

CLASSIFICAÇÃO MULTIESPECTRAL UTILIZANDO O CONCEITO DE MISTURA

Vitor Haertel

Instituto de Pesquisas Hidráulicas
Universidade Federal do Rio Grande do Sul
Caixa Postal 530, 90001 - Porto Alegre, RS, Brasil

Yosio E. Shimabukuro

José Carlos Moreira

Instituto de Pesquisas Espaciais
Ministério da Ciência e Tecnologia
Caixa Postal 515, 12201 - São José dos Campos, SP, Brasil

RESUMO

Em imagens digitais é frequente um pixel conter duas ou mais classes. Esse fato causa uma resposta espectral não representativa e resulta em erros no processo de classificação. Esse trabalho sugere a implementação do conceito de mistura no processo de classificação.

ABSTRACT

In digital imagery it often happens for a pixel to include in it different types of land cover. As a result, the spectral response is not representative of any of the individual components. This mixture is a source of errors in the classification procedure. This paper deals with the implementation of this concept in order to improve the classification accuracy.

1. INTRODUÇÃO

Diferentes sistemas sensores, a bordo de satélites para Sensoriamento Remoto da superfície da Terra são hoje disponíveis, cada sistema apresentando características próprias de resolução espectral e espacial. A resolução espacial está relacionada ao tamanho do pixel, que apresenta variações de um sistema para outro.

O problema da mistura está relacionado ao tamanho do pixel. Quanto maior for o pixel, maior será a probabilidade de duas ou mais classes ocorrerem no mesmo pixel. Neste caso, a energia refletida não é representativa de nenhuma das classes envolvidas, mas da mistura. Assim, no processo de classificação, os pixels que contêm mistura são classificados erroneamente ou não classificados.

Diferentes situações podem contribuir para gerar mistura de classes em um pixel. Uma das mais comuns é o efeito de borda que ocorre quando um pixel cai sobre os limites entre dois ou mais tipos de classes no terreno. A resposta espectral obtida não será representativa de nenhuma das classes individualmente.

Outro tipo de mistura pode ser causada pela variação na densidade de uma

cobertura vegetal uniforme (floresta plantada, por exemplo). Esse fato resulta na mistura da vegetação e outro elemento, em geral solo exposto ou substrato inferior.

2. O PROBLEMA DA MISTURA

O problema da mistura está extensivamente estudado em Shimabukuro (1987). O modelo linear de mistura pode ser expresso da maneira seguinte:

$$r_i = \sum_{j=1}^n (a_{ij} \cdot x_j) + e_i \quad (2.1)$$

onde:

- r_i = reflectância espectral média na banda espectral 'i' para um determinado pixel.
- x_j = fração da área total do pixel ocupada pela componente j.
- a_{ij} = reflectância espectral da componente 'j', medida na banda espectral 'i'.
- e_i = componente de erro na banda espectral 'i'.
- j = 1, 2, ..., n (n=número de componentes no pixel).
- i = 1, 2, ..., m (m=número de bandas espectrais utilizadas).

Normalmente $m > n$. Neste caso, o sistema é resolvido através do processo dos mínimos quadrados, sob as condições seguintes (Shimabukuro, 1987):

- para qualquer pixel, as frações x_j devem ser não negativas;
- em cada pixel, a soma das frações x_j deve ser igual à unidade.

Normalmente, os algoritmos para classificação de imagens digitais, a partir de dados obtidos em CCTs fazem diretamente uso dos contadores digitais. A implementação do modelo linear de mistura exige a conversão do contador digital em reflectância. A solução desse problema é apresentada em Robinove (1982) e resumidamente pode ser descrito por duas equações:

- 1) Obtenção da radiancia a partir dos contadores digitais:

$$L_\lambda = LMIN_\lambda + (LMAX_\lambda - LMIN_\lambda) * (QCAL / QCALMAX) \quad (2.2)$$

onde:

