



UNESP
UNIVERSIDADE ESTADUAL PAULISTA
"JULIO DE MESQUITA FILHO"

ENGFOTO

opto

SBD

BMAC

FAPESP

VI Workshop de Visão Computacional
04-07/Julho/2010 - Presidente Prudente, SP - Brasil
ISSN 2175-6120

RECNPq

FAPESP

engcomp

CAETS

Condições Climáticas:
 Temperatura: 30°C/86°F
 Umidade: 62%

[Home](#)

Home
Local
Programa
Comitês
Tópicos de Interesse
WVC Anteriores
Datas Importantes
Inscrições/Taxas
Submissão de Trabalhos
Parceiros
Apoio aos Participantes
Contatos

Bem-vindos ao site oficial do VI Workshop de Visão Computacional.

A área de Visão Computacional, fortemente consolidada em vários países, tem-se desenvolvido de maneira intensa nos últimos anos, desenvolvimento em grande parte motivado por seu alto poder de geração de novas tecnologias (produtos, processos). No Brasil, essa situação não é diferente, sendo possível observar inúmeros grupos de pesquisadores das principais Universidades e Centros de Pesquisa envolvidos com estudos nesse âmbito do conhecimento.

A área de Visão Computacional é altamente multidisciplinar e caracteriza-se primordialmente pela utilização de imagens digitais associadas a técnicas de Reconhecimento de Padrões, Processamento de Imagens, Fotogrametria, entre outras. Além disso, o estudo de métodos cognitivos, de processos biológicos, processos físicos e estatísticos tem gerado soluções significativas para problemas de vital importância do mundo real.

O conjunto de técnicas oriundas desse novo campo do saber possui aplicações em diferentes áreas do conhecimento humano e com impactos relevantes nas vidas de cidadãos comuns. Entre elas, podemos citar: auxílio no diagnóstico de doenças graves a partir da análise de imagens médicas, navegação autônoma de robôs, de veículos aéreos e terrestres, biometria, sensoriamento remoto e automatização do controle de qualidade em processos industriais (inspeção industrial).

O propósito principal do Workshop de Visão Computacional (WVC) é possibilitar a integração de pesquisadores brasileiros que atuam nessa área, de modo a facilitar a apresentação, divulgação e discussão de trabalhos desenvolvidos ou em desenvolvimento, fomentar e disseminar a nucleação de novos grupos de pesquisas, assim como estimular alunos de graduação a iniciarem seus estudos nessa área por meio de ações específicas de formação que são materializadas por minicursos introdutórios. Nesse sentido, o WVC constitui-se num importante espaço de integração para a troca de experiências acadêmico-científicas, objetivando o desenvolvimento da ciência e da tecnologia brasileira nesse domínio do conhecimento.

A primeira edição do evento ocorreu em 2005 na cidade de Piracicaba, SP. Nos anos subsequentes, o evento aconteceu, respectivamente, nas cidades de São Carlos, São José do Rio Preto, Bauru e São Paulo. Apesar de ter sido, até o presente momento, realizado apenas em cidades paulistas, tem conseguido atrair trabalhos de pesquisadores de diferentes partes do país e, desse modo, está se consolidando como um evento de referência em âmbito nacional. Há interesses para que o evento expanda suas fronteiras, alcançando outros estados, assim, na Assembléia Geral realizada durante o evento, grupos de pessoas representando Universidades e Centros de Pesquisas poderão se candidatar para realizarem as próximas edições. Um histórico dos eventos já efetuados pode ser obtido a partir do **site** <<http://iris.sel.eesc.usp.br/wvc>>, onde é possível acessar o conjunto de informações relacionadas às edições anteriores, bem como o nível qualitativo dos respectivos programas.

O WVC é um evento científico que já está devidamente institucionalizado e tem contado com o suporte financeiro da FAPESP e da CAPES, além do apoio da SBC, das Universidades e dos Departamentos onde o evento acontece. Sua importância tem crescido no cenário nacional, sendo o único evento a tratar especificamente de pesquisas envolvendo Visão Computacional. No ano de 2010 irá alcançar a 6ª edição e será realizado na Faculdade de Ciências e Tecnologia da Universidade Estadual Paulista "Júlio de Mesquita Filho" - UNESP, localizada na cidade de Presidente Prudente. Reafirmamos os objetivos enumerados acima e esperamos contar com sua presença entre os participantes do VI WVC. Desde já nos colocamos a sua inteira disposição.

