

# ***SatImagExplorer: tornando a mineração de dados de sensores orbitais mais flexível*** \*

**Daniel Y. T. Chino<sup>1</sup>, Bruno F. Amaral<sup>1</sup>, Luciana A. S. Romani<sup>2</sup>,  
Elaine P. M. Sousa<sup>1</sup>, Agma J. M. Traina<sup>1</sup>**

<sup>1</sup>Departamento de Ciência de Computação – ICMC/USP – São Carlos – SP – Brasil

{chinodyt|brunoslash}@grad.icmc.usp.br, {parros|agma}@icmc.usp.br

<sup>2</sup>Embrapa Informática Agropecuária – Campinas – SP – Brasil

luciana@cnptia.embrapa.br

**Resumo.** *A agricultura é uma das principais fontes de recursos do Brasil, o que torna importante a análise das safras agrícolas. Devido ao amplo território, imagens obtidas por sensores orbitais têm sido usadas para obtenção de longas séries de dados, cuja análise pode ser realizada por meio de técnicas de mineração de dados. Porém, atualmente as ferramentas de manipulação dessas imagens não possuem funcionalidades para análise, tornando necessária a utilização de ferramentas distintas. Nesse contexto, este artigo apresenta o sistema SatImagExplorer, que integra a extração de dados a partir dessas imagens e técnicas de mineração de dados, auxiliando o agrometeorologista na análise e acompanhamento de safras por longos períodos de tempo.*

**Abstract.** *Agriculture is Brazil's main economic sector. Hence, crop analysis has become a central need in agrobusiness practices. Particularly, due to its large territory, images obtained from orbital sensors can be extremely useful in acquiring and analyzing data. In this context, efficient data mining techniques are inherently required. However, currently tools for satellite images do not provide analysis features. In this context, this paper describes the SatImagExplorer system, which integrates time series generation from satellite images and data mining techniques. The system aids agrometeorologists to evaluate an agricultural crop development along large periods of time.*

## **1. Introdução**

As imagens e dados obtidos por meio de sensores orbitais têm se tornado uma fonte valiosa de informação para apoiar as pesquisas em agricultura, especialmente em países de dimensões territoriais como o Brasil. A diversidade de satélites disponíveis atualmente permite seu uso no monitoramento agrícola de diferentes culturas e na previsão de safras, o que é estratégico para o país. O agronegócio brasileiro corresponde a um terço do Produto Interno Bruto e absorve 40% da mão-de-obra do país. Desse modo, é muito importante tanto para o governo brasileiro quanto para as empresas agrícolas monitorarem e analisarem o desempenho das safras agrícolas, consolidando e ampliando o seu ganho.

---

\*Este trabalho foi realizado com o apoio do CNPq, FAPESP, Microsoft Research, Santander e CAPES. Agradecemos ao CEPAGRI-UNICAMP pelo conjunto de imagens AVHRR/NOAA.

No entanto, para que o gestor ou empresário possa tomar as decisões mais adequadas sobre a produção agrícola, é de extrema importância o uso de sistemas computacionais que o auxiliem a analisar as informações obtidas de várias fontes, como sensores em campo, modelos agrometeorológicos ou mesmo sensores orbitais. Com a melhoria da tecnologia de construção dos sensores, de hardware e software, o volume de dados coletados e armazenados aumenta a cada dia. Logo, é fundamental que ferramentas computacionais especializadas sejam desenvolvidas a fim de tornar viável o trabalho dos especialistas. Uma das características fundamentais desses novos sistemas é a possibilidade de integração de dados provenientes de imagens de sensores orbitais com módulos de análise, elaboração de relatórios e visualização espacial dos resultados.

