

# Métodos hierárquicos para redução de dimensões e classificação de imagens AVIRIS

Denis Altieri de Oliveira Moraes<sup>1</sup>  
Victor Haertel<sup>1</sup>

<sup>1</sup> Universidade Federal do Rio Grande do Sul - Centro Estadual de Pesquisas em  
Sensoriamento Remoto e Meteorologia - UFRGS/CEPSRM  
Caixa Postal 15044 - CEP 91501-970 - Porto Alegre – RS, Brasil  
d\_altieri@yahoo.com.br, vitor@ufrgs.br

**Abstract.** In this paper we investigate the performance of a tree structured classifier in the labeling of high dimensional image data. The aim is to mitigate the effects of the Hughes phenomenon in hyperspectral image data classification. The use of a multi-stage classifier, analyzing a sub-set of classes at each stage rather than the full set at once, allows for a more efficient way to reduce the data dimensionality. The extracted features can then be selected in order to maximize the discrimination between the sub-set of classes under consideration. In a binary tree approach, only two classes are considered at each node, allowing for the implementation of the Bhattacharyya distance as a criterion for feature extraction at each tree node. Experiments are performed using AVIRIS image data. The performance of the proposed methodology is compared against the more traditional methods for feature selection and extraction.

**Key words:** pattern recognition, hyperspectral image, decision trees, Bhattacharyya distance, feature selection, feature extraction, reconhecimento de padrões, imagem hiperespectral, árvores de decisão, distância de Bhattacharyya, seleção de variáveis, extração de variáveis.

## 1. Introdução

Avanços mais recentes na área de tecnologia de sistemas sensores, têm disponibilizado um número crescente de bandas espectrais em dados de Sensoriamento Remoto. Os dados em alta dimensionalidade, conhecidos como *hyper-spectrais*, abrem novas possibilidades no processo de classificação de cenas naturais. Sabe-se que classes muito semelhantes entre si, isto é, classes que apresentam vetores de médias muito semelhantes, podem ser separadas com uma razoável precisão, desde que as respectivas matrizes de covariância sejam suficientemente distintas (Fukunaga, 1990). Neste contexto o problema que surge reside na estimação de um crescente número de parâmetros no classificador, com um número geralmente limitado de amostras de treinamento. Soluções visando minorar este problema vêm sendo investigadas segundo três linhas gerais: (1) métodos de análise discriminante regularizada, (2) incremento no número de amostras de treinamento pela inclusão das chamadas amostras *semi-rotuladas* e (3) redução na dimensionalidade dos dados sob a condição de minimizar a perda de informação. Técnicas de análise discriminante regularizada foram investigadas por vários autores, como por exemplo, Friedman (1989), Aeberhardt *et al* (1994), entre outros. O uso de amostras semi-rotuladas foi investigado por Shashahani e Landgrebe (1994), Jackson e Landgrebe (2001), entre outros autores.

As técnicas de redução da dimensão dos dados vêm sendo objeto de investigação por um número grande de autores. Estas técnicas podem ser agrupadas em duas categorias gerais: extração de variáveis e seleção de variáveis. No primeiro caso, transformações lineares ou não-lineares são aplicadas aos dados originais, concentrando o poder de discriminação entre as classes em um número menor de variáveis transformadas, permitindo desta forma reduzir a dimensionalidade dos dados originais, com uma perda de informação desprezível. No processo de seleção de variáveis, critérios de separabilidade entre as classes são utilizados para selecionar um subconjunto ótimo dos dados originais. Assim, esse estudo investiga técnicas de seleção e extração de variáveis para fins de redução da dimensionalidade dos dados a serem implementados em classificadores estágio único e estágio múltiplo. O critério

