

DETERMINAÇÃO DE MEMBROS DE REFERÊNCIA EM DADOS HIPERESPECTRAIS PELO MÉTODO DA MÁXIMA ENTROPIA

Endmembers selection in hyperspectral images by maximum entropy method

Marco Antonio Pizarro^{1,2}
David Fernandes¹

¹ Instituto Tecnológico de Aeronáutica - ITA
Comando-Geral de Tecnologia Aeroespacial - CTA
Praça Marechal Eduardo Gomes, 50
12228-900 - São José dos Campos - SP, Brasil
{pizarro, david}@ita.br

² Instituto Nacional de Pesquisas Espaciais - INPE
Caixa Postal 515 - 12245-970 - São José dos Campos - SP, Brasil
pizarro@dea.inpe.br

RESUMO

Neste trabalho é proposta uma nova metodologia supervisionada de busca de Membros de Referência (MR) em imagens hiperespectrais. O método proposto baseia-se no conceito da Entropia máxima calculada sobre os autovalores normalizados da matriz de covariância de um conjunto de espectros candidatos a MR. Estabelecesse ainda dois limitantes para o número máximo de MR, um deles baseia-se na configuração relativa dos espectros pertencentes a um dado conjunto e o outro em um limitante inferior para o valor da Entropia do conjunto de espectros. Foi avaliado, com sucesso, através do modelo linear de mistura espectral (MLME), os MR selecionados pelo método utilizando-se imagens hiperespectrais obtidas pelo sensor AVIRIS (Airborne Visible/InfraRed Imaging Spectrometer).

Palavras-chave: sensoriamento remoto, análise de dados hiperespectrais, imagens hiperespectrais, membros de referência, similaridade entre espectros, entropia.

ABSTRACT

This paper proposes a new method to find the Endmembers in hyperspectral images. The methodology is based on the maximum entropy. The entropy is calculated through the normalized eigenvalues obtained by the covariance matrix of a set of spectra. Two bounds for the number of Endmembers is also given, one of them is based on the internal configuration of a set of spectra and the other in a minimum entropy value. The Linear Model of Spectral Mixture was applied successfully in AVIRIS (Airborne Visible/InfraRed Imaging Spectrometer) images to test the selection of Endmembers.

Keywords: remote sensing, hyperspectral data analysis, hyperspectral images, endmember, spectral similarity, entropy.

1. INTRODUÇÃO

A alta resolução espectral dos sensores imageadores hiperespectrais tornou possível a identificação mais precisa dos espectros associados aos diversos materiais presentes na cena imageada, propiciando assim o desenvolvimento de inúmeras aplicações em sensoriamento remoto.

A Fig. 1 ilustra a formação do espectro de três *pixels* na imagem através do conjunto de imagens hiperespectrais. Para o sensor aerotransportado AVIRIS, utilizado neste trabalho, há 224 imagens

correspondentes a faixa espectral óptica de $383\text{nm} \leq \lambda \leq 2.510\text{nm}$, com cada imagem tendo uma faixa aproximada de 10nm.

Existe o interesse, para fins de classificação, que sejam identificados os espectros de locais na cena que apresentam um único material dominante. O conjunto destes espectros, relativamente puros ou menos contaminados, que podem representar a variedade espectral existente na cena imageada são denominados Membros de Referência (MR) (SCHOWENGERDT, 1997).

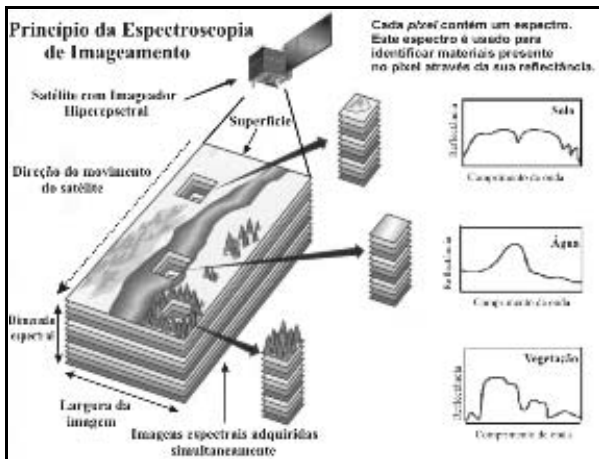


Figura 1 - Espectroscopia de imageamento.
Fonte: adaptação de Shaw e Monolakis (2002, p. 12).

Um dos possíveis empregos dos MR é a sua utilização em modelos de mistura espectral, sendo o mais utilizado o Modelo Linear de Mistura Espectral (MLME). Com o MLME determina-se a abundância dos MR em cada *pixel* da cena, *pixel* esse representado por um espectro, obtendo-se assim uma imagem classificada segundo a preponderância (abundância) dos MR (BOARDMAN; KRUSE e GREEN, 1995).

Para a seleção dos MR, aqueles que explicam melhor a variabilidade espectral da cena, ou para uma mera comparação de similaridade entre espectros podem ser utilizados vários procedimentos (PIZARRO e FERNANDES, 2003 e 2005; CHANG, 2003, p. 15; SWEET et al., 2000; FARAKLIOTI e PETROU, 2000; SCHOWENGERDT, 1997).

Neste trabalho é proposta a utilização da Entropia como uma nova medida de similaridade entre um conjunto qualquer de espectros. A diferença entre a Entropia e as demais medidas de similaridade espectral, tais como a Distância Euclidiana (ROBILA e GERSHMAN, 2005), Coerência Espectral (PIZARRO e FERNANDES, 2005) e Divergência (PIZARRO e FERNANDES, 2003), é que estas últimas são aplicáveis somente a pares de espectros, enquanto a Entropia pode ser aplicada a um conjunto qualquer de espectros, incluindo pares de espectros. O método proposto também pode ser utilizado como um indicador do limitante superior do número de MR presentes em um conjunto de espectros.

2. MEDIDA DE SIMILARIDADE ESPECTRAL PELA ENTROPIA

Seja o vetor espectro N -dimensional $\vec{I}(i_k, j_k)$ relativo ao *pixel* (i_k, j_k) , associado a amostra k candidata a MR de um conjunto de K amostras:

$$\vec{I}(i_k, j_k) = [I(i_k, j_k, 1) \ I(i_k, j_k, 2) \ \dots \ I(i_k, j_k, N)]^T \quad k=1, 2, \dots, K \quad (1)$$

onde T denota o operador transposto.

