

Associação de mineração de dados e imagens do sensor AVHRR/NOAA na busca de padrões para identificação de áreas com cana-de-açúcar no Estado de São Paulo

Cristina Rodrigues Nascimento^(1,2)
Jurandir Zullo Junior^(1,2)
Luis Henrique Antunes Rodrigues⁽¹⁾

¹ Universidade Estadual de Campinas – Faculdade de Engenharia Agrícola
UNICAMP/FEAGRI
Caixa Postal 6011 - 13083-875 - Campinas - SP, Brasil
cristina.nascimento@agr.unicamp.br

² Centro de Pesquisas Meteorológicas e Climáticas Aplicadas à Agricultura –
CEPAGRI/UNICAMP
Caixa Postal 6011 - 13083-875 - Campinas - SP, Brasil
jurandir@cpa.unicamp.br
lique@feagri.unicamp.br

Abstract: Brazil is the first producer of sugar cane in the world. Despite the economic and social importance of agribusiness to Brazil, it is still very difficult to estimate the harvest of the its main crops with the precision and anticipation needed, justifying the study and development of new methods based on the use of remote sensing data. The use of time series of AVHRR/NOAA images presented a very satisfactory outcome, allowing an increase of the objectivity of the methods of agricultural monitoring. The study aims to evaluate the application of harmonic analysis in the time series of images AVHRR/NOAA-17 aimed at identifying areas with sugar cane in São Paulo state. In an analysis was first tested the feasibility of the methodology and use of the technique of decision trees in search of patterns that can be applied to characterize the growing season 2006/2007. the images generated from the application of harmonic analysis in the time series show promise for the identification of areas with sugar cane in the state. The methodology discussed in the patterns of discrimination, from the application of algorithms, decision tree, has proved efficient, but it is limited as the difficulty in locating space of the pixels with such characteristics.

Palavras-chave: Image processing, Harmonic Analysis, NDVI, Processamento de imagens, Análise Harmônica, NDVI.

1. Introdução

O Brasil é, hoje, o principal produtor de cana-de-açúcar do mundo. Seus produtos são largamente utilizados na produção de açúcar, álcool combustível e, mais recentemente, biodiesel. Segundo a CONAB (Companhia Nacional de Abastecimento) a safra 2007/2008 de cana-de-açúcar no Estado de São Paulo aponta aumento do volume produzido, totalizando 387,5 milhões de toneladas, 18,3% a mais em relação à safra anterior (2006/2007), que foi de 327,7 milhões de toneladas. A área em produção está em 4,5 milhões de hectares e outros 727 mil hectares em áreas novas, isto é, que ainda não tiveram o primeiro corte. Apesar da evolução ao logo dos anos e da importância econômica e social do agronegócio para o Brasil, existe, ainda, uma grande dificuldade em estimar com a precisão e a antecipação necessárias à safra das principais culturas agrícolas do país, o que justifica o estudo e a elaboração de metodologias empregando o sensoriamento remoto na mensuração dessas áreas.

A agilidade e a disponibilidade de informações precisas podem facilitar a elaboração de estratégias referentes ao armazenamento, comercialização e suporte às decisões. Dentre os vários sistemas orbitais existentes atualmente, os satélites da série NOAA/AVHRR têm grande potencial de aplicação em métodos operacionais e objetivos de previsão de safras, pois

possibilitam a obtenção de coberturas globais diárias a partir de pelo menos dois satélites devidamente sincronizados e com resolução temporal de doze horas. O Cepagri (Centro de Pesquisas Meteorológicas e Climáticas Aplicada à Agricultura) opera um sistema de recepção destas imagens desde 1995, totalizando, aproximadamente, 5TB (Terabyte), em julho de 2008. O tamanho desta série de imagens tem, sem dúvida, um valor inestimável para as pesquisas na agricultura, devido à diversidade de condições climáticas registradas desde abril de 1995 e à grande evolução da agricultura no país.