Neste contexto, este artigo apresenta o sistema *SatImagExplorer*, que recebe como entrada imagens de satélite de diversas fontes e calcula automaticamente índices e medidas de sequências de imagens, gerando séries temporais relativas a tais índices para uma região de interesse (ROI - *Region of Interest*) especificada pelo usuário. Tal ferramenta permite que o especialista possa focar sua análise em pontos estratégicos e de interesse, além de facilitar a análise de longas séries de dados. A partir das séries temporais geradas, é possível aplicar técnicas de mineração de dados como algoritmos de detecção de agrupamentos (*clustering*). Também é possível aplicar técnicas de mineração visual por meio da visualização dos dados extraídos. Na Seção 2 são apresentadas as técnicas e trabalhos correlatos, o sistema *SatImagExplorer* é descrito em detalhes na Seção 3 e as considerações finais são apresentadas na Seção 4.

## 2. Técnicas e Trabalhos Correlatos

O sensor *Advanced Very High Resolution Radiometer* (AVHRR) a bordo dos satélites *National Oceanic and Atmospheric Administration* (NOAA) tem sido amplamente utilizado nos estudos de ecossistemas devido à disponibilidade de longas séries de dados, além de possuir uma cobertura global e ter seus dados distribuídos gratuitamente [Kampel 2004]. Os sensores AVHRR possuem canais no espectro visível e frequências no infravermelho próximo com resolução espacial de  $1\text{km}^2$  de área por pixel<sup>1</sup>. A partir dos canais dos sensores AVHRR é possível calcular índices como o NDVI (*Normalized Difference Vegetation Index*) [Rouse et al. 1973], que indica o vigor vegetativo da planta. Os satélites NOAA, por serem de alta resolução temporal, permitem a geração de séries de imagens que vêm sendo usadas para auxiliar na previsão de índices agrometeorológicos [Gonçalves et al. 2009], na estimativa de produtividade [Nascimento et al. 2009], dentre outras atividades de pesquisa.

Séries temporais extraídas de imagens NDVI podem ser analisadas por meio de técnicas de mineração de dados, como análise de agrupamento, para acompanhar e avaliar o comportamento de culturas de cana-de-açúcar ao longo de safras [Romani et al. 2011]. A detecção de agrupamento visa, por meio da similaridade entre os elementos de um conjunto de dados, determinar agrupamentos de modo não supervisionado, como as técnicas *K-Means* e *K-Medoids* [Han and Kamber 2001].

Ferramentas como o ENVI<sup>2</sup> (*The Environment for Visualizing Images*) permitem construir séries temporais a partir de imagens, porém esse processo é feito manualmente

---

<sup>1</sup><http://satellite.cptec.inpe.br/home>

<sup>2</sup><http://www.ittvis.com/ProductServices/ENVI.aspx>

para cada imagem, demandando um grande tempo do especialista. Os softwares Idrisi<sup>3</sup> e ArcGis<sup>4</sup> também são muito usados para manipulação de imagens georeferenciadas, mas o custo deles muitas vezes é proibitivo. Na literatura, também são encontradas ferramentas que implementam diversos algoritmos de mineração de dados, entre as quais estão Weka<sup>5</sup> e RapidMiner<sup>6</sup>. No entanto, tais ferramentas implementam abordagens genéricas para diversos tipos de dados. Diferentemente das ferramentas citadas, o sistema *SatImagExplorer* permite realizar a extração e análise de séries temporais extraídas de imagens de satélite integradas em uma mesma ferramenta.

### 3. O Sistema *SatImagExplorer*

O sistema *SatImagExplorer* apresenta uma interface simples e intuitiva, na qual o usuário pode abrir uma ou mais imagens de satélite no formato GeoTIFF<sup>7</sup>, definir uma ROI, extrair os índices contidos nas imagens e utilizar técnicas de mineração de dados. Além disso, é possível visualizar arquivos contendo os contornos de municípios (vetores de municípios). O sistema possui uma arquitetura modular, que permite um desenvolvimento incremental de novas técnicas. Na Figura 1 são ilustrados os módulos do *SatImagExplorer* que serão explicados com maiores detalhes nas seções seguintes.

