciou a relevância prática dessas técnicas em aplicações reais, como câmeras de smartphones, softwares de fotografia panorâmica e ambientes de realidade virtual, além de reforçar a importância da compreensão teórica de cada etapa. Adicionalmente, consolidou conhecimentos em processamento digital de imagens, álgebra linear e estatística aplicada, mostrando como diferentes áreas convergem para a solução de problemas concretos.

Dessa forma, o projeto não apenas valida a eficácia de pipelines de *image stitching*, mas também fornece uma base sólida para futuras implementações mais avançadas, como )otimizações de desempenho, fusões sofisticadas, remoção de fantasmas em cenários mais complexos (com iluminação não uniforme nas imagens e etc) e adaptação a cenários dinâmicos.)



**Figura 7 – Resultado final do panorama**