

EXTRAÇÃO DE FEIÇÕES EM DADOS DE IMAGENS HIPERESPECTRAIS POR OTIMIZAÇÃO DA DISTÂNCIA DE BHATTACHARYYA EM UM CLASSIFICADOR ÁRVORE DE DECISÃO

*On the feature extraction in hyperspectral data image using the optimization of
bhattacharyya distance in a decision tree classifier*

DENIS ALTIERI DE OLIVEIRA MORAES

VITOR HAERTEL

Universidade Federal do Rio Grande do Sul

Centro Estadual de Pesquisas em Sensoriamento Remoto e Meteorologia

Caixa Postal 15044 - CEP 91501-970 - Porto Alegre – RS, Brasil

d_altieri@yahoo.com.br, vitor.haertel@ufrgs.br

RESUMO

Neste estudo é investigado um método para fins de classificação de dados imagem em alta dimensionalidade, por meio de um classificador em estágio múltiplo. Esse classificador, estruturado na forma de árvore binária, emprega um número reduzido de feições em cada nó visando reduzir os efeitos do fenômeno de Hughes. O método para redução de feições consiste na otimização da distância de Bhattacharyya em cada nó individual da árvore. O classificador hierárquico estruturado em árvore binária oferece as condições adequadas para implementação deste método, pois a distância de Bhattacharyya está definida para um par de classes, permitindo desta forma a extração de um subconjunto ótimo de feições em cada nó individual. Foram desenvolvidos experimentos empregando dados da imagem AVIRIS, envolvendo diferentes números de amostras de treinamento e de feições extraídas. Os resultados mostraram que sob certas condições a metodologia investigada apresenta resultados mais acurados do que aquelas mais comumente utilizadas para fins de redução na dimensionalidade dos dados.

Palavras chaves: Reconhecimento de Padrões; Imagem Hiperespectral; Árvores de Decisão; Distância de Bhattacharyya.

ABSTRACT

In this paper we investigate a method for the classification of high-dimensional image data using a multi-stage classifier structured as a binary tree, and employing a reduced number of features at each node, in order to mitigate the Hughes phenomenon. The selected method for feature reduction is based on the optimization of Bhattacharyya distance at each individual node of the tree. As the Bhattacharyya distance is defined for a pair of classes, the binary tree approach allows the extraction of an optimal sub-set of features at each individual node. Experiments were performed using an AVIRIS image data set, varying the number of training samples and also the number of selected features at each node. The results have shown an improvement in the accuracy of the thematic image, as compared to more traditional methods for feature selection and extraction.

Keywords: Pattern Recognition; Hyperspectral Image; Decision Trees; Bhattacharyya Distance.

1. INTRODUÇÃO

Os classificadores mais tradicionalmente utilizados no processo de classificação de imagens digitais em sensoriamento remoto tem sido os do tipo estágio único. Por outro lado, com o advento de dados imagem em alta dimensionalidade (imagens hiperespectrais) tornou mais evidente o problema ocasionado pela necessidade de estimar-se um número crescente de parâmetros a partir de um número geralmente limitado de amostras de treinamento. Nestes casos, observa-se que a acurácia obtida no processo de classificação inicialmente aumenta, na medida em que novas bandas espectrais, i.e., informação adicional é acrescentada ao classificador. Neste processo a acurácia obtida atinge eventualmente um máximo para em seguida voltar a diminuir, na medida em que novas bandas espectrais são acrescentadas. Esta característica, freqüentemente observada nos processos de classificação utilizando dados de imagens em alta dimensionalidade, é conhecida como o *Fenômeno de Hughes* e resulta da dificuldade em se obter uma estimativa confiável para os parâmetros do classificador a partir de um número geralmente limitado de amostras de treinamento (HUGHES, 1968).