Define-se o vetor espectro médio $\bar{I}_m(i_k, j_k)$ para cada amostra k como a média dos vetores espectros $\vec{I}(p, q)$ contidos em uma vizinhança $(M+1) \times (M+1)$, com M par, do *pixel* (i_k, j_k) :

$$\bar{I}_m(i_k, j_k) = \frac{1}{(M+1)^2} \sum_{p=i_k-M/2}^{i_k+M/2} \sum_{q=j_k-M/2}^{j_k+M/2} \vec{I}(p, q) \quad (2)$$

Em algumas aplicações é conveniente trabalhar com a média da derivada dos espectros presentes na vizinhança da amostra k . Este procedimento é utilizado com o intuito de se enfatizar as transições nos espectros provocadas pelas absorções oriundas dos materiais presentes na superfície, eliminar o nível médio nos espectros originais das imagens e minimizar os efeitos de fundo na composição do sinal, devido ao solo, quando se analisa áreas com vegetação (TSAI e PHILPOT, 2002; DEMETRIADES-SHAH et al., 1990). Por motivos de simplicidade de notação será utilizada a mesma notação $\bar{I}_m(i_k, j_k)$ para representar a média dos espectros, com ou sem derivada desses espectros, na vizinhança $(M+1) \times (M+1)$ da amostra k .

Consideremos o espectro

$$\vec{\tilde{I}}_m(i_k, j_k) = [\tilde{I}_m(i_k, j_k, 1) \ \tilde{I}_m(i_k, j_k, 2) \ \dots \ \tilde{I}_m(i_k, j_k, N)]^T \quad (3)$$

denominado espectro médio normalizado, com suas componentes descritas por:

$$\tilde{I}_m(i_k, j_k, n) = \frac{I_m(i_k, j_k, n) - \bar{I}_m(i_k, j_k)}{\sqrt{\sum_{u=1}^N [I_m(i_k, j_k, u) - \bar{I}_m(i_k, j_k)]^2}} \quad k=1, 2, \dots, K \quad (4)$$

com a média do espectro $\bar{I}_m(i_k, j_k)$ dada por

$$\bar{I}_m(i_k, j_k) = \frac{1}{N} \sum_{u=1}^N I_m(i_k, j_k, u) \quad e$$

$$\vec{I}_m(i_k, j_k) = [I_m(i_k, j_k, 1) \ I_m(i_k, j_k, 2) \ \dots \ I_m(i_k, j_k, N)]^T$$

A matriz covariância relativa a um conjunto qualquer de $R \leq K$ espectros médios normalizados, diferentes entre si, identificados por $e_1, e_2, \dots, e_R \in \{1, 2, \dots, K\}$ é definida como:

$$L_R = \begin{bmatrix} L_{e_1, e_1} & L_{e_1, e_2} & \dots & L_{e_1, e_R} \\ L_{e_2, e_1} & L_{e_2, e_2} & \dots & L_{e_2, e_R} \\ \dots & \dots & \dots & \dots \\ L_{e_R, e_1} & L_{e_R, e_2} & \dots & L_{e_R, e_R} \end{bmatrix}_{R \times R} \quad (5)$$

onde $L_{p,q} = L_{q,p}$ e

$$L_{p,q} = \frac{1}{N} \sum_{n=1}^N [\tilde{I}_m(i_p, j_p, n) \tilde{I}_m(i_q, j_q, n)] \quad p, q \in \{e_1, e_2, \dots, e_R\} \quad (6)$$

Por construção L_R é uma matriz simétrica, ou seja $L_R = (L_R)^T$, positiva definida e portanto, os seus autovalores calculados através da equação característica $|L_R - \lambda I| = 0$, sendo I a matriz identidade, são números reais não negativos, (STRANG e BORRE, 1997). Sejam estes autovalores $\lambda_r \geq 0$, $r = 1, 2, \dots, R$ dispostos em ordem decrescente, tal que: $\lambda_r \leq \lambda_{r-1}$, $r = 2, \dots, R$.

Pode-se, ainda, calcular os R autovetores \vec{V}_r , $r = 1, 2, \dots, R$, de dimensão $R \times 1$, associados a cada autovalor λ_r pela expressão $L_R \vec{V}_r = \lambda_r \vec{V}_r$. Define-se então a matriz V contendo na sua r -ésima coluna o autovetor \vec{V}_r , $r = 1, 2, \dots, R$. Sendo L_R uma matriz simétrica positiva definida os autovetores \vec{V}_r , correspondentes a diferentes autovalores, são ortogonais entre si, ou seja $V^T V = I$ e em conseqüência $V^{-1} = V^T$ (STRANG e BORRE, 1997).

Para a situação em que um ou mais autovalores são nulos os autovetores correspondentes a eles estão no Espaço Nulo da matriz L_R . Isso implica que o Posto da Matriz L_R (número máximo de linhas/colunas independentes) é menor que R (BIRKHOFF e MACLANE, 1977).

Uma relação conveniente que pode avaliar a separabilidade dos R MR é a Entropia definida em função dos autovalores normalizados:

$$H(e_1, e_2, \dots, e_R) = - \sum_{r=1}^R p_r \log_R p_r \quad (7)$$

onde p_r é o autovalor λ_r normalizado:

$$p_r = \frac{\lambda_r}{\sum_{u=1}^R \lambda_u}, \quad r = 1, 2, \dots, R \quad (8)$$

Devido a normalização tem-se que $\sum_{r=1}^R p_r = 1$.

Como $\lambda_r \geq 0$ então $p_r \geq 0$ e em virtude da ordenação decrescente dos autovalores tem-se que $p_r \leq p_{r-1}$, $r = 2, \dots, R$. A entropia $H(e_1, e_2, \dots, e_R)$ é tal que: $0 \leq H(e_1, e_2, \dots, e_R) \leq 1$. Em particular se $H(e_1, e_2, \dots, e_R) = 1$ então todos autovalores são iguais, ou seja, $p_r = 1/R$, $r = 1, 2, \dots, R$, indicando que os R espectros que geraram a matriz L_R estão bem alocados no espaço de dimensão R . Se $H(e_1, e_2, \dots, e_R) = 0$ então um dos autovalores normalizados é unitário e os demais são nulos, indicando que os espectros utilizados para a

geração da matriz L_R são todos iguais, ou podem ser alocados em uma única dimensão.

Para rejeitar esta má configuração dos espectros candidatos a MR faz-se um teste para evitar que dois espectros médios quaisquer, do conjunto de R espectros, sejam muito similares. Define-se, segundo o critério da Entropia, que um conjunto de R amostras $e_1, e_2, \dots, e_R \in \{1, 2, \dots, K\}$ é bem configurada se:

$$H(p, q) \geq \eta_H \quad \text{para } \forall p, q \in \{e_1, e_2, \dots, e_R\} \quad (9)$$

onde η_H é o limiar de configuração para a Entropia.

