MONITOR do DEBATE POLÍTICO no MEIO DIGITAL

NOTA TÉCNICA · EVOLUÇÃO DO MÉTODO

1 à mão

GIMP, Inkscape ou Photoshop. Seleção e corte feitos a olho: o gargalo de tempo e fonte de sobreposição a mais ou a menos

2 metadados

GPS, altura, yaw e campo de visão. Posicionam as fotos de forma aproximada, mas não as encaixam

3 + SIFT

Casamento visual entre fotos vizinhas. Seleciona, encaixa e corta: cada pixel vem de uma só foto

VERSÃO ATUAL

Márcio Moretto Ribeiro
1 de novembro de 2026

Seleção e corte automáticos de fotos de drone

O método seleciona as fotos e define o corte de cada uma sem edição manual: as informações gravadas em cada foto a posicionam e a comparação com as fotos vizinhas a encaixa, de modo que cada ponto do mosaico vem de uma só foto.

Para contar um público a partir de fotos de drone é preciso juntar dezenas de imagens que se sobrepõem em uma só, de modo que ninguém seja contado duas vezes nem fique de fora. Até aqui essa montagem era feita à mão e era o maior gargalo de tempo do processo de contagem.

Esta nota apresenta o método automático e o caminho até ele: uma primeira ideia, que usa só os metadados das fotos, e a versão final, que os combina com o casamento visual de pontos (SIFT). Como exemplo, a nota acompanha um voo sobre a orla de Copacabana: os números dele mostram o método em ação e não medem seu desempenho em geral.

Os números vêm de um voo de exemplo, em Copacabana, processado pela versão 3.2.2 do script de mosaico, e valem só para ele. Não medimos o tempo do processo manual nem o erro do mosaico em metros contra um mapa de referência; ver a seção "Como os números foram feitos".

SEÇÃO 01

01

O problema e como fazíamos

A montagem manual era o gargalo da contagem e podia contar a mesma pessoa duas vezes ou deixar parte da multidão de fora.

FIGURA 1

Exemplo: mosaico de Copacabana montado automaticamente a partir de 20 fotos

Voo de exemplo, em 16/08/2026, das 12h34 às 12h37



Fonte: fotos de drone DJI do Monitor · mosaico gerado por mosaico_dji.py v3.2.2, reduzido para exibição · 20 fotos, 10.305 × 4.887 px no arquivo original.

Um drone a 70 metros de altura enxerga um trecho de avenida por foto. Para cobrir uma multidão, o piloto faz uma passada com dezenas de fotos que se sobrepõem em boa parte. Contar foto a foto conta duas vezes quem aparece em duas fotos, e por isso as imagens precisam virar uma só.

Até aqui, isso era manual: selecionávamos as fotos à mão e as cortávamos no GIMP, no Inkscape ou no Photoshop, até formar um único mosaico. Era a etapa mais lenta de todo o processo de contagem.

Era também uma etapa sujeita a erro. O corte feito a olho podia deixar mais sobreposição do que devia, e parte da multidão aparecia duas vezes, ou deixar menos, e parte dela sumia. O erro ia nos dois sentidos e dependia de quem montava.

