



VII Workshop de Visão Computacional

Anais do Evento

www.wvc2011.ufpr.br

22 a 25 de maio de 2011 - UFPR - Curitiba

ISSN - 2175-6120

WVC 2011

Anais do VII Workshop de Visão Computacional

Universidade Federal do Paraná

ISSN 2175-6120

Editores:

Luiz Antônio Pereira Neves

Lucas Ferrari de Oliveira

Hugo Vieira Neto

Curitiba, 22 a 25 de Maio de 2011

Mosaico Automático de Imagens Agrícolas Aéreas

André de Souza Tarallo¹; Maria Stela Veludo de Paiva¹; Lúcio André de Castro Jorge²

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

²Pesquisador, Embrapa Instrumentação Agropecuária, São Carlos – SP.

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

Abstract

The generation of automatic mosaics from high-resolution digital images is a very important tool, because it increases the efficiency in obtaining results and helps fast decision-making. In agriculture, the automatic image mosaics of forest and crop areas, helps different applications in precision farming that needs to answer in a very fast way. A problem of this area is joining two or more images in a new image without great distortion. In this paper it is presented a tool for automatic mosaic reconstruction, using a method based on SIFT transform to join sequential images. Satisfactory results were obtained using two image databases. Image pairs were bonded with 100% accuracy in one of the bases. At another base, this percentage was 66.6%, but the reasons are discussed. It was concluded that tool minimize the time to obtain results (compared with the manual method executed on the field).

Keywords: mosaics, SIFT, agriculture

1. Introdução

Até o presente momento, a construção de mosaicos na agricultura vem sendo feita de maneira semi-automática, necessitando obter o modelo digital do terreno, fazer a ortoretificação de imagens e colocação manual de bandeirinhas, que são usadas como marcadores para que um software possa construir um mosaico deste terreno. Utilizando este tipo de implementação, a construção de mosaicos é demorada e trabalhosa.

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 a controle de pragas, doenças ou queimadas.

Este fato levou à proposta deste projeto, da construção automática de mosaicos para a agricultura a partir de imagens digitais 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.

O problema em utilizar imagens digitais de grande dimensão e alta qualidade de imagens, é que a geração dos mosaicos se torna lenta, devido ao custo computacional. Outro problema é que imagens sequenciais, com mesmas características (mesma textura), dão a impressão de uma única área contínua e, podem dificultar a detecção de pontos comuns entre pares de imagens para formar um mosaico.

Pode ser observado nos trabalhos de [1], [2], [3] e [4], exemplos de aplicações com mosaicos automáticos de imagens aéreas. Um dos problemas desta área é unir duas ou mais imagens, em uma nova imagem com a menor distorção possível.

A estratégia para desenvolver este projeto foi baseada na pesquisa de [5], [6], [7] e [8], com algumas modificações.

Na metodologia empregada foi utilizada a técnica SIFT (*Scale Invariant Features Transform*) [6] para extrair as características de cada uma das imagens (Seção 3), uma *K-d tree* modificada, denominada BBF (*Best-Bin-First*) para encontrar as características correspondentes entre pares de imagens (Seção 4) e o algoritmo RANSAC (*Random Sample Consensus*) para filtrar os pontos falsos correspondentes (Seção 5). Após estes passos, é realizada a junção entre um par de imagens sequenciais, fazendo uma transformada em perspectiva de uma imagem em relação à outra, ou seja, o cálculo da matriz homográfica.

2. Base de Imagens

As bases de imagens utilizadas neste projeto envolvem áreas de vegetação, pastagens e áreas urbanizadas. Foram utilizadas duas bases de imagens, com 48 imagens em cada base. Uma das bases de imagens possui padronização para a obtenção das imagens e a outra não.

As imagens da primeira base de imagens (base 1) foram capturadas em Santa Rita do Sapucaí – MG. As imagens possuem dimensão de 3504 x 2336 pixels, com 24 bits por pixel e 72 *dpi*.

A base 1 foi adquirida de acordo com o seguinte procedimento: a câmera foi acoplada ao suporte do avião de maneira bem horizontal à superfície da terra; o avião percorreu sempre a mesma distância, fazendo movimentos horizontais na área demarcada, para obtenção das imagens; a obtenção entre uma imagem e a próxima imagem foi feita com um tempo determinado e síncrono, entre a imagem anterior e a posterior (3 segundos). Na figura 1 é apresentada uma imagem exemplo da base 1.



Figura 1. Exemplo de imagem utilizada na base 1.

