

Mosaicos de Imagens Aéreas Sequenciais Construídos Automaticamente

André de Souza Tarallo,* Francisco Assis da Silva, Alan Kazuo Hiraga,
Maria Stela Veludo de Paiva, Lúcio André de Castro Jorge

Resumo: A geração automática de mosaicos de imagens aéreas agrícolas aumenta a eficiência na análise das áreas agrícolas e nas tomadas de decisão relacionadas a controle de pragas, doenças e desmatamento. Neste capítulo é apresentada uma ferramenta para construção automática de mosaicos de imagens sequenciais. As principais características que interferem no desempenho da ferramenta são: construir mosaicos sem distorção e o custo computacional. Dez mosaicos foram obtidos a partir de 200 imagens agrícolas e comparados com aqueles obtidos com os softwares comerciais, mostrando melhor qualidade e menor distorção. Posteriormente, eles foram visualmente inspecionados por um profissional, que confirmou a qualidade da ferramenta desenvolvida.

Palavras-chave: Custo computacional, Imagens aéreas agrícolas, imagens de alta resolução, Mosaicos automáticos.

Abstract: *The automatic generation of mosaics of aerial agricultural images increases efficiency in the analysis of agricultural areas and in decisions-making related to pest control, diseases and deforestation. This chapter presents a tool for automatic construction of mosaics from sequential images. The main features that interfere with the performance of the tool are: building mosaics without distortion and the computational cost. Ten mosaics were obtained from 200 agricultural images and compared with those obtained with commercial software, showing better quality and less distortion. After, they were visually inspected by a professional, which confirmed the quality of the tool developed.*

Keywords: *Computational cost, Aerial agricultural images, High-resolution images, Automatic mosaics.*

*Autor para contato: andre.tarallo@gmail.com

1. Introdução

Até o presente momento, a construção de mosaicos de imagens na agricultura vem sendo feita de maneira semiautomática, necessitando obter o modelo digital do terreno, fazer a ortorretificação das imagens e colocação manual de bandeirinhas (marcadores), para que um software possa reconstruir esta área e gerar um mosaico deste terreno. Deste modo, a construção de mosaicos é demorada e trabalhosa, podendo demorar um dia todo ou mais dias. Isto também envolve uma grande demanda de pessoal para fazer as marcações na área em questão, além de poder gerar um mosaico com pouca precisão.

Com os mosaicos, é possível direcionar vistorias de campo durante o ciclo do cultivo ou em datas posteriores à colheita, possibilitando fornecer um diagnóstico preciso da área de cultivo. A partir daí, podem ser elaborados os mapas de recomendações: descompactação, fertilidade e aplicação de insumos em taxa variável.

Mais recentemente, aplicações na agricultura passaram a exigir maior rapidez na construção destes mosaicos, para possibilitar a obtenção mais rápida de informações para tomada de decisões relativas ao controle de pragas, doenças ou queimadas.

Estes fatos levaram à construção de uma metodologia apresentada neste capítulo, para a construção automática de mosaicos de imagens digitais na agricultura. As imagens são fotos aéreas, obtidas com uma câmera de alta resolução acoplada em um avião. A alta resolução da câmera minimiza possíveis problemas de distorção nas imagens, causados pela distância entre a aeronave e o solo.

Para a implementação desta metodologia, foi utilizada a transformada SIFT (*Scale Invariant Feature Transform*) para a extração de características das imagens, o algoritmo BBF (*Best-Bin-First*) para determinar pontos correspondentes entre pares de imagens e o algoritmo RANSAC (*Random Sample Consensus*) para filtrar os falsos pontos correspondentes entre os pares de imagens. Por fim, após a aplicação destas técnicas, os pares de imagens foram unidos.

O uso da SIFT para a implementação de mosaicos pode ser encontrado nos trabalhos de [Bei & Haizhen \(2009\)](#) e [Li & Geng \(2010\)](#). Outra técnica usada para este fim é a PCA-SIFT, encontrada no trabalho de [Ke & Sukthankar \(2004\)](#). SURF é uma técnica mais recente que a SIFT e tem se destacado por ter menor custo computacional ([Hong et al., 2009](#)). No entanto, de acordo com o trabalho de [Juan & Gwun \(2009\)](#), que faz uma comparação destas técnicas (SIFT, PCA-SIFT e SURF), a transformada SIFT apresenta melhor estabilidade quanto à invariância à escala e rotação, apesar desta transformada ter um custo computacional mais elevado do que as outras técnicas. De acordo com [Lowe \(2004\)](#), as técnicas BBF e

RANSAC, após aplicação da SIFT, são as mais indicadas para encontrarem os pontos correspondentes entre pares de imagens.

O restante deste capítulo está estruturado da seguinte maneira: na Seção 2 é apresentada a fundamentação teórica, na Seção 3 é detalhada a metodologia utilizada, na Seção 4 são apresentados os resultados, na Seção 5 são apresentadas as discussões e conclusões.

2. Fundamentação Teórica

Nesta seção será apresentado como foi construída a base de dados de imagens aéreas bem como a descrição das principais técnicas utilizadas neste artigo.