geralmente adotado para se estimar a separabilidade entre as classes consiste em distâncias estatísticas, como a distância de Bhattacharyya ou a de Jeffries-Matusita. A dificuldade que surge neste particular reside no fato de que estas distâncias estatísticas são definidas para um par de classes. No caso de várias classes estarem presentes, torna-se necessário adotar um critério que leve em consideração o conjunto de todas as classes. No caso de classificadores em estágio único, onde todas as classes presentes são tratadas simultaneamente, emprega-se frequentemente a média das distâncias estimadas para os vários pares de classes. Embora prática, essa abordagem pode não ser adequada sempre que os valores de separação entre as classes no espaço das variáveis apresente valores muito díspares. Neste caso, o subconjunto de variáveis selecionado poderá apresentar-se pouco adequado na separação de algumas classes presentes, resultando em uma perda de acurácia na imagem temática produzida pelo classificador. Uma possível solução para estes casos consiste na utilização de classificadores em estágio múltiplo. Nesse caso, um único subconjunto de classes é tratado a cada etapa, o que permite uma seleção mais adequada das variáveis a serem utilizadas no processo de classificação.

## 2. Metodologia

Neste estudo, é proposto um classificador em múltiplo estágio, estruturado como árvore binária (Breiman *et al*, 1984). Em uma árvore binária, cada nó dá origem a dois nós descendentes, cada um caracterizado individualmente pelos parâmetros de uma única classe. O classificador inicia pelo nó raiz, no qual fica alocada a totalidade das classes em consideração, representadas na etapa inicial de construção do classificador, pelas respectivas amostras de treinamento. Um critério para estimação da separação entre classes é adotado para definir individualmente em cada nó da árvore o par de classes que apresenta a separação máxima entre si. As duas classes assim selecionadas darão origem aos dois nós descendentes. Os parâmetros associados a cada um dos nós descendentes são estimados a partir das amostras de treinamento de cada uma destas duas classes, e serão posteriormente utilizados no processo de classificação. Este processo continua até que se atinja os nós terminais, isto é, nós ocupados por uma única classe. Completada a etapa de treinamento do classificador, o processo de classificação pode ser iniciado, com cada pixel individual a ser rotulado, entrando no classificador pelo nó raiz e percorrendo a estrutura em árvore até atingir um nó terminal, o qual vai definir o rótulo a ser atribuído ao pixel. O critério de separabilidade entre as classes adotado neste estudo é a distância de Bhattacharyya, a qual tem expressão geral definida em Fukunaga (1990):

$$B = -\ln \left[ \int_{-\infty}^{+\infty} \sqrt{p(x/\omega_1) p(x/\omega_2)} dx \right] \quad (1)$$

Sendo  $p(x/\omega_1)$  e  $p(x/\omega_2)$  as probabilidades *a posteriori* associadas as classes  $\omega_1$  e  $\omega_2$  respectivamente. Supondo-se uma distribuição normal para os dados, a distância de Bhattacharyya fica:

$$B = \frac{1}{8} (\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)^T \left( \frac{\boldsymbol{\Sigma}_1 + \boldsymbol{\Sigma}_2}{2} \right)^{-1} (\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2) + \frac{1}{2} \ln \left( \frac{|(\boldsymbol{\Sigma}_1 + \boldsymbol{\Sigma}_2)/2|}{|\boldsymbol{\Sigma}_1|^{1/2} |\boldsymbol{\Sigma}_2|^{1/2}} \right) \quad (2)$$

Sendo  $\boldsymbol{\mu}_1$  e  $\boldsymbol{\mu}_2$  os vetores de médias das classes  $\omega_1$  e  $\omega_2$  respectivamente,  $\boldsymbol{\Sigma}_1$  e  $\boldsymbol{\Sigma}_2$  as matrizes de covariância. A distância de Bhattacharyya é uma medida bastante conveniente para estimação da separabilidade entre classes. As duas parcelas em (2) podem ser facilmente

interpretadas: a primeira estima a contribuição dos vetores de médias para a separabilidade entre as classes e a segunda a contribuição das matrizes de covariância.

## 2.1 Extração de variáveis

O método utilizado para extração de variáveis nesse estudo implementa em cada nó da árvore binária, a otimização da distância de Bhattacharyya (2). Entretanto, esse processo não se constitui em uma tarefa trivial, sendo soluções sub-ótimas geralmente adotadas: otimização apenas da primeira parcela (vetores de médias), otimização apenas da segunda parcela (matrizes de covariância), otimização de ambas as parcelas, mas na suposição de que a contribuição das médias é predominante, e a última, considerando que a contribuição das covariâncias é predominante. Uma descrição mais completa destes processos pode ser encontrada em Fukunaga (1990). O caso aplicado nesse estudo utiliza a suposição de que a contribuição dominante é das matrizes de covariâncias.