Métodos para amenizar os efeitos do Fenômeno de Hughes vem sendo objeto de estudo por vários autores. Métodos, como análise discriminante regularizada, foram investigados por alguns autores, mostrando resultados promissores (FRIEDMAN, 1989, AEBERHARD *et al*, 1994). Métodos implementando o conceito de amostras semi-rotuladas foram propostos por SHASHAHANI e LANDGREBE (1994) e JACKSON e LANDGREBE (2001). Uma terceira alternativa consiste na redução da dimensionalidade dos dados, minimizando a perda de informação. Neste contexto, duas abordagens principais vem sendo objeto de estudo: *seleção de feições* e *extração de feições* (RICHARDS,

1998; SERPICO *et al*, 2003; WEBB, 2002; THEODORIDIS, 2003). No primeiro caso (seleção de feições), busca-se encontrar um subconjunto de bandas espectrais que ofereça boa separabilidade entre as classes envolvidas. Desta forma, o processo de seleção de feições é geralmente implementado por meio da otimização de um critério de separabilidade entre as classes. No processo de extração de feições, a redução na dimensionalidade dos dados é obtida aplicando-se transformações, lineares ou não, aos dados originais e selecionando-se neste espaço transformado as feições que proporcionam uma maior separabilidade entre as classes em consideração.

Os processos tanto de seleção de feições como os de extração de feições são classe-dependentes, isto é, as feições ótimas dependem do conjunto das classes sendo analisadas. Neste particular, classificadores árvore de decisão (CAD) são potencialmente mais vantajosos com relação aos classificadores em estágio único (CEU) (SAFAVIAN e LANDGREBE, 1991) pelas seguintes razões: 1 - Os CAD's processam subconjuntos de classes a cada estágio em lugar de processar o conjunto completo de classes em um único estágio, como ocorre na metodologia dos CEU's; 2 - A presença de um subconjunto de classes (e não da totalidade das classes) a cada estágio permite uma escolha mais adequada com relação às feições a serem adotadas; 3 - No caso de um CAD estruturado em forma binária, somente duas classes são consideradas em cada nó, o que permite uma otimização do processo de seleção/extração de feições em cada nó individual.

2. METODOLOGIA

Neste estudo é investigado um CAD com estrutura binária (BREIMAN *et al*, 1984). Nesse classificador, cada estágio de classificação constitui-se em um nó da árvore, e cada nó originando dois nós descendentes. Nesta abordagem, em cada nó da árvore binária é selecionado um par de classes que irá definir os parâmetros associados a cada um dos dois nós descendentes. As duas classes selecionadas são aquelas que apresentam a maior separabilidade entre si. Neste estudo o critério adotado para estimar a separabilidade entre as duas classes integrantes de um par é a distância de Bhattacharyya (B). A expressão geral da distância de Bhattacharyya entre duas classes ω_1 e ω_2 é dada por (THERRIEN, 1989, FUKUNAGA, 1990):

$$B = -\ln \left[\int_{-\infty}^{\infty} \sqrt{P(\mathbf{X}/\omega_1)P(\mathbf{X}/\omega_2)} d\mathbf{X} \right] \quad (1)$$

Onde, $P(\mathbf{X}/\omega_i)$ é a probabilidade *a priori* do pixel \mathbf{X} , dado que pertence à classe i . Para dados normalmente distribuídos a expressão (1) torna-se (MORAES e HAERTEL, 2005):

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

sendo $\boldsymbol{\mu}_1$ e $\boldsymbol{\mu}_2$ os vetores de médias das classes ω_1 e ω_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 pares de classes. Para dados normalmente distribuídos essa distância fornece um limite superior para o erro de Bayes (FUKUNAGA, 1990). Em (2) a primeira parcela estima a contribuição dos vetores de médias para a separabilidade entre as classes e a segunda a contribuição das matrizes de covariância.

