

Investigação sobre métodos para redução da dimensionalidade dos dados em imagens hiperespectrais

Maciel Zortea
Victor Haertel
Robin Thomas Clarke

Centro Estadual de Pesquisas em Sensoriamento Remoto e Meteorologia
Programa de Pós-Graduação em Sensoriamento Remoto - UFRGS
Av. Bento Gonçalves, 9500 - Caixa Postal 15044 - 91501-970 - Porto Alegre - RS - Brasil
mzortea@yahoo.com - victor.haertel@ufrgs.br - clarke@iph.ufrgs.br

Abstract. In the present study, we propose a new simple approach to reduce the data dimensionality in hyperspectral image data. The basic assumption here consists in assuming that a pixel's curve of spectral response, as defined in the spectral space by the recorded digital numbers (DN's) at the available spectral bands, can be segmented and each segment can be replaced by a smaller number of statistics: mean and variance, describing the main characteristics of a pixel's spectral response. It is expected that this procedure can be accomplished without significant loss of information. The DN's at every spectral band are here used to calculate a few statistics that will then replace them in a classifier. For the pixel's spectral curve segmentation, we propose tree sub-optimal algorithms that are easy to implement and also computationally efficient. Using a top-down strategy, the length of the segments along the spectral curves can or not be adjusted sequentially. Experiments using a parametric classifier are performed on an AVIRIS data set. Encouraging results have been obtained in terms of classification accuracy and execution time, suggesting the effectiveness of the proposed algorithms.

Palavras-chave: feature extraction, feature selection, data dimensionality reduction, hyperspectral image data, remote sensing, extração de feições, seleção de feições, redução da dimensionalidade, imagens hiperespectrais, sensoriamento remoto.

1. Introdução

Recentes avanços na tecnologia de sensores no âmbito de sensoriamento remoto vem disponibilizando para a comunidade científica internacional uma nova geração de sensores, capazes de fornecer imagens em alta dimensionalidade. A análise e a interpretação destes dados apresentam novas possibilidades e também novos desafios. Dados coletadas em um grande número de bandas espectrais contêm informações mais detalhadas a respeito das classes a serem identificadas, permitindo teoricamente uma acurácia no processo de classificação maior do que aquela obtida com a utilização de dados fornecidos por sistemas sensores mais tradicionais com Landsat-TM e SPOT.

A utilização de dados em alta dimensionalidade apresenta ao analista alguns desafios. Neste contexto, uma das maiores dificuldades reside na estimativa, a partir de um número geralmente limitado de amostras de treinamento, dos parâmetros de um classificador estatístico. Em um processo de classificação, tipicamente a acurácia aumenta a medida em que informação adicional, i.e., novas bandas espectrais, vão sendo adicionadas. A acurácia do classificador atinge um máximo em um determinado ponto, para decrescer com a introdução de bandas adicionais. Este é o bem conhecido fenômeno de Hughes, resultado da incerteza nos valores estimados para os parâmetros do classificador, efeito este causado pelo pequeno número de amostras de treinamento comparado à dimensionalidade dos dados.

Diversas alternativas têm sido propostas na literatura para amenizar este problema. Alternativas baseadas na regularização da matriz de covariância foram investigadas por Friedman (1989), Aeberhard et al. (1994), Hoffbeck e Landgebe (1996) entre outros. Uma segunda possibilidade consiste na introdução de amostras ditas não rotuladas para aumentar o número de amostras de treinamento. Para atingir este objetivo, Shahshahani e Landgrebe

(1994) propuseram o uso do algoritmo Expectation Maximization (EM). Jackson e Landgrebe (2001) propuseram um classificador iterativo que faz uso de amostras classificadas (designadas como semi-rotuladas) em adição às amostras de treinamento existentes. Outra possível alternativa que tem recebido grande quantidade de atenção de muitos autores consiste em métodos para representação dos dados numa dimensão reduzida [Ref.6]-[Ref.14]. Duas alternativas têm sido propostas por nesta abordagem: seleção de feições e extração de feições. Ambas são essencialmente classe-dependentes.