A utilização de um grande volume de dados é útil, pois reduz a limitação do AVHRR/NOAA quanto à sua resolução espacial, em aplicações de real interesse prático, como o apoio à estimativa de safras agrícolas com maior objetividade e antecipação. Além disso, o uso de técnicas de sensoriamento remoto aplicadas a sensores de moderada e baixa resolução espacial e alta frequência temporal é fundamental para a obtenção de informação atualizada de regiões extensas, o que tem dado origem à procura de novas técnicas de processamento e análise para este tipo de informação (Pardi Lacruz, 2006). O trabalho tem como objetivo avaliar a aplicação da análise harmônica na série temporal de imagens AVHRR/NOAA-17 visando à identificação das áreas com cana-de-açúcar no estado de São Paulo. Numa primeira análise foi testada a viabilidade da metodologia e a utilização da técnica de árvores de decisão na busca de padrões que possam ser característicos a cultura aplicada à safra 2006/2007.

2. Metodologia de Trabalho

A Figura 1 apresenta a metodologia da pesquisa. Os itens no canto inferior direito dos símbolos indicam posterior seqüência descritiva das etapas.

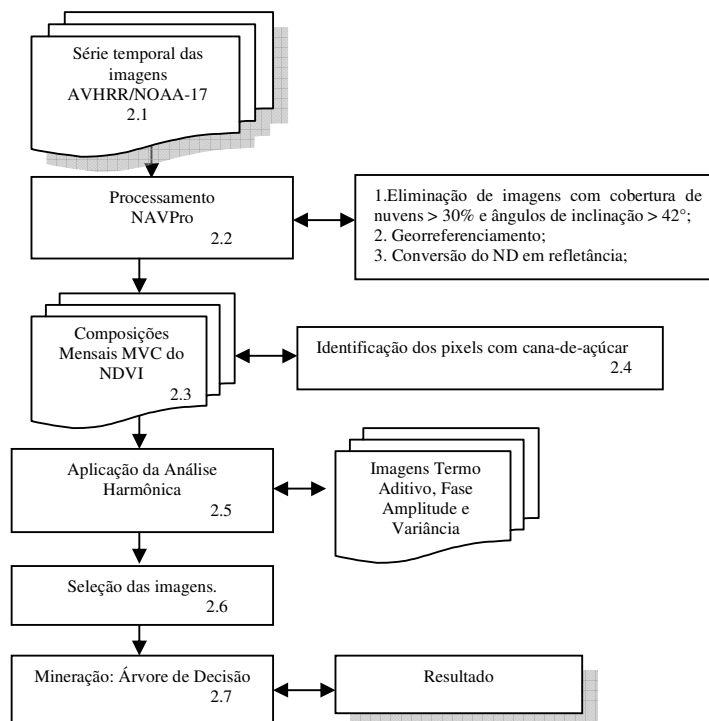


Figura 1. Metodologia adotada na obtenção de padrões para identificação de pixels com cana-de-açúcar.

2.1 Área de Estudo: Série Temporal de Imagens AVHRR/NOAA

Foi utilizada uma série temporal contendo 12 imagens do sensor AVHRR/NOAA, para o ano-safra 2006/2007 (período de abril/06 a março/07). Cada imagem representa a composição mensal (MVC) do NDVI e apresenta como limite as seguintes coordenadas geográficas: 19° 37' 45,12" e 25° 27' 0,72" de Latitude Sul e 44° 1' 47,17" e 53° 15' 3,34" de Longitude Oeste (Figura 2).