Prof. Dr. Marco Antônio Piteri
 Coordenador Geral do VI WVC

piteri@fct.unesp.br
 (018) 9784 7767
 [Skype] marco.piteri1

Mosaico Automático de Imagens Agrícolas através da Transformada SIFT

André de Souza Tarallo¹; Roberta Vendramini Gonçalves²; Maria Stela Veludo de Paiva¹; Lúcio André de Castro Jorge³

- 1- Escola de Engenharia de São Carlos - Universidade de São Paulo - USP, Depto. de Engenharia Elétrica, São Carlos-SP
- 2- Universidade Federal de São Carlos – UFSCAR, Depto. de Engenharia de Computação, São Carlos – SP
- 3- Pesquisador, Embrapa Instrumentação Agropecuária, São Carlos – SP.

andre.tarallo@gmail.com, mstela@sc.usp.br, rinoaroberta@yahoo.com.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 with accurate mosaics, 3.4% differences among each pairs of sequential images. It was concluded that tool minimize the time to obtain results (compared with the manual method executed on the field, which has no tool to obtain results quickly) and can help precision farming operations.

Keywords: mosaics, SIFT, agriculture

1. Introdução

A construção automática de grandes mosaicos a partir de imagens digitais de alta resolução é uma área de grande importância, pois aumenta a eficiência na obtenção dos resultados. Porém, em geral, a precisão cartográfica impede o uso destas tecnologias em algumas aplicações. Na agricultura, os requisitos de precisão dos mosaicos de culturas anuais ou perenes não é tão alta, porém a rapidez na sua obtenção é o fator mais crítico. A eficiência nos resultados está ligada à obtenção mais rápida e precisa de informações para tomada de decisão, principalmente no controle de pragas, doenças ou controle de queimadas.

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.

Técnicas de mosaicos de imagens podem ser divididas em duas categorias: métodos baseados em características e métodos não baseados em características. Nos métodos baseados em características, assume-se que características correspondentes entre pares de imagens estão disponíveis, e utiliza esta correspondência para achar as transformadas (que contém as características que unirá duas imagens sequenciais) que relacionará os pares de imagens. A maior dificuldade destes métodos é a aquisição e localização das características na imagem. Boas características são geralmente selecionadas à mão, e a confiabilidade da localização das características é um problema devido a ruídos e oclusões na imagem.

Por outro lado, métodos não baseados em características descobrem transformadas para registro em imagens pela minimização da função soma do quadrado da diferença que envolve a soma de parâmetros. Desde que os métodos não baseados em características não dependam das correspondentes características explícitas, eles não têm problemas associados com a aquisição e localização de características. Contudo, métodos dessa categoria tipicamente requerem que a mudança (translação, rotação etc.) a partir de uma imagem para outra seja pequena, e que as boas chances para os parâmetros de transformada sejam dados como valores iniciais para o programa.

A base de imagens utilizada no projeto contém 390 imagens aéreas, envolvendo áreas de vegetação e pastagens e, áreas urbanizadas; mais informações serão mostradas na Seção 2.

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

A técnica empregada neste projeto baseia-se nas características das imagens, e é denominada como SIFT (*Scale Invariant Feature Transform*), apresentada na Seção 3. A SIFT é capaz de transformar uma imagem, em uma coleção de vetores de características locais, e cada

um desses vetores são invariantes para escala, translação, e rotação; e também é parcialmente invariante para mudanças de iluminação, transformação afim e à visão 3D [6].

Após obtenção das características das imagens é utilizado uma *k-d tree* (*K Dimensional Tree*) para o casamento das características extraídas, depois a utilização do RANSAC (*Random Sample Consensus*), apresentado na Seção 4 - para a extração dos *inliers* (conjunto de características que tem maior probabilidade de estarem corretamente interconectados - relacionados) e, a utilização dos *inliers* para o cálculo da matriz homográfica entre as imagens [7]. A matriz homográfica é utilizada para a transformação em perspectiva em uma das imagens, e finalmente a realização do mosaicamento das imagens, conforme Seção 5.