2.1 A base de imagens

As imagens aéreas utilizadas neste projeto de doutorado foram fornecidas pela EMBRAPA (Empresa Brasileira de Pesquisas Agropecuárias) Instrumentação, situada em São Carlos - SP. As imagens contêm áreas de pastagens, lavouras e áreas urbanas, obtidas na região de Santa Rita do Sapucaí - MG em Setembro de 2007, com o auxílio de um avião de pequeno porte, contendo uma câmera acoplada a um suporte. A base de imagens é composta por 200 imagens sequenciais. As imagens foram adquiridas com dimensão de 3504×2336 *pixels*, com 24 bits por *pixel* com 72 dpi (pontos por polegada) no formato JPG, utilizando 8 *Megapixels* de resolução. A Figura 1 ilustra um exemplo de imagem usada no projeto.



Figura 1. Exemplo de imagem usada no projeto.

As imagens foram obtidas seguindo um padrão de aquisição (para evitar distorções), que inclui os seguintes itens:

- A câmera foi posicionada horizontalmente em relação ao solo;
- O avião percorreu sempre a mesma distância, fazendo movimentos horizontais na área demarcada, para obtenção das imagens (Figura 2);
- A obtenção da sequência de imagens foi feita com um tempo determinado e síncrono, entre a imagem anterior e a posterior, para gerar regiões de sobreposição.

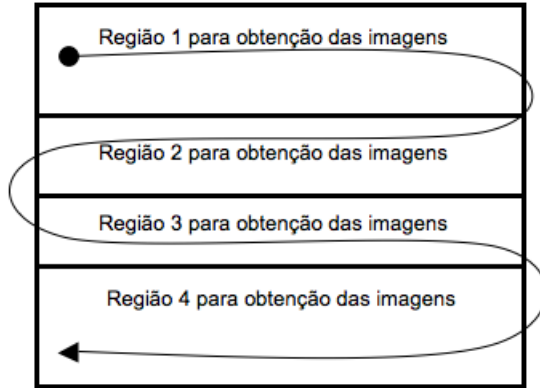


Figura 2. Padrão de rota do avião para obtenção das imagens.

Para uma construção sem muita distorção de um mosaico, é de extrema importância que seja seguido o padrão apresentado para a obtenção das imagens.

2.2 A transformada SIFT

A transformada SIFT (Lowe, 2004) é capaz de transformar uma imagem, em uma coleção de vetores de características locais (descritores de características), e cada um destes vetores são invariantes à escala, rotação e parcialmente invariante à mudanças de iluminação e ponto de vista.

As características fornecidas pela transformada SIFT são bem localizadas em ambos os domínios, o da frequência e o do espaço, reduzindo assim a probabilidade de não haver correspondência das características por oclusão ou ruído. As características são altamente distintas, permitindo que uma simples característica seja corretamente correspondida com alta probabilidade diante de um grande banco de dados de características, possibilitando assim, uma base para o reconhecimento de objetos e cenas.

O custo de extrair essas características é minimizado por meio de uma abordagem de filtragem em cascata, na qual as operações com maior custo operacional são executadas apenas em locais que passaram em testes iniciais.

Os quatro principais estágios que compõem a transformada SIFT para gerar o conjunto de características de imagens são:

- **Detecção de extremos no Espaço Escala:** Neste primeiro estágio é feito a procura por todas as escalas e locais de uma imagem. Para isto é utilizada uma função conhecida como Diferença da Gaussiana, para identificação dos potenciais pontos de interesse que são invariantes à escala e orientação. Esta é a parte mais custosa do algoritmo.
- **Localização dos Pontos Chave:** Para cada local candidato, é determinada a sua posição e escala. Os pontos chave são selecionados baseados em medidas de sua estabilidade.
- **Definição da Orientação:** Uma ou mais orientações são atribuídos para cada ponto chave localizado, baseado em direções do gradiente. Todas as operações posteriores são realizadas sobre os dados da imagem que foram considerados ponto chave e que foram transformados em relação à orientação, escala e localização, proporcionando invariância a estas transformações.
- **Descritor dos Pontos Chave:** Os gradientes da imagem são medidos na escala selecionada, na região ao redor de cada ponto chave, sendo criados histogramas de orientações para compor o descritor.

Com as características extraídas a partir de todas as imagens, as mesmas devem ser pareadas. Na Figura 3 pode ser observado um exemplo da localização de características através da transformada SIFT.

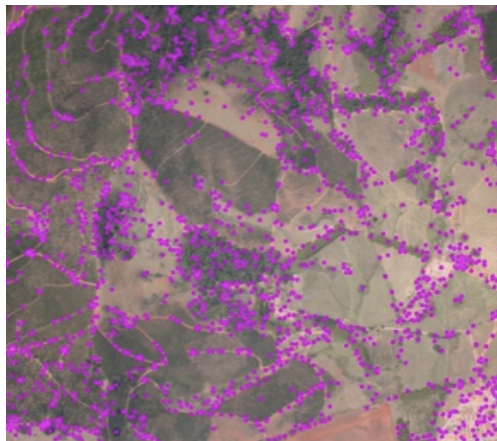


Figura 3. Localização de características com a transformada SIFT.

A transformada SIFT converte dados da imagem em coordenadas invariantes à escala, relativas às características locais. Um aspecto importante é o grande número de características geradas, que cobrem densamente toda a imagem (Lowe, 2004).

2.3 BBF

Uma vez aplicada a transformada SIFT sobre as imagens, é possível encontrar a correspondência entre duas imagens de acordo com os pontos chaves detectados. Há a comparação dos descritores das duas imagens, encontrando os melhores candidatos a serem seus equivalentes na outra imagem.

O melhor candidato correspondente para cada ponto chave é encontrado, identificando os seus vizinhos mais próximos na base de dados dos pontos chave a partir de imagens de entrada. O vizinho mais próximo é definido como o ponto chave, com distância euclidiana mínima entre os descritores em questão.