SEÇÃO 02

02

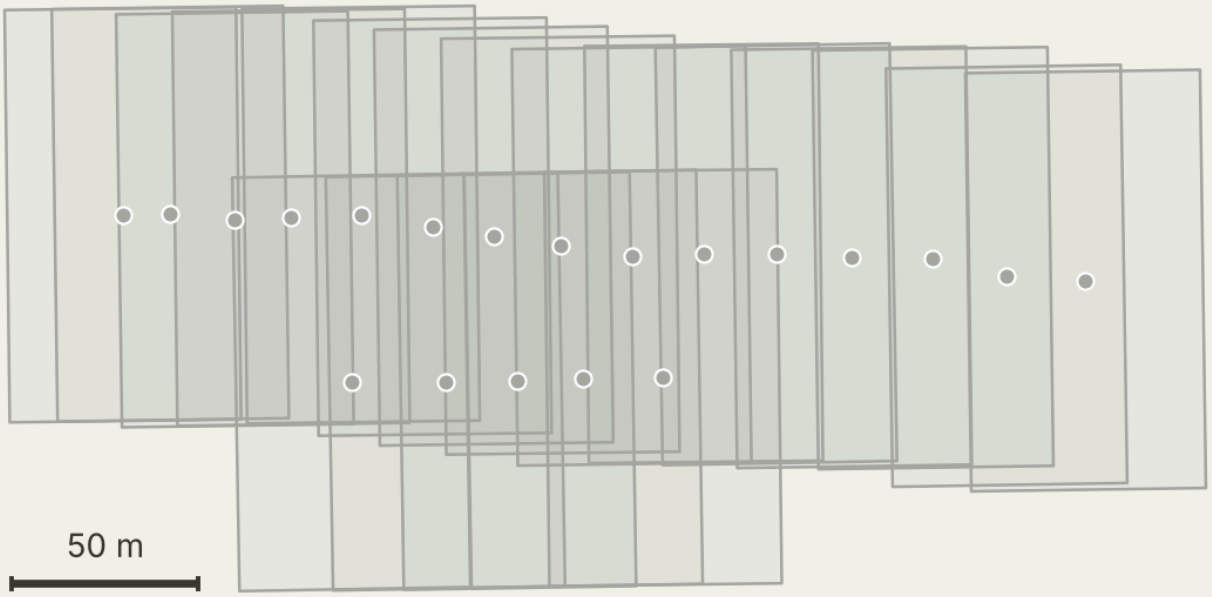
Primeira ideia: usar só os metadados

Os metadados da DJI bastam para posicionar as fotos de forma aproximada, mas não para encaixá-las.

FIGURA 2

Onde os metadados dizem que cada foto cai no chão

Voo de exemplo, em Copacabana: retângulos previstos por GPS, altura, yaw e campo de visão, sem olhar a imagem



Fonte: elaboração própria a partir dos metadados DJI das 20 fotos de Copacabana (16/08/2026) · escala: a barra mede 50 m · nenhuma imagem foi usada.

Cada foto traz, no EXIF e no XMP da DJI, a posição do drone (GPS), a altura relativa à decolagem, o yaw e a inclinação da câmera; o modelo da câmera dá o campo de visão. Com isso se calcula, para cada foto, o retângulo que ela cobre no chão, e todas podem ser projetadas num mesmo sistema de coordenadas (UTM) sem olhar para a imagem. O resultado é a Figura 2. Esse cálculo continua no método final, como ponto de partida.

Como método único, ele esbarra em três limites que aparecem nos dados. O primeiro é a escala: a prevista por altura e campo de visão errava cerca de 6% no voo de exemplo, o que desalinha a sobreposição entre fotos vizinhas. O segundo é a altura: a altitude relativa da DJI é medida a partir do ponto de decolagem e não do terreno. Em outro voo, na Avenida Paulista, ela valeu −8,2 m, impossível para uma foto aérea, e o script teve de recorrer à altura padrão de 70 m.

O terceiro é o GPS. Depois do ajuste visual, as fotos do voo de exemplo se moveram em mediana 0,6 m e até 1,4 m em relação ao ponto de GPS. A 39 pixels por metro, 1,4 m são cerca de 55 pixels, o suficiente para cortar ou duplicar algumas pessoas numa emenda.

SEÇÃO 03

03

Ideia final: metadados mais SIFT

Os metadados dizem onde cada foto deve estar; o SIFT confere olhando as imagens, e o corte passa a ser uma regra.

O método final mantém o cálculo da seção anterior como ponto de partida e acrescenta uma checagem visual. Primeiro, ele escolhe as fotos: descarta as que não são zenitais, isto é, aquelas em que a câmera não aponta para baixo. No voo de exemplo, em Copacabana, 3 das 23 fotos saíram por isso e 20 entraram. Os números a seguir vêm desse mesmo voo; em outro voo eles mudam, e o método é o mesmo.

Segundo, decide quais fotos comparar. Não adianta comparar fotos que ficaram longe uma da outra, então os metadados indicam os pares que provavelmente se sobrepõem. No exemplo, foram 119 pares.

Terceiro, cada par é conferido olhando as imagens. O SIFT procura, em cada foto, pontos fáceis de reconhecer, como o canto de um quiosque, um desenho do calçadão ou uma árvore, e liga os que aparecem nas duas. Uma etapa de votação (o RANSAC) mantém só as ligações que concordam entre si: se todas indicam o mesmo deslocamento entre as fotos, o par está bem encaixado; se cada uma aponta para um lado, o par é descartado. No exemplo, 30 pares foram aceitos, 10 ficaram como fracos e 79 foram rejeitados.

Quarto, com os pares aceitos o método ajusta todas as fotos de uma vez, sem se afastar do GPS, e corrige o que os metadados erraram, como a escala. No exemplo, os pares aceitos ligaram as 20 fotos num só conjunto. Os pares fracos servem apenas para regularizar a escala.