Representando por  $\mathbf{X}$  o espaço dos dados originais e por  $\mathbf{Y}$  o espaço dos dados transformados busca-se uma transformação que leve de  $\mathbf{X}$  para  $\mathbf{Y}$  sob a condição de maximização da distância de Bhattacharyya entre duas classes  $\omega_1$  e  $\omega_2$  na suposição de que a contribuição dominante à separação entre as classes é devida à diferença entre as matrizes de covariância. Pode-se demonstrar, que nestas condições, os eixos em  $\mathbf{Y}$  são definidos pelos autovetores de  $\Sigma_{2X}^{-1} \Sigma_{1X}$ , onde  $\Sigma_{1X}$  e  $\Sigma_{2X}$  representam as matrizes de covariância de  $\omega_1$  e  $\omega_2$  no espaço original  $\mathbf{X}$  (Fukunaga, 1990). Representando por  $\Phi$  e  $\Lambda$  respectivamente as matrizes de autovetores e autovalores de  $\Sigma_{2X}^{-1} \Sigma_{1X}$  a transformação do espaço  $\mathbf{X}$  para o espaço  $\mathbf{Y}$  é implementada por:

$$\mathbf{Y} = \Phi^T \mathbf{X}$$

No espaço  $\mathbf{Y}$ , a distância de Bhattacharyya fica:

$$B = \sum_{i=1}^n \left[ \frac{1}{4} \frac{\{\varphi_i^T (\boldsymbol{\mu}_2 - \boldsymbol{\mu}_1)\}^2}{1 + \lambda_i} + \frac{1}{4} \left\{ \ln \left( \lambda_i + \frac{1}{\lambda_i} + 2 \right) - \ln 4 \right\} \right] \quad (3)$$

sendo  $n$  a dimensão dos dados. Pode-se então selecionar como as  $m$  variáveis extraídas desejadas, selecionando-se os  $m$  autovetores de  $\Sigma_{2X}^{-1} \Sigma_{1X}$  correspondentes as  $m$  maiores parcelas no somatório em (3). Esta é, portanto, a abordagem adotada neste estudo.

## 2.2 Seleção de variáveis

Para fins de comparação, a redução na dimensionalidade dos dados foi também implementada em cada nó do classificador via seleção de variáveis. Com esta finalidade foi utilizado o método conhecido como *Sequential Forward Selection* (SFS) (Serpico et al., 2003). Neste método, parte-se do conjunto das variáveis originais  $\mathbf{X}$ , com dimensionalidade  $n$  e de um conjunto  $S$ , inicialmente vazio ( $S=\emptyset$ ), que irá conter as variáveis selecionadas. A cada etapa do processo iterativo, uma variável em  $\mathbf{X}-S$ , aquela que maximiza um critério de separabilidade, é acrescentada a  $S$ . O processo continua até que o número de variáveis em  $S$  atinja o valor desejado  $m$  ( $m < n$ ). Neste estudo adotou-se como critério para estimar a separabilidade entre as duas classes, a distância de Jeffries-Matusita. Deve-se observar, entretanto, que nesta metodologia uma variável de  $\mathbf{X}$  uma vez inserida em  $S$ , não tem possibilidade de ser removida, o que torna este método sub-ótimo. Neste estudo a abordagem SFS é aplicada em cada nó individual do classificador na seleção das  $m$  variáveis em  $S$  mais adequadas à separação do par de classes em consideração.