A maneira mais eficaz de identificar o melhor ponto candidato é obtida através da comparação da distância do vizinho mais próximo ao de um segundo vizinho mais próximo. Quando se procura classificar uma imagem em um extenso banco de dados de descritores para vários objetos, a busca exaustiva de vizinho mais próximo pode ser demorada e para tal existe a técnica BBF (Beis & Lowe, 1997) para acelerar a busca.

O algoritmo BBF usa uma busca ordenada modificada de um algoritmo *k-d tree* de modo que as posições no espaço das características são procuradas na ordem de suas distâncias mais próximas a partir do local investigado.

Uma razão para o algoritmo BBF (Beis & Lowe, 1997) funcionar bem é que somente são consideradas correspondências nas quais o vizinho mais próximo é menor do que 0,8 vezes a distância do segundo vizinho mais próximo e, portanto não é necessário resolver os casos mais difíceis, nos quais muitos vizinhos têm muitas distâncias similares.

2.4 RANSAC

Após a correspondência dos pontos chave, os mesmos são usados para calcular uma transformada que mapeia as posições dos pontos de uma imagem para as posições dos pontos correspondentes, na outra imagem, de um par de imagens.

Às vezes acontece de pares encontrados corresponderem a falsas correspondências, sendo necessário identificar estas falsas correspondências e de removê-las. A solução para este problema envolve o conceito da geometria epipolar (Oram, 2001) e homografia (Hartley & Zisserman, 2004). Com isso será reduzido o número de falsas correspondências e calculado uma transformação para juntar duas imagens sequenciais.

A correspondência de imagens fornece um conjunto de vetores de deslocamento relativo às características de um par de imagens obtidas, ou seja, cada vetor representa as coordenadas da mesma característica em ambas imagens. Com isto, é possível determinar o movimento entre tais imagens através da homografia.

Como a etapa de correspondência fornece um conjunto de n pontos correlacionados, estes pontos podem ser usados para se achar a matriz H . A matriz homográfica H é determinada, permitindo estimar o movimento entre as imagens.

2.5 Geometria epipolar

Os seres humanos têm a capacidade de distinguir quais objetos estão mais um próximos dos outros quando olham para eles, por possuímos visão estéreo. Ou seja, cada um dos nossos olhos observa o mundo de pontos de vista diferentes e, a partir disto, o nosso cérebro consegue extrair várias relações geométricas entre as imagens formadas em cada retina. Assim, é capaz de reconstruir o ambiente 3D de forma que possamos perceber as diferenças de profundidade dos objetos que compõem a cena observada. A simulação computacional deste processo de visão que recria o ambiente 3D a partir de duas imagens é baseada na área da geometria, denominada geometria epipolar. Ela depende apenas dos parâmetros da câmera, independente da estrutura da cena (Roberto et al., 2009).

A modelagem da visão estéreo pode ser realizada usando duas câmeras, como mostrado na Figura 4. Pode-se observar que cada câmera possui o seu próprio centro e orientação. Deste modo, cada uma possui também o seu próprio sistema de coordenadas de câmera (Pollefeys, 1999).

Dentre as várias relações possíveis entre pares de imagens, algumas são bastante importantes, pois ocorrem em todos os casos de visão estéreo. A primeira delas é a reta que liga o centro C_1 da primeira câmera com o centro C_2 da segunda, chamada de *baseline*.

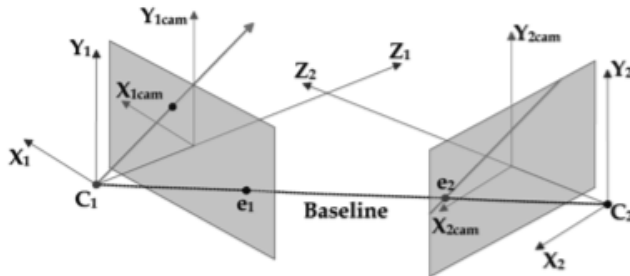


Figura 4. Esquema computacional da visão estéreo (Roberto et al., 2009).

O ponto de intersecção desta reta com o plano de imagem é chamado de epipolo. Para a primeira câmara têm-se o epipolo e_1 e para a segunda câmara tem-se o epipolo e_2 .

Se existem dois pontos m_1 e m_2 na primeira e na segunda imagem respectivamente, que são a projeção de um ponto M em coordenadas reais, pode-se dizer que M , C_1 , C_2 são coplanares, formando o plano epipolar, como visto na Figura 5. Este plano intersecta com o plano de imagem de cada uma das câmeras formando as linhas epipolares. A Figura 5 ilustra a relação que as linhas epipolares possuem entre uma imagem e outra. Por uma análise, usando o ponto m_1 como referência, é possível definir um raio que parte de C_1 até m_1 . A partir deste raio, pode-se perceber que m_1 na realidade não é apenas a projeção de M , mas sim de todos os pontos que pertencem ao raio. Isto significa que é impossível determinar exatamente a posição espacial de um ponto projetado numa imagem sem que haja uma outra imagem, capturada por uma segunda câmara em uma outra posição. Neste exemplo, m_2 seria este segundo ponto de vista de M . Desta forma, a intersecção dos raios que vão de C_1 à m_1 e de C_2 à m_2 ocorreria no ponto M (Oram, 2001).