Dado um conjunto de medidas na dimensionalidade p , o objetivo das técnicas de seleção de feições consiste em selecionar um subconjunto de variáveis d ($d < p$) que mais contribua para discriminação das classes de interesse. Seguindo essa abordagem, subconjuntos com d bandas são testados, optando-se por aquele que maximiza uma dada função de otimização adotada como critério de seleção. As técnicas de extração de feições promovem a redução da dimensionalidade dos dados através da aplicação de uma transformação (linear ou não) dos dados originais.

A metodologia proposta neste estudo insere-se na categoria geral de extração de feições, havendo apresentado resultados satisfatórios nos testes executados, além de ser de fácil implementação e computacionalmente eficiente.

A suposição básica nesta proposta é que a curva de resposta espectral de cada pixel pode ser segmentada, e cada um dos segmentos gerados pode ser representado por um número menor de estatísticas simples sem perda significativa de informação. Entre as várias possíveis estatísticas a serem utilizadas, optou-se pela média aritmética e variância dos contadores digitais das bandas contidas em cada segmento. Desta forma, cada segmento passa assim a ser representado por dois parâmetros estatísticos: um de posição e outro de forma.

Para fins de segmentação da curva de resposta espectral, são propostos três algoritmos, possibilitando a geração de segmentos de comprimentos constante ou variável. No caso da geração de segmentos de comprimento variável, as partições são definidas sequencialmente seguindo uma estrutura em árvore do tipo *top-down*. O critério utilizado na partição da curva de resposta espectral em segmentos utiliza uma medida de separabilidade estatística entre classes, baseada na distância de Bhattacharyya. Uma vez definida uma partição, esta não pode mais ser alterada, o que faz com que os algoritmos propostos sejam sub-ótimos.

2. Metodologia

Numa imagem hiperespectral e para fins de classificação, cada pixel pode ser representado por um vetor $\mathbf{X} = [x_1 \ x_2 \ . \ . \ . \ x_p]$. As variáveis x_i ($i = 1, \dots, p$), caracterizados por uma alta correlação entre valores adjacentes, são as respostas individuais fornecidas em cada uma das p bandas espectrais disponíveis.

O passo inicial para implementação da metodologia aqui proposta, conforme descrito na seção anterior, consiste na partição de toda a curva de resposta espectral em segmentos consecutivos s_j , $j = 1, \dots, k$, de bandas adjacentes, sendo k o número total de segmentos. Cada segmento é composto de um conjunto contínuo de bandas $s_j = [x_l \ . \ . \ . \ x_u]$, onde cada variável no vetor representa a resposta do pixel em uma banda espectral, sendo $l \geq 1$ e $(l+2) \leq u \leq p$. A segmentação da curva de resposta espectral é gerada de tal forma que $s_1 \cup s_2 \cup \dots \cup s_k = \mathbf{X}$ e $s_i \cap s_j = \emptyset$ para cada $i \neq j$. Nesta metodologia, os contadores digitais em cada segmento s_j são substituídos pela correspondente média aritmética (1) e variância (2):

$$\bar{x}_{l,u}(x) = \frac{1}{u-l+1} \sum_{i=l}^u x_i \quad (1)$$

$$\sigma^2_{l,u}(x) = \frac{1}{u-l} \sum_{i=l}^u (x_i - \bar{x}_{l,u}(x))^2 \quad (2)$$

Designando por $f_j = [\bar{x}_{l,u}(x) \ \sigma^2_{l,u}(x)]$ as duas feições provenientes de cada segmento s_j , cada pixel passará a ser representado por um vetor, de dimensionalidade reduzida, com componentes $Y = [f_1 \ f_2 \ f_3 \ \dots \ f_{2k}]$. Desta forma, a redução de \mathbf{X} para \mathbf{Y} equivale a uma passagem do espaço de $i^N \rightarrow i^{2k}$.

Neste trabalho, são propostos e testados três algoritmos simples para segmentação da curva de resposta espectral dos pixels (vetor \mathbf{X}), conforme ilustrado na **Figura 1**.