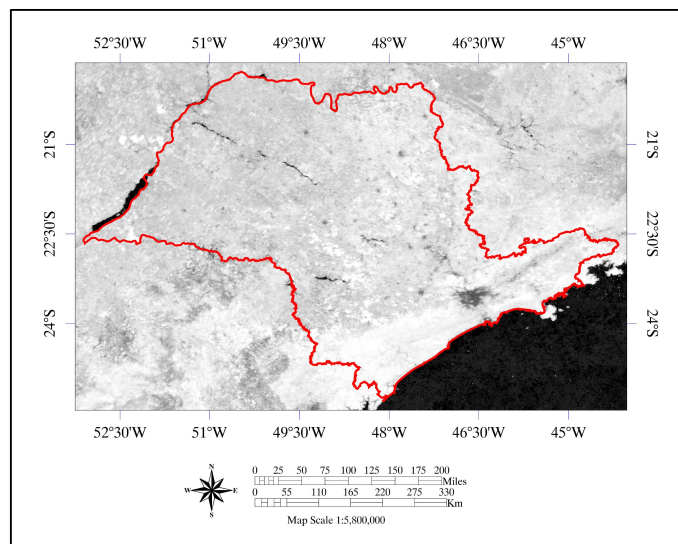


Figura 2. Imagem composição MVC-NDVI do mês de abril de 2006 delimitando a área de estudo.

2.2 Processamento NAVPro: Georreferenciamento e Geração das imagens NDVI no Programa NAVPro (*Navigation Products*)

A questão geométrica das imagens é um fator preponderante quando se pretende explorar a alta resolução temporal do sensor AVHRR/NOAA. Neste sentido Esquerdo et al (2006), em parceria com a universidade do Colorado (Colorado Center for Astrodynamics Research - CCAR) adaptou um sistema automático para o georreferenciamento preciso dessas imagens e posterior geração de produtos. O pacote é composto por uma série de programas executados via linha de comando no IDL (*Interactive Data Language*), necessitando de um conjunto de imagens (série temporal) para geração dos produtos finais. O processamento no NAV inclui: Conversão das imagens do formato HRPT (*High Resolution Picture Transmission*) “raw data” para o formato CCAR, bastante similar ao Level 1B; transformação dos níveis de cinza das imagens dos canais 1 e 2 para valores de refletância no topo da atmosfera; mascaramento de nuvens; geração de imagens NDVI e geração das composições de máximo valor ou MVC do índice de vegetação NDVI.

2.3 Composições mensais (MVC) do índice de vegetação NDVI

As composições de máximo valor no intervalo mensal foram geradas automaticamente por meio da rotina IDL acima descrita. O processamento excluiu imagens com mais de 30% de cobertura de nuvens em razão do sombreamento dos pixels, que proporciona um “mascaramento” do valor do NDVI. Esse parâmetro tem sido utilizado com sucesso na estimativa de produtividade em diversas culturas (Taylor et al, 1998; Qi et al.,1994). Foram descartadas também, imagens com ruídos de aquisição e as obtidas em passagens com baixo grau de elevação (ângulo zenital solar maior que 70° e ângulos de varredura maiores que 42°,

em relação à estação de recepção, recebendo valores de -1,1) do satélite devido à dificuldade na realização da correção geométrica.

2.4 Identificação dos pixels com cana-de-açúcar: Referência Terrestre

Foi utilizado como referência terrestre na identificação das áreas com cana-de-açúcar o mapeamento que vem sendo realizado pelo CTC (Centro de Tecnologia Canavieira) em parceria com o CANASAT/INPE para a safra 2006/2007 analisada. Os vetores considerados para análise correspondem à órbita/ponto 220/75 das imagens do satélite LANDSAT, onde concentram-se grande parte dos principais municípios produtores. A Figura 2 mostra a metodologia empregada na separação dos pixels tomando-se como exemplo o município de Sertãozinho. Esta análise é de fundamental importância, pois na cena é possível encontrar pixels com diferentes estádios fenológicos da cultura. Devido à resolução espacial do sensor AVHRR em relação ao sensor ETM+ foi considerado para análise apenas os pixels puros, ou seja, aqueles que representam exclusivamente áreas com cana-de-açúcar.