L_λ = radiância espectral

QCAL = contador digital

QCALMAX = intervalo de variação de QCAL

LMIN $_\lambda$ = radiância espectral para QCAL=0

LMAX = radiância espectral para QCAL = QCALMAX

- 2) Obtenção da reflectância a partir da radiância:

$$\rho = \pi \cdot L_\lambda \cdot d^2 / (ESUN_\lambda \cdot \cos \theta_s) \quad (2.3)$$

onde:

ρ = reflectância medida no topo de atmosfera

d = distância sol-terra em unidades astronômicas

ESUN $_\lambda$ = irradiância solar média no topo da atmosfera ($m W \cdot cm^{-2} \cdot \mu m^{-1}$)

θ_s = ângulo zenital do sol.

A utilização conjunta dessas duas equações permite a conversão de contadores digitais em reflectância e vice-versa possibilitando portanto a utilização do modelo (2.1) a partir de dados fornecidos por CCTs (contadores digitais).

3. IMPLEMENTAÇÃO DO CONCEITO DE MISTURA NO PROCESSO DE CLASSIFICAÇÃO

O problema da mistura, constitui-se em um fator que contribui para o decréscimo da exatidão no processo de classificação. Variações de densidade numa cobertura vegetal, que de outra forma seria considerada homogênea, além do problema de bordas, constituem-se em fonte de erros.

O modelo (2.1) pode ser facilmente implementado, utilizando-se processos determinísticos de classificação multi-espectral (clusterização). Como exemplo,

pode ser citado o método K-medias. Esse método encontra-se bastante bem descrito na literatura especializada (Tou e Gonzalez, 1974) e baseia-se na minimização da distância euclidiana de cada ponto (pixel) ao centro do aglomerado (cluster). O processo exige que seja fornecido inicialmente o número de aglomerados (classes espectrais) e valores aproximados para os respectivos centros. Em um processo iterativo, a posição dos centros dos aglomerados é corrigida e os aglomerados formados. Um problema desse algoritmo é a escolha do número de aglomerados (K) para cada conjunto de dados em particular, sendo muitas vezes difícil decidir o número ótimo K. Para os casos gerais, algumas técnicas para estimar K foram sugeridas (Jain e Moreau, 1987). No processo de classificação de cenas naturais, um conhecimento prévio da área pode auxiliar essa estimativa (número de classes presentes). O modelo (2.1) poderá então ser utilizado para simular assinaturas espectrais de misturas de classes, que servirão como centros de aglomerados (clusters).

Um procedimento similar pode ser adotado nos processos estatísticos de classificação multiespectral. Os parâmetros da distribuição para cada classe espectral (vetor médio e matriz variância) podem ser estimados a partir do modelo (2.1).

4. EXPERIMENTOS E CONCLUSÕES

Experimentos iniciais foram realizados, utilizando a área teste Itapevã. Trata-se de uma área de reflorestamento localizada no Estado de Mato Grosso do Sul. Florestas de eucalipto, de idades distintas estão presentes nessa área, apresentando portanto densidades diferentes, que por sua vez resultam em proporções distintas de mistura eucalipto e solo nos pixels. Os canais 3, 4 e 5 de uma cena LANDSAT-TM, obtida em 18 de julho de 1984 são utilizados aqui (figura 1).

Os testes foram realizados utilizando o sistema SITIM-150, instalado no Centro Estadual de Pesquisas em Sensoriamento Remoto e Meteorologia, em Porto Alegre.

Para fins de teste do modelo (2.1), cinco classes foram consideradas: 100% eucalipto, 75% eucalipto e 25% solo, 50% eucalipto e 50% solo, 25% eucalipto e 75% solo, 100% solo. Existem portanto duas classes "puras" (eucalipto e solo) e três classes "mistura", contendo proporções distintas de eucalipto e solo.

A classe "100% eucalipto" corresponde a floresta homogênea e densa, na

qual a cobertura vegetal é completa. A classe "100% solo" corresponde à ausência total de cobertura florestal.