Ainda, para testar a boa configuração de R espectros candidatos a MR faz-se dois testes adicionais. Um deles baseado na Distância Euclidiana (DE) entre dois vetores espectros e outro baseado na Coerência Espectral (CE), ou coeficiente de correlação normalizado, entre dois vetores espectros.

Para o mesmo conjunto de R amostras estabelecido anteriormente defini-se, segundo o critério da DE , que este conjunto é bem configurado se:

$$DE(p, q) \geq \eta_{DE} \quad \text{para } \forall p, q \in \{e_1, e_2, \dots, e_R\} \quad (10)$$

onde η_{DE} é o limiar de configuração para a DE e

$$DE(p, q) = \sqrt{\sum_{n=1}^N [I_m(i_p, j_p, n) - I_m(i_q, j_q, n)]^2} \quad (11)$$

onde $DE(p, q) \geq 0$ e $DE(p, p) = 0$.

Um exemplo de amostras mal configuradas, segundo o critério da DE , são duas amostras muito próximas entre si, mas que estão distantes de todas as demais.

Ainda para o mesmo conjunto de R amostras defini-se, segundo o critério da CE , que este conjunto é bem configurado se:

$$CE(p, q) \leq \eta_{CE} \quad \text{para } \forall p, q \in \{e_1, e_2, \dots, e_R\} \quad (12)$$

onde η_{CE} é o limiar de configuração para a CE e

$$CE(p, q) = \frac{\left| \sum_{n=1}^N [\tilde{I}_m(i_p, j_p, n) \tilde{I}_m(i_q, j_q, n)] \right|}{\sqrt{\sum_{n=1}^N (\tilde{I}_m(i_p, j_p, n))^2 \sum_{n=1}^N (\tilde{I}_m(i_q, j_q, n))^2}} \quad (13)$$

onde

$$\tilde{I}_m(i_p, j_p, n) = I_m(i_p, j_p, n) - \bar{I}_m(i_p, j_p)$$

$$\tilde{I}_m(i_q, j_q, n) = I_m(i_q, j_q, n) - \bar{I}_m(i_q, j_q)$$

sendo $0 \leq |CE(p, q)| \leq 1$ e $CE(p, p) = 1$.

Um exemplo de amostras mal configuradas, segundo o critério da CE , são duas amostras muito coerentes entre si, mas pouco coerentes com todas as demais.

Define-se o conjunto U_R de amostras espectrais bem configuradas como aquele formado por todos os subconjuntos de R amostras (e_1, e_2, \dots, e_R) que satisfazem pelo menos um dos três critérios de configuração, dados por (9), (10) e (12):

$$U_R = \left\{ \begin{array}{l} (e_1, e_2, \dots, e_R) \text{ com } e_1, e_2, \dots, e_R \in \{1, 2, \dots, K\} \\ \text{tal que, para } \forall p, q \in \{e_1, e_2, \dots, e_R\} \text{ tem-se} \\ H(p, q) \geq \eta_H \vee DE(p, q) \geq \eta_{DE} \vee CE(p, q) \leq \eta_{CE} \end{array} \right\} \quad (14)$$

Estabelece-se então a função entropia modificada como:

$$\Delta_{H_R}(e_1, e_2, \dots, e_R) = \begin{cases} H(e_1, e_2, \dots, e_R), & \text{se } (e_1, e_2, \dots, e_R) \in U_R \\ 0, & \text{se } (e_1, e_2, \dots, e_R) \notin U_R \end{cases} \quad (15)$$

O critério de busca dos MR formado pelo conjunto de R espectros médios bem configurados de um conjunto de K espectros médios fica então definido por:

$$\left(e_{h_1}, e_{h_2}, \dots, e_{h_R} \right) = \arg \left[\max_{(e_1, e_2, \dots, e_R)} \left(\Delta_{H_R}(e_1, e_2, \dots, e_R) \neq 0 \right) \right] \quad (16)$$

Segundo este critério os espectros médios escolhidos como MR são dados por: $\bar{I}_m(i_{e_{h_r}}, j_{e_{h_r}})$ $r = 1, 2, \dots, R$.

3. ESTIMATIVA DO NÚMERO DE MR

Para um dado conjunto de K espectros médios normalizados $\bar{I}_m(i_k, j_k)$ $k = 1, 2, \dots, K$, qual é o número ideal R de MR?

Há na literatura algumas tentativas de solução deste difícil problema (GROSS e SCHOTT, 1996; PIZARRO e FERNANDES, 2005; BIOUCAS-DIAS, 2005; CHANG e DU, 2004). Neste trabalho, propõem-se dois limitantes superiores para o número de MR presentes na cena. Um deles baseia-se no conceito de amostras bem configuradas e depende dos limiares η_H , η_{DE} e η_{CE} . O outro baseia-se em um limitante inferior para a Entropia calculada para o conjunto de R espectros escolhidos como MR.

O cálculo do primeiro limitante considera para o número de MR igual a $2, 3, \dots, \mathfrak{R}_1 \leq K$ as seguintes escolhas possíveis, bem configuradas, de MR:

$$(e_1^2, e_2^2), (e_1^3, e_2^3, e_3^3), \dots, (e_1^{\mathfrak{R}_1}, e_2^{\mathfrak{R}_1}, \dots, e_{\mathfrak{R}_1}^{\mathfrak{R}_1}),$$

com $e_1^r, e_2^r, \dots, e_r^r \in \{1, 2, \dots, K\}$, $r = 2, \dots, \mathfrak{R}_1$

e que não foi possível achar um conjunto de $\mathfrak{R}_1 + 1$ MR que fossem bem configuradas.

Com estas condições estabelece-se que o limitante superior do número de MR é o valor \mathfrak{R}_1 , ou seja, $R \leq \mathfrak{R}_1$.

O cálculo do segundo limitante estabelece um valor mínimo, H_{min} , para o valor da Entropia do conjunto de MR, bem configurado, escolhido. Consideremos que para o número de MR igual a $2, 3, \dots, \mathfrak{R}_2 \leq K$ tem-se as seguintes escolhas de MR:

$$(e_1^2, e_2^2), (e_1^3, e_2^3, e_3^3), \dots, (e_1^{\mathfrak{R}_2}, e_2^{\mathfrak{R}_2}, \dots, e_{\mathfrak{R}_2}^{\mathfrak{R}_2}),$$

com $e_1^r, e_2^r, \dots, e_r^r \in \{1, 2, \dots, K\}$, $r = 2, \dots, \mathfrak{R}_2$,

tal que $\Delta_{H_R}(e_1, e_2, \dots, e_{\mathfrak{R}_2}) \geq H_{min}$ e

$$\Delta_{H_R}(e_1, e_2, \dots, e_{\mathfrak{R}_2+1}) < H_{min}.$$

Para estas condições estabelece-se que o limitante superior do número de MR é o valor \mathfrak{R}_2 , ou seja $R \leq \mathfrak{R}_2$.

