

PONTIFÍCIA UNIVERSIDADE CATÓLICA  
DO RIO DE JANEIRO



**Paula Beatriz Cerqueira Leite**

**Identificação de Tipos de Culturas Agrícolas a  
partir de Seqüências de Imagens Multitemporais  
Utilizando Modelos de Markov Ocultos**

**Dissertação de Mestrado**

Dissertação apresentada como requisito parcial para  
obtenção do grau de Mestre pelo Programa de Pós-  
graduação em Engenharia Elétrica do Departamento de  
Engenharia Elétrica da PUC-Rio.

Orientador: Prof. Raul Queiroz Feitosa  
Co-orientador: Prof. Antônio Roberto Formaggio

Rio de Janeiro  
Agosto de 2008



**Paula Beatriz Cerqueira Leite**

**Identificação de Tipos de Culturas Agrícolas a  
partir de Seqüências de Imagens Multitemporais  
Utilizando Modelos de Markov Ocultos**

Dissertação apresentada como requisito parcial para obtenção do grau de Mestre pelo Programa de Pós-Graduação em Engenharia Elétrica do Departamento de Engenharia Elétrica do Centro Técnico Científico da PUC-Rio. Aprovada pela Comissão Examinadora abaixo assinada.

**Dr. Raul Queiroz Feitosa**  
**Orientador**

Departamento de Engenharia Elétrica - PUC-Rio

**Dr. Antônio Roberto Formaggio**  
INPE

**Dr. Álvaro de Lima Veiga Filho**  
Departamento de Engenharia Elétrica - PUC-Rio

**Dr. Cláudia Maria de Almeida**  
INPE

**Prof. José Eugenio Leal**  
Coordenador Setorial do Centro  
Técnico Científico

Rio de Janeiro, 15 de agosto de 2008

Todos os direitos reservados. É proibida a reprodução total ou parcial do trabalho sem autorização da universidade, da autora e do orientador.

### **Paula Beatriz Cerqueira Leite**

Graduou-se em Engenharia Elétrica com ênfase em sistemas de apoio à decisão na PUC-Rio (Pontifícia Universidade Católica do Rio de Janeiro) em 2006.

#### Ficha Catalográfica

Leite, Paula Beatriz Cerqueira

Identificação de tipos de culturas agrícolas a partir de seqüências de imagens multitemporais utilizando modelos de Markov ocultos / Paula Beatriz Cerqueira Leite ; orientador: Raul Queiroz Feitosa ; co-orientador: Antônio Roberto Formaggio. – 2008.

79 f. ; 30 cm

Dissertação (Mestrado em Engenharia Elétrica)– Pontifícia Universidade Católica do Rio de Janeiro, Rio de Janeiro, 2008.

Inclui bibliografia

1. Engenharia elétrica – Teses. 2. Identificação de culturas agrícolas. 3. Modelos de Markov ocultos. 4. Análise multitemporal. 5. Sensoriamento remoto. 6. Landsat 5 TM. 7. Landsat 7 ETM. I. Feitosa, Raul Queiroz. II. Formaggio, Antônio Roberto. III. Pontifícia Universidade Católica do Rio de Janeiro. IV. Departamento de Engenharia Elétrica. V. Título.

CDD: 621.3

Aos meus pais.

## Agradecimentos

Ao meu orientador Dr. Raul Queiroz Feitosa pela instrução, força e ajuda ao longo deste trabalho.

Ao meu co-orientador Dr. Antônio Roberto Formaggio pelo auxílio constante.

Ao CNPq e à PUC-Rio, pelos auxílios concedidos, sem os quais este trabalho não poderia ter sido realizado.

Aos meus pais e a minha avó Adelina pelo carinho, amor e apoio incondicional em tudo.

Ao meu avô Cerqueira pelos conselhos sábios que guardarei para sempre comigo.

Ao Leo pela compreensão, incentivo e amor em todos os momentos.