As imagens da segunda base de imagens (base 2) foram capturadas em Araraquara – SP. As imagens foram obtidas manualmente, por um engenheiro de dentro do avião. As imagens possuem dimensão de 3888 x 2592 pixels, com 24 bits por pixel e 72 *dpi*.

A base 2 não seguiu um padrão para a obtenção de suas imagens, como: as imagens não foram obtidas de maneira horizontal em relação à superfície da terra, não tendo como fixar um ângulo de obtenção, já que a obtenção foi manual. Não houve tempo sincronizado de obtenção entre a imagem anterior e posterior; o avião nem sempre fez os movimentos horizontais na área demarcada para obtenção das imagens.



Figura 2. Exemplo de imagem utilizada na base 2.

Estes fatores dificultaram a construção do mosaico, uma vez que muitas características da imagem anterior não aparecem na posterior, e também porque o ângulo de obtenção entre a imagem anterior e posterior era diferente, ocasionando a perda de informações nas imagens. Na figura 2 é possível observar uma imagem exemplo da base 2 de imagens.

Podem ser observadas na Figura 1, a presença de uma borda preta ao redor da imagem. Esta borda é o suporte onde foi acoplada a câmera. Em todas as imagens existe a presença desta borda preta, mas através de um pré-processamento esta borda foi tratada para não afetar o resultado do mosaicamento das imagens.

3. SIFT

A SIFT [5], [6] 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.

As características fornecidas pela 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 SIFT para gerar o conjunto de características de imagens são:

- **Deteção de extremos no Espaço Escala:** Neste primeiro estágio é feita 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, baseada 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 SIFT.

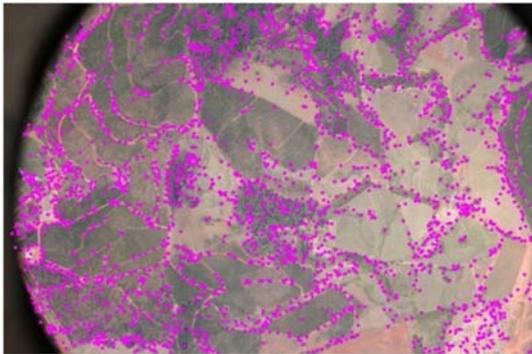


Figura 3. Localização de características com a SIFT.

A técnica SIFT transforma 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.

A quantidade de características é importante para reconhecimento de objetos, sendo que para pequenos objetos em imagem de fundo complexo (imagem cujo fundo contém informação não relevante), são necessárias pelo menos três características corretamente correspondidas de cada objeto para uma identificação confiável.

4. BBF

Uma vez aplicada a técnica 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 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 [9] 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 tem muitas distâncias similares.

5. Ransac

Após a correspondência dos pontos chave, os mesmos são usados para encontrar uma transformada que mapeie as posições dos pontos de uma imagem para as posições dos pontos correspondentes, na outra imagem.

Primeiramente, é usado o RANSAC para excluir falsas correspondências, ou seja, selecionar um conjunto de *inliers* (características consistentes de um conjunto de características) que são compatíveis com a homografia entre as imagens. Após o RANSAC, haverá o cálculo das homografias entre esses pares sequenciais.

RANSAC [10] é um procedimento de estimação robusto que usa um conjunto mínimo de correspondências amostradas randomicamente para estimar os parâmetros de transformação da imagem e achar a solução que tem o melhor consenso com os dados. Para o cálculo das homografias [11] devem ser selecionados conjuntos de $r=4$ correspondências de características; o cálculo da homografia H é realizado utilizando o método da transformação linear direta (DLT). Isto deve ser repetido n vezes ($n = 200$) e após, selecionar a solução que tem o maior número de *inliers* (cujas projeções são consistentes com H com uma tolerância de x pixels) com o uso do RANSAC.

A figura 4 apresenta um exemplo de aquisição dos *inliers* entre duas imagens sequenciais.

5.1. Matriz Homográfica

Existem vários fatores a serem considerados ao fazer mosaicos de partes de imagem de um mesmo panorama. Entre esses fatores estão as transformações que podem ocorrer entre uma imagem e outra devido às distorções causadas pelas lentes da câmera. Para o mosaicamento de imagens, é então, necessário o tratamento dessas imagens, para a correção de tais distorções.

A matriz homográfica [11] mapeia uma transformação do plano projetivo real para o plano projetivo transformado, fazendo as correções geradas por essa transformação.