Levando-se que o conjunto de amostras deve ser bem configurada tem-se:

$$R \leq \mathfrak{R}_2 \leq \mathfrak{R}_1 \quad (17)$$

4. TESTE DA METODOLOGIA DE ESCOLHA DOS MR

Para o teste da metodologia proposta foram utilizados espectros conhecidos, extraídos de dados hiperspectrais AVIRIS de uma área de estudo situada próximo à cidade de Campo Grande, MS, nas coordenadas $19^\circ 55' S$ e $53^\circ 30' W$. A área é dominada pela presença Neossolo Quartzarênico (RQ) e Latossolo Vermelho (LV), que representam uma transição química de baixo para altos teores de ferro, óxidos de titânio e fração argila (NOVAES et al., 1983). Também é caracterizada por atividades agrícolas como pasto e alguns talhões de colheitas sazonais, por exemplo a soja. A cobertura de vegetação natural está composta de pequenos bosques de cerrado (PIZARRO, 1999).

Os dados AVIRIS utilizados foram adquiridos em 20 de agosto de 1995, praticamente no ápice da estação regional seca, em 224 faixas estreitas (≈ 10 nm de largura), posicionadas na faixa espectral óptica de $383 \text{ nm} \leq \lambda \leq 2.510 \text{ nm}$ e com resolução espacial nominal de 20×20 m. A subcena 4 selecionada para o teste é formada por uma área retangular de 12 por 10 km (614×512 pixels), aproximadamente. Os dados, originalmente fornecidos em valores de radiância, foram convertidos para valores de reflectância de superfície, por um método de transferência radiativa baseado no MODTRAN (GREEN, 1991).

Do conjunto de imagens excluem-se as bandas de 1 a 8 (383 a 451 nm) por ausência de dados e 218 a 224 (2.450 a 2.510 nm) por estarem extremamente ruidosas. Nas bandas 78 a 85 (1.095 a 1.162 nm), 95 a 123 (1.258 a 1.512 nm) e 148 a 175 (1.761 a 2.021 nm) o MODTRAN introduziu correções excessivas, transformando as bandas de absorção causadas pelas respectivas influências do vapor d'água e do CO₂ em picos de reflectância. Os valores destas bandas foram suprimidos e novos valores foram estimados utilizando-se a função *spline* cúbica.

Na Fig. 2 mostra-se uma das imagens, banda 192 (2.271 nm) da subcena 4 obtida pelo sensor AVIRIS. Nessa Figura salienta-se a localização, aproximada, dos *pixels* de 14 amostras espectrais ($K=14$) escolhidas como candidatas a MR e que poderão representar classes e/ou subclasses de alvos. Cinco amostras foram escolhidas por um especialista com o conhecimento da área e utilizando o Método Manual para Seleção de MR, conhecido como MEST (Manual Endmember Selection Tool) (BATESON e CURTIS, 1996). Essas amostras representam as classes: vegetação verde (*vv*), vegetação com pouca atividade fotossintética (*vs*), Latossolo Vermelho (*LV*), Neossolo Quartzarênico (*RQ*) e água (*a*). Alvos contendo sombra não foram encontrados na cena. Adicionalmente a estas cinco amostras foram escolhidas pela amostragem AESNA - Amostragem Estratificada Sistemática não Alinhada - (CONGALTON, 1988) mais nove amostras espectrais na cena que foram rotuladas como: *amo1*, *amo2*, ..., *amo9*. A amostragem AESNA consiste em escolher, aleatoriamente, amostras ao longo da imagem de modo a ter-se uma boa dispersão destas amostras, evitando assim a concentração de amostras em alguma região particular da imagem. A Fig. 3a mostra os vetores espectros médios $\bar{I}_m(i_k, j_k)$, calculados por (2), para cada classe de alvo escolhido pelo especialista, considerando-se no cálculo do espectro médio uma vizinhança de 5×5 *pixels*. A Fig. 3b mostra alguns outros espectros relevantes para a análise de resultados que será realizada.

Salienta-se que as cinco classes (*vv*, *vs*, *a*, *LV*, *RQ*) foram consideradas pelo especialista como bem representativas da variabilidade espectral na subcena 4 e, assim, os seus espectros configuram uma boa escolha de MR.

Utilizando-se a média, da derivada dos espectros originais, conforme descrito na expressão (2) em uma vizinhança $(M+1) \times (M+1) = 5 \times 5$ *pixels* da amostra que será testada, obteve-se com o cálculo da expressão (16) os MR listados na Tabela 1. Nessa Tabela foi suprimido o prefixo *amo* das amostras da AESNA.

Na Tabela 1 nota-se que a medida que R aumenta as amostras escolhidas anteriormente como MR vão se mantendo. Para $R=9$ todas as combinações das $K=14$ amostras, agrupadas em conjuntos de 9 amostras, foram consideradas mal configuradas e, deste modo, não houve escolha de nenhum MR. Logo, o limitante superior para o número de MR da cena é

$R_1 = 8$. Para $R=5$ obteve-se como resposta todos os espectros considerados bons MR pelo especialista. Para ($2 \leq R \leq 4$) também foram escolhidas amostras pertencentes ao grupo de amostras consideradas boas pelo especialista.

