

Construção automática de mosaico de imagens agrícolas aéreas sequenciais**

André de Souza Tarallo^{1*}, Alan Kazuo Hiraga^{2*}, Maria Stela Veludo de Paiva^{3*},
Lúcio André de Castro Jorge^{4*}, Francisco Assis da Silva¹

¹ Pós-graduando, Departamento de Engenharia Elétrica, Escola de Engenharia de São Carlos - EESC, Universidade de São Paulo - USP, Av. Trabalhador São-carlense, 400, CEP 13566-590, São Carlos, SP, Brasil

² Pesquisador do Departamento de Ciência da Computação, Universidade do Oeste Paulista – Unoeste, Rua José Bongiovani, 700, Cidade Universitária, CEP 19050-680, Presidente Prudente, SP, Brasil

³ Docente do Departamento de Engenharia Elétrica, Universidade de São Paulo – EESC, Av. Trabalhador São-carlense, 400, CEP 13566-590, São Carlos, SP, Brasil

⁴ Pesquisador da Embrapa Instrumentação, Rua XV de Novembro, 1452, São Carlos, SP, Brasil

*e-mail: andre.tarallo@gmail.com; alanhiraga@hotmail.com; mstela@sc.usp.br; lucio@cnpdia.embrapa.br

**Parte da tese de doutorado do primeiro autor

Resumo: A geração de mosaicos automáticos a partir de imagens digitais de alta resolução é uma ferramenta muito importante, porque aumenta a eficiência na obtenção dos resultados e ajuda nas tomadas de decisão. Na agricultura, mosaicos automático de imagens de áreas de floresta e de cultivo, ajudam diferentes aplicações na agricultura de precisão que necessitam obter respostas de maneira rápida. Um problema desta área é juntar duas ou mais imagens em uma nova imagem com a menor distorção possível. Neste artigo foi apresentado uma ferramenta para construção automática de mosaicos, usando um método baseado na transformada SIFT para juntar imagens sequenciais. Para gerar os resultados, foram utilizadas 50 imagens aéreas sequencias, obtendo 100% de junção das imagens. Os resultados foram inspecionados visualmente por um profissional da área, que confirmou a boa qualidade das junções.

Palavras-chave: agricultura, automático, imagens, mosaicos, sift.

Automatic construction of agricultural aerial image mosaic sequential

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 construction, using a method based on SIFT transform to join sequential images. Satisfactory results were obtained using two image databases. To generate the results, we used 50 aerial sequences, obtaining 100% of junction of the images. The results were visually inspected by a professional, which confirmed the quality of the joints.

Keywords: agriculture, automatic, images, mosaics, sift

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 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, podendo demorar um dia todo ou mais dias.

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.

O problema em utilizar imagens digitais de grande dimensão e alta qualidade, 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. Finalmente não é trivial a união de duas ou mais imagens, em uma nova imagem com a menor distorção possível.

Este fato levou à implementação de uma ferramenta, apresentada neste artigo, para a 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. Para resolver o problema da lentidão, uma técnica de processamento paralelo vai ser aplicada.

2. Material e métodos

Primeiramente, as 50 imagens que formam a base de imagens aéreas possuem dimensão de 3504×2336 pixels, com 24 bits por pixel e 72 dpi e foram obtidas em Santa Rita do Sapucaí – MG. A base de imagens foi dividida em 5 subclasses de 10 imagens para acelerar o processamento. A base de imagens 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 s). Na Figura 1 é apresentada uma imagem exemplo da base de imagens.

Neste projeto foram utilizadas as técnicas SIFT (*Scale Invariant Feature Transform*) (LOWE, 2004) BBF (*Best-Bin-First*) (BEIS; LOWE, 1997) e RANSAC (*Random Sample Consensus*) (FISCHLER; BOLLES, 1981). A SIFT é utilizada para a extração de características, a BBF para encontrar pontos correspondentes entre pares de imagens, a partir das características extraídas e o RANSAC é utilizado para filtrar os falsos pontos correspondentes entre imagens.

A 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. 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 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



Figura 1. Exemplo de imagem da base de imagens aéreas.

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 2 pode ser observado um exemplo da localização de características através da SIFT.

Para encontrar correspondências entre duas imagens, conhecendo-se os descritores, é necessário localizar os melhores candidatos a serem seus equivalentes na outra imagem. Isto é feito procurando-se o vizinho mais próximo ou *nearest neighbor* do descritor do ponto entre todos os possíveis candidatos. Quando se procura classificar um ponto em um extenso banco de dados de descritores para vários objetos, a busca exaustiva do vizinho mais próximo pode ser demorada, assim Lowe propõe o uso de uma variação do K-D Tree (FRIEDMAN; BENTLEY; FINKEL, 1977), o BBF (BEIS; LOWE, 1997) para acelerar a busca (LOWE, 2004). Para determinar-se o vizinho mais próximo, é utilizada a mínima distância euclidiana entre os descritores. Porém, nem sempre o vizinho mais próximo é o ponto

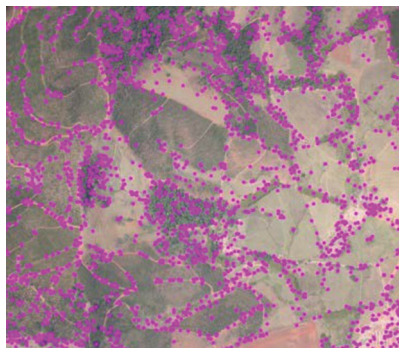


Figura 2. Localização de características com a SIFT.

procurado, assim é necessária a utilização da distância euclidiana com segundo vizinho mais próximo, sendo excluídas correlações que tem a razão entre o vizinho mais próximo e o segundo maior que um determinado limiar, Lowe propõe 0,8, removendo assim 90% de falsos verdadeiros (*outliers*) e apenas 5% de correlações corretas (LOWE, 2004).

