



Anais do VIII Workshop de Visão Computacional - ISSN 2175-6120

WVC 2012 Centro-Oeste

UFG

INSTITUTO DE
INFORMÁTICA
UFG

Goiânia-GO
Brasil

27 a 30 de maio
2012

Construção de Mosaicos de Imagens Aéreas Agrícolas e Comparação com Outras Metodologias

André de Souza Tarallo¹, Francisco Assis da Silva¹, Alan Kazuo Hiraga², Maria Stela Veludo de Paiva¹, Lúcio André de Castro Jorge³

¹Escola de Engenharia de São Carlos - Universidade de São Paulo - USP, Depto. de Engenharia Elétrica, São Carlos-SP.

²Universidade Federal de São Carlos – UFSCar, Depto. de Computação, São Carlos – SP.

³Embrapa Instrumentação Agropecuária, São Carlos – SP.

andre.tarallo@gmail.com, chfran@hotmail.com, alanhiraga@hotmail.com,
mstela@sc.usp.br, lucio@cnpdia.embrapa.br

Abstract

The automatic generation of mosaics of aerial agricultural images increases efficiency in the analysis of agricultural areas and in decisions-making related to pests control, diseases and deforestation. This paper presents a tool for automatic construction of sequential images mosaics. 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 two hundred 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.

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 artigo, 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 e Haizhen (2009) e Li e Geng (2010). Outra técnica usada para este fim é a PCA-SIFT, encontrada no trabalho de Sukthankar e Ke (2004). Uma técnica mais recente que a SIFT, a SURF, tem-se destacado por ter menor custo computacional (HONG et.al., 2009). No entanto, de acordo com o trabalho de Juan e 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 artigo 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 O Banco de Dados de Imagens Aéreas

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 x 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.

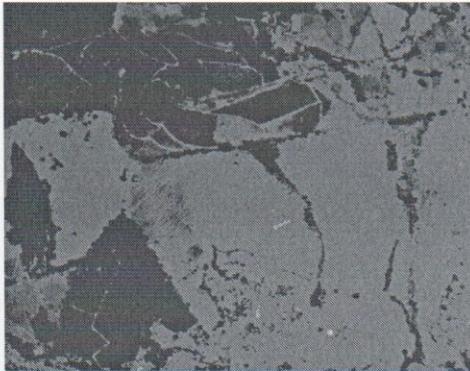


Figura 1 – Exemplo de imagem usada no projeto.

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.

Para gerar um mosaico com a mínima distorção possível, é de extrema importância que seja seguido o padrão apresentado para a obtenção das imagens.

2.2. 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 desses vetores são invariantes à escala, rotação e

parcialmente invariante à mudanças de iluminação e ponto de vista.

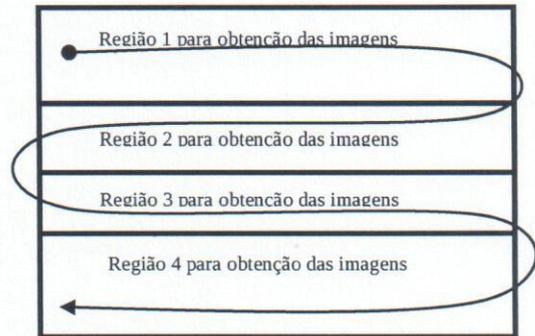


Figura 2 – Padrão de rota do avião.

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:

- **Deteção 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 atribuídas 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 mensurados 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 chave 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 e 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 de características são procuradas na

ordem de suas distâncias mais próximas a partir do local investigado.

Uma razão para o algoritmo BBF (LOWE, 2004) 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 e ZISSERMAN, 2003). Com isso será reduzido o número de falsas correspondências e calculada 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 as imagens. Com isso, é possível determinar o movimento entre tais imagens através da matriz homográfica.

2.4.1. 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 desses 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 e 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* (pontos falsos) 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 calculada, 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* (pontos corretos).