2. Base de Imagens

As imagens aéreas adquiridas para este projeto foram capturadas em Santa Rita do Sapucaí - MG, com o auxílio de um avião de pequeno porte, contendo uma câmera acoplada em um suporte. As imagens foram adquiridas com tamanho de 800 x 533 pixels, com 24 bits por pixel.

A câmera utilizada é do modelo "Canon EOS 30D" com 12 Megapixels de resolução, com as seguintes características:

- Velocidade do obturador e tempo de exposição: 1/5000 s,
- Abertura da lente e abertura do diafragma: f/3,5,
- Comprimento focal: 18 mm.

Para este artigo foram selecionadas 50 imagens seqüenciais da base de imagens, para realização dos testes e comprovação dos resultados. Neste artigo, as imagens foram pareadas, gerando 25 pares. Na figura 1 é apresentada uma imagem utilizada.



Figura 1. Exemplo de imagem utilizada.

Pode ser observado na Figura 1a 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

As características obtidas pela técnica SIFT são bem localizadas tanto no domínio do espaço quanto no domínio da frequência, reduzindo a probabilidade de erro por oclusão, variações ou ruídos. Além disso, essas características são bem distintas, o que permite que uma única característica seja corretamente identificada com alta probabilidade, quando comparada com um grande banco de dados de características, possibilitando uma base para reconhecimento de objetos e cenários. 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 nos testes iniciais. As etapas que compõem o algoritmo SIFT para obter o conjunto de características de uma imagem são [6]:

- **Deteção espacial-escalar:** Esta primeira etapa faz uma busca em todos os locais e escalas da imagem. A implementação é feita por uma função de diferença de Gauss, também conhecida como *blur* Gaussiano, para identificar potenciais pontos de interesse que são invariantes à escala e orientação. Esta é a parte mais custosa do algoritmo.

- **Localização de pontos-chave:** A cada local candidato, um modelo detalhado é relacionado para determinar a escala e sua localização. Pontos-chave são selecionados baseados em medidas de sua estabilidade.

- **Atribuição de orientação:** Uma ou mais orientações são atribuídas para cada localização de ponto-chave, baseada em direções do gradiente da imagem local. Todas as operações futuras são realizadas nos dados da imagem que foi transformada relativamente à orientação, escala e localização atribuídas para cada característica, fornecendo invariância a essas transformações.

- **Descritor de ponto-chave:** Os gradientes locais da imagem são medidos na escala selecionada, na região ao redor de cada ponto-chave. Estes são transformados em uma representação que, permite níveis significantes de distorção de forma e mudanças na iluminação.

Com as características extraídas a partir de todas as imagens, as mesmas devem ser pareadas. Conjuntos de imagens conectadas irão futuramente se tornar panoramas. Na figura 2 pode ser observado um exemplo da localização de características através da SIFT.

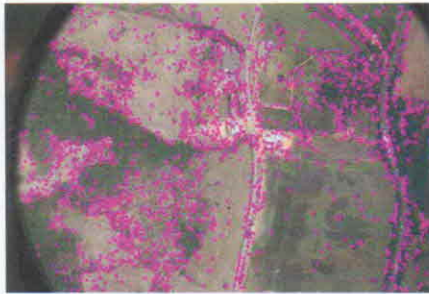


Figura 2. Características localizadas pela utilização da SIFT.

O melhor pareamento de características de um ponto-chave é encontrado ao se identificar seu vizinho mais próximo no conjunto de características. O vizinho mais próximo pode ser definido como o ponto-chave com mínima distância euclidiana, para o vetor de descritores invariantes.

Desde que múltiplas imagens podem ser sobrepostas sobre um simples raio, cada característica é relacionada com seus k vizinhos mais próximos no espaço das características. Isto pode ser feito utilizando-se uma k -*d tree* – Árvore Binária de Pesquisa - para achar os vizinhos mais próximos [6], [7].