Algumas vezes, pode acontecer do par definido pela correlação não se o par correto, assim, o sistema deve ser capaz de identificar e remover essas falsas correlações. Uma solução para o problema é através da geometria epipolar, que será explicada brevemente a seguir, visto que a única restrição geométrica entre dois pares de imagens não calibrados que existe é a geometria epipolar. Considerando a geometria epipolar, é possível calcular a matriz fundamental. Sendo, $m = [x, y, 1]$ um ponto sobre o plano imagem L e $n = [x', y', 1]$ um ponto sobre o plano imagem L' (DUBROFSKY, 2009). Assim a Equação 1 define tal matriz.

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

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

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

A estimação robusta da matriz fundamental é feita pesando o residual 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.

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

Muitas funções peso diferentes foram propostas, cada uma sendo uma nova variação do método. Os resultados obtidos são muito bons na presença de *outliers*, mas são ruins quando os pontos não estão bem localizados. Uma das funções mais utilizadas é o RANSAC que é baseada na seleção randômica de um conjunto de pontos que são usados para a estimação linear da matriz



Figura 3. Resultados aplicando transformações.

fundamental. O RANSAC calcula o número de *inliers*, ou verdadeiras correlações, para cada matriz F e escolhe a que maximiza esse número. Tendo eliminado os *outliers*, a matriz F é recalculada com o objetivo de obter melhoramento da estimação. Uma vez que os pares de correspondências foram determinados entre as imagens, com as devidas transformações entre as imagens, é possível conectar sequências de imagens.

3. Resultados e discussão

Foi aplicado as transformações planar, cilíndrica e esférica (OPENCV, 1999; SZELISKI; SHUM, 1999) para achar as correspondências entre duas imagens sequenciais, testando o algoritmo desenvolvido. Cada uma das transformações tem a desvantagem de que com o aumento da quantidade de imagens sendo inseridas na composição do mosaico, as imagens vão sendo reduzidas ao fazer as junções. Vale lembrar que a construção do mosaico se inicia pelo meio da imagem resultante e cresce no sentido das duas extremidades da imagem resultante. A Figura 3 apresenta uma comparação dos resultados gerados pelas três transformações. De acordo com a aplicação das três transformadas em cada uma das 5 subclasses da base de imagens, foi possível observar por inspeção visual por um profissional da área que a transformação cilíndrica gerou imagens finais com menos distorção ao formar o mosaico.

Para gerar os resultados e diminuir o custo computacional, uma vez que a técnica de processamento paralelo ainda não foi implementada, as 50 imagens tiveram sua dimensão reduzida em 70% da dimensão original apresentada. O tempo de processamento médio para a construção de um mosaico contendo 10 imagens reduzidas é de 10 min.

4. 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 devido à redução do tempo de construção do mosaico, da ordem de horas (semiautomático) para minutos (automático). A ferramenta pode ser modificada para atender outras necessidades na área, relacionadas à análise do mosaico, necessitando para isso somente adaptações para a nova aplicação em questão.

Referências

- LOWE, D. Distinctive image features from scale-invariant keypoints. **International Journal of Computer Vision**, v. 60, n. 2, p. 91-110, 2004. <http://dx.doi.org/10.1023/B:VISI.0000029664.99615.94>
- BEIS, J.; LOWE, D. G. Shape indexing using approximate nearest-neighbor search in high dimensional spaces. In: CONFERENCE ON COMPUTER VISION AND PATTERN RECOGNITION, 1997, Washington. **Proceedings...** Washington, 1997. p. 1000-1006.
- 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**, v. 24, p. 381- 395, 1981.
- FRIEDMAN, J.; BENTLEY, J.; FINKEL, R. An algorithm for finding best matches in logarithmic expected time. **ACM Transactions on Mathematical Software**, v. 3, n. 3, p. 209-226, 1977. <http://dx.doi.org/10.1145/355744.355745>
- DUBROFSKY, E. **Homography Estimation**. 2009. Dissertação (Mestrado)-Universidade de Columbia, Vancouver, 2009.
- OPENCV. **The open computer vision library**. 1999. Disponível em: <<http://opencv.willowgarage.com/wiki/Welcome>>. Acesso em: 14 abr. 2011.
- SZELISKI, R.; SHUM, H.-Y. Creating Full View Panoramic Image Mosaics and Environment Maps. In: INTERNATIONAL CONFERENCE ON COMPUTER GRAPHICS AND INTERACTIVE TECHNIQUES, 24., 1997, Los Angeles. **Proceedings...** ACM, 1997. p. 251-258.