Aos meus amigos do LVC pelo companheirismo e incentivo: Dário, Gilson, Marcelo, Vanessa, Ilana, Henrique, Cecília e Edmundo.

As minhas irmãs do ballet pelos momentos de descontração.

A Renata e a Fabiola por estarem sempre comigo, prontas para me ouvir e ajudar.

A minha família por estar, de um jeito ou de outro, presente em mais esta etapa da minha vida.

Aos meus queridos amigos pelo apoio e incentivo nos momentos difíceis.

A Deus.

## Resumo

Leite, Paula Beatriz Leite; Feitosa, Raul Queiroz (Orientador). **Identificação de tipos de culturas agrícolas a partir de seqüências de imagens multitemporais utilizando modelos de Markov ocultos**. Rio de Janeiro, 2008. 79p. Dissertação de Mestrado – Departamento de Engenharia Elétrica, Pontifícia Universidade Católica do Rio de Janeiro.

Esta dissertação propõe uma metodologia baseada em Modelos de Markov Ocultos (*Hidden Markov Models – HMM*) para a classificação de culturas agrícolas, explorando informações de seqüências temporais de imagens dos sensores TM e ETM<sup>+</sup>/Landsat. O método reconhece os diferentes tipos de culturas agrícolas analisando os perfis espectrais em uma seqüência temporal de imagens de satélite de média resolução espacial ( $\cong 30m$ ). Nesta abordagem, o comportamento temporal de cada classe de cultura é modelado por um HMM específico. A classificação é feita segmento-a-segmento, descritos por um vetor de atributos calculado como as médias espectrais dos pixels contidos no segmento em cada banda da imagem. Os vetores de atributos do segmento em cada imagem da seqüência de imagens são subseqüentemente submetidos aos HMMs de cada classe de cultura. O segmento é então associado à cultura cujo HMM correspondente gera a maior probabilidade de emitir a seqüência de valores espectrais observada. Os experimentos para análise foram conduzidos utilizando-se um conjunto de 12 imagens LANDSAT co-registradas e corrigidas radiometricamente. As imagens cobrem uma área do estado de São Paulo, Brasil, com aproximadamente 124.100ha, entre 2002 e 2004. As seguintes coberturas vegetais foram consideradas: cana de açúcar, soja, milho, pastagem e mata-galeria. A avaliação do desempenho do método foi efetuada utilizando-se um conjunto de dados classificado visualmente por dois especialistas e validado por um extenso trabalho de campo. O desempenho do método de classificação multitemporal proposto foi comparado com o de um classificador monotemporal de máxima verossimilhança, e os resultados mostraram a superioridade notável do método baseado em HMM, o qual alcançou uma acurácia média de nada menos que 91% na identificação do tipo correto de cultura agrícola, para seqüências de dados contendo apenas uma única classe de cultura.

## Palavras-chaves

Identificação de culturas agrícolas, modelos de Markov ocultos, análise multitemporal, sensoriamento remoto, Landsat 5 TM, Landsat 7 ETM+.

## Abstract

Leite, Paula Beatriz Leite; Feitosa, Raul Queiroz (Advisor). **Crop type identification based on Hidden Markov Models using multitemporal image sequences**. Rio de Janeiro, 2008. 79p. MSc Dissertation – Departamento de Engenharia Elétrica, Pontifícia Universidade Católica do Rio de Janeiro.