### 3 Experimentos

#### 3.1 Imagem AVIRIS

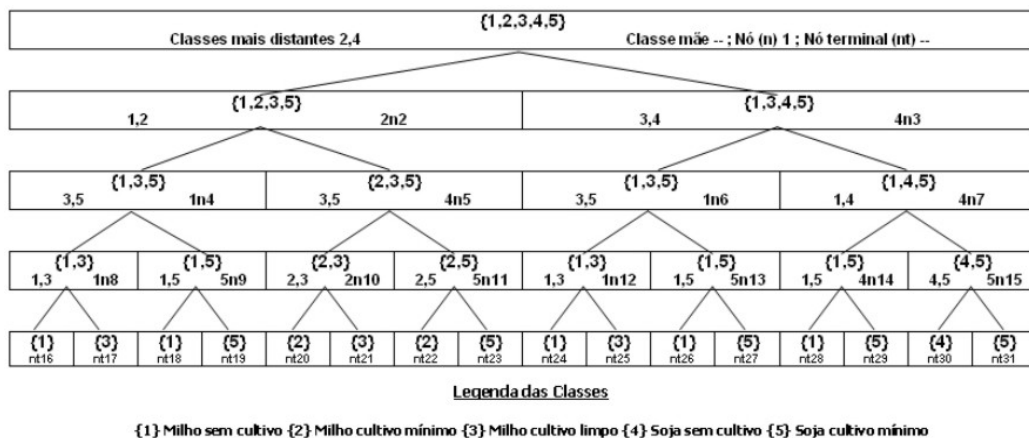
Os experimentos foram realizados utilizando-se dados imagem coletados pelo sensor AVIRIS, que possui 224 bandas espectrais de largura 10nm, cobrindo toda a região do visível até o infra-vermelho termal. Das 224 bandas, foram excluídas aquelas ruidosas, devido à absorção causada pelo vapor de água na atmosfera, restando um total de 190 bandas.

A imagem foi tomada em uma região agrícola no estado de *Indiana*, EUA, com dados de verdade terrestre disponíveis. Em especial, as cinco classes de vegetação selecionadas foram escolhidas por apresentarem alta semelhança espectral constituindo-se, portanto, em um desafio maior para o classificador. As classes se constituem de três tipos de manuseio do solo para o plantio de milho e dois tipos de manuseio para o plantio de soja.

Todas as amostras de treinamento do classificador e as amostras de teste foram tomadas com igual tamanho em cada experimento. Assim, os experimentos foram realizados em várias etapas, utilizando amostras com tamanhos iguais a 200, 350, 500, 650 e 800, de forma a observar o comportamento dos classificadores em diferentes níveis. Todos os algoritmos para os experimentos foram programados em MATLAB<sup>®</sup>, obtendo-se os resultados apresentados a seguir.

#### 3.2 Árvore de decisão binária

Inicialmente, procurou-se identificar a estrutura do classificador em árvore binária (CDA) maximizando a estabilidade do processo de classificação e a acurácia dos resultados. As estruturas em árvore são geradas a partir das amostras de treinamento, partindo-se do nó raiz, de número 1, estendendo-se através dos ramos até os nós terminais.



**Figura 1. Classificador de decisão em árvore**

Moraes e Haertel (2005) investigaram uma estrutura binária para o CDA, que mostrou uma maior estabilidade nos resultados de classificação das amostras de teste através do maior número possível de comparações entre os pares de classes. Os autores mostraram que esta estrutura fornece também o maior grau de acurácia média obtida entre todas as classes (Figura 1). Essa estrutura do CDA foi igualmente utilizada nos experimentos hierárquicos envolvendo o método SFS e o método ODB, sendo designados respectivamente por SFS-CDA e ODB-CDA.

Especificamente para o método ODB, a alta correlação verificada entre bandas espectrais adjacentes ( $r \approx 0,99$ ) permitiu uma redução inicial na dimensão dos dados, por meio de uma seleção sistemática de bandas, contribuindo desta forma para uma estimação mais confiável

dos parâmetros requeridos nesse método. Desta forma, a dimensão inicial dos dados foi reduzida para 95 bandas, o que permitiu uma estimação mais precisa dos vetores de médias e das matrizes de covariâncias, aumentando conseqüentemente a acurácia final de classificação, além de reduzir expressivamente o tempo de processamento. Essa seleção sistemática foi implementada exclusivamente nos experimentos ODB-CDA, sendo assim considerada como uma modificação do algoritmo ODB-CDA original. Desse modo, esse algoritmo modificado é denominado de ODB-CDA com Seleção Sistemática de bandas (ODB-CDA-SS).