Se o primeiro raio for projetado na segunda imagem ele formará uma reta no plano projetivo, que é a linha epipolar correspondente ao ponto m_1 e esta reta contém o ponto m_2 . O mesmo acontece se o raio de C_2 à m_2 for projetado na primeira imagem. Desta análise pode ser extraída mais uma importante conclusão: para todos os pontos de uma imagem, seu correspondente na outra figura estará na sua respectiva linha epipolar (Roberto et al., 2009).

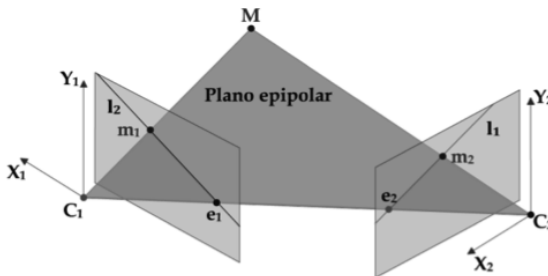


Figura 5. Geometria epipolar e seus principais elementos (Roberto et al., 2009).

Todas as linhas epipolares passam pelo epipolo da imagem e, independente da coordenada espacial do ponto M , todos os planos epipolares passarão pela *baseline*, como mostra a Figura 6. A partir de um ponto m_1 na primeira imagem, a linha epipolar l_1 na segunda imagem, que conterà o ponto m_2 , pode ser achada a partir da seguinte relação:

$$l_1 = Fm_1 \tag{1}$$

Sendo F a matriz fundamental (uma representação algébrica da geometria epipolar entre duas imagens). Ela é uma matriz 3×3 que pode ser encontrada a partir da seguinte relação:

$$m_2^T = Fm_1 = 0 \tag{2}$$

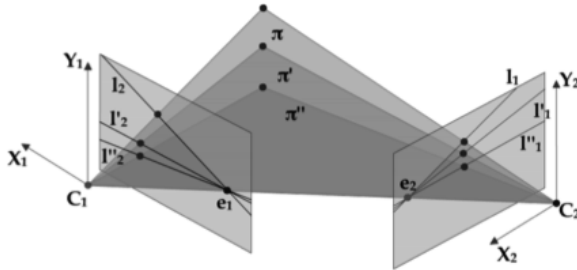


Figura 6. Vários planos epipolares, todos passando pela *baseline*, formando várias linhas epipolares, todas passando pelos epipolos (Roberto et al., 2009).

Para casos onde o objetivo é encontrar a linha epipolar l_2 na primeira imagem, correspondente ao ponto m_2 da segunda, a matriz fundamental também pode ser calculada:

$$l_2 = F^T m_2 \tag{3}$$

Outro papel importante da matriz fundamental é calcular os epipolos da imagem. Eles são definidos como os núcleos da matriz fundamental:

$$F e_1 = 0 \quad e \quad F^T e_2 = 0 \tag{4}$$

Também é possível encontrar linhas epipolares correspondentes. Ou seja, dado que a equação da linha l_1 na segunda imagem é conhecida, encontrada a partir do ponto m_1 na primeira imagem, é possível encontrar a linha epipolar l_2 que contém o ponto m_1 , mesmo sem conhecer o ponto m_2 na segunda imagem. Isto é possível porque existe uma matriz homográfica H que mapeia todos os pontos e retas da primeira imagem na segunda, assim como o contrário. Esta matriz é encontrada a partir da matriz fundamental e dos epipolos da imagem pela seguinte equação:

$$H = [e_2]_x F + e_2 a^T = 0 \tag{5}$$

sendo a um vetor qualquer não nulo, usado para garantir que a matriz H tenha uma inversa e $[e_2]_x$ é a matriz anti-simétrica do epipolo da segunda imagem, definida por:

$$[e_2]_x = \begin{bmatrix} 0 & -e_2z & e_2y \\ e_2z & 0 & -e_2x \\ -e_2y & e_2x & 0 \end{bmatrix} \quad (6)$$

Assim, conhecendo a matriz homográfica, as linhas epipolares correspondentes podem ser facilmente calculadas usando:

$$I_2 = H^{-T} I_1 \quad e \quad I_1 = H^T I_2 \quad (7)$$

Homografias são estimadas entre imagens para detectar características correspondentes nessas imagens. Dentre os algoritmos capazes de estimar a matriz fundamental é possível citar o RANSAC.

Para a estimação da matriz fundamental, o RANSAC calcula as verdadeiras correspondências (*inliers*) para cada matriz H e escolhe a que maximiza esse número. Tendo eliminado os *outliers*, a matriz H é recalculada com o objetivo de melhorar a estimação.

Mesmo com a matriz homográfica calculada, para fazer a junção de um par de imagens é necessário retificar as imagens no sentido de minimizar distorções e suavizar as junções. As retificações utilizadas neste artigo foram: Planar e Cilíndrica. O cálculo da homografia em si com retificação, juntamente com a junção das imagens é feito pelo algoritmo da Transformação Linear Direta – DLT (Hartley & Zisserman, 2004).

2.5.1 Estimando a matriz fundamental

RANSAC é um procedimento de estimação robusto que usa um conjunto mínimo de correspondências amostradas, para estimar os parâmetros de transformação da imagem e achar a solução que tem o melhor consenso com os dados. Os métodos clássicos procuram utilizar o maior número de pontos para obter uma solução inicial e, então, eliminar os pontos inválidos. O RANSAC, ao contrário destes métodos, utiliza apenas o número mínimo e suficiente de pontos necessários para uma primeira estimativa, aumentando o conjunto com novos pontos consistentes sempre que possível (Fischler & Bolles, 1981).