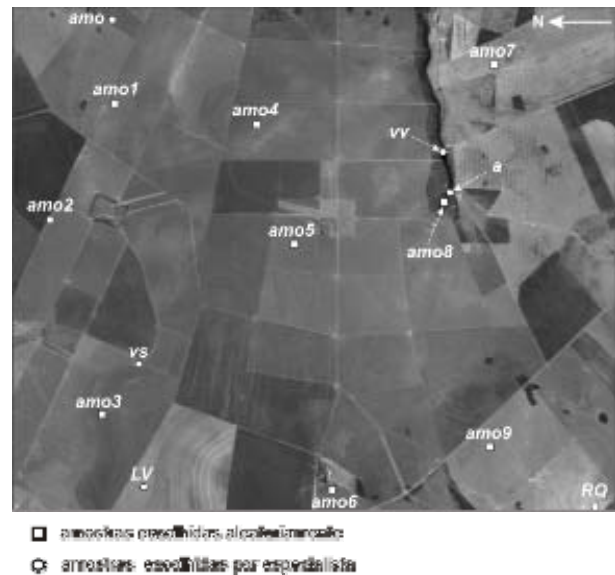
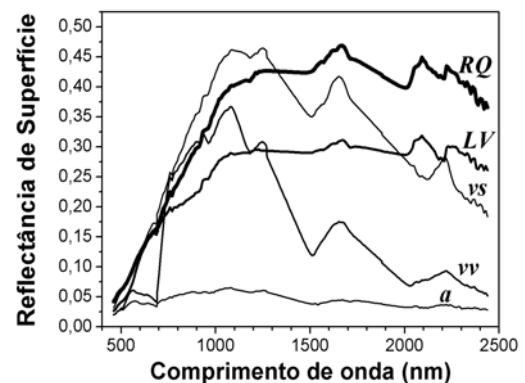
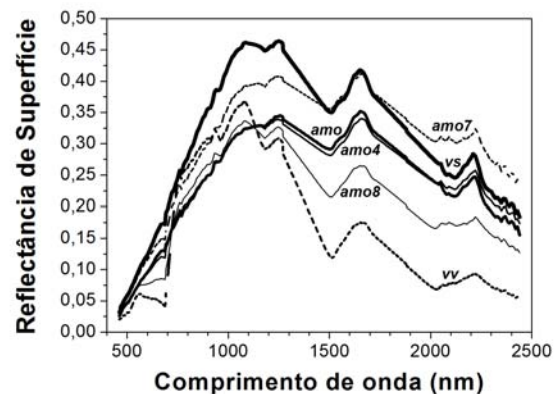


Figura 2 - Imagem da subcena 4, banda 192 (2.271 nm), com as amostras espectrais utilizadas nos testes.



3a - Espectros das amostras (*vv*, *vs*, *a*, *LV*, *RQ*) escolhidas por um especialista.



3b - Espectros das amostras (*amo*, *amo4*, *amo7* e *amo8*) em conjunto com *vv* e *vs*.

Figura 3 - Espectros médios das amostras na subcena 4.

TABELA 1 - MR SELECIONADOS NA SUBCENA 4 COM LIMIARES DE CONFIGURAÇÃO

$$\eta_H = 0,052, \eta_{DE} = 547,87 \text{ E } \eta_{CE} = 0,98.$$

MR (R):	Amostras selecionadas como MR:	Entropia	Imagem-erro	
			média	desvio padrão
2	<i>a, LV</i>	0,91	33,75	8,86
3	<i>a, LV, vv,</i>	0,80	22,33	5,77
4		0,71	6,93	1,98
5	<i>a, LV, vv, vs, RQ</i>	0,63	5,13	1,47
6	<i>a, LV, vv, vs, RQ, 4</i>	0,54	3,12	0,90
7	<i>a, LV, vv, vs, RQ, 4, 8</i>	0,47	2,93	0,81
8	<i>a, LV, vv, vs, RQ, 4, 8, 7</i>	0,42	2,74	0,75
9	Nenhuma amostra foi selecionada $\Rightarrow \mathfrak{R}_1 = 8$			

Para $R=6$ foi escolhida como MR a amostra rotulada como *amo4*. Analisando-se, na Fig. 2b, o espectro médio dessa amostra conclui-se que ela pode ser considerada uma outra espécie de vegetação seca, distinta de *vs*. Para $R = 7$, além de *amo4*, foi escolhida *amo8*. Por inspeção na Fig. 2b o espectro da *amo8* pode ser considerada um tipo de vegetação verde distinto de *vv*. Para $R = 8$, além da *amo4* e *amo8*, foi selecionada a amostra *amo7*, que pode ser também considerada um outro tipo de vegetação seca distinto de *vs* e de *amo4*.

Quanto à escolha final do número de MR, a ser utilizado em um processo de classificação, dependerá da finalidade da classificação. Se houver interesse na classificação de estágios intermediários entre vegetação verde e seca então $R = 6, 7$ ou 8 podem ser recomendados.

Os limiares $\eta_H = 0,052$ e $\eta_{DE} = 547,87$ foram escolhidos, respectivamente, como sendo o valor de Entropia e *DE* correspondentes ao quartil inferior do conjunto formado pelas medidas dos conjuntos:

$$\{H(p, q) \mid p, q \in \{1, 2, \dots, K\}, p \neq q\} \text{ e}$$

$$\{DE(p, q) \mid p, q \in \{1, 2, \dots, K\}, p \neq q\},$$

Quanto ao limiar $\eta_{CE} = 0,98$ foi escolhido como sendo o valor de *CE* correspondentes ao quartil superior do conjunto:

$$\{CE(p, q) \mid p, q \in \{1, 2, \dots, K\}, p \neq q\}.$$

A Tabela 1 mostra também os valores da Entropia, da média e do desvio padrão da imagem-erro, calculada pelo MLME, utilizando-se os MR selecionados. Se for estabelecido que o valor mínimo da Entropia $H_{min} = 0,50$, então o limitante superior para

o número de MR é $\mathfrak{R}_2 = 6$. Nota-se que para $R = 4$ a seleção de *vs* como MR fez o erro diminuir consideravelmente, isto porque *vs* é uma classe com alta presença espacial na cena. As Figs. 6a e 6b mostram as imagens pós-classificadas da subcena 4, para $R = 5$ e 6 . Na imagem pós-classificada são atribuídos aos *pixels* da imagem as classes dos MR que tem mais de 50% na composição espectral do *pixel* segundo o MLME.

Na Fig. 6a percebe-se a predominância da classe vegetação seca (correspondente a amostra *vs*). Como *amo4* é um tipo diferente de vegetação seca observa-se na Fig. 6b a subdivisão da classe vegetação seca em duas subclasses, uma correspondente a amostra *vs* e a outra a amostra *amo4*.

Para complementar o teste realizado escolheu-se ainda um novo conjunto de 47 amostras pela AESNA, além das amostras (*vv, vs, a, LV, RQ*), selecionadas pelo especialista, totalizando assim um conjunto de $K = 52$ amostras candidatas a MR. Utilizando-se a expressão (16) determinou-se para $R = 5$ o seguinte conjunto de MR (*a, LV, vv, amo, RQ*). Analisando-se na Fig. 2b o espectro da amostra rotulada por *amo* verifica-se que ela representa um tipo de vegetação seca, muito semelhante a *amo4*. Portanto, com um conjunto maior de amostras candidatas o método escolheu *amo* no lugar de *vs*.