### 3.3 Acurácia média dos classificadores

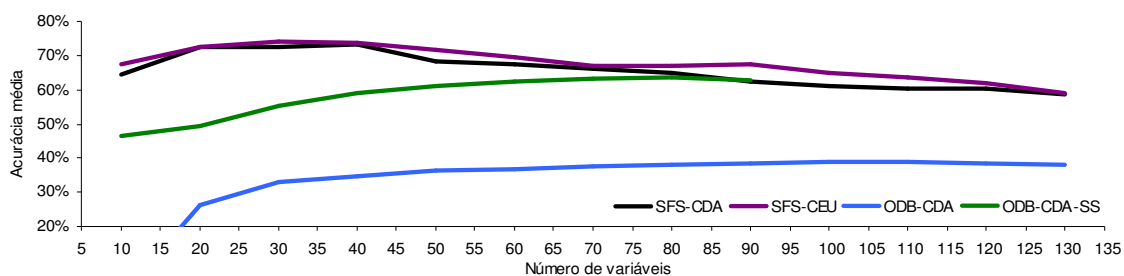
Para comparação dos métodos hierárquicos foi utilizado o método *Sequential Forward Selection* com a abordagem de Classificação em Estágio Único (SFS-CEU), e assim foram obtidas as acurácias médias dos quatro métodos de classificação estudados (SFS-CEU, SFS-CDA, ODB-CDA e ODB-CDA-SS). Para todos os métodos, foram empregados diferentes números de amostras de treinamento e teste. Também para todos os métodos, a seleção/extração de variáveis foi limitada entre 10 e 130, com intervalos de 10 variáveis entre cada incremento, com exceção do método ODB-CDA-SS, o qual pode extrair no máximo 90 variáveis. Os experimentos limitaram-se a este espectro de variáveis devido ao fato de que o tempo total de processamento cresce exponencialmente e também porque os efeitos do fenômeno de Hughes tornam-se cada vez mais acentuados a partir de 130 variáveis, não adicionando assim nenhuma contribuição para o acréscimo na acurácia média final de classificação.

Conforme mencionado, em cada nó do CDA, a seleção do par de classes que irá caracterizar os dois nós descendentes é realizada neste estudo pelo critério da distância de Bhattacharyya máxima. A estimação desta distância emprega a dimensão total dos dados (190 bandas espectrais). O experimento com 200 amostras de treinamento para cada classe (Figura 2), constituiu-se, portanto, em um desafio maior para os classificadores. Nesse particular experimento, as técnicas SFS-CEU e SFS-CDA apresentaram diferenças significativas (para um  $p\text{-value} < 0,001$ ) entre os valores estimados para a acurácia fornecida pelos vários métodos para diferentes valores na dimensionalidade dos dados (variáveis). Pode-se notar que, embora as curvas de acurácia sejam muito próximas, a acurácia no método SFS-CEU é superior ao método SFS-CDA para qualquer número de variáveis utilizado. Ambas as técnicas mostram inicialmente um acréscimo na acurácia média de classificação na medida em que novas bandas são selecionadas, atingindo um pico entre 20 e 40 bandas. A partir deste ponto, os efeitos do fenômeno de Hughes começam a aparecer, resultando em um decréscimo na acurácia final com o aumento da dimensionalidade.