As classes "mistura" correspondem à diferentes densidades da floresta de eucalipto, causada entre outros possíveis fatores, pela diferença de idade da floresta.

Amostras de 100 pixels serviram como base para a definição do centro de cada aglomerado (classe), para o processo K-medias.

Nas duas classes "puras" os 100 pixels corresponderam à áreas homogêneas de eucalipto e solo respectivamente. Os centros das classes "mistura" foram obtidos, tomando-se proporções correspondentes de pixels "puros" de eucalipto e solo.

Essa abordagem é equivalente à implementação do modelo (2.1) ao algoritmo de classificação K-medias.

O resultado desse experimento está ilustrado na figura 2, e são coerentes com as informações disponíveis dessa área.

Ainda dentro dessa abordagem, uma outra componente, a sombra, é certamente de importância (Shimabukuro, 1987) e sua inclusão deverá contribuir para um acréscimo na exatidão do processo de classificação proposto.

5. REFERÊNCIAS BIBLIOGRÁFICAS

- JAIN, A.K. e MOREAU, J.V. Bootstraap Technique in Cluster Analysis, Pattern Recognition, 20(5): -47,567, 1987.
- ROBINOVE, C.J. Computation with Physical Values from Landsat Digital Data, Photogrammetric Engineering and Remote Sensing, 48(5): 781-784.
- SHIMABUKURO, Y.E. Shade Images Derived from Linear Mixing Models of Multi-spectral Measurements of Florested Areas, Dissertação de Doutorado, Colorado State University, Fort Collins Colorado, U.S.A., 1987.
- TOU, J.T. e GONZALEZ, R.C. Pattern Recognition Principles. Addison-Wesley, 1974.

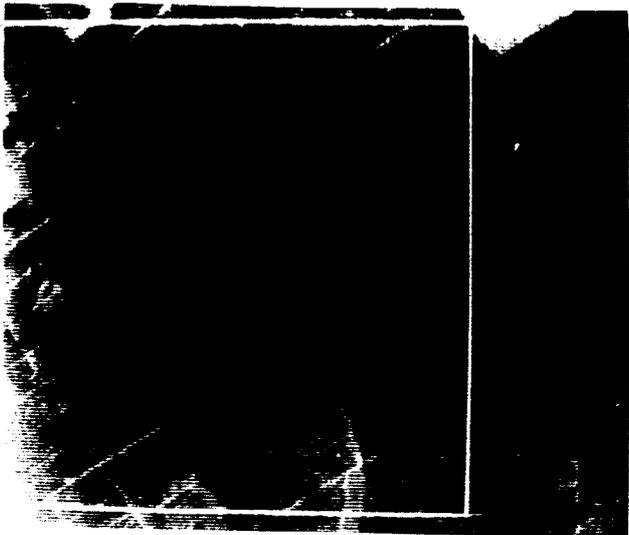


Fig. 1 - Área Teste

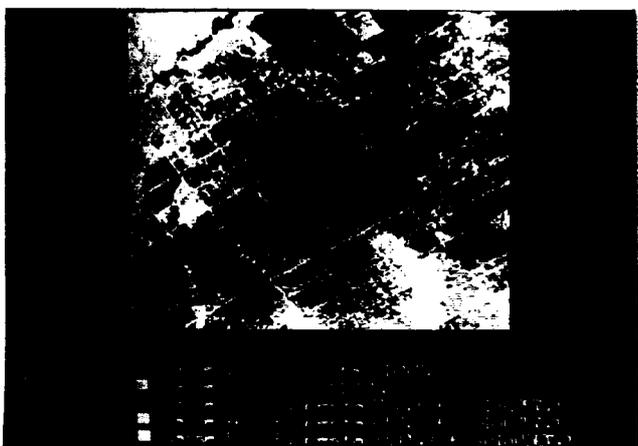


Fig. 2 - Resultado