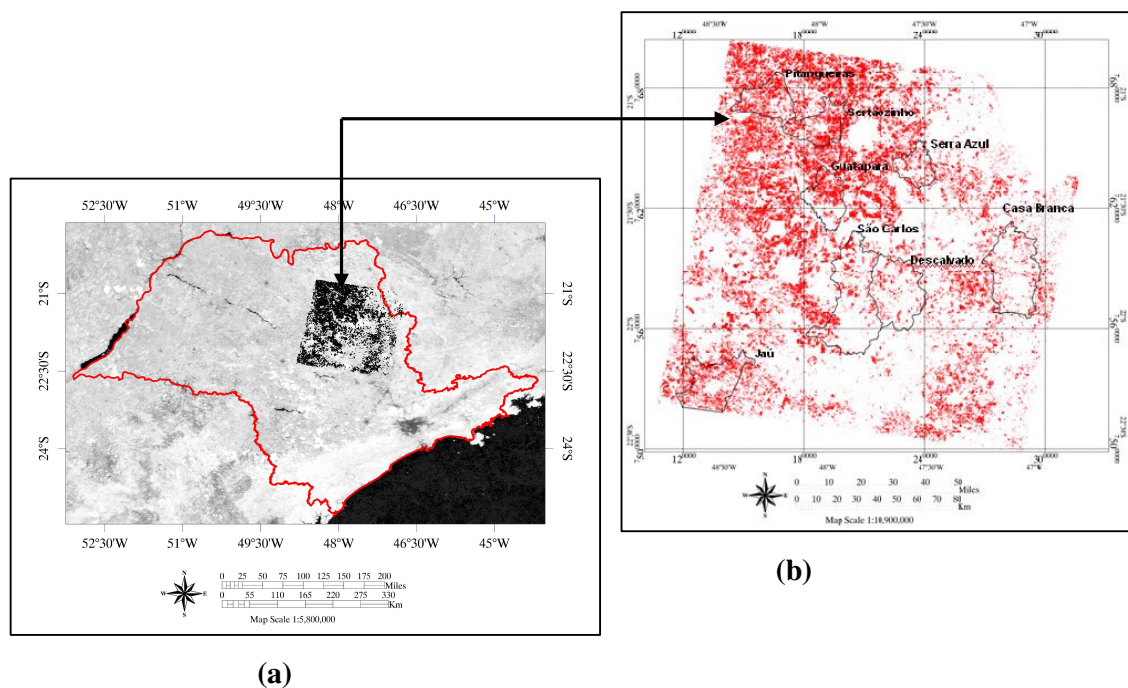


Figura 3. Cobertura espacial da cena 220/75 LANDSAT-7, sensor ETM+, utilizada para seleção dos pixels puros de cana-de-açúcar na safra 2006/07.

2.5 Análise Harmônica

A análise harmônica consiste na representação de flutuações ou variações em uma série temporal que se originou da adição de uma série de funções seno e co-seno. Estas funções trigonométricas denominam-se harmônicos e são frequências mútuas da frequência fundamental (Wilks, 1995). Ainda de acordo com Wilks (1995), uma dada série temporal de dados de N amostras equidistantes de uma variável y , pode ser representada por uma função harmônica, conforme descrito pela Equação 1.

$$y_i = \bar{y} + \sum_{j=1}^{N/2} C_j \cdot \cos(\omega_j t - \phi_j) \quad (1)$$

Onde \bar{y} representa a média dos valores da série de dados, C_j a amplitude, ω_j a frequência, t o tempo e ϕ_j o ângulo de fase. Na prática a análise harmônica permite representar uma função complexa como a somatória de ondas (termos) senoidais e cossenoidais, onde cada onda é definida por um valor único de amplitude e de fase. A amplitude corresponde à metade do valor onde a função é maximizada e a fase (ou ângulo de fase) é o deslocamento entre a origem e o pico da onda no intervalo de 0 até 2π (DAVIS, 1986). Cada termo designa o número de ciclos completados por uma onda num determinado intervalo. O primeiro harmônico ou harmônico fundamental tem um período T (onda anual) igual ao período total em estudo. O segundo harmônico tem um período igual à metade do período do primeiro harmônico $T/2$ (onda semestral) e assim sucessivamente.