Uma vez definido o par de classes que irá gerar os dois nós descendentes, o passo seguinte consiste em construir as duas funções de decisão, definidas pelos parâmetros das classes integrantes do par selecionado no nó ascendente. Tais funções de decisão serão utilizadas para fins de classificação das amostras de treinamento das demais classes presentes neste nó, em um dos dois nós descendentes. Neste estudo foi utilizada a função de decisão da Máxima Verossimilhança Gaussiana (MVG), a qual requer a estimação dos parâmetros vetor de médias ($\boldsymbol{\mu}_1, \boldsymbol{\mu}_2$) e matriz de variâncias e covariâncias ($\boldsymbol{\Sigma}_1, \boldsymbol{\Sigma}_2$), associados a cada uma das duas classes.

2.1 Otimização da distância de Bhattacharyya

Nessa fase, entra em questão o problema da redução na dimensionalidade dos dados, isto é, do número de feições a serem utilizadas nas funções de decisão para o processo de classificação. Conforme mencionado anteriormente, este passo é necessário para minorar os efeitos do fenômeno de Hughes na acurácia da imagem classificada. Nesse estudo foi adotado um método para redução na dimensionalidade dos dados via extração de feições. Neste processo, os dados originais (\mathbf{X}) são mapeados por meio de transformações, lineares ou não, em um outro espaço (\mathbf{Y}), no qual a maior parte da informação do espaço original concentra-se em um número menor de feições extraídas, possibilitando desta forma uma redução na dimensionalidade dos dados, sem que se incorra em uma perda significativa de informação. Neste caso, o critério da transformação adotada consistiu na extração de feições que otimizam a separação entre as duas classes, adotando-se a distância de Bhattacharyya como critério de separabilidade. Essa transformação pode ser representada por:

$$\mathbf{Y} = \mathbf{A}\mathbf{X}$$

Sendo \mathbf{A} , a matriz de transformação que concentra a separação entre as duas classes, conforme estimado pela distância de Bhattacharyya, em um número menor de feições em \mathbf{Y} .

A implementação desta transformação implica, portanto, em um processo

de otimização da distância de Bhattacharyya (ODB). Entretanto, devido à forma da expressão que define a distância de Bhattacharyya (2), tal processo de otimização não se constitui em uma tarefa trivial, sendo necessário adotar-se abordagens sub-ótimas, no sentido de que as duas parcelas em (2) não são otimizadas simultaneamente (MORAES e HAERTEL, 2005; FUKUNAGA, 1990):

- a) Otimização unicamente em função das matrizes de covariância (Σ_1, Σ_2). Neste caso se está assumindo igualdade entre os vetores de médias ($\mu_1 = \mu_2$);
- b) Otimização unicamente em função dos vetores de médias (μ_1, μ_2). Neste caso se está assumindo igualdade entre as matrizes de covariância ($\Sigma_1 = \Sigma_2$);
- c) Otimização conjunta, em função dos vetores de médias e das matrizes de covariância, com peso maior à diferença entre os vetores de médias (μ_1, μ_2);
- d) Otimização conjunta, em função dos vetores de médias e das matrizes de covariância, com peso maior à diferença entre as matrizes de covariância (Σ_1, Σ_2).

Uma descrição mais completa destes processos pode ser encontrada em FUKUNAGA (1990). Em TADJUDIN e LANDGREBE (1998) o critério adotado foi o da otimização conjunta com peso maior atribuído à diferença entre os vetores de médias. Neste estudo o critério de otimização testado é aquele que também leva em consideração as duas parcelas em (2), mas assume que a contribuição predominante é aquela devida à diferença entre as matrizes de covariância. A razão desta escolha reside na natureza do problema em estudo. A presença de classes espectralmente muito semelhantes entre si, isto é, com vetores de médias muito próximas, normalmente resulta em uma baixa acurácia no processo de classificação quando são utilizados dados em baixa ou média resolução espectral, como, por exemplo, dados imagem provenientes dos sensores Landsat-TM ou SPOT. Pode-se mostrar, entretanto, que é possível separar com boa acurácia classes espectralmente muito semelhantes entre si, ou mesmo idênticas, isto é, classes com vetores de médias muito semelhantes ou mesmo iguais entre si, desde que se utilizem dados em dimensionalidade alta e que as matrizes de covariância das classes (FUKUNAGA, 1990) sejam suficientemente distintas entre si. Desta forma, neste estudo implementou-se a ODB que prioriza a 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 Σ_{1X} e Σ_{2X} representam as matrizes de covariância de ω_1 e ω_2 respectivamente, no espaço original (\mathbf{X}) (FUKUNAGA, 1990). Representando por Φ e Λ as matrizes de autovetores e autovalores de