O primeiro algoritmo proposto é o *SCC - Segmentação Comprimento Constante*. Esse algoritmo implementa a alternativa mais simples para segmentação, isto é, o vetor \mathbf{X} é particionado em segmentos de igual comprimento (**Figura 1(a)**). Por ser uma solução extremamente simples, torna-se atraente do ponto de vista computacional, pois além da facilidade para implementação, a execução do algoritmo é veloz. Este processo de partição não leva entretanto em consideração as características locais das curvas de resposta espectral nem critérios de separabilidade entre as classes. Na formação dos segmentos, pode ocorrer ainda que o número de bandas espectrais não seja um múltiplo inteiro do número de segmentos. Neste caso ajusta-se levemente o comprimento de cada segmento de modo que as bandas “excedentes” sejam convenientemente distribuídas ao longo da totalidade de segmentos gerados. Isto representa uma pequena alteração na maneira como foi implementado o algoritmo em Zortea e Haertel (2004b).

Esta metodologia adotando segmentos de comprimento fixo obviamente não deve conduzir aos melhores resultados. Idealmente, o comprimento e a posição de cada segmento deveria ser ajustado com base nas características locais da curva de resposta espectral utilizando-se um critério de separabilidade entre classes.

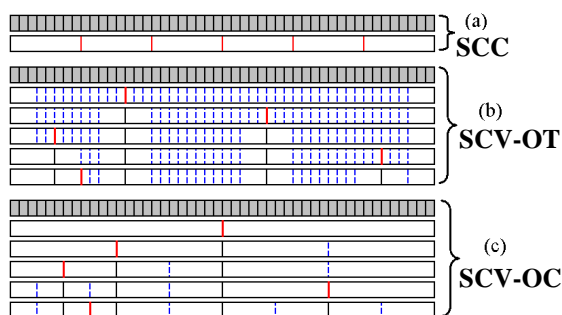


Figura 1. Segmentação da curva de resposta espectral de cada pixel, segundo os três algoritmos propostos neste trabalho, para a geração de seis segmentos. As linhas pontilhadas representam todas as posições testadas a cada “nível hierárquico”. As linhas sólidas representam as posições selecionadas: prévias, invariáveis, e nova, aquela que maximiza a função critério $J(m)$.

Para tanto, propõem-se dois algoritmos adicionais em estrutura “*top-down*” que possibilitam a geração de segmentos de largura variável. Tratam-se dos algoritmos *SCV-OT: Segmentação Comprimento Variável com Otimização Total* e uma simplificação deste representada pelo algoritmo *SCV-OC: Segmentação Comprimento Variável com Otimização Central*. Nessas abordagens, a otimização dos segmentos é feita seqüencialmente. Ao final de cada “nível hierárquico” (**Figura 1(b)(c)**), gera-se uma partição adicional à(às) pré-existente(s), numa posição m da curva original ($1 < m < p$). Esta posição é definida como sendo o ponto que maximiza uma função critério $J(m)$. Esta função critério estima a separabilidade entre as classes.

A avaliação é feita em termos de separabilidade estatística entre classes, estando a partição posicionada em m . Assumindo que as classes apresentem distribuição de probabilidades normal, o critério $J(m)$ explorado neste trabalho para otimização das partições geradas é baseado na minimização do limiar superior de Bhattacharyya (*upper bound*) do erro

bayesiano ε_u que ocorre no processo de classificação (Fukunaga (1990), Bruzzone *et al.* (1995)). Desta forma, define-se:

$$J(m) = -\sum_{i=1}^c \sum_{j>i}^c \sqrt{p(\omega_i)p(\omega_j)} e^{-b_{ij}}, \text{ onde} \quad (3)$$

$$b_{ij} = \frac{1}{8}(M_i - M_j)^T \left(\frac{\Sigma_i + \Sigma_j}{2} \right)^{-1} (M_i - M_j) + \frac{1}{2} \ln \left(\frac{\left| \frac{\Sigma_i + \Sigma_j}{2} \right|}{\sqrt{|\Sigma_i| |\Sigma_j|}} \right) \quad (4)$$

onde b_{ij} é a distância de Bhattacharyya entre as classes ω_i e ω_j , M_i e M_j o vetor média e Σ_i e Σ_j as matrizes de covariância das duas classes, obtidas a partir de \mathbf{Y} , estando a partição na posição m , $p(\omega_i)$ e $p(\omega_j)$ são as probabilidades de ocorrência a priori das classes. A utilização de um critério baseado num limiar justifica-se pelo fato de estar diretamente relacionado ao patamar superior do erro de classificação que será obtido posteriormente.