A função coseno é geralmente utilizada, pois o ângulo de fase pode ser facilmente determinado como correspondendo ao tempo onde a função harmônica é maximizada, pela Equação 2.

$$t = \frac{\phi_j N}{2\pi} \quad (2)$$

Para estimar a amplitude utiliza-se a Equação 3.

$$C_j = \sqrt{A_j^2 + B_j^2} \quad (3)$$

Onde A_j é B_j são as componentes de coseno e seno, respectivamente, do vetor amplitude (C_j), sendo obtidas pelas Equações 4 e 5.

$$A_j = \frac{2}{N} \sum_{i=1}^N y_i \cdot \cos\left(\frac{2\pi i}{N}\right) \quad (4)$$

e

$$B_j = \frac{2}{N} \sum_{i=1}^N y_i \cdot \sin\left(\frac{2\pi i}{N}\right) \quad (5)$$

Já, o ângulo de fase ϕ_j é dado pela Equação 7.

$$\left\{ \begin{array}{lll} \phi_j = \tan^{-1} \frac{B_j}{A_j} & \text{se} & A_j > 0 \\ \phi_j = \tan^{-1} \frac{B_j}{A_j} \pm \pi \quad \text{ou } 180^\circ & \text{se} & A_j < 0 \\ \frac{\pi}{2} \quad \text{ou } 90^\circ & \text{se} & A_j = 0 \end{array} \right. \quad (6)$$

2.6 Seleção das Imagens

A análise harmônica dos dados foi realizada a partir de uma série de algoritmos implementados em linguagem *IDL* e executados em ambiente *ENVI* (Pardi La cruz, 2006). O arquivo de entrada (meta-arquivo) contém as doze imagens correspondentes ao ano-safra 2006/2007. O resultado final do processamento consistiu em uma imagem termo aditivo, assim como as imagens amplitude, fase e variância para os seis harmônicos. A partir do resultado do processamento foram selecionadas as imagens para a geração da composição RGB. A seleção das imagens está diretamente relacionada com o padrão de resposta da cultura, ou seja, áreas agrícolas, apresentam altos valores de amplitude em comparação com o restante das tipologias de uso e cobertura do solo, e se manifestam até o quarto harmônico (Pardi Lacruz, 2006).

2.7 Mineração de Dados: Árvore de Decisão

A técnica de árvores de decisão foi utilizada para a busca de padrões nos pixels com cana-de-açúcar na cena do LANDSAT. Para a construção da árvore foi usado o algoritmo C4.5 que não depende de suposições sobre a distribuição dos valores das variáveis ou da independência entre si das variáveis. Isto é importante quando se utiliza dados de SIG juntamente com dados de imagem (Araki, 2005). Foi utilizado o software de domínio público, denominado *Waikato Environment for Knowledge Analysis (Weka)*, da Universidade de Waikato, Nova Zelândia (Weka, 2006). O pacote *Weka* consiste de uma coleção de algoritmos de aprendizado de máquina para tarefas de mineração de dados. No trabalho ele foi usado no processo de aprendizado a um conjunto de dados (pixels puros com cana-de-açúcar) para extração de padrões.

3. Resultados e Discussões

A Figura 4 apresenta a distribuição mensal do total de 170 imagens utilizadas na geração das composições mensais MVC-NDVI para a safra 2006/2007.

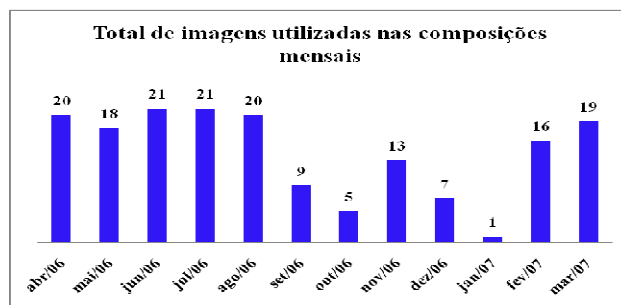


Figura 4. Distribuição mensal do total de imagens utilizadas na geração das composições mensais do NDVI.