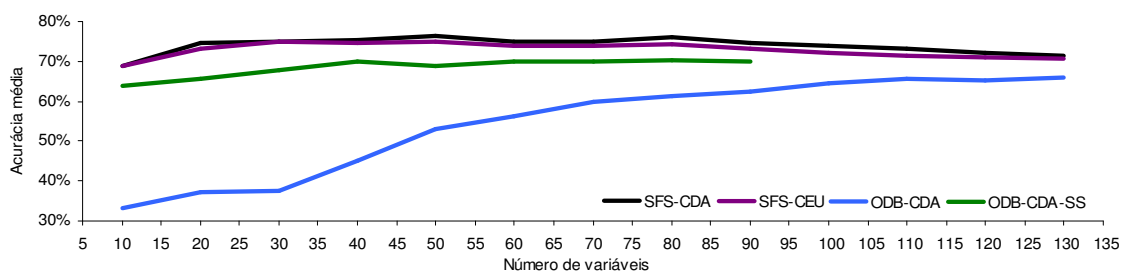
A técnica ODB-CDA, para esse tamanho de amostra, apresentou nos experimentos, o pior desempenho entre todos os métodos testados. Esse comportamento se deve principalmente ao fato de que o número de amostras disponível é muito pequeno comparado com o número de parâmetros a serem estimados. Com o algoritmo modificado, isto é, com a pré-seleção sistemática de metade das bandas originais, estas dificuldades ficam reduzidas. Este fato deve-se ao aumento na confiabilidade dos valores estimados para os parâmetros utilizados pelo método ODB, resultando em um aumento no valor estimado para a acurácia média produzida pelo classificador ODB-CDA-SS. Conforme comentado anteriormente, nota-se que no método ODB-CDA-SS, o número total de bandas passa a ser igual a 95 e por essa razão o número de variáveis extraídas fica limitado entre 10 e 90 variáveis.

O segundo conjunto de testes (Figura 3) utiliza 350 amostras de treinamento para cada classe. Nesse teste, observa-se uma inversão entre o comportamento da acurácia média produzida pelos dois métodos que utilizam o SFS. Aqui, a utilização da técnica para seleção de variáveis (SFS) no classificador em árvore (SFS-CDA) mostrou resultados superiores à

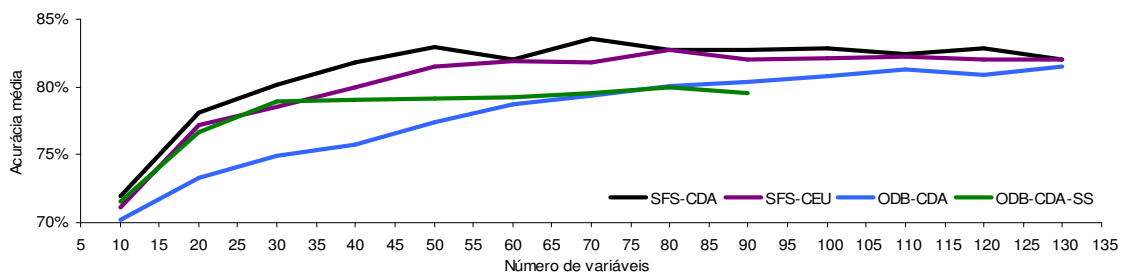
aqueles obtidos pelo classificador em estágio único (SFS-CEU) em praticamente todo o espectro de variáveis analisado, sendo significativamente superior ( $p\text{-value} < 0001$ ) no teste envolvendo a comparação entre médias pareadas. Com relação às técnicas que utilizam a otimização da distância de Bhattacharyya, nota-se um salto significativo na acurácia média de classificação para valores baixos na dimensionalidade dos dados (10 variáveis), passando de aproximadamente zero para 33,2% com o método ODB-CDA e de 46,6% para 63,8% com o método ODB-CDA-SS.



**Figura 2. Acurácia média dos classificadores – 200 amostras por classe**



**Figura 3. Acurácia média dos classificadores – 350 amostras por classe**



**Figura 4. Acurácia média dos classificadores – 800 amostras por classe**

Com 500 amostras para cada classe, notou-se que as acurácias médias produzidas pelos métodos que utilizam seleção de variáveis apresentam valores estimados praticamente idênticos para a acurácia média, não apresentando diferenças significativas no teste  $t$  para amostras pareadas ao nível de 5% de significância. Este fato sugere que as diferenças constatadas anteriormente com menos de 500 amostras por classe podem ser o resultado de flutuações aleatórias entre as classes envolvidas.