Note que cada posição m testada altera o conjunto de segmentos $s_1 \dots s_k$, obtidos. Cada segmento é substituído pelas correspondentes feições, sendo $J(m)$ avaliado em dimensionalidade reduzida, ou seja, em \mathbf{Y} .

No algoritmo *SCV-OT*, dado um conjunto de segmentos iniciais $s_1 \dots s_k$, testam-se *todas as possibilidades* para geração de um novo segmento. Neste caso, a partição de cada segmento s_j $j=1, \dots, k$, definido no intervalo $x_{l,u} = [x_l, \dots, x_u]$, em dois novos segmentos $[x_l, \dots, x_m] \cup [x_{m+1}, \dots, x_u]$, é feita variando-se m no intervalo $l+2 \leq m < u-2$. Em outras palavras, esta procura é exaustiva, no sentido que testam-se todas as possíveis posições para posicionar m (**Figura 1(b)**).

O algoritmo *SCV-OC* é uma simplificação do anterior. Ao invés de testarmos todas as possíveis alternativas para a definição da nova posição de partição m , testa-se apenas a *posição central* de cada um dos segmentos s_j previamente existentes (**Figura 1(c)**). Neste acaso, m é definido como sendo o valor inteiro da divisão $(u - l + 1)/2$. Obviamente, a execução deste algoritmo é mais rápida do que o *SCV-OT*, mantendo-se ainda uma certa possibilidade de adaptação à resposta espectral do conjunto de pixels da cena de estudo.

3. Resultados experimentais

A. Dados utilizados

A fim de testar a metodologia utilizando os algoritmos propostos neste trabalho, realizou-se uma série de experimentos. Utilizou-se uma imagem hiperespectral de 220 bandas obtidas pelo sensor aerotransportado *Airborne Visible Infrared Imaging Spectrometer* (AVIRS) da NASA. A imagem cobre, de maneira aproximadamente continua, o espectro eletromagnético na região de comprimentos de onda dos 400–2500 nm. Trata-se de uma imagem com 145 x 145 pixels, (**Figura 2**) com resolução espacial de aproximadamente 20 m, adquirida em junho de 1992, numa região agrícola de testes, no norte de Indiana (EUA) [15].



Figura 2. Imagem hiperespectral obtida pelo sensor AVIRS. Apresenta-se uma composição colorida RGB utilizando as bandas correspondentes aos comprimentos de onda 860, 657 e 557 nm.

A data de aquisição corresponde à época de início do cultivo, onde apenas uma pequena fração da superfície ($\cong 5\%$) era coberta pelas cultivares. A maior contribuição provém da superfície, isto é, solo exposto e detritos. Esta situação resulta em classes com respostas espectrais muito semelhantes, tornando o problema de classificação bastante desafiador.

Utilizando-se a verdade terrestre, foram selecionadas seis classes de interesse. Conforme procedimento usual, os pixels pertencentes a cada uma das classes foram separados em dois conjuntos: pixels de treinamento (utilizados para geração da segmentação e estimativa dos parâmetros do classificador) e teste (utilizados para verificação da acurácia de classificação obtida) (Tabela 1). As bandas ruidosas, principalmente devido à absorção causada pela presença de vapor d'água na atmosfera, foram descartadas. Desta forma, das 220 bandas originais disponíveis, utilizaram-se 190.

Tabela 1. Número total de pixels de treinamento e teste disponíveis para cada uma das classes de interesse

Classe:	Treinamento	Teste	Total
(ω_1) Milho - cultivo direto	717	717	1434
(ω_2) Milho - cultivo mínimo	417	417	834
(ω_3) Soja - cultivo direto	484	484	968
(ω_4) Soja - cultivo mínimo	1234	1234	2468
(ω_5) Soja - cultivo limpo	307	307	614
(ω_6) Floresta	647	647	1294

Do conjunto de pixels disponíveis (Tabela 1), foram extraídos, de maneira aleatória, cinco subconjuntos contendo 100, 200 e 300 pixels para cada uma das seis classes de treinamento e teste. Desta forma, espera-se melhor avaliar o desempenho dos algoritmos frente à utilização de diferentes tamanhos de amostra para pixels de treinamento. O processo de seleção de pixels foi repetido cinco vezes para melhor testar a estabilidade dos algoritmos quando variadas as amostras de ingresso.