Inicialmente, está sendo utilizado $k = 2$ entre duas imagens. Com isso, é feita a busca pela correspondência entre as características e dois vizinhos mais próximos. Encontrados os dois vizinhos, é realizada uma verificação para constatar se a distância euclidiana entre o vizinho mais próximo é menor que a metade da distância entre o ponto-chave e seu segundo vizinho mais próximo. Essa constatação dá maior segurança (em probabilidade), de correto pareamento entre ponto-chave e o vizinho mais próximo.

Encontrados os pares de características é então utilizado o RANSAC para obter os *inliers*.

4. RANSAC

RANSAC é um método de estimação robusto para a extração dos *inliers* do conjunto total de características e então cálculo da matriz homográfica H , calculada através do método direto de transformação linear (DLT). Os *inliers* aumentam a probabilidade de se encontrar uma correta matriz homográfica [7],[8]. A figura 3 apresenta um exemplo dos *inliers* obtidos em uma imagem utilizada.



Figura 3. Exemplo de aquisição dos *inliers* entre duas imagens sequenciais.

4.1. 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 e, para isto utilizam-se técnicas de calibração de câmera.

Na calibração de câmera os fatores que se referem a tais distorções são os fatores intrínsecos e extrínsecos. Os fatores intrínsecos denotam uma projeção das coordenadas do mundo real para coordenadas de imagem (pixel), ou seja, projeção de 3D para 2D. Os fatores extrínsecos denotam um sistema de coordenadas de transformações do mundo real (3D) para um de coordenadas de câmera. Os fatores intrínsecos podem ser subdivididos entre parâmetros intrínsecos e parâmetros de distorção [9]. Os parâmetros intrínsecos (foco da lente, ponto principal e direção dos pontos e parâmetro de inclinação), são utilizados na determinação da matriz intrínseca da câmera que transforma as coordenadas 3D da imagem, em coordenadas 2D. Os parâmetros de distorção determinam os coeficientes de distorção (distorção radial e tangencial) utilizados na correção das distorções da imagem. Os parâmetros extrínsecos são utilizados na determinação do vetor de rotação e translação e junto com a matriz intrínseca da câmera, determinam a matriz homográfica [8], [9].

A matriz homográfica mapeia uma transformação do plano projetivo real para o plano projetivo, fazendo as

correções geradas por essa transformação. 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 [8].

5. Transformação em Perspectiva

Após obtenção da matriz homográfica, uma transformação em perspectiva é utilizada para compor o mosaico de imagens (inicialmente neste projeto, entre imagens pareadas).

A transformação em perspectiva projeta a primeira imagem em relação à segunda de modo que foram corrigidos seus fatores intrínsecos, rotação e translação em relação à segunda imagem, realizando assim uma correta junção entre as imagens, quando projetadas no mesmo plano [8]. Na figura 4 é possível visualizar um exemplo do pareamento entre duas imagens.

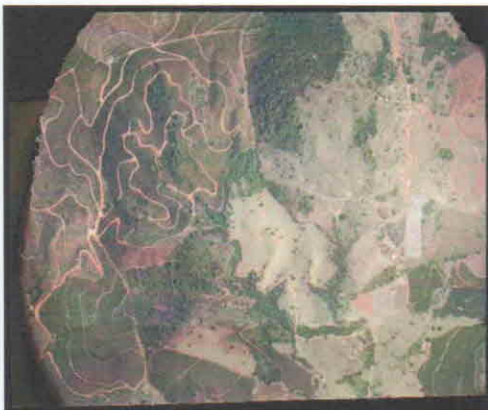


Figura 4. Exemplo de duas imagens unidas.

Utilizou-se a função *cvWarpPerspective* da biblioteca "free" desenvolvida pela Intel, OpenCV [10], para realizar a transformação da imagem (projeção) e assim realizar a mosaicagem da imagem, através da seguinte estratégia:

$$x' = H * x \quad (1)$$

com H = matriz homográfica, x = imagem no plano projetivo real e x' = imagem no plano projetivo.

6. Considerações Sobre a Metodologia Empregada

A metodologia utilizada neste projeto baseou-se na pesquisa desenvolvida por [5], [6], [7], constando para tanto somente as etapas citadas na seção 3,4 e 5. Na pesquisa de Lowe há outras etapas para verificação de concordância de pontos, probabilidades das relações entre

pontos, entre outras. Neste projeto estas etapas intermediárias foram desconsideradas por motivos de simplificação do algoritmo e, por não afetarem a obtenção do resultado final.