Uma visão mais detalhada da relação do RANSAC com a geometria epipolar vai ser 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 1 define a matriz fundamental.

$$m^T F n = 0 \quad (1)$$

Diversos métodos para estimação da matriz fundamental são encontrados na bibliografia, contudo o método mais conhecido é o algoritmo de 8 pontos. Tal método estima a matriz fundamental de forma linear, considerando um conjunto com $n \geq 8$ correspondências e solucionando a equação 2.

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

A estimação robusta da matriz fundamental é feita ponderando o resíduo para cada ponto. O resíduo é mostrado pela equação 3, sendo r o resíduo e i o número do par de pontos na lista de pontos correlacionados.

$$r_i = m_i^T F n_i \quad (3)$$

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.

3. A Metodologia

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 mosaico), do respectivo grupo. As junções são feitas por pares de imagens, ou seja, "imagem1" com "imagem2", que na sequência 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 desse primeiro par, fazer as correspondências destas características (pelo método do vizinho mais próximo, com auxílio da BBF para acelerar este processo). Na sequência 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 4).

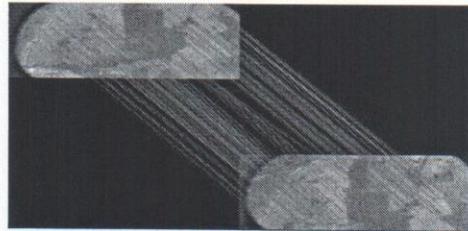


Figura 4 – Exemplo de pontos correspondentes após aplicação do RANSAC.

Com isso 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 anterior de cada junção ser efetivada é aplicado o algoritmo *Blend Feathering* para suavizar a região de junção. A Figura 5 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, incluindo as 20 imagens de cada grupo.

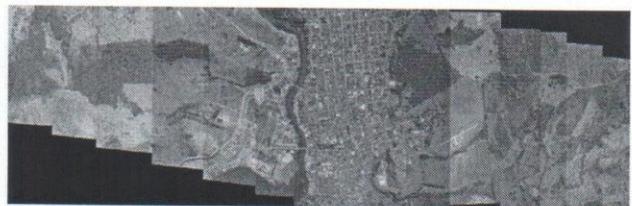


Figura 5 – Exemplo de mosaico sem aplicação do Blend Feathering nas junções.

A Figura 6 apresenta um diagrama com a metodologia empregada neste artigo.

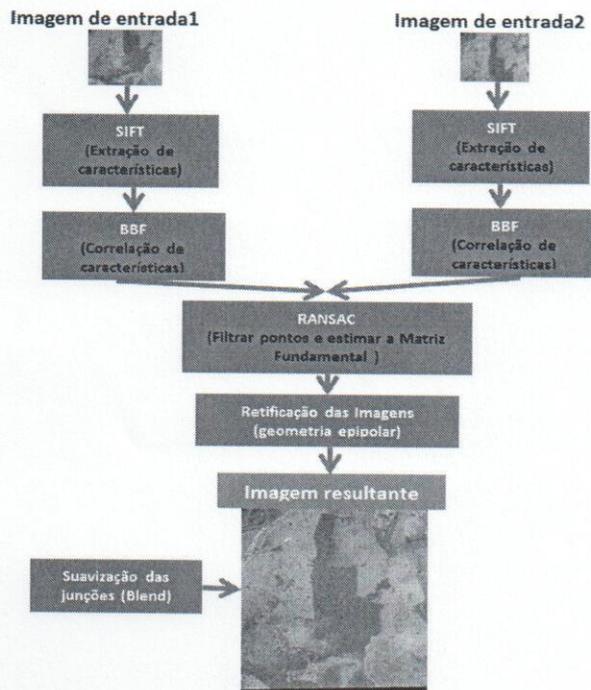


Figura 6 – Diagrama da metodologia empregada.

4. Resultados

Para gerar os resultados, inicialmente 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 (2011). Por fim, os resultados finais foram comparados para verificar a qualidade das junções e o tempo de processamento.

A Figura 7 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 8 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 8; 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.

Tabela 1 – Tempo de Processamento das metodologias usadas.