This work proposes a Hidden Markov Model (HMM)-based methodology to classify agricultural crops, exploring information of temporal image sequences from TM and ETM<sup>+</sup>/Landsat sensors. HMMs are used to relate the varying spectral response along the crop cycle with plant phenology for different crop classes. The method recognizes different agricultural crops by analyzing their spectral profiles over a temporal sequence of medium resolution satellite images ( $\cong$  30m). In our approach the temporal behaviour of each crop class is modelled by a specific HMM. A segment-based classification is performed using the average spectral values of the pixels in each image segment across an image sequence, which is subsequently submitted to the HMMs of each crop class. The image segment is assigned to the crop class, whose corresponding HMM delivers the highest probability of emitting the observed sequence of spectral values. Experiments were conducted upon a set of 12 co-registered and radiometrically corrected LANDSAT images. The images cover an area of the State of São Paulo, Brazil with about 124.100ha, between the years 2002 and 2004. The following classes were considered: sugarcane, soybean, corn, pasture and riparian forest. Performance assessment was carried out upon a data set classified visually by two analysts and validated by extensive field work. The performance of the proposed multitemporal classification method was compared to that of a monotemporal maximum likelihood classifier, and the results indicated a remarkable superiority of the HMM-based method, which achieved an average of no less than 91% accuracy in the identification of the correct crop, for sequences of data containing a single crop class.

## Keywords

Crop class identification, hidden Markov models, multitemporal analysis, remote sensing, Landsat 5 TM, Landsat 7 ETM+.

## Sumário

1	Introdução	13
1.1.	Motivação	13
1.2.	Objetivos	15
1.2.1.	Objetivo Geral	15
1.2.2.	Objetivos Específicos	16
1.3.	Organização do restante da dissertação	16
2	Trabalhos anteriores	17
2.1.	Interpretação multitemporal de imagens de sensoriamento remoto	17
2.2.	Estimativas agrícolas a partir de imagens de sensoriamento remoto	19
2.3.	Fenologia	20
2.4.	Aplicações de Modelos de Markov Ocultos	21
3	Modelos de Markov Ocultos	23
3.1.	Processos Estocásticos	23
3.2.	Processos Markovianos	23
3.3.	Cadeias de Markov	24
3.4.	Modelos de Markov Ocultos	25
3.4.1.	Elementos de um Modelo de Markov Oculto	26
3.4.2.	Problemas básicos associados a HMMs	27
3.4.3.	Soluções para os problemas 1 e 2 associados a HMMs	28
3.4.3.1.	Solução do problema 1	28
3.4.3.2.	Solução do problema 2	31
4	Método Proposto	34
4.1.	Descrição geral do modelo	34
4.2.	Ajustando o modelo à aplicação – Estimativa de parâmetros	36
4.2.1.	Estimativa da matriz de transição de estado	38
4.2.2.	Estimativa das probabilidades a priori de estados	39
4.2.3.	Estimativa da emissão de símbolo	39
4.3.	Classificação	41
5	Análise experimental	42
5.1.	Culturas e respectivos ciclos fenológicos	42
5.1.1.	Cana-de-açúcar	42

5.1.2. Culturas de ciclo curto (cereais)	43
5.1.3. Pastagem	46
5.1.4. Outras coberturas vegetais	46
5.2. Área geográfica	46
5.3. Imagens	48
5.4. Pré-processamento das imagens	48
5.5. Dados de referência	49
5.5.1. Segmentação	49
5.5.2. Processo de classificação visual	49
5.6. Experimentos	54
5.6.1. Experimento 1: Conjunto ótimo de atributos espectrais	54
5.6.2. Experimento 2: Desempenho do método para identificar as culturas agrícolas	55
5.6.3. Experimento 3: Desempenho do método para identificar os estágios fenológicos	56
5.6.4. Experimento 4: Avaliação da influência do comprimento das seqüências sendo utilizadas no desempenho do método	58
5.6.5. Experimento 5: Determinação dos estágios fenológicos que melhor caracterizam cada cultura	59
5.6.6. Experimento 6: Comparação do método multitemporal baseado em HMMs com uma abordagem monotemporal de máxima verossimilhança	62
6 Conclusão	64
7 Referências Bibliográficas	66
Apêndice I	70
Apêndice II	71
Apêndice III	78