A estratégia adotada para unir cada par de imagens, usando as 50 imagens da base, foi a de unir a imagem 1 com a imagem 2, imagem 3 com a imagem 4 e assim por diante. Outras estratégias vão ser adotadas no decorrer do desenvolvimento do projeto.

7. Resultados e Discussão

Uma das estratégias para obter resultados através dos pareamentos obtidos, foi mensurar a quantidade de *inliers* e *outliers* em cada par obtido, quantidade de *inliers* encontrados entre as duas imagens do par mosaicado, bem como a quantidade total de características encontrada em cada imagem.

Para os 25 pares de mosaicos gerados, com tamanhos médios de 800x1066 ou 1600x533 pixels obtiveram-se os seguintes resultados iniciais:

- Relação dos *inliers* com os *outliers* de cada par de imagem formado: 17,4%,
- Relação da diferença de pixels encontrados entre a 1ª imagem e a 2ª imagem de cada par pelos *inliers* encontrados para cada par: 3,4%,
- Quanto à diferença média de características encontradas entre a 1ª imagem e a 2ª imagem de cada par: 4%.

Uma questão a considerar nesta área de pesquisa é o tempo de processamento e, com este fim foram feitos testes de desempenho, utilizando as mesmas 50 imagens, em dois computadores diferentes. Um dos computadores é um Intel i7, modelo 920 - 2.67 GHz com 12Gb de memória RAM; o outro é um Dual Core E5200 -2.5 GHz com 4 Gb de memória RAM. O resultado de desempenho (em segundos), usando as mesmas 50 imagens para obter os 25 pares de imagens, em cada um dos computadores foram os seguintes:

- Média de 29,2 segundos para obter cada um dos 25 pares de imagens, utilizando o Intel i7,
- Média de 41,1 segundos para obter cada um dos 25 pares de imagens, utilizando o Dual Core.

A relação de *inliers* com *outliers* obteve porcentagem com um valor capaz de não interferir na geração dos mosaicos com eficiência, pois na seqüência das 50 imagens obtidas, o cenário muda de característica, mudando a relação das informações adquiridas para cada imagem. Já para a relação entre os *inliers* e de características para cada par de imagens, a porcentagem foi reduzida (comparado com a porcentagem anterior),

pois cada par está associado a duas imagens seqüenciais; com isso as características sofrem pouca variação entre as respectivas imagens.

A seção 4.1 fala sobre a calibração da câmera e de como estes parâmetros podem ajudar a tratar a questão da distorção da câmera, que por consequência trata as distorções da imagem. Este projeto ainda não fez a calibração da câmera, não dispondo assim desta informação, mas mesmo assim os resultados foram satisfatórios.

Quanto ao tempo de processamento, fica explícito a diferença quando é aumentado o poder computacional.

8. Conclusões

De acordo com os resultados obtidos, verifica-se que a ferramenta funciona de maneira satisfatória, capaz de gerar resultados confiáveis e de maneira automática, auxiliando no controle de pragas, desmatamento e queimadas com mais rapidez. A idéia está sendo bem aceita pelos profissionais da área agrícola, uma vez agiliza o trabalho e as tomadas de decisões.

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.

9. Alterações Futuras

Nas etapas seguintes do projeto estão previstas as seguintes implementações:

- Fazer o mosaico com mais de duas imagens simultaneamente é o próximo passo de desenvolvimento do projeto,
- Testar outras maneiras de união entre as imagens seqüências, diferenciando da estratégia adotada na seção 6,
- Aplicar processamento paralelo no algoritmo, para otimizar a performance de tempo e desempenho na obtenção dos mosaicos é um passo futuro que será implementado neste projeto,
- Obter os parâmetros de calibração de câmera e, aplicar estes valores na execução do algoritmo para melhorar os resultados é uma idéia viável e que deve ser colocada em prática neste projeto.

10. Referência 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]- Dubrofsky, E. Homography Estimation, dissertação de mestrado, Universidade de Columbia-Vancouver, 2009.

[10]-<http://opencv.willowgarage.com/wiki/>, acessado em 12 de maio de 2010