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

5. Discussões e Conclusões

Os testes de retificações realizados, conforme exemplo da Figura 7, tiveram uma alta porcentagem de resultados similares devido ao fato de que 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 tamanho infinito, 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 considerada a distorção na geração dos mosaicos, é possível verificar que os resultados das 3 metodologias diferem, como pode ser observado na Figura 9. 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 8 dada como exemplo. Pode-se observar que o ponto inicial e ponto final sempre foram marcados na mesma localização e que o ponto médio tem sua posição alterada.

Na Figura 9 dada como exemplo, usando o Google Earth, foram feitas as mesmas marcações, nas mesmas posições que foram feitas na Figura 8 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 3 metodologias utilizadas para comparação. Com isso, 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 9, a linha vermelha corresponde às coordenadas do Google Earth, a linha azul à 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 9.

Pelas informações da Tabela 2, é possível verificar que os pontos médios pertencentes a cada metodologia variam e isto pode ser comprovado na Figura 9. As metodologias AutoStitch e PTGui distorcem as

imagens para que as junções ocorram, não se preocupando com a precisão do mosaico, mas visualmente aparentam ser mosaicos de boa qualidade.

Tabela 2 – Coordenadas GPS dos pontos marcados.

Coordenadas 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 foi possível verificar qual

metodologia aproxima-se mais da coordenada ideal do Google Earth (Figura 9 exemplo). Com todas essas verificações e inspeções, foi possível comprovar que em 85% dos casos é a Metodologia Proposta que se aproxima mais da referência ideal do Google Earth.

Com isto foi possível concluir 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.

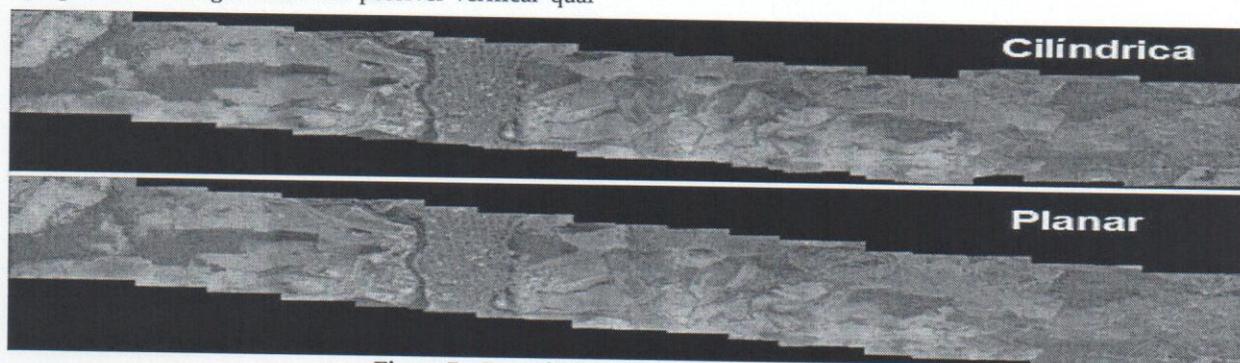


Figura 7 – Exemplo da retificação cilíndrica e planar.