$\Sigma_{2X}^{-1} \Sigma_{1X}$, a transformação do espaço X para o espaço Y se dá por:

$$Y = \Phi^T X$$

Desse modo, MORAES e HAERTEL (2005) provam que no espaço Y , a distância de Bhattacharyya (2) é dada por:

$$B = \sum_{i=1}^n \left[\frac{1}{4} \frac{\{\Phi_i^T (\mu_2 - \mu_1)\}^2}{1 + \lambda_i} + \frac{1}{4} \left\{ \ln \left(\lambda_i + \frac{1}{\lambda_i} + 2 \right) - \ln 4 \right\} \right] \quad (3)$$

sendo n a dimensão dos dados originais (X).

Para se obter as m direções oferecendo a maior separação entre as duas classes ($m < n$) no espaço transformado (Y), pelo critério da ODB, podem então ser tomados os m autovetores de $\Sigma_{2X}^{-1} \Sigma_{1X}$ correspondentes as m maiores parcelas no somatório em (3). Esta é, portanto, a abordagem para redução de dimensionalidade através da ODB adotada neste estudo.

3. EXPERIMENTOS E ANÁLISE DOS RESULTADOS

3.1 Dados Imagem AVIRIS

Os experimentos foram realizados utilizando-se dados coletados pelo sensor AVIRIS, cobrindo uma área teste no estado de Indiana, Estados Unidos (Figura 1). Esta área apresenta campos agrícolas com culturas espectralmente muito semelhantes entre si, o que a torna adequada à investigação proposta neste estudo. Em especial, cinco classes foram consideradas nesse trabalho, as quais por apresentarem alta semelhança espectral entre si, se constituem em um desafio maior para o classificador. Estas classes caracterizam três diferentes tipos de manuseio do solo para o plantio de milho e dois tipos de manuseio para o plantio de soja. Nesse caso, classificadores empregando dados tradicionais em baixa ou média dimensionalidade como, por exemplo, dados Landsat-TM, irão fornecer resultados com baixo grau de acurácia. Uma separação entre estas cinco classes, com um razoável grau de exatidão poderá, entretanto, ser obtida utilizando-se dados espectrais em alta dimensionalidade e explorando as diferenças entre as matrizes de covariância.

Os dados imagem obtidos pelo sensor AVIRIS contém 224 bandas espectrais de largura 10nm, cobrindo o espectro eletromagnético nas regiões do visível, infravermelho próximo e infravermelho médio ($0.4\mu\text{m} - 2.4\mu\text{m}$) (LANDGREBE, 2003). Das 224 bandas originais, foram excluídas as ruidosas (vapor de água) restando 190 bandas. Além disso, foi verificado que para o caso dos dados utilizados, existe uma forte correlação positiva entre bandas adjacentes ($r \approx 0,99$), o que sugere a possibilidade de uma redução preliminar na dimensionalidade dos dados por meio de uma seleção sistemática de uma banda

espectral a cada intervalo de três, reduzindo a dimensão original dos dados para 64. Essa seleção inicial visa oferecer condições mais confiáveis para estimativas dos parâmetros requeridos para a transformação proposta (do espaço X para o espaço Y), dado o número limitado de amostras de treinamento.

Figura 1 - Imagem hiperespectral *AVIRIS*



Figura 2 - Verdade terrestre das classes



Nos experimentos realizados, foi tomado um número igual de amostras de treinamento e teste para todas as classes. Assim, os conjuntos testados foram iguais a 70, 150, 300 e 500 amostras para cada uma das classes em consideração. A Figura 2 ilustra a correspondente verdade terrestre da imagem hiperespectral em estudo. Para a implementação do experimento, os algoritmos foram programados em MATLAB[®], obtendo-se os resultados descritos na seção seguinte.