Os experimentos empregando número de amostras maiores, ilustrados na Figura 4 para 800 amostras por classe, mostram que o comportamento médio do método SFS-CDA é superior ao método SFS-CEU ( $p\text{-value} < 0,0001$ ), sendo a diferença entre as técnicas cada vez maior a favor do classificador em árvore, principalmente quando a dimensão usada para classificação é pequena, inferior a 60, e diminuindo na medida em que novas bandas são adicionadas.

Aumentando-se o número de amostras de treinamento, nota-se que os resultados obtidos empregando as técnicas para extração de variáveis tendem a se aproximar daqueles obtidos quando se utilizam técnicas para seleção de variáveis, atingindo resultados próximos, no patamar mínimo de 70%, quando são utilizadas 800 amostras por classe e um número de variáveis igual a 10. Nota-se também que o método ODB é menos sensível ao efeito do fenômeno de Hughes do que o método SFS, apresentando um comportamento crescente na acurácia média final de classificação na medida em que novas variáveis são adicionadas ao classificador.

### 3.4 Tempo de processamento dos classificadores

Os experimentos desenvolvidos nesse estudo mostraram diferenças significativas entre os métodos investigados, em termos de desempenho computacional. Para a comparação do desempenho dos métodos investigados, são apresentados a seguir os resultados para o maior número de amostras utilizado, isto é, amostras iguais a 800 para todas as classes. Para números inferiores de amostras, o desempenho dos algoritmos é similar, sendo reduzido por um fator de escala constante.

A Figura 5 apresenta o tempo total de processamento para treinamento e teste dos classificadores dos quatro métodos estudados. Nessa figura, observa-se a superioridade do desempenho da ODB em relação aos métodos que utilizam o SFS. Esse comportamento é esperado, visto que a ODB aplica em cada nó da árvore uma única transformação linear nas bandas originais enquanto o método SFS precisa varrer todo o conjunto de bandas seqüencialmente, em busca daquele que mais contribui à separação das classes. Mantendo-se as devidas proporções devido à quantidade original de bandas haver sido reduzida pela metade, o algoritmo modificado ODB-CDA-SS é superior ao método ODB-CDA. Observou-se que a ODB apresenta praticamente o mesmo comportamento em ambos os casos, crescendo exponencialmente, contudo mais lentamente do que os métodos de classificação que utilizam seleção de variáveis.

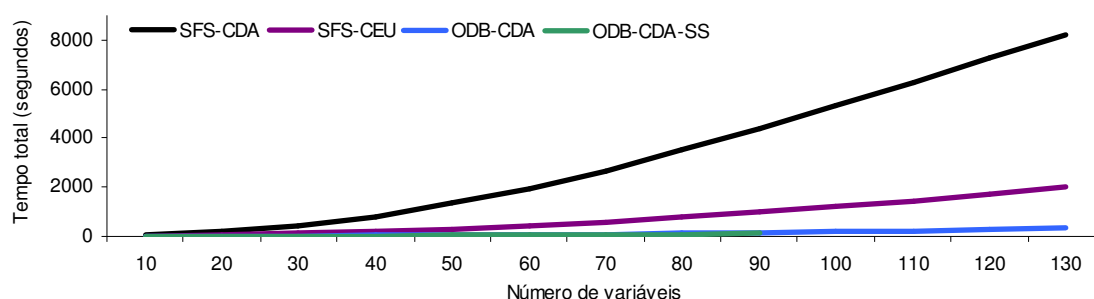


Figura 5. Tempo de processamento dos classificadores

## 4. Conclusões

Nesse trabalho, foram realizados experimentos para classificação de dados imagem em alta dimensionalidade (hiper-espectrais) testando-se quatro diferentes algoritmos. Foram estudados métodos de classificação baseados em seleção de variáveis em estágio único (SFS-CEU) e seleção de variáveis em estágio múltiplo através de um classificador de decisão com estrutura em árvore (SFS-CDA). Também foi utilizada uma técnica para extração de variáveis baseada na otimização da distância de Bhattacharyya em um classificador de decisão em árvore (ODB-CDA) e a mesma técnica com pré-seleção sistemática de bandas (ODB-CDA-SS).