Uma vantagem do RANSAC é a sua habilidade de realizar a estimativa de parâmetros de um modelo de forma robusta, ou seja, ele pode estimar parâmetros com um alto grau de acerto mesmo quando um número significativo de *outliers* esteja presente nos dados analisados. Uma desvantagem do algoritmo é que ele tem de possuir uma quantidade pré-estabelecida de iterações e com isso a solução obtida pode não ser a melhor existente.

Para o problema específico de remoção de *outliers* na correspondência de imagens, a Matriz Fundamental (H) pode ser determinada da seguinte maneira:

- Selecionar randomicamente um subconjunto de oito pontos correlacionados, retirados do conjunto total de pontos correlacionados;
- Para cada subconjunto, indexado por j , calcular a matriz fundamental F_j através do algoritmo de oito pontos;
- Para cada matriz F_j computada, determinar o número de pontos com distância até a linha epipolar, ou residual, menor que um limiar;
- Selecionar a matriz F que apresenta o maior número de pontos com residual inferior ao máximo definido;
- Recalcular a matriz F considerando todos os pontos *inliers*.

Uma visão mais detalhada da relação do RANSAC com a geometria epipolar é apresentada a seguir. Pela geometria epipolar, é possível calcular a matriz fundamental entre dois pares de imagens. Considere $m = [x, y, 1]$ um ponto sobre o plano da imagem L e $n = [x', y', 1]$ um ponto sobre o plano da imagem L' . Assim, a equação 8 define a matriz fundamental.

$$m^T F n = 0 \quad (8)$$

Diversos métodos para estimação da matriz fundamental são encontrados na literatura, contudo o método mais conhecido é o algoritmo de 8 pontos. Tal método, dado um conjunto com $n \geq 8$ correspondências, estima a matriz fundamental de forma linear, solucionando a Equação 9.

$$\sum_{i=1}^n \|m_i^T F n_i\|^2 \quad (9)$$

A estimação robusta da matriz fundamental é feita pesando o residual para cada ponto. O resíduo é mostrado pela equação 10, sendo r o resíduo e i o número do par de pontos na lista de pontos correlacionados.

$$r_i = m_i^T F n_i \quad (10)$$

Para o cálculo das homografias neste projeto é selecionado um conjunto mínimo de $S = 4$ correspondências de características e o processo é repetido N vezes ($N = 200$) para um limiar t (distância máxima do modelo que um dado pode estar para ser considerado um *inlier*) de 4 *pixels*.

2.6 Retificação de imagens

Dois maneiras de retificação são conhecidas na literatura e ambas determinam que o par de imagens a ser retificadas deva ser reorganizado a partir de uma projeção. Os algoritmos diferem basicamente na forma como as imagens serão reprojadas (Roberto et al., 2009).

O método tradicional de retificar um par de imagens consiste em re-projetar as imagens num plano em comum paralelo à *baseline*. Desta forma, quando a imagem for mapeada numa região em comum deste plano, têm-se a garantia que linhas epipolares correspondentes estarão na mesma altura (Fusiello et al., 2000). Conhecida como retificação planar, esta abordagem é relativamente simples de ser implementada. Porém, ela falha com alguns movimentos de câmara. Isto se deve ao fato de que, quanto mais próximo da imagem o epipolo estiver, maior será o tamanho da imagem retificada, culminando no caso extremo, onde o epipolo está localizado dentro da imagem, que resultaria numa imagem de tamanho infinito.

A segunda maneira, chamada de retificação cilíndrica, consegue tratar esses casos. Ela se diferencia da retificação planar principalmente por, ao invés de usar um plano em comum, usar um cilindro em comum para re-projetar o par de imagens. O método consiste em determinar um cilindro de raio unitário que tem a *baseline* com eixo de revolução e, em seguida, mapear cada *pixel* da imagem numa coordenada (z, θ) de um sistema de coordenadas cilíndricas, que pode ser usado normalmente, como um ponto (x, y) na imagem (Roy et al., 1997). Apesar de mais geral, esta técnica é bem mais complexa de ser implementada e possui um alto custo computacional, pois todos os cálculos realizados para cada *pixel* da imagem são feitos num espaço tridimensional.

A retificação cilíndrica, entretanto, pode ser simplificada se as informações das geometrias projetiva e epipolar forem usadas. Desta forma, é possível realizar todos os cálculos no plano de imagem, evitando assim operações tridimensionais. A ideia da retificação cilíndrica simplificada é muito semelhante à anterior, ou seja, reparametrizar a imagem num sistema de coordenadas cilíndrico. Entretanto, ela difere no cilindro escolhido. Enquanto na implementação convencional o cilindro é centrado na *baseline*, na forma simplificada as transformações ocorrem ao redor dos epipolos e, como estes estão no mesmo plano da imagem, nenhuma operação ocorrerá no espaço tridimensional (Pollefeys et al., 1999).

Como mostrado na Figura 7, cada linha epipolar possui um ângulo θ em relação ao epipolo, assim como cada *pixel* dela está a uma distância r deste mesmo ponto. Desta forma, as linhas epipolares são reescritas horizontalmente na nova imagem. No final, o par estará retificado porque linhas epipolares correspondentes possuem o mesmo ângulo em relação ao epipolo, já que elas estão no mesmo plano epipolar.