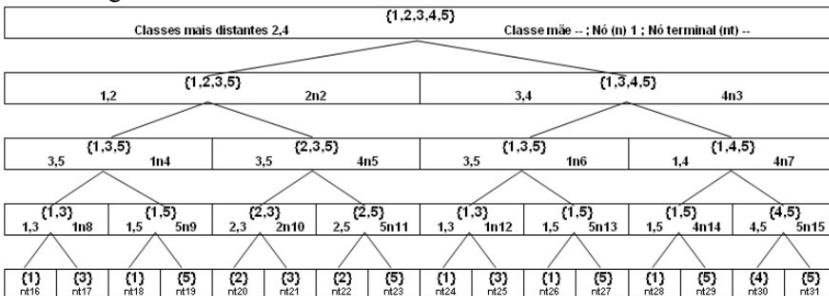
3.2 Classificador árvore de decisão em estrutura binária

Inicialmente, procurou-se identificar a estrutura do CAD binário, isto é, o número de estágios ou nós que forneceria um resultado mais acurado e estável. As árvores são criadas a partir das amostras de treinamento, partindo-se do nó raiz e estendendo-se através dos ramos até atingir os nós terminais. Em uma primeira

etapa, identifica-se em cada nó individual o par de classes que apresenta a maior separabilidade, conforme o critério da distância de Bhattacharyya. Esta estimativa é feita no espaço X, o qual nestes experimentos é composto por 64 bandas espectrais. O processo de extração de feições (espaço Y) com base na otimização da distância de Bhattacharyya é realizado então com base neste par de classes. Neste espaço de dimensão reduzida (Y) são estimados os parâmetros das duas classes (μ_{Y1}, μ_{Y2} e Σ_{Y1}, Σ_{Y2}) para construção das respectivas funções de decisão, e que servirão para classificar as amostras de treinamento presentes neste nó (nó ascendente), em um dos dois nós descendentes. A função decisão adotada neste estudo é a da Máxima Verossimilhança Gaussiana (MVG).

Neste processo, uma fração das amostras de treinamento pertencente a uma classe é geralmente classificada em um dos nós descendentes e o restante das amostras em outro nó descendente. Desta forma torna-se necessário estabelecer um critério para decidir em que condições uma determinada classe será integralmente atribuída a um único nó descendente e em que condições esta classe será atribuída a ambos os nós descendentes. Para esta finalidade é proposto neste estudo o conceito de *Limiar de Verossimilhança (LV)*. Este limiar serve para definir a fração mínima das amostras de treinamento que deve ser classificada em um dos dois nós descendentes, para que a classe em questão seja atribuída integralmente e unicamente a este nó. Caso o LV não seja alcançado, a classe em questão é atribuída igualmente a ambos os nós descendentes, nos quais o processo se repetirá até que sejam alcançados os nós terminais, os quais definirão os rótulos das classes aos pixels que serão classificados.

Figura 3 - Classificador árvore de decisão em estrutura binária



Legenda das Classes

{1} Milho sem cultivo {2} Milho cultivo mínimo {3} Milho cultivo limpo {4} Soja sem cultivo {5} Soja cultivo mínimo

Um exemplo da metodologia proposta está ilustrado na Figura 3. Nesta figura, em cada nó do CAD são apresentadas as informações relevantes do processo. Os números no primeiro conjunto {1, 2, 3, 4, 5} se referem à legenda numérica das classes a que pertencem as amostras nesse nível inicial da árvore. O

par “2,4” à esquerda, expressa o par de classes que apresenta a maior separação entre si. O número situado à direita do nó refere-se à classe cujos parâmetros (vetor de médias e matriz de covariância) foram utilizados na seleção das amostras que o compõe. O algarismo inicial à direita do nó, por exemplo, “4n3”, refere-se à classe cujos parâmetros foram utilizados no processo de seleção do subconjunto de classes pertencentes ao nó, nesse caso, a classe “4” (esta convenção não se aplica ao nó raiz porque este contém o conjunto completo de classes). O número que define o nó encontra-se indicado após a letra “n” (“n3”) e os nós terminais ficam identificados pelo símbolo “nt”.