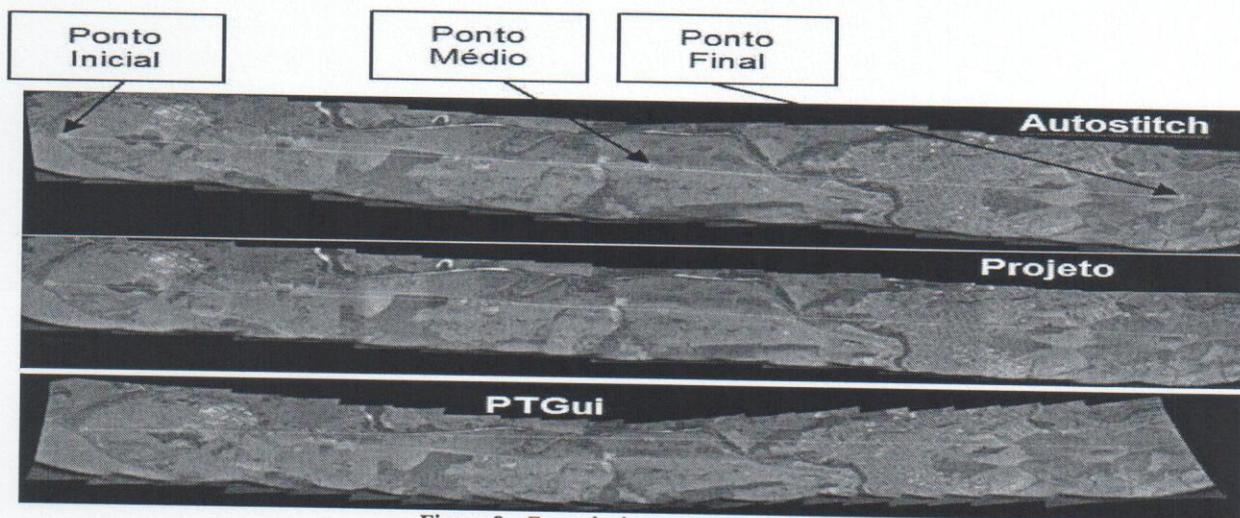


Figura 8 – Exemplo de mosaicos gerados.



Figura 9 – Comparação das metodologias utilizadas.

6. Referências Bibliográficas

Beis, J. & Lowe, D.G., **Shape indexing using approximate nearest-neighbour search in highdimensional spaces**. Conference on Computer Vision and Pattern Recognition, Puerto Rico, pp. 1000-1006, 1997.

Bei, L.; Haizhen, Z. **An algorithm of fabric image mosaic based on SIFT feature matching**. IEEE International Conference on Artificial Intelligence and Computational Intelligence. p. 435 – 438, 2009

Brown, M. & Lowe, D. G., **Automatic Panoramic Image Stitching using Invariant Features**. International Journal of Computer Vision, 2007.

Fischler, M. A. & Bolles, R. C., **Random Sample Consensus: A Paradigm for Model Fitting with Applications to Image Analysis and Automated Cartography**. Conference of the ACM, v. 24, p. 381-395, 1981.

Li, L.; Geng; N. **Algorithm for Sequence Image Automatic Mosaic based on SIFT Feature**. IEEE – International Conference on Information Engineering. P. 203 -206, 2010.

Lowe D. G., **Distinctive Image Features from Scale Invariant**. International Journal of Computer Vision, 2004.

Pollefeys, M., **Self-calibration and metric 3d reconstruction from uncalibrated image sequences**, Ph.D. dissertation, 1999.

Fusiello, A.; Trucco, E. & Verri, A. **A compact algorithm for rectification of stereo pairs**, Machine Vision and Applications, no. 12, 2000.

Hartley, R. & Zisserman, A., **Multiple View Geometry in Computer Vision**. Cambridge University Press, 2ª. ed, 2003.

Hong, J.; Lin, W.; Zhang, H; Li, L. **Image Mosaic Based on SURF Feature Matching**, IEEE - The 1st International Conference on Information Science and Engineering ICISE, 2009

Juan, L.;Gwun, O., **A Comparison of SIFT, PCA-SIFT and SURF**. International Journal of Image Processing (IJIP) Volume(3), Issue(4), 2009

Ke, Y.; Sukthankar, R. **PCA-SIFT: A More Distinctive Representation for Local Image Descriptors**. In IEEE Conference on Computer Vision and Pattern Recognition, 2004.

Oram, D., **Rectification for any epipolar geometry**, in British Machine Vision Conference, 2001, pp. 653–662.

Pollefeys, M.; Koch, R. & Van Gool, L. **A simple and efficient rectification method for general motion**, in Computer Vision, 1999. The Proceedings of the Seventh IEEE International Conference on, vol. 1, 1999, pp. 496–501 vol.1.

PTGui. **Graphical User Interface for Panorama Tools**. Panoramic stitching software, 2011.