3. A Metodologia Empregada

Para se ter um padrão e reduzir o custo computacional foram selecionadas 20 imagens por vez para a construção dos mosaicos, resultando em 10 grupos de 20 imagens. A construção do mosaico se inicia pela primeira imagem (a esquerda do mosaico) em direção à última imagem (a direita do

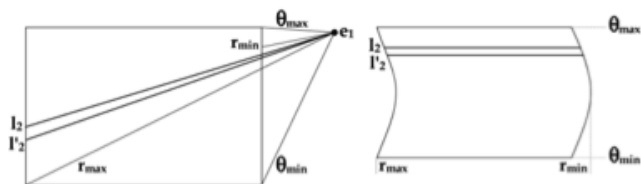


Figura 7. Retificação cilíndrica simplificada (Roberto et al., 2009).

mosaico), do respectivo grupo. As junções são feitas por pares de imagens, ou seja, imagem1 com imagem2, que na sequencia é juntada à imagem3 e assim por diante.

Como as imagens foram obtidas de maneira sequencial, a primeira etapa consiste em extrair as características do primeiro par de imagens pela SIFT, criar os descritores deste primeiro par, fazer as correspondências destas características (pelo método do vizinho mais próximo, com auxílio da BBF para acelerar este processo). Na sequencia é aplicado o RANSAC que estima a matriz homográfica, que é responsável por corresponder partes comuns da primeira imagem com a segunda do par de imagens, eliminando falsos pontos correspondentes (Figura 8).

Com isto é possível realizar a retificação (planar ou cilíndrica) no par de imagens para corrigir possíveis distorções de ângulo ou movimentação entre as imagens e fazer a interpolação nas imagens para que as mesmas possam ser unidas. Enfim, as imagens são unidas pelos pontos correspondentes restantes, após a aplicação do RANSAC e estimação da matriz homográfica. Um momento antes de cada junção ser efetivada é aplicado o algoritmo *Blend Feathering* para suavizar a região de junção. A Figura 9 mostra um exemplo de mosaico sem aplicação do *Blend Feathering*. Todo este processo é repetido para cada par de imagens, até formar um mosaico completo,

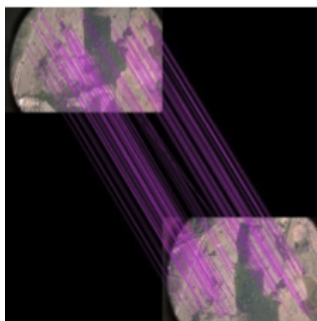


Figura 8. Exemplo de pontos correspondentes após aplicação do RANSAC.



Figura 9. Exemplo de mosaico sem aplicação do *Blend Feathering* nas junções.

incluindo as 20 imagens de cada grupo. Todo este processo é repetido para cada par de imagens, até formar um mosaico completo, incluindo as 20 imagens de cada grupo.

A Figura 10 apresenta um diagrama com a metodologia empregada neste projeto.

4. Resultados

Para gerar os resultados, primeiro foi feita uma análise nas retificações (planar e cilíndrica) para determinar qual é a mais indicada para este projeto. Depois foi realizada a montagem dos mosaicos utilizando a Metodologia Proposta (Projeto), a metodologia comercial livre desenvolvida por [Brown & Lowe \(2007\)](#) (*Autostitch*), e a metodologia comercial livre PTGui. Por fim, os resultados finais foram comparados para verificar a qualidade das junções e o tempo de processamento.

A Figura 11 apresenta os resultados gerados para um mesmo grupo de imagens (contendo 20 imagens), utilizando a retificação cilíndrica e planar.

Observando as retificações realizadas por um profissional da área, foi possível verificar que em 90% dos mosaicos gerados, as retificações planar e cilíndricas geraram resultados similares.

A Figura 12 apresenta um mosaico completo, composto por 20 imagens, que foi construído pelas metodologias citadas anteriormente. Os mosaicos gerados pelo *Autostitch* e pela Metodologia Proposta são visualmente semelhantes, como pode ser observado na Figura 12, já o mosaico gerado pelo PTGui mostrou-se diferente.

A Tabela 1 apresenta os tempos médios em segundos de processamento para cada metodologia gerar um mosaico contendo 20 imagens.

5. Discussão e Conclusões

Os testes de retificações realizados, conforme exemplo da Figura 11, tiveram uma alta porcentagem de resultados similares devido ao fato de que

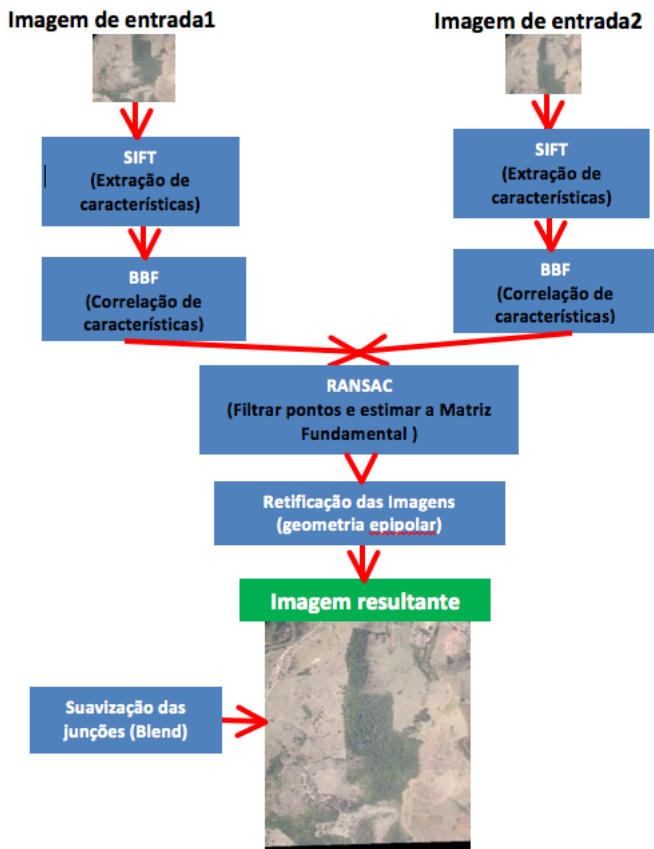


Figura 10. Diagrama da metodologia empregada.