B. Resultados e discussão

O desempenho dos algoritmos foi avaliado em termos de acurácia de classificação. Para tal, utilizou-se o classificador supervisionado paramétrico Máxima Verossimilhança Gaussiana, sob hipótese de igual probabilidade de ocorrência a priori das classes.

Os desempenhos dos três algoritmos propostos neste trabalho é comparado com dois algoritmos clássicos: um de seleção e outro de extração de feições. No âmbito de seleção de feições, a comparação é feita utilizando-se o algoritmo *Sequential Forward Selection* (SFS). Tal escolha é justificada pelo simples fato desse algoritmo ser um dos mais simples, e conseqüentemente, rápido, para seleção de feições. Adicionalmente, os algoritmos SCV aqui propostos para extração de feições utilizam o mesmo tipo de critério de otimização que é geralmente utilizado no âmbito de seleção de feições. O algoritmo SFS adota uma estratégia *bottom-up*. Utilizando uma função (neste caso utilizou-se a mesma $J(m)$, equação (3)) que visa otimizar a separabilidade entre classes, identifica o melhor conjunto de feições que pode ser obtido iterativamente, dado um conjunto inicialmente vazio de feições, adicionando uma feição a cada iteração. Uma vez selecionada, a feição não pode mais ser descartada, ou seja, o processo é sub-ótimo, da mesma maneira que os algoritmos aqui propostos, pois uma vez definida uma partição, essa não é mais alterada. Outras alternativas interessantes surgiram para aperfeiçoar o SFS, por exemplo Jain e Zongker (1997), Serpico e Bruzzone (2001), Serpico *et al.* (2002), possibilitando melhores resultados em termos de acurácia de classificação, porém às custas de uma procura mais elaborada, e por conseqüência, com um custo computacional mais elevado.

Os valores correspondentes ao número de vezes que deve ser calculado $J(m)$ são apresentados na **Figura 3**. Para os dois algoritmos SCV, constituem-se em limites superiores para chegar-se a dada dimensionalidade, visto que durante o processo de segmentação alguns

segmentos gerados podem vir a conter apenas três bandas, impedindo que sejam posteriormente particionados, conseqüentemente diminuindo ainda mais o número de avaliações de $J(m)$. Experimentos preliminares indicam que o algoritmo SCV-OC é de execução mais rápida do que o SFS. Em certas circunstâncias, observou-se que o SCV-OT também supera o SFS (Zortea e Haertel 2004b). O incremento a cada nova partição é de duas feições, ao contrário do aumento unitário propiciado pelo SFS. Já o algoritmo SCC não necessita do cálculo de $J(m)$, conferindo grande vantagem em termos de execução em relação aos demais testados.

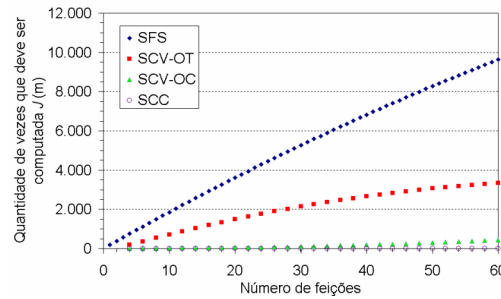


Figura 3. Representação, para o caso de 190 bandas originais disponíveis, do número de vezes que deve ser calculado o funcional $J(m)$ para atingir-se uma dada dimensionalidade. Nos três algoritmos propostos neste trabalho o incremento é de duas feições a cada nova partição, ao contrário do clássico SFS.

Para fins de comparação no âmbito de extração de feições, optou-se por utilizar a bem conhecida transformação de componentes principais (PCT), neste caso, aplicada aos dados na dimensionalidade original (190 bandas), utilizando-se as d primeiras componentes.