Um ponto 2D (x,y) em um imagem pode ser representado por um vetor 3D $x=(x_1,x_2,x_3)$, sendo $x=x_1/x_3$ e $y=x_2/x_3$. Isto é denominado de representação homogênea de um ponto e encontra-se no plano projetivo P^2 . Uma homografia é um mapeamento invertível de pontos e linhas no plano projetivo P^2 .

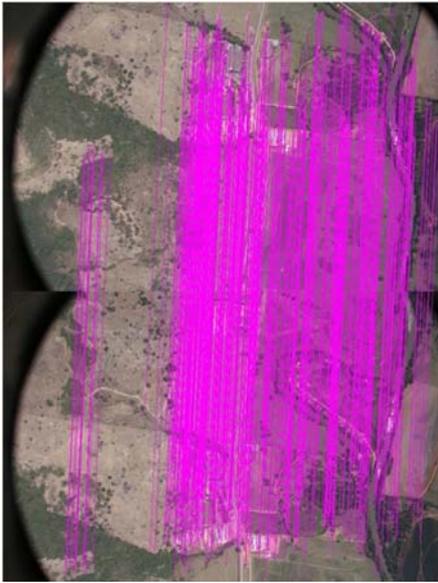


Figura 4. *Inliers* entre duas imagens sequenciais.

Um mapeamento a partir de $P^2 \rightarrow P^2$ é uma projetividade se e somente se existe uma matriz H 3×3 não singular tal que para qualquer ponto em P^2 representado por um vetor x é verdade que seu ponto mapeado é igual a Hx . Isso mostra que para calcular a homografia que mapeia cada x_i com o seu correspondente x_i' , é suficiente calcular a matriz homografia H 3×3 , conforme equação 1.

$$\begin{pmatrix} x_1' \\ x_2' \\ x_3' \end{pmatrix} = \begin{bmatrix} h_1 & h_2 & h_3 \\ h_4 & h_5 & h_6 \\ h_7 & h_8 & h_9 \end{bmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} \quad (1)$$

Sendo $x'=Hx$. É possível notar que H pode ser modificado, multiplicando por uma constante não zero (fator de escala) sem alterar a transformação projetiva. Consequentemente é, possível dizer que H é uma matriz fundamental. Há oito independentes relações, mesmo H tendo nove elementos, formalizando oito graus de liberdade. Estes oito parâmetros precisam ser calculados.

A matriz homográfica possui 8 graus de liberdade no que diz respeito a isometria, transformação de similaridade, transformação afim, transformação projetiva

e projeção em perspectiva. Há transformações simples que contém restrições específicas, reduzindo o número de graus de liberdade.

Homografias são estimadas entre imagens para detectar características correspondentes nessas imagens. Dentre os algoritmos capazes de estimar a homografia é possível citar o DLT (Transformação Linear Direta) e o RANSAC.

Neste projeto é utilizada a transformada em perspectiva, que transforma a imagem de origem através de uma matriz específica, conforme a equação 2.

$$\begin{pmatrix} x \\ y \\ 1 \end{pmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix} = \begin{pmatrix} x' \\ y' \end{pmatrix} \quad (2)$$

Utilizou-se a função *cvWarpPerspective* da biblioteca livre desenvolvida pela Intel, OpenCV [12], para realizar a transformação da imagem (projeção) e assim realizar o mosaico da imagem, através da seguinte estratégia:

$$x' = H * x \quad (3)$$

sendo H a matriz homográfica, x a imagem no plano projetivo real e x' a imagem no plano projetivo.

Após ser realizada a transformação em perspectiva, as imagens estão prontas para serem unidas. Na figura 5 é possível visualizar um exemplo do pareamento entre duas imagens.

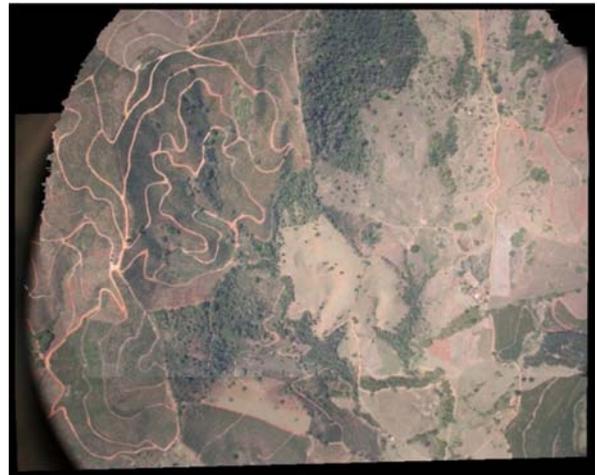


Figura 5. Exemplo de duas imagens unidas.