Para a aplicação do algoritmo da análise harmônica foi criado um meta-arquivo contendo as 12 composições MVC-NDVI, correspondente a safra 2006/2007. O resultado da aplicação do algoritmo consistiu em uma imagem termo aditivo, assim como as imagens amplitude, fase e variância para os seis harmônicos. As imagens termo aditivo representam a média do NDVI ao longo da safra, as imagens amplitude indicam a variação máxima do NDVI na série, já as imagens variância representam o tempo ao longo da série em que acontece o mínimo e o máximo valor da amplitude e finalmente a variância indica a porcentagem de informação

contida em cada harmônico. A Figura 5 apresenta a resposta dos harmônicos em relação às imagens amplitude e a variância, para a cultura da cana-de-açúcar. É possível observar que a cana-de-açúcar, apresenta um padrão unimodal, ou seja, a variação ocorre basicamente no primeiro harmônico, apresentando pouca variação no segundo harmônico e quase nula nos harmônicos restantes.

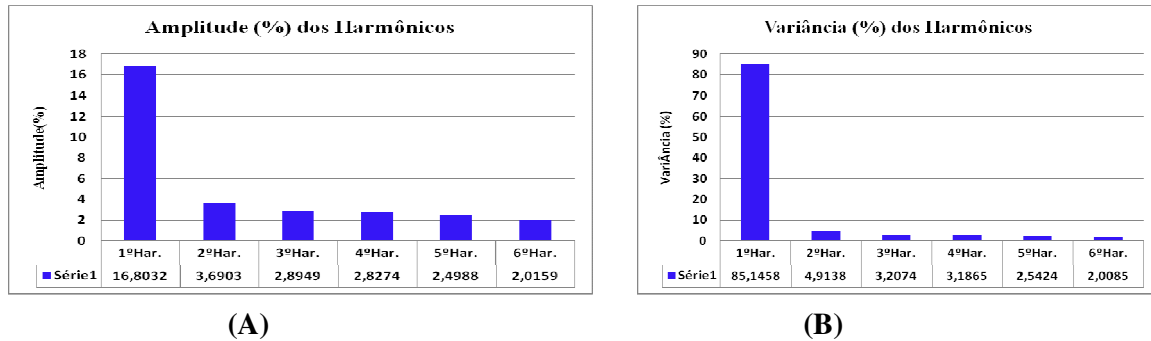


Figura 5. Resposta dos Harmônicos para as imagens Amplitude e Variância.

A Figura 6 apresenta a composição RGB (R: Imagem amplitude 1º Harmônico, G: Imagem termo aditivo e B: Imagem fase) gerada a partir da resposta do primeiro harmônico das imagens amplitude e fase de acordo com os resultados da Figura 5. É possível observar que em comparação com os vetores da Figura 6B, as áreas de cana-de-açúcar apresentam-se em tons magenta na Figura 6A. Para discriminar o comportamento desses pixels foi gerada uma árvore de decisão, onde os parâmetros de entrada foram: Os valores de amplitude, fase e a relação entre a amplitude e o termo aditivo para cada pixel mapeado de cana-de-açúcar. A relação entre a amplitude e o termo aditivo foi proposta para testar a viabilidade da junção dessas duas variáveis.