Os resultados para a acurácia obtidos no processo de classificação são apresentados na **Figura 4(a)**. Nesta figura estão ilustrados os valores médios da acurácia obtida para as seis classes utilizadas no experimento, em função da dimensionalidade dos dados. Na **Figura 4(a)** são apresentados os resultados utilizando-se amostras de treinamento e teste com 100 pixels cada, em **4(b)** 200 pixels, em **4(c)** 300 pixels e por fim em **4(d)** são utilizados todos os pixels disponíveis.

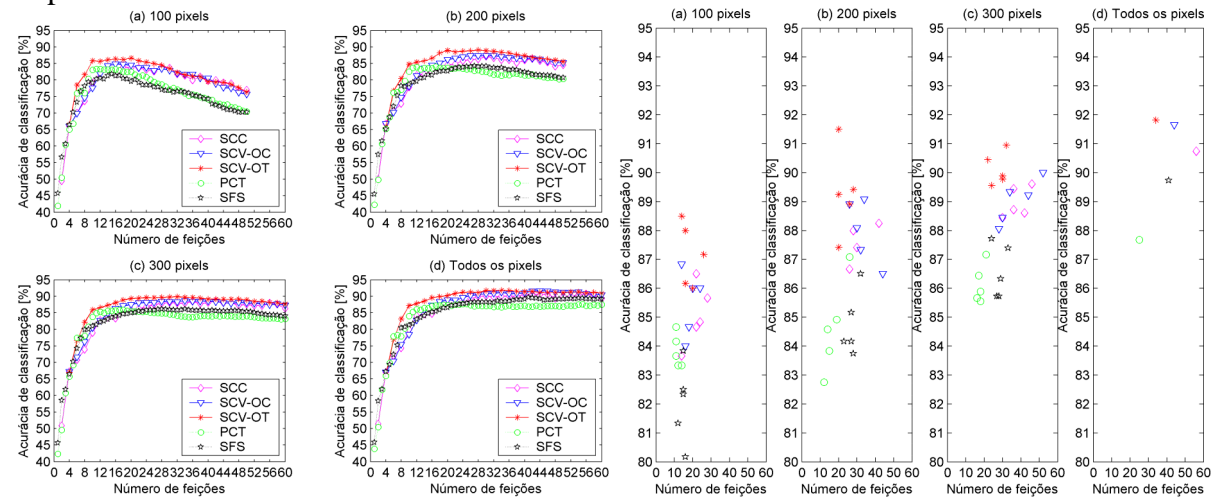


Figura 4. (a) [à esquerda] Acurácias médias para as seis classes segundo o algoritmo, número de feições (dimensionalidade) e tamanho das amostras de treinamento e teste utilizadas. Resultados correspondem à média das cinco repetições de seleção aleatória de n pixels por classe. (b) [à direita] Com a mesma legenda, são apresentados os correspondentes picos de acurácia média de classificação em uma das cinco repetições realizadas. Para cada seleção de n pixel, aplicaram-se todos os algoritmos.

Os valores de pico da acurácia obtida no processo de classificação, e o correspondente número de feições para o qual este valor foi obtido são apresentados na **Figura 4(b)**. Na média, observa-se na **Figura 5** que os três algoritmos propostos superam os valores de pico de

acurácia fornecidos pelas duas abordagens clássicas testadas, com desempenho crescente para os algoritmos SCC, SCV-OC, SCV-OT respectivamente. Para um número reduzido de feições, houveram situações em que os algoritmos clássicos apresentaram melhor desempenho (**Figura 4(a)**), porém, num processo de classificação, em geral estamos interessados na obtenção de um conjunto de feições que propicie o máximo de acurácia.

Os efeitos do fenômeno de Hughes aparecem claramente nos experimentos, sendo mais intenso quanto menor a relação entre o tamanho da amostra de treinamento e número de feições utilizadas. Este fato independente do algoritmo utilizado (**Figura 4(a)**).

Na condição de realização dos experimentos, verificou-se que à medida que o número de pixels utilizados aumentou, o desempenho do algoritmo SFS tende a superar os resultados obtidos utilizando-se a transformação PCT (**Figura 5**).