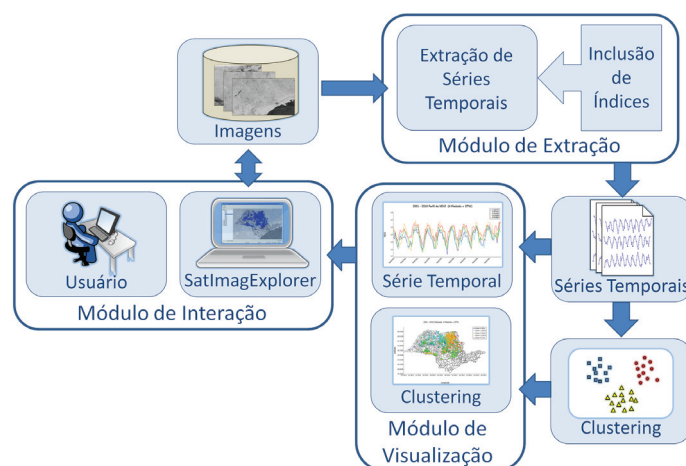


Figura 1. Arquitetura modular do *SatImagExplorer*.

#### 3.1. Módulo de Interação

O usuário interage diretamente no sistema, que é constituído por janelas e menus permitindo uma manipulação intuitiva e ágil. Na Figura 2 são apresentadas telas do sistema em operação. Após abertas, as imagens são exibidas à direita e informações como latitude, longitude e seu respectivo valor (índice) são exibidas na parte inferior. O usuário indica os arquivos de imagens que serão usados como base para a geração das séries temporais e especifica regiões de interesse a serem processadas.

Os métodos de seleção da ROI são ilustrados na Figura 2. Arrastando o mouse, o usuário indica a região poligonal que será utilizada, a região selecionada é então marcada

<sup>3</sup><http://www.clarklabs.org/>

<sup>4</sup><http://www.esri.com/software/arcgis/index.html>

<sup>5</sup><http://www.cs.waikato.ac.nz/ml/weka/>

<sup>6</sup><http://rapid-i.com/content/view/181/190/>

<sup>7</sup><http://trac.osgeo.org/geotiff/>



**Figura 2. Métodos de seleção da ROI.**

na imagem (Figura 2(a)). Alternativamente, pode ser utilizada a opção de especificação das ROIs por meio de listas de coordenadas (latitude e longitude) e da leitura de arquivos texto (Figura 2(b)). Desse modo, pode-se indicar com mais precisão regiões que se deseja analisar. No exemplo, estão marcados locais onde há cultivo de cana-de-açúcar no estado de São Paulo. Outro modo de selecionar uma ROI é por meio de um vetor de município (Figura 2(c)). Assim, o especialista pode realizar análises focadas em municípios. No exemplo, estão selecionados os municípios do estado de São Paulo.

### 3.2. Módulo de Extração de Séries Temporais

Após receber a indicação da ROI, o *SatImagExplorer* extrapola a ROI para todas as imagens da sequência para a geração das séries temporais, calculando as medidas ou índices especificados pelo usuário. Desse modo, são obtidas séries temporais correspondendo ao índice calculado segundo a indicação de tempo da imagem capturada pelo satélite. As séries temporais podem ser armazenadas em um arquivo de dados permitindo que o especialista de domínio os utilize de forma adequada para um processamento posterior.

Outra funcionalidade presente no sistema é a possibilidade do usuário definir índices por meio de fórmulas matemáticas utilizando os canais presentes em uma imagem. Podem ser inseridas expressões matemáticas parentisadas num formato quase linguístico, como em  $a * (b + c)$ , utilizando os canais presentes em cada imagem, constantes, as quatro operações básicas, operadores unários e potenciação.

Deve-se ressaltar que, desse modo, evita-se pré-processar a imagem para incluir um índice experimental ou para calibrar constantes em índices já existentes na imagem. A possibilidade de incluir novos índices e medidas, de maneira automática e simples, traz flexibilidade ao processo de análise e compreensão do comportamento de ROIs em um período de tempo específico, tornando a ferramenta útil na compreensão de fenômenos e no apoio a tomada de decisões.