Tabela 1. Tempo de processamento das metodologias usadas.

Metodologia	Tempo (s)
AutoStitch	48
PTGui	39
Metodologia Proposta	60

as transformações se comportaram de maneira estável, pelo motivo de que as imagens obtidas sequencialmente possuem pouca distorção angular entre uma e outra imagem; isto foi comprovado por inspeção visual, por um profissional da área. Para evitar o possível surgimento de uma imagem de

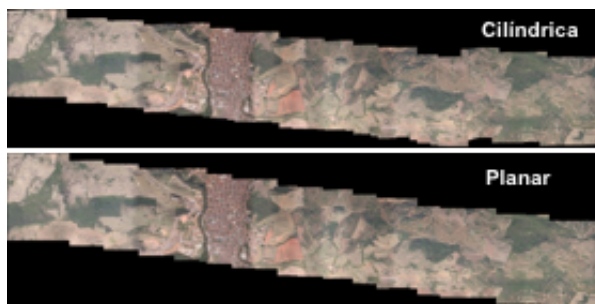


Figura 11. Exemplo da retificação cilíndrica e planar.

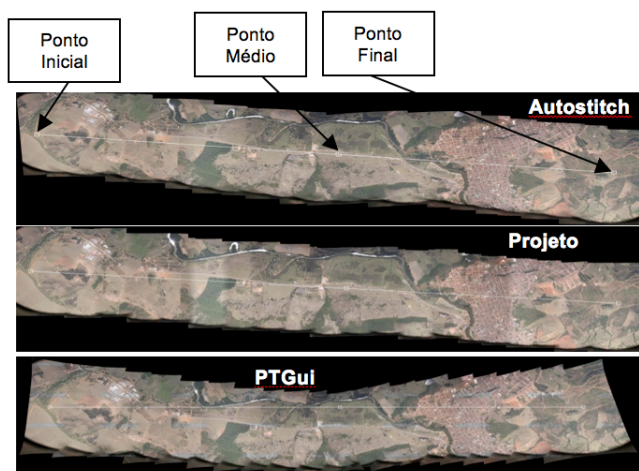


Figura 12. Exemplo de mosaicos gerados.

tamanho infinito (Seção 2.6) que reduz a precisão e qualidade do mosaico resultante, neste projeto foi utilizada a retificação cilíndrica.

Os mosaicos gerados pelas três metodologias (Autostitch, PTGui e Metodologia Proposta) apresentaram resultados visualmente similares, mas se for considerado questões de distorção para gerar os mosaicos, é possível verificar que os resultados das três metodologias diferem muito, como pode ser observado na Figura 13.

Para comprovar as distorções de cada metodologia, foram marcados 3 pontos comuns de junção (ponto inicial, ponto médio e ponto final), conforme pode ser observado pela linha branca na Figura 12 dada como exemplo. O ponto inicial e ponto final sempre foram marcados na mesma localização.

Na Figura 13 dada como exemplo, usando o *Google Earth*, foram feitas as mesmas marcações, nas mesmas posições que foram feitas na Figura 12 e mais uma marcação pelas coordenadas ideais do *Google Earth*. Pelas coordenadas de GPS, foi possível ter uma referência ideal dos pontos marcados do *Google Earth* em relação aos pontos marcados pelas três metodologias utilizadas para comparação. Com isto, foi possível comparar qual metodologia se aproxima mais da marcação do *Google Earth* (ideal) e, verificar qual metodologia gerou mais distorção. De acordo com a Figura 13, a linha vermelha corresponde às coordenadas do *Google Earth*, a linha azul a Metodologia Proposta (Projeto), a linha amarela ao AutoStitch e a linha verde ao PTGui. Na Tabela 2 são apresentadas as coordenadas GPS dos pontos marcados na Figura 13.

Pelas informações da Tabela 2, é possível verificar que os pontos médios pertencentes a cada metodologia variam e isto pode ser comprovado na Figura 13. As metodologias AutoStitch e PTGui distorcem as imagens para que as junções ocorram, não se preocupando muito com a precisão do mosaico, mas visualmente aparentam ser mosaicos de boa qualidade.

Tabela 2. Coordenadas GPS dos pontos marcados.

Coordenada GPS		Ponto	Metodologia
-22,269857	-45,771812	Inicial	
-22,252313	-45,729144	Médio	Autostich
-22,252913	-45,729219	Médio	Projeto
-22,247868	-45,730388	Médio	PTGui
-22,239154	-45,691300	Final	

Foi feita uma inspeção visual por um profissional da área nos 10 mosaicos gerados por cada metodologia empregada neste artigo. Então verificou-se qual metodologia se aproxima mais da coordenada ideal do *Google Earth* (Figura 9). Com todas estas verificações e inspeções, comprovou-se que em 85% dos casos é a Metodologia Proposta que se mais aproxima da referência ideal do *Google Earth*.