Aplicando-se o MLME para o conjunto (*a, LV, vv, vs, RQ*) e (*a, LV, vv, amo, RQ*) gerou-se as imagens-erro para os dois conjuntos de amostras escolhidos como MR. Para o primeiro conjunto obteve-se um erro com média de 5,13 e desvio padrão de 1,47. Para o segundo conjunto de MR, no qual esta presente *amo* no lugar de *vs*, obteve-se o erro com média de 3,78 e desvio padrão de 1,52. Essas medidas de erro mostram que para a subcena 4 da imagem AVIRIS a escolha da amostra *amo* para MR foi mais acertada que a amostra *vs*, escolhida a priori pelo especialista, tendo-se como critério a média da imagem-erro.

5. AVALIAÇÃO DOS MR SELECIONADOS

O método proposto foi utilizado para a determinação dos MR de uma outra área, subcena 5, da

cena imageada pelo sensor AVIRIS. Escolheu-se pela AESNA $K = 900$ amostras espectrais candidatas a MR. A Fig. 4 mostra a posição das amostras candidatas a MR na subcena 5.

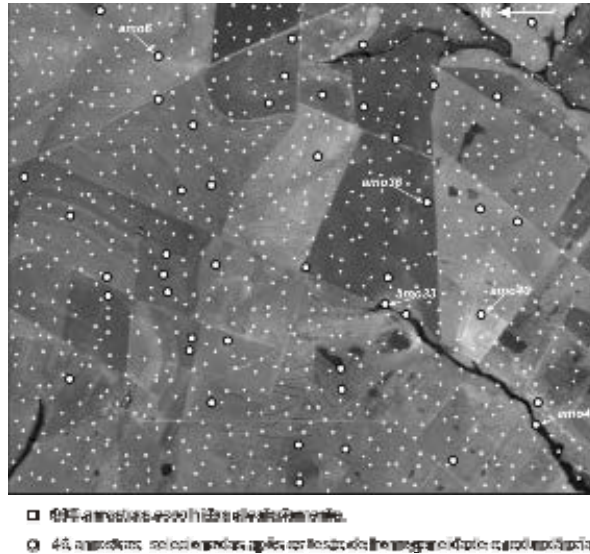


Figura 4 - Posicionamento das amostras candidatas a MR na subcena 5.

Para se reduzir o número de amostras candidatas a MR e em conseqüência diminuir-se o tempo computacional do processo de escolha dos MR utilizou-se o critério de eliminação de amostras proposto por Pizarro e Fernandes (2007). O critério de eliminação de amostras verifica inicialmente a homogeneidade da vizinhança das amostras candidatas a MR, descartando as amostras com vizinhanças muito heterogêneas (vizinhanças com muita mistura de materiais). Após o descarte pelo critério da homogeneidade é testada a redundância entre os espectros médios das amostras candidatas com vizinhança homogênea. O teste da redundância escolhe um espectro médio, calculado em uma vizinhança homogênea, e rejeita os outros que possam ser considerados redundantes com o escolhido. Das 900 amostras iniciais o teste de homogeneidade escolheu 624 amostras e destas o teste de redundância escolheu 46 amostras candidatas a MR ($K = 46$). As amostras candidatas rotuladas como (*amo1, amo2, ..., amo46*) estão em destaque na Fig. 4.

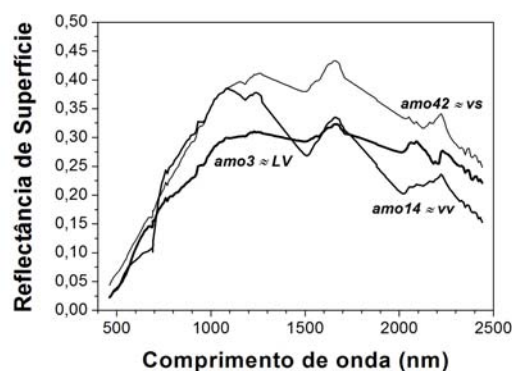
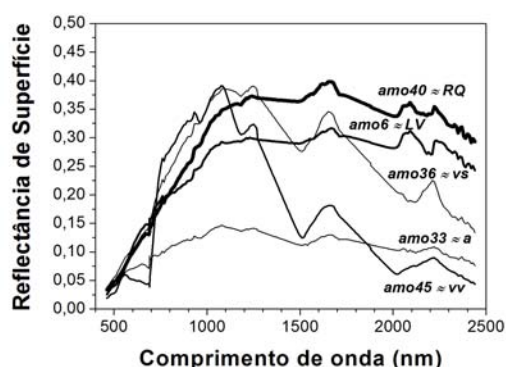
Utilizando-se a média, da derivada dos espectros originais, conforme descrito pela expressão (2), em uma vizinhança $(M + 1) \times (M + 1) = 5 \times 5$ pixels da amostra que será testada, obteve-se com o cálculo da expressão (16) os MR listados na Tabela 2, que também mostra a equivalência das amostras escolhidas com as classes básicas dominantes na subcena 5 (*a, LV, vv, vs, RQ*). Nessa Tabela foi suprimido o prefixo *amo* das amostras.

Os limiares constantes na Tabela 2 foram escolhidos segundo o mesmo critério de escolha dos limiares da Tabela 1. Os espectros dos MR relativos as amostras constantes na Tabela 2 são mostrados na Fig. 5. Por inspeção associa-se nessa Figura as amostras escolhidas (*amo6, amo45, amo33, amo36, amo42, amo3, amo14, amo40*) com as classes básicas (*vv, vs, a, LV, RQ*).

TABELA 2 - MR SELECIONADOS NA SUBSENA 5 COM $H_{MIN} = 0,50$ E LIMIARES DE CONFIGURAÇÃO $\eta_H = 0,044$, $\eta_{DE} = 307,78$ E $\eta_{CE} = 0,99$.

MR (R):	Amostras selecionadas como MR:	Entropia	Imagem-erro	
			média	desvio padrão
2	6, 45 ~ (LV, vv)	0,86	18,76	6,84
3	6, 45, 33 ~ (LV, vv, a)	0,79	18,36	6,86

4	6, 45, 33, 36	0,67	6,05	2,00
	~ (LV, vv, a, vs)			
5	6, 45, 33, 36, 40	0,57	4,35	1,26
	(LV, vv, a, vs, RQ)			
6	6, 45, 33, 36, 42, 3	0,49	3,65	1,08
	~ (LV, vv, a, vs, vs, LV) $\Rightarrow \mathfrak{R}_2 = 5$			
7	6, 45, 33, 36, 42, 3, 14	0,45	3,02	0,94
	~ (LV, vv, a, vs, vs, LV, vv)			
8	6, 45, 33, 36, 42, 3, 14, 40	0,41	2,70	0,76
	~ (LV, vv, a, vs, vs, LV, vv, RQ)			



5a - Espectros selecionados para $R = 5$.