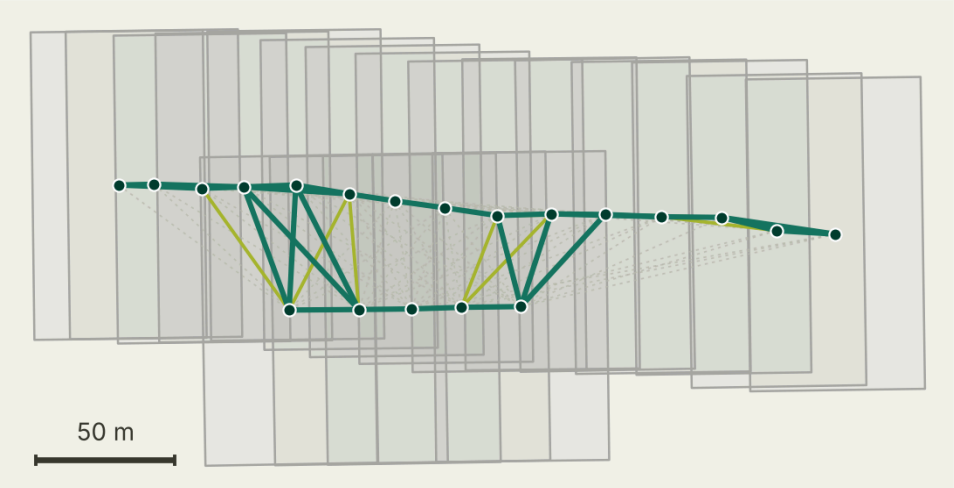
Por fim vem o corte, que antes era feito à mão e agora é uma regra: cada ponto do mosaico vem da foto cujo centro está mais próximo, sem misturar fotos, e um mapa registra de qual foto veio cada ponto. O mosaico é então girado para a horizontal.

FIGURA 3

No voo de exemplo, 30 dos 119 pares testados foram aceitos e ligam as 20 fotos

Pares de fotos candidatos e classificação do alinhamento visual

ok (30) fraco (10) rejeitado (79, pontilhado)



Fonte: elaboração própria a partir de registro_visual_rede.csv · centros das fotos nas posições finais ajustadas · a barra mede 50 m.

SEÇÃO 04

04

Por que isso importa para a contagem

O mapa de origem dos pixels permite contar cada pessoa em uma só foto.

Além do mosaico, o script grava o mapa de qual foto é dona de cada pixel e uma tabela com os pixels que cada foto contribuiu. No voo de exemplo, as 20 fotos entraram no mosaico, somando 40,6 milhões de pixels; a Figura 4 mostra as emendas entre elas.

No sistema de contagem, esse mapa é usado para contar nas fotos originais, em resolução nativa, e não no mosaico reduzido. A documentação do sistema registra que o mosaico reduzido tem cerca de 10 pixels por metro contra cerca de 34 nas fotos, o que deixa cada pessoa com 5 a 8 pixels e faz o detector de pessoas perder parte delas. Um ponto detectado só conta se cair num pixel do mosaico que pertence à foto em que foi detectado, e isso evita contar duas vezes quem aparece em fotos vizinhas.

FIGURA 4

Cada pixel vem de uma só foto: as linhas finas brancas são as emendas

Voo de exemplo, em Copacabana: detalhe do mosaico na resolução do arquivo (cerca de 3 cm por pixel)



Fonte: mosaico_costuras.jpg do voo de Copacabana (16/08/2026) · recorte de 2.600 × 1.400 px do arquivo original, reduzido para exibição.

SEÇÃO 05

Como os números foram feitos

05

Os números do voo de exemplo saem dos arquivos de diagnóstico que o script de mosaico grava; nada foi recalculado ou ajustado para esta nota.

1 **COBERTURA**

Um único voo, em Copacabana, usado como exemplo: 23 fotos, das quais 20 foram selecionadas. Os números ilustram o método e não medem seu desempenho em geral. O script é o mosaico_dji.py, versão 3.2.2. O voo da Avenida Paulista só entra na seção 02, como exemplo da altitude relativa.

2 **AMOSTRA**

Copacabana, 16/08/2026, fotos das 12h34 às 12h37, pelo horário gravado nos nomes dos arquivos.

3 **LIMITES**

O ajuste global é ancorado no GPS; por isso o deslocamento de 0,6 m mede a diferença entre o resultado e o GPS, não o erro absoluto do mosaico. No voo da Avenida Paulista, o ajuste visual aceitou só 1 de 41 pares testados e as posições vieram quase só dos metadados; esse caso não é mostrado aqui. Por ser um único voo, os números valem como exemplo e não como avaliação do método. Não houve mapa de referência nem pontos de controle, não medimos o tempo do processo manual e não geramos o mosaico só com metadados para comparar; a Figura 2 mostra os retângulos previstos, não uma imagem.

4 **FONTE DOS DADOS**

Arquivos de saída do script (mosaico_info.json, registro_visual_rede.csv, poses_finais_utm.csv, mosaico_fontes.csv, imagens do mosaico) e metadados das fotos lidos pelo ExifTool. As figuras 2 e 3 foram redesenhadas a partir desses arquivos.