Com isto conclui-se que as metodologias comerciais dão ênfase a um baixo tempo de processamento ao invés da precisão do mosaico a ser gerado. De acordo com os resultados da Tabela 1 e pela análise apresentada anteriormente, é possível concluir que a Metodologia Proposta apresenta mosaicos com melhor qualidade e menor distorção em relação às metodologias comerciais, mas com tempo de processamento mais elevado.

Como trabalhos futuros, pode-se citar a implementação de processamento paralelo na etapa de extração de características para ganho de desempenho, uma vez que esta etapa é a mais custosa computacionalmente, a implementação de processamento paralelo para unir vários pares de imagens simultaneamente e a construção de mosaicos de imagens georreferenciados.



Figura 13. Comparação das metodologias utilizadas.

Referências

- Bei, L. & Haizhen, Z., An algorithm of fabric image mosaic based on SIFT feature matching. In: *Proceedings of IEEE International Conference on Artificial Intelligence and Computational Intelligence*. Piscataway, USA: IEEE Press, v. 3, p. 435–438, 2009.
- Beis, J.S. & Lowe, D.G., Shape indexing using approximate nearest-neighbour search in high-dimensional spaces. In: *Proceedings of Conference on Computer Vision and Pattern Recognition*. Washington, USA: IEEE Computer Society, p. 1000–1006, 1997.
- Brown, M. & Lowe, D.G., Automatic panoramic image stitching using invariant features. *International Journal of Computer Vision*, 74(1):59–73, 2007.
- Fischler, M.A. & Bolles, R.C., Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. *Communications of the ACM*, 24(6):381–395, 1981.
- Fusiello, A.; Trucco, E. & Verri, A., A compact algorithm for rectification of stereo pairs. *Journal Machine Vision and Applications*, 12(1):16–22, 2000.
- Hartley, R. & Zisserman, A., *Multiple View Geometry in Computer Vision*. Cambridge, UK: Cambridge University Press, 2004.
- Hong, J.; Lin, W.; Zhang, H. & Li, L., Image mosaic based on SURF feature matching. In: *Proceedings of 1st International Conference on Information Science and Engineering*. Piscataway, USA: IEEE Press, p. 1287–1290, 2009.
- Juan, L. & Gwon, O., A comparison of SIFT, PCA-SIFT and SURF. *International Journal of Image Processing*, 3(4):143–152, 2009.
- Ke, Y. & Sukthankar, R., PCA-SIFT: A more distinctive representation for local image descriptors. In: *Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition*. Los Alamitos, USA: IEEE Computer Society, v. 2, p. 506–513, 2004.
- Li, L. & Geng, N., Algorithm for sequence image automatic mosaic based on SIFT feature. In: *Proceedings of WASE International Conference on Information Engineering*. Los Alamitos, USA: IEEE Computer Society, v. 1, p. 203–206, 2010.
- Lowe, D.G., Distinctive image features from scale invariant keypoints. *International Journal of Computer Vision*, 60(2):91–110, 2004.
- Oram, D., Rectification for any epipolar geometry. In: *Proceedings of British Machine Vision Conference*. Manchester, UK: BMVA, p. 653–662, 2001.

- Pollefeys, M., *Self-calibration and metric 3D reconstruction from uncalibrated image sequences*. Phd. thesis, Departement Elektrotechniek, Katholieke Universiteit Leuven, Leuven, Belgium, 1999.
- Pollefeys, M.; Koch, R. & van Gool, L., A simple and efficient rectification method for general motion. In: *Proceedings of Seventh IEEE International Conference on Computer Vision*. Piscataway, USA: IEEE Press, v. 1, p. 496–501, 1999.
- Roberto, R.A.; Teichrieb, V. & Kelner, J., Retificação cilíndrica: um método eficiente para retificar um par de imagens. In: Conci, A.; Silva, L. & Lewiner, T. (Eds.), *Proceedings of XXII SIBGRAPI Workshops – Undergraduate Work*. Rio de Janeiro, RJ: SBC, 2009.
- Roy, S.; Meunier, J. & Cox, I., Cylindrical rectification to minimize epipolar distortion. In: *Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition*. Washington, USA: IEEE Computer Society, p. 393–399, 1997.

Notas Biográficas

André de Souza Tarallo é graduado em Engenharia de Computação (UNIFEV, 2004) e mestre em Engenharia Elétrica na área de processamento digital de imagens (USP São Carlos, 2007). Atualmente é doutorando na mesma área do mestrado na USP São Carlos e Embrapa Instrumentação.

Francisco Assis da Silva é graduado em Ciência da Computação (UNOESTE, 1998) e mestre em Computação na área de Processamento Digital de Imagens (UFRGS, 2002). Atualmente é doutorando na mesma área na USP São Carlos.

Alan Kazuo Hiraga é graduado em Ciência da Computação (UNOESTE, 2011) e atualmente é mestrando em Computação na UFSCar.

Maria Stela Veludo de Paiva é graduada em Engenharia Elétrica (USP, 1979), mestre e doutor em Física Aplicada (USP São Carlos, 1984 e 1990, respectivamente) e tem pós-doutorado (University of Southampton, 1992). Atualmente é Professor Associado do Departamento de Engenharia Elétrica da USP Carlos.

Lúcio André de Castro Jorge é graduado em Engenharia Elétrica (Faculdade de Engenharia de Barretos, 1987), mestre em Ciência da Computação (USP São Carlos, 2001) e doutor em Engenharia Elétrica na área de processamento digital de imagens (USP São Carlos, 2011). Atualmente é pesquisador da EMBRAPA Instrumentação em São Carlos.