5b - Espectros selecionados para $R \geq 6$.

Figura 5. Espectros dos MR escolhidos para a subcena 5 em correspondência com (vv, vs, a, LV, RQ).

Na Tabela 2 nota-se que a medida que R aumenta as amostras escolhidas anteriormente como MR vão se mantendo, com exceção da amostra *amo40*, que foi substituída pela amostra *amo42* para $R = 6$ e 7, mas foi escolhida novamente para $R = 8$, juntamente com *amo42*. Observa-se que para $R = 6$ *amo36* e *amo42* representam tipos de amostras próximas à classe vegetação seca e que a classe relativa a *RQ* está ausente, indicando assim, que esta escolha de MR não é adequada. A ausência de *RQ* também pode ser notada para $R = 7$, onde estão presentes dois tipos de vegetação verde e dois tipos de vegetação seca.

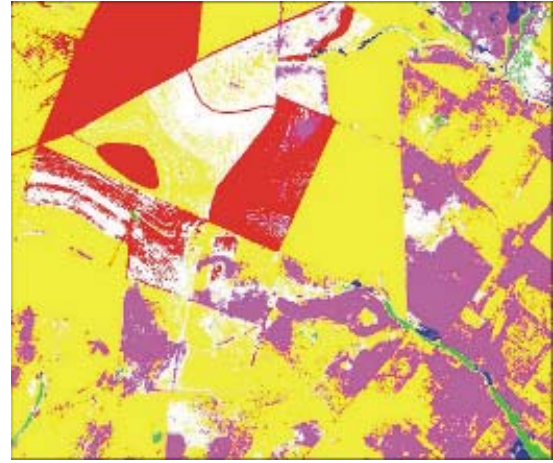
A Tabela 2 mostra também os valores da Entropia, da média e do desvio padrão da imagem-erro, calculada pelo MLME utilizando-se os MR listados. Nota-se que a escolha de *amo36* como MR, para $R = 4$, diminui consideravelmente o erro, isto porque a *amo36* é um tipo de vegetação seca (*vs*) que é predominante na área imageada. No caso de $R = 6, 7, 8$ o valor da Entropia também ficou inferior ao valor mínimo estabelecido ($H_{min} = 0,50$), assim, o limitante superior de MR é $\mathfrak{R}_2 = 5$.

As imagens pós-classificadas da subcena 5 para $R = 5$ e 6 encontram-se nas Fig. 7a e 7b. Analisando essas Figuras nota-se que a qualidade da

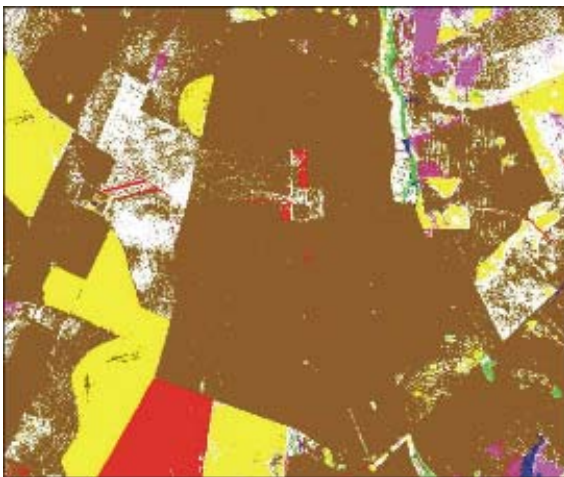
classificação para $R = 5$, como era esperado, é melhor, uma vez que, como comentado anteriormente quando $R = 6$ a classe *RQ* está ausente.



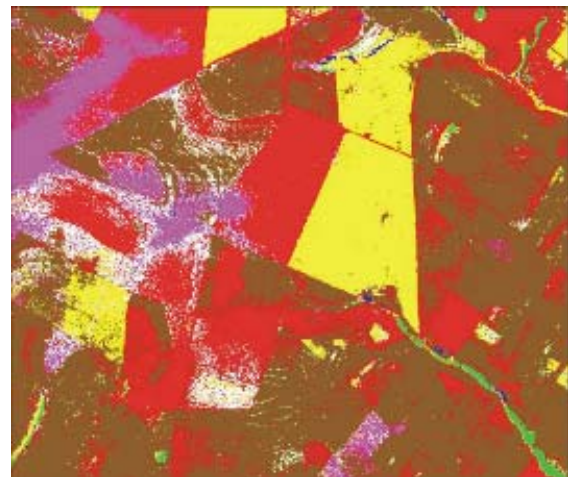
6a - subcena 4 classificada com $R=5$.



7a - subcena 5 classificada com $R=5$.



6b - subcena 4 classificada com $R=6$.



7b - subcena 5 classificada com $R=6$.

Figura 6 - Imagens pós classificadas da subcena 4. Em cores as classes com *pixels* tendo valores de abundância superiores a 50% da fração total e em branco as áreas não-classificadas (intensa mistura espectral).

Figura 7 - Imagens pós classificadas da subcena 5. Em cores as classes com *pixels* tendo valores de abundância superiores a 50% da fração total e em branco as áreas não-classificadas (intensa mistura espectral).

6 – CONSIDERAÇÕES FINAIS

Apresentou-se uma nova metodologia para a determinação dos MR em dados hiperespectrais baseada no conceito de Entropia. A Entropia é calculada em função dos autovalores normalizados obtidos da matriz de covariância entre os espectros candidatos a MR.

Foi proposto ainda o estabelecimento de limitantes superiores para o número de MR a serem escolhidos. Um destes limitantes é estabelecido testando-se a configuração do conjunto de espectros que podem ser escolhidos como MR. Quando um conjunto de R espectros tem espectros muito similares, segundo a medida de similaridade baseada na DE , CE e Entropia, o conjunto de espectros é classificado como mal configurado e não pode ser escolhido como MR. O outro limitante é função de um valor de Entropia mínima que pode ter um conjunto de R espectros escolhidos como MR.

A metodologia proposta foi testada com êxito utilizando-se dados hiperespectrais de uma subcena obtida pelo sensor aerotransportado AVIRIS. Foi realizada também a aplicação do método para a escolha dos MR e a determinação da proporção com que cada MR participa dos *pixels* da imagem, utilizando-se uma outra subcena do AVIRIS.

Na metodologia proposta destacam-se dois resultados importantes:

- a) os $R+1$ MR escolhidos contém em geral os R MR escolhidos anteriormente, mostrando assim que o processo de busca dos MR produz resultados coerentes entre si;
- b) no teste realizado na subcena 4 o método proposto trocou, com sucesso, uma amostra de vegetação seca (*vs*) selecionada pelo MEST por uma outra amostra da mesma classe, mostrando assim um melhor desempenho no processo de seleção de MR.