Durante a realização dos experimentos, percebeu-se que a classe ω_4 (*Soja - cultivo mínimo*) é a que apresenta maior dificuldade para uma correta classificação. Em geral, todos os pixels da classe ω_6 (*Floresta*) são corretamente classificados, visto ser uma classe com características bem diversa das demais. De fato, verificou-se que para esta classe, apenas duas feições são suficientes para obtenção de uma acurácia de 100%.

O valor máximo para a acurácia média obtida nos experimentos utilizando-se o algoritmo SFS, e com amostras de treinamento e teste de tamanho 100 (condição mais desfavorável), foi 82,03%. O algoritmo PCT nas mesmas condições apresentou acurácia de 83,83%. A utilização dos novos algoritmos propostos SCC, SCV-OC e SCV-OT proporcionaram um acréscimo médio de 3,03%, 3,47% e 5,13%, respectivamente em relação ao algoritmo SFS (**Figura 5(b)**). Resultados favoráveis também são obtidos para os demais conjuntos de pixels testados.

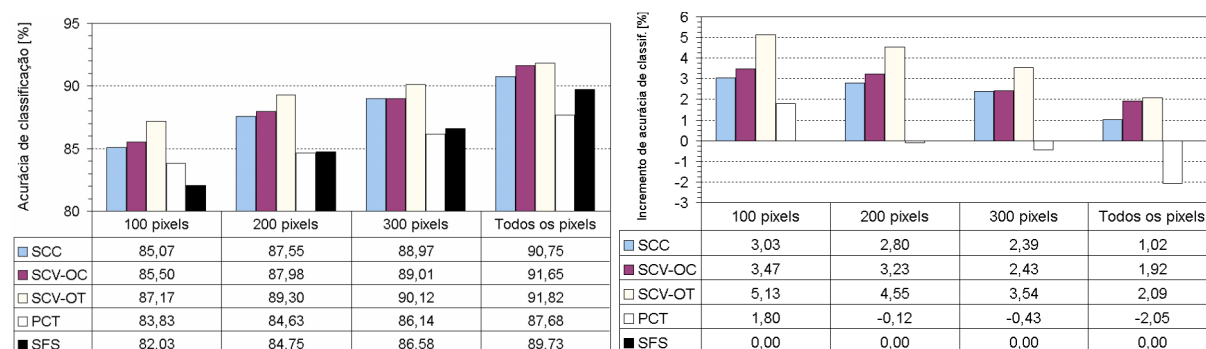


Figura 5. (a) [à esquerda] Valores médios dos picos de acurácia de classificação para os algoritmos testados. (b) [à direita], acréscimo de acurácia de classificação em relação ao algoritmo SFS.

4. Conclusões

Os novos algoritmos propostos neste trabalho, para redução da dimensionalidade dos dados, foram testados utilizando-se uma imagem hiperespectral de uma área agrícola, provida de verdade terrestre. Os experimentos realizados sugerem que os algoritmos propostos além de serem computacionalmente eficientes, também fornecem acurácias médias superiores àquelas obtidas por métodos tradicionais como *Sequential Forward Selection* (SFS) – normalmente apresentado na literatura como um dos métodos mais simples e velozes no âmbito de redução da dimensionalidade através da seleção de feições. Resultados também superiores, em termos de acurácia de classificação, são obtidos se comparados à extração de feições utilizando a transformação de componentes principais (PCT).

Os resultados obtidos nos experimentos sugerem que a suposição inicial de que a curva de resposta espectral de cada pixel pode ser particionada em segmentos, sendo os DN's originais em cada segmento substituídos por um número reduzido de estatísticas simples, sem perda

significativa de informação, parece razoável. Atualmente investiga-se a possibilidade de utilização de outros tipos de feições.

Considera-se que esses resultados animadores sugerem a eficácia dos algoritmos propostos.