6. Resultados e Discussão

Para a geração dos resultados, por motivos de custo computacional as imagens da base 1 e base 2 tiveram suas dimensões reduzidas em cinco vezes.

De acordo com os padrões adotados na obtenção das imagens da base 1, foi possível parear todas as imagens desta base sem nenhuma falha (24 pares gerados). A tabela 1 mostra alguns valores médios obtidos com o total de imagens da base 1. Esses valores são: quantidade de pontos detectados na imagem 1 (img1) dos respectivos pares, quantidade de pontos detectados na imagem 2 (img2), quantidade de pares de pontos correspondentes detectados antes da aplicação do RANSAC (antes_RANSAC) e a quantidade de pontos correspondentes (*inliers*) após aplicação do RANSAC (após_RANSAC) e o tempo médio gasto para gerar os 24 pares de imagens, em segundos.

Tabela 1 – Resultados da base 1 de imagens.

	img1	img2	antes_RANSAC	após_RANSAC	tempo (s)
média	4234,6	4182,5	628,5	419,3	840

Já para a base 2 de imagens, que não teve uma padronização na obtenção de suas imagens, foi possível fazer o pareamento de apenas 16 pares de imagens, de um total de 24 (66,6% de acerto). A tabela 2 mostra os resultados da base 2 de imagens, utilizando os mesmos parâmetros dos resultados da base 1.

Tabela 2 – Resultados da base 2 de imagens.

	img1	img2	antes_RANSAC	após_RANSAC	tempo (s)
média	3152,3	3189,8	505,6	478,1	612

É possível observar, comparando os resultados da tabela 1 com o da tabela 2, que os resultados da tabela 2 foram inferiores ao da tabela 1 devido à falta de padronização na obtenção das imagens que geraram os resultados da tabela 2 (base 2).

7. Conclusões

De acordo com os resultados obtidos, verifica-se a possibilidade desta ferramenta funcionar de maneira satisfatória, com geração de resultados confiáveis e de maneira automática, auxiliando com rapidez no controle de pragas, desmatamento e queimadas.

A ferramenta pode ser modificada para atender outras necessidades na área, necessitando para isso somente

adaptações para a nova aplicação em questão, baseada na análise das características que se deseja analisar.

8. Referências Bibliográficas

[1]- Liu, J.; Chen, Z.; Guo, R. A Mosaic Method for Aerial Image Sequence by R/C Model. International Conference on Computer Science and Software Engineering – IEEE, 2008.

[2]- Abedini, A., Hahn, M., Samadzadegan, F. An Investigation Into the Registration of Lidar Intensity Data and Aerial Images Using the SIFT. Proceedings of The International Society for Photogrammetry, and Remote Sensing - ISPRS, 2008.

[3]- Goltz, G. A. M.; SHIGUEMORI, E. H. . Aplicação do Algoritmo SIFT em Imagens de Navegação Autônoma. Workshop Anual de Pesquisa e Desenvolvimento do IEAv, 2008.

[4]- Lin, Q. Yu, and G. Medioni, Map-Enhanced UAV Image Sequence Registration. IEEE Workshop on Applications of Computer Vision (WACV), 2007.

[5]- Lowe D. G. Object Recognition from Local Scale-Invariant Features. Proceeding of ICCV. Institute of Electrical and Electronics Engineers Inc, 1999.

[6]- Lowe D. G. Distinctive Image Features from Scale Invariant Keypoints. International Journal of Computer Vision, 2004.

[7] - Brown, M. and Lowe D. G. Recognising Panoramas, International Conference on Computer Vision, 2003.

[8] - Brown, M. and Lowe D. G. Automatic Panoramic Image Stitching using Invariant Features, International Conference on Computer Vision, 2007.

[9] - 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

[10] - FISCHLER, M. A.; BOLLES, R. C. Random Sample Consensus: A Paradigm for Model Fitting with Applications to Image Analysis and Automated Cartography. Comm. of the ACM, v. 24, p. 381- 395, 1981.

[11]- Dubrofsky, E. Homography Estimation, dissertação de mestrado, Universidade de Columbia-Vancouver, 2009.

[12]- <http://opencv.willowgarage.com/wiki/>, acessado em 12 de Janeiro de 2011.