### 3.3. Módulo de Detecção de Agrupamentos

Foram implementadas e incorporadas ao sistema técnicas de detecção de agrupamentos (*clustering*) para análise e descoberta de padrões sobre as séries temporais extraídas das imagens. O especialista pode aplicar os métodos de *clustering* sobre a área da imagem determinada na etapa de extração das séries, obtendo agrupamentos dos pontos da ROI de acordo com valores do índice utilizado. Os algoritmos já implementados no *SatImagExplorer* são o *K-Means* e o *K-Medoids*. O *K-Means* foi implementado por ser rápido, simples e demonstrar bons resultados, além de ser muito usado em diversas aplicações. O

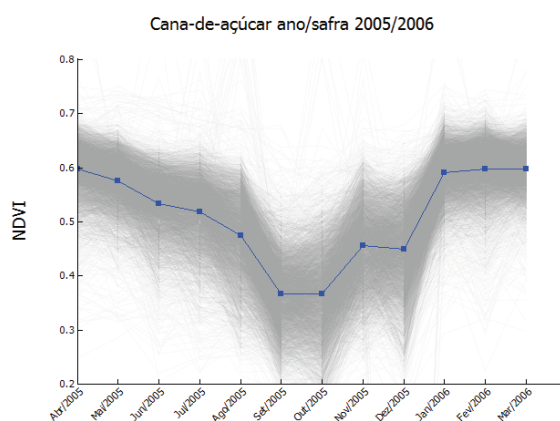
*K-Medoids*, apesar do maior custo computacional, é menos sensível à presença de ruídos e *outliers* no conjunto de dados, assim como à escolha aleatória dos representantes iniciais.

Os métodos de *clustering* realizam cálculos de distâncias entre elementos do conjunto de dados utilizando a distância Euclideana ou a métrica DTW (*Dynamic Time Warping*) [Berndt and Clifford 1994], permitindo duas abordagens na análise dos dados. Em uma delas, utilizando a distância Euclideana, os elementos são definidos por apenas um valor, referente ao índice de um mês de uma safra em uma determinada área, visando análises mensais da ROI. A outra abordagem visa trabalhar com as séries de valores do índice, em que cada elemento do conjunto é definido por uma série correspondente a uma ou várias safras de cana-de-açúcar. Nesse caso, é utilizada a métrica DTW, por ser uma função de distância amplamente aplicada para estudo de séries temporais.

### 3.4. Módulo de Visualização

O módulo de visualização de dados permite tanto o acompanhamento visual do perfil das séries temporais quanto a visualização do resultado das técnicas de *clustering*. As séries temporais podem ser visualizadas em um gráfico. No eixo das abscissas é representada a escala de tempo e no eixo das ordenadas os valores dos índices utilizados.

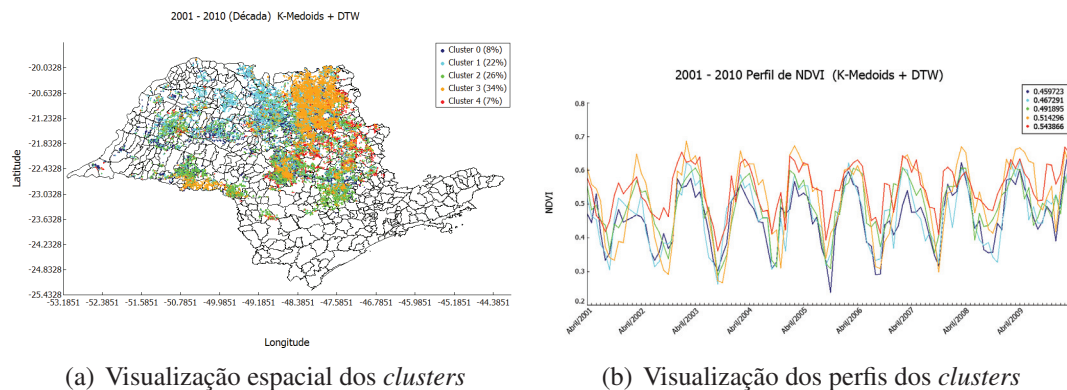
Para cada ROI, o gráfico exibe os valores para todas as coordenadas selecionadas, representadas por linhas em cinza, e os valores médios da região, representados pela linha contínua em azul. Na Figura 3 é ilustrado o gráfico com o perfil NDVI do comportamento anual da cultura de cana-de-açúcar, para o ano/safra de 2006/2007. Os valores de NDVI começam a crescer entre os meses de outubro e novembro, com o início da temporada de chuva, até alcançarem seu pico no mês de março. A colheita da cana-de-açúcar começa no mês de abril e a partir de então os valores de NDVI caem gradativamente até alcançarem os valores iniciais do começo do ciclo.