Para as técnicas baseadas em seleção de variáveis (SFS), os experimentos evidenciaram que na medida em que o tamanho das amostras de treinamento aumenta, o classificador de decisão em árvore (CDA) tende a apresentar uma acurácia média de classificação superior à acurácia obtida pelo método tradicional do classificador em estágio único (CEU).

Os métodos que utilizam extração de variáveis baseados na ODB foram inferiores aos métodos baseados no algoritmo SFS. Os experimentos desenvolvidos utilizando o algoritmo modificado ODB-CDA-SS mostraram que a principal dificuldade do uso da ODB reside na baixa confiabilidade dos valores estimados para os parâmetros, devido a um número limitado de amostras de treinamento disponíveis. Assim, foi observado que na medida em que o tamanho amostral aumenta, ocorre um aumento progressivo na acurácia produzida pelo método ODB, apresentando inclusive uma menor sensibilidade ao efeito do fenômeno de Hughes.

Um outro aspecto a ser mencionado, é a suposição de normalidade multivariada dos dados para a ODB. Embora esta suposição seja geralmente aceita para dados de imagens digitais obtidas por sensores remotos (Richards e Jia, 1998), ao analisar individualmente as distribuições das classes nas bandas originais, verifica-se que, pelos resultados dos testes para verificar normalidade de dados (Lilliefors, Jarque-Bera, Kolmogorov-Smirnov) praticamente nenhuma banda atende individualmente esta suposição. Com isso, conclui-se que a utilização da ODB para extração de variáveis sob a suposição de normalidade pode estar sendo seriamente prejudicada pela violação dessa hipótese.

Vale ainda salientar que a ODB escolhida é apenas um dos casos sub-ótimos, o qual utiliza a contribuição predominante das matrizes de covariância. Em um outro estudo, Tadjudin (1998) mostra que outra opção de otimização sub-ótima, que utiliza a contribuição predominante dos vetores de médias ao invés das matrizes de covariância, apresenta resultados mais favoráveis para a ODB, inclusive superiores aos outros métodos utilizados para comparação naquele estudo.

Dessa forma, propõe-se para estudos futuros, que todas as opções para a otimização da distância de Bhattacharyya sejam utilizados para classificar os mesmos conjuntos de classes, a fim de comprovar e comparar consistentemente os resultados obtidos nesses trabalhos.

## Referências

- Aeberhard, S., Coomans, D., De Vel, O. Comparative analysis of statistical pattern recognition methods in high dimensional settings. **Pattern Recognition**. v. 27, n. 8, p. 1065-1077, 1994.
- Breiman, L., Friedman, J. H., Olshen, R. A., and Stone, C. J. **Classification and regression trees (CART)**. Belmont, CA: Wadsworth Int., 1984. 358 p.
- Fukunaga, K. **Introduction to statistical pattern recognition**. 2nd. Ed. Boston: Academic Press, 1990. 591 p.
- Friedman, J. H. Regularized discriminant analysis, **Journal of the American Statistical Association**, v. 84, n. 405, p. 165-175, 1989.
- Jackson, Q., Landgrebe, D. An adaptive classifier design for high-dimensional data analysis with a limited training data set. **IEEE Transactions on Geoscience and Remote Sensing**, v. 39, p. 2664-2679, 2001.
- Moraes, D. A. O., Haertel, V. **Extração de feições em dados imagem com alta dimensão por otimização da distancia de Bhattacharyya em um classificador de decisão em árvore**. Dissertação (Mestrado em Sensoriamento Remoto) - PPGSR, Universidade Federal do Rio Grande do Sul, 99 p., 2005.
- Richards, J. A. and Jia, X. **Remote sensing digital image analysis: an introduction**, 3<sup>rd</sup>. ed. Springer, 1998.
- Shashahani, B.; Landgrebe, D. The effect of unlabeled samples in reducing the small sample size problem and mitigating the hughes phenomenon. **IEEE Transactions on Geoscience and Remote Sensing**, v. 32, n. 5, p. 1087-1095, 1994.
- Tadjudin, S. **Classification of high-dimensional data with limited training samples**, Ph.D. dissertation and Tech. Rep. TR-ECE-98-8, School of Elec. Eng. Purdue University, W. Lafayette, IN, 1998.