## Lista de figuras

- Figura 1 - Cadeia de Markov com 3 estados ( $S_1$  a  $S_3$ ), onde as setas indicam as transições possíveis entre os estados. 24
- Figura 2 - Exemplo de um modelo de Markov Oculto ( $S_i \rightarrow$  estados,  $v_k \rightarrow$  símbolos observáveis,  $a_{ij} \rightarrow$  probabilidade de transição de estado,  $b_{jk} \rightarrow$  probabilidade de emissão de símbolo). 27
- Figura 3 - Ilustração das seqüências de operações necessárias para o cálculo da variável direta  $\alpha_{t+1}(j)$ . 29
- Figura 4 - Procedimento de cálculo de  $\alpha_{t+1}(j)$  em função dos instantes das observações  $t$  e dos estados  $S_i$  30
- Figura 5 - Ilustração das seqüências de operações necessárias para o cálculo da variável reversa  $\beta_{t+1}(j)$ . 31
- Figura 6 - HMM usado neste trabalho para cana de açúcar, soja e milho (PP = pré-plantio, CR = crescimento, AD = fase adulta e PA = palhada). 34
- Figura 7 - HMM usado neste trabalho para pastagem e mata-galeria (AD = fase adulta). 35
- Figura 8 - Exemplo de áreas de pastagem e mata-galeria nas imagens de satélite utilizadas. 35
- Figura 9 - Exemplo de áreas de soja, milho e cana de açúcar nas imagens de satélite utilizadas. 36
- Figura 10 - Processo de estimativa dos valores dos parâmetros do HMM de cada cultura. 38
- Figura 11 - Processo de classificação. 41
- Figura 12 - Ciclos fenológicos da cana de açúcar de 12 meses (cana-de-ano) e 18 meses (cana-de-ano-e-meio). Figura obtida em (Gleriani, 2004). 43
- Figura 13 - Ciclo fenológico da cultura de milho. Figura adaptada de (Gleriani, 2004). 44
- Figura 14 - Ciclo fenológico da cultura de soja. Figura adaptada de (Gleriani, 2004). 45
- Figura 15 - Área de estudo, estado de São Paulo, Brasil. Os pontos amarelos correspondem a 316 pontos amostrais de referência. 47

Figura 16 - Recorte de imagem Landsat na composição colorida RGB 453, e foto digital de campo de talhões de milho em diferentes estágios de crescimento. FONTE: (Sanches, 2004). 51

Figura 17 - Recorte de imagem Landsat na composição colorida RGB 453, e foto digital de campo de talhões de cana-de-açúcar em diferentes estágios de crescimento. FONTE: (Sanches, 2004). 52

Figura 18 - Recorte de imagem Landsat na composição colorida RGB 453, e foto digital de campo de talhões de soja em diferentes estágios de crescimento. FONTE: (Sanches, 2004). 53

## Lista de tabelas

Tabela 1 - Imagens disponíveis no conjunto de dados utilizado no presente trabalho.	47
Tabela 2 - Conjunto ótimo de n atributos espectrais e o desempenho da classificação correspondente.	55
Tabela 3 - Taxas de acerto para a classificação de coberturas vegetais.	55
Tabela 4 - Matriz de confusão para a classificação de coberturas vegetais.	56
Tabela 5 - Taxas de acerto para a classificação de estados.	57
Tabela 6 - Matriz de confusão para a classificação de estados.	57
Tabela 7 - Taxas de acerto para seqüências de tamanhos variados.	58
Tabela 8 - Taxas de acerto para seqüências de soja com certos estados (indicados por ✓) obrigatoriamente presentes.	60
Tabela 9 - Taxas de acerto para seqüências de milho com certos estados (indicados por ✓) obrigatoriamente presentes.	61
Tabela 10 - Taxas de acerto para seqüências de cana de açúcar com certos estados (indicados por ✓) obrigatoriamente presentes.	62
Tabela 11 - Desempenho da abordagem monotemporal.	63