**Figura 3. Visualização de uma série temporal de NDVI.**

O módulo de visualização para as técnicas de *clustering* é ilustrado na Figura 4. As séries temporais de NDVI são representadas por meio de um gráfico de dispersão de acordo com sua posição geográfica (latitude/longitude) (Figura 4(a)). É possível visualizar o perfil de valores NDVI referentes a cada *cluster* ao longo dos meses no gráfico da Figura 4(b). As cores frias (próximas do azul) representam as áreas com valores mais





**Figura 4. K-Medoids aplicado a áreas de plantio de cana-de-açúcar no estado de São Paulo entre os anos de 2001 a 2010.**

baixos do índice utilizado, enquanto as mais quentes (próximas do vermelho) se referem às séries com valores elevados de NDVI. Analisando os gráficos da Figura 4(b) é possível notar que o *cluster* de cor laranja possui picos mais acentuados em abril e outubro, perfil que condiz com os períodos de maturidade e poda da cana-de-açúcar. Além disso, as áreas pertencentes ao *cluster* laranja se localizam na região nordeste do estado, como mostrado na Figura 4(a), uma região importante na produção de cana-de-açúcar do estado.

#### 4. Conclusões

O sistema *SatImagExplorer* se apresenta como uma solução de código aberto para trabalhar com imagens de satélites e sua arquitetura modular permite estendê-lo com novas funcionalidades. Devido aos processos automáticos e integração com algoritmos de mineração de dados, o sistema também permite realizar análises que apoiam o especialista na tomada de decisão no agronegócio.

#### Referências

- Berndt, D. J. and Clifford, J. (1994). Using dynamic time warping to find patterns in time series. In *KDD'1994*, pages 359–370. ACM Press.
- Gonçalves, R. R. V., Nascimento, C. R., Zullo Jr., J., and Romani, L. A. S. (2009). Relationship between the spectral response of sugar cane, based on avhrr/noaa satellite images, and the climate condition, in the state of são paulo (brazil), from 2001 to 2008. In *Multitemp'2009*, pages 315–322.
- Han, J. and Kamber, M. (2001). *Data Mining - Concepts and Techniques*. Morgan Kaufmann Publishers.
- Kampel, M. (2004). Características gerais dos satélites NOAA: histórico, instrumentos e comunicação de dados. Oficina de textos, São Paulo.
- Nascimento, C. R., Zullo Jr., J., Romani, L. A. S., and Rodrigues, L. H. A. (2009). Identification of sugar cane fields in the state of sao paulo using a time series of avhrr/noaa satellite images. In *Multitemp'2009*, pages 104–111.
- Romani, L. A. S., Gonçalves, R. R. V., Amaral, B. F., Chino, D. Y. T., Zullo Jr., J., Traina Jr, C., Sousa, E. P. M., and Traina, A. J. M. (2011). Clustering analysis applied to ndvi/noaa multitemporal images to improve the monitoring process of sugarcane crops. In *MultiTemp'2011*, pages 5–8.
- Rouse, J. W., Haas, R. H., Schell, J. A., and Deering, D. W. (1973). Monitoring vegetation systems in the great plains with erts. In *ERTS*, pages 309–317. NASA.