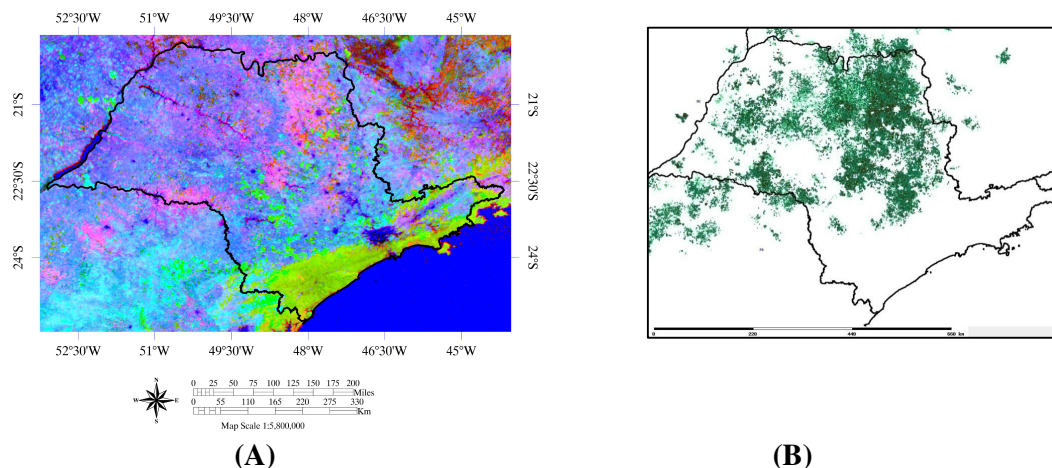


Figura 6. Composição RGB (R: Imagem amplitude 1º Harmônico, G: Imagem termo aditivo e B: Imagem fase 1º Harmônico) da safra 2006/2007 (a) e vetores de cana-de-açúcar encontrados no projeto CANASAT.

Os resultados gerados são apresentados abaixo, é possível notar que a amplitude dos pixels com cana-de-açúcar variam entre 0,0978 e 0,1119 e o termo aditivo 0,56. Para a relação termo aditivo e amplitude os valores foram de 0,223669.

```
Amplitude <= 0.0978: N (642.0/36.0)
Amplitude > 0.0978
| TA <= 0.56
| | Amplitude <= 0.1119
| | | AT_A <= 0.223669: S (125.0/32.0)
| | | AT_A > 0.223669: N (8.0/1.0)
| | Amplitude > 0.1119: S (491.0/29.0)
| TA > 0.56
| | TA <= 0.5705: S (11.0/3.0)
| | TA > 0.5705: N (20.0/5.0)

=== Confusion Matrix ===

  a  b  <-- classified as
206 18 | a = S
 26 191 | b = N
```

Figura 7. Resultado da aplicação do algoritmo de árvore de decisão na imagem composição RGB.

4. Conclusão

Pode-se observar que as imagens geradas da aplicação da análise harmônica na série temporal, apresentam potencial para a identificação das áreas com cana-de-açúcar no estado. A metodologia analisada na discriminação dos padrões, a partir da aplicação de algoritmos de árvore de decisão, mostrou-se eficiente, porém apresenta como limitação a dificuldade na localização espacial dos pixels com tais características.

Referência Bibliográfica

Araki H. **Fusão de Informações Espectrais, Altimétricas e de dados auxiliares na classificação de Imagens de Alta Resolução Espacial**. 2005. 136 p. Tese (Doutorado em Ciências Geodésicas) Universidade Federal do Paraná, Curitiba, 2005.

Davis, J.C. **Statistics and data analysis in geology**. New York: J. Wiley and Sons, 1986. 646p.

Esquerdo, J. C. D. M.; Antunes, J. F. G.; Emery, W.; Baldwin, D. G; Zullo Júnior, J. An Automatic System for AVHRR Land Surface Product Generation. **International Journal of Remote Sensing**, v.27, n.18, p. 3925 - 3942, 2006.

Pardi Lacruz, M.S. Análise harmônica de séries temporais de dados *MODIS* como uma nova técnica para a caracterização da paisagem e análise de lacunas de conservação. 2006. 130 p.

Tese (Doutorado)- INPE, Instituto Nacional de Pesquisas Espaciais, São José dos Campos, São Paulo.

Qi, J.; Chehbouni, A.; Huete, A.R.; Kerr, Y.H.; Sorooshian, S. A modified soil adjusted vegetation index. *Remote Sensing of Enviroment*, v. 48, p. 119-126, 1994.

Weka 3 - Data Mning Software in Java The University of Waikato, disponível em: <http://www.cs.waikato.ac.nz/ml/weka/>> ultimo acesso em: junho,2008.

WILKS, D.S. **Statistical methods in the atmospheric sciences: an introduction**. San Diego, California: Academic Press, 1995. 467 p.