5. Referências

- [1] J. H. Friedman, “Regularized Discriminant Analysis”, **Journal of American Statistical Association**, vol 84, No. 405, pp. 165-175, March 1989.
- [2] J. F. Hoffbeck and D. A. Landgrebe, “Covariance Matrix Estimation with Limited Training Data” , **IEEE Transactions on Pattern Analysis and Machine Intelligence**, vol. 18, pp. 763-767, July 1996.
- [3] S. Aeberhard, D. Coomans and O. De Vel, “Comparative Analysis of Statistical Pattern Recognition Methods in High Dimensional Settings”, **Pattern Recognition**, vol. 27, No. 8, pp. 1065-1077, 1994.
- [4] B. Shahshani, D. A. Landgrebe. “The Effect of Unlabeled Samples In Reducing The Small Sample Size Problem And Mitigating The Hughes Phenomenon”, **IEEE Transaction on Geoscience and Remote Sensing**, vol. 32, No. 5, pp. 1087-1095, Sept. 1994.
- [5] Q. Jackson, D. A. Landgrebe. “An Adaptive Classifier Design for High-Dimensional Data Analysis with a Limited Training Data Set”, **IEEE Transaction on Geoscience and Remote Sensing**, vol. 39, pp.2664-2679, 2001.
- [6] K. Fukunaga “**Introduction to Statistical Pattern Recognition**”, 2ª edição, Academic Press, New York, 1990.
- [7] X. Jia, J. A. Richards, “Segmented Principal Components Transformation for Efficient Hyperspectral Remote-Sensing Image Display and Classification” **IEEE Trans. on Geoscience and Remote Sensing**, vol. 37. no. 1 pp.538-542, 1999.
- [8] C. Lee and D. A. Landgrebe, “Decision boundary feature extraction for neural networks,” **IEEE Trans. Neural Networks**, vol. 8, pp. 75–83, Jan. 1997.
- [9] C. Lee and D. A. Langrebe, “Feature Extraction Based on Decision Boundaries”, **IEEE Transactions on Pattern Analysis and Machine Intelligence**, vol. 15, pp 388-400, 1993.
- [10] S. Kumar, J. Ghosh, and M. M. Crawford, “Best-Bases Feature Extraction Algorithms for Classification of Hyperspectral Data”, **IEEE Transaction on Geoscience and Remote Sensing**, vol. 39, pp. 1368-1379, 2001.
- [11] A. Jain, D. Zongker, D. “Feature Selection: Evaluation, Application and Small Sample Performance”, **IEEE Transaction on Pattern Analysis and Machine Intelligence**, vol. 19, No. 2, pp. 153-158, Feb. 1997.
- [12] S. B. Serpico e L. Bruzzone, “A New Search Algorithm for Feature Selection in Hyperspectral Remote Sensing Images”, **IEEE Transaction on Geoscience and Remote Sensing, Special Issue on Analysis of Hyperspectral Image Data**, vol. 39, No. 7, pp. 1360-1367, Julho. 2001.
- [13] S. B. Serpico, M. D’inca, F. Melgani, G. Moser, “A comparison of feature reduction techniques for classification of hyperspectral remote-sensing data” *Image and Signal Processing of Remote Sensing VIII*, Sebastiano B. Serpico, Editor, **Proceedings of SPIE**, Vol. 4885, 2002.
- [14] L. Bruzzone, F. Roli, S. B. Serpico. “An extension to multiclass cases of the Jeffries-Matusita Distance”, **IEEE Transaction on Geoscience and Remote Sensing**, vol. 33, pp. 1318-1321, 1995.
- [15] Multispec homepage, Purdue University, [Online]. Disponível em:
<<http://dynamo.ecn.purdue.edu/~biehl/MultiSpec/documentation.html>>, Acesso em: 12 jan. 2004
- [16] M. Zortea, “**Investigação Sobre Métodos para Redução da Dimensionalidade dos Dados em Imagens Hiperespetrais**”, Dissertação de Mestrado, Centro Estadual de Pesquisas em Sensoriamento Remoto e Meteorologia, Universidade Federal do Rio Grande do Sul, Porto Alegre, Brasil, 107 p., Abril 2004.
- [17] M. Zortea, V. Haertel, “Experiments on Feature Extraction in Remotely Sensed Hyperspectral Image Data”, **Proceedings of International Geoscience and Remote Sensing Symposium (IGARSS)**, 2004.