Neste experimento foi empregado um LV igual a 1 (100%), o qual produz para os dados utilizados a maior estrutura possível e conseqüentemente um resultado mais acurado e estável possível, isto é, com menor variabilidade. Os experimentos que levaram a essa conclusão em relação ao LV e à estrutura ideal do CAD em estrutura binária foram demonstrados por MORAES e HAERTEL (2005).

Contudo, estudos mais recentes têm demonstrado que uma modificação no algoritmo do CAD para produzir uma árvore com o número máximo de decisões tem o mesmo efeito de se utilizar um LV igual a 100%. Esta modificação consiste em identificar o par de classes mais distante entre si em cada nó da árvore para compor respectivamente as classes descendentes e a seguir atribuir todos os pixels das demais classes diretamente aos dois nós descendentes. Este procedimento elimina a necessidade da utilização do LV, pois não requer a classificação das amostras de treinamento das classes restantes para identificar a qual dos nós descendentes estas devem ser atribuídas, e ao mesmo tempo aumenta significativamente o desempenho computacional do algoritmo proposto, tornando-o mais eficiente que a metodologia tradicional em estágio único, conforme será mostrado a seguir.

Os resultados dos experimentos em relação a acurácia de classificação e ao desempenho dos métodos estudados são descritos a seguir, onde em cada nó da árvore e para cada classe, foi empregado um número igual de amostras para fins de treinamento e para fins de teste.

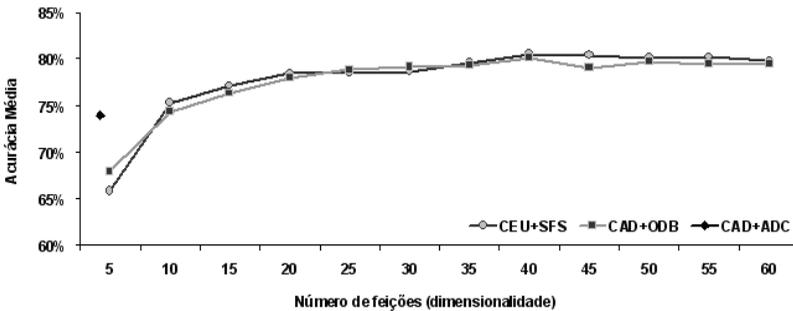
3.3 Acurácia média dos classificadores

Para comparação dos resultados obtidos, um CAD semelhante foi também utilizado em conjunto com o método para redução de dimensionalidade através da Análise Discriminante Canônica (ADC). Contudo, deve-se observar que devido às restrições impostas pela ADC, a acurácia de classificação desse método é representada apenas por um ponto. Esta restrição é devida ao fato de que o número de variáveis canônicas é igual ao número de classes menos um ou ao valor da dimensão dos dados menos um, o menor entre os dois casos. Assim, como nos experimentos foram empregadas cinco classes, o máximo de feições extraídas pela técnica CAD+ADC é igual a quatro. Desse modo, na fase de experimentos foram

estimadas as acurácias produzidas com os métodos para extração de feições ODB e ADC aplicados em conjunto com o CAD.

Para fins de comparação do CAD com métodos mais tradicionais de classificação em estágio único e redução da dimensionalidade dos dados, foi implementado também o método *Sequential Forward Selection* (SFS) em um Classificador em Estágio Único (CEU), sendo este procedimento designado por CEU+SFS. Dentre os experimentos realizados com número de amostras igual a 70, 150, 300 e 500, a Figura 4 ilustra apenas os resultados obtidos no melhor dos casos, sendo este o que utiliza o maior