AGRADECIMENTOS

Os autores agradecem ao Laboratório de Tratamento de Imagens (LTID) do INPE, pela concessão das imagens do sensor AVIRIS utilizadas neste trabalho, à Divisão de Eletrônica Aeroespacial do INPE pelo apoio e ao Projeto FINEP-CAPTAER em desenvolvimento no ITA.

REFERÊNCIAS

BATESON, A.; CURTISS, B. A method for manual endmember selection and spectral unmixing. **Remote Sensing of Environment**, v. 55, n. 1, p. 229-243, Jan. 1996.

BIOUCAS-DIAS, J. M.; NASCIMENTO, J.M.P. Estimation of Signal Subspace on Hyperspectral Data. In: Image and Signal Processing for Remote Sensing XI, Oct. 2005. **Proceedings: SPIE**, Bellingham, Wash., 2005, p. 191-198 (SPIE, 5982).

BIRKHOFF, G.; MACLANE, S. **Álgebra moderna básica**. Rio de Janeiro, RJ: Editora Guanabara Dois S.A., 1977. 484 p.

BOARDMAN, J.W.; KRUSE, F.A.; GREEN, R.O. Mapping Target Signatures via Partial Unmixing of AVIRIS data. In: Annual JPL Airborne Geoscience Workshop, 5. **Anais**. Pasadena: NASA.JPL, Jan. 23, 1995. v. 1, p. 23-26. (JPL Publication 95-1).

CHANG, C.-I. **Hyperspectral Imaging: Techniques for Spectral Detection and Classification**. New York: Kluwer Academic/Plenum Publishers, 2003. XVIII+370p.

CHANG, C.-I.; DU, Q. Estimation of number of spectrally distinct signal sources in hyperspectral imagery. **IEEE Transactions on Geoscience and Remote Sensing**, v. 42, n. 3, p. 608-619, March 2004.

CONGALTON, R.G. A comparison of sampling schemes used in generating error matrices for assessing the accuracy of maps generated from remotely sensed data. **Photogrammetric Engineering and Remote Sensing**, v. 54, n. 5, p. 593-600, May 1988.

DEMETRIADES-SHAH, TANVIR H.; STEVEN, M.D.; CLARK, J.A. High resolution derivative spectra in remote sensing. **Remote Sensing of Environment**, v. 33, n. 1, p. 55-64, July 1990.

FARAKLIOTI, M.; PETROU, M. Recovering more classes than available bands for sets of mixed pixels in satellite images. **Image and Vision Computing**, v. 18, n. 9, p. 705-713, 2000.

GREEN, R.O. Retrieval of reflectance from AVIRIS-measured radiance using a radiative transfer code. In: Airborne Visible/Infrared Imaging Spectrometer (AVIRIS) Workshop, 3. Pasadena, CA, U.S.A.: JPL, 20-21 maio, 1991. **Proceedings**. NASA.JPL, 1991. p.200-210. (JPL Publication 91-28).

GROSS, H.N.; SCHOTT, J.R. Application of Spatial Resolution Enhancement and Spectral Mixture Analysis to Hyperspectral Images. In: Hyperspectral Remote Sensing and Applications, Nov. 2006. **Proceedings: SPIE**, Bellingham, Wash., 2006, p. 30-41 (SPIE, 2821).

NOVAES, A.S.S.; AMARAL FILHO, Z.P.; VIEIRA, P.C.; FRAGA, A.G.C. **Pedology**, Technical Report of the Radam Brasil Project, v. 31, Section SE.22, p. 413-576, 1983.

PIZARRO, M.A.; FERNANDES, D. Seleção de amostras espectrais candidatas a membro de referência em imagens hiperespectrais. In: Simpósio Brasileiro de Sensoriamento Remoto (SBSR), 13., 2007,

Florianópolis, SC, Brasil **Anais...** São José dos Campos: INPE.

PIZARRO, M.A.; FERNANDES, D. Avaliação de métodos de estimativa de coerência espectral para seleção de membros de referência em imagens hiperespectrais. **Revista Brasileira de Cartografia**, v. 57, n. 2, p. 113-120, 2005.

PIZARRO, M.A.; FERNANDES, D. aplicação da medida de divergência na determinação de membros de referência em imagens hiperespectrais . In: Simpósio Brasileiro de Sensoriamento Remoto (SBSR). 11., 2003, Belo Horizonte, MG, Brasil **Anais...** José dos Campos: INPE, p. 1083-1090. 1 85-17-00017-X).

PIZARRO, M.A.; **Sensoriamento remoto hiperespectral para a caracterização e identificação mineral em solos tropicais**. Jun. 1999. 185 p. (INPE-7249-TDI/693). Dissertação (Mestrado em Sensoriamento Remoto) - Instituto Nacional de Pesquisas Espaciais, São José dos Campos. 1999. Disponível na biblioteca digital *URLib*: <<http://iris.sid.inpe.br:1912/rep-/sid.inpe.br/deise/1999/10.22.17.45>>. Acesso em: 2004, Nov. 18.

ROBILA, S.A.; GERSHMAN, A. Spectral matching accuracy in processing hyperspectral data. In: International Symposium on Signals, Circuits and Systems (ISSCS), 2005, Iasi, Romenia, **Proceedings...** Romenia: 2005, 14-15 July, p. 163-166, v. 1 (ISBN: 0-7803-9029-6).

SHAW, G.; MANOLAKIS, D. Signal processing for hyperspectral image exploitation. **IEEE Signal Processing Magazine**, v. 19, n. 1, p. 12-16, Jan.. 2002.

SCHOWENGERDT, R.A. **Remote Sensing: models and methods for image processing**. San Diego, California: Academic Press, 1997. XLV+522 p. (ISBN: 0-12-628981-6).

STRANG, G.; BORRE, K. **Linear algebra, geodesy, and GPS**. Wellesley, MA: Wellesley-Cambridge Press, 1997. XVI+624 p. (ISBN: 0-9614088-6-3).

SWEET, J.; GRANAHAN, J.; SHARP, M. An objective standard for hyperspectral image quality. In: Proceedings IX Nasa JPL Airborne Earth Science Workshop. **Anais**. Pasadena, CA; U.S.A.: JPL, 2000.

TSAI, F.; PHILPOT, W.D. A derivative-aided hyperspectral image analysis system for land-cover classification. **IEEE Transactions on Geoscience and Remote Sensing**, v. 40, n. 2, p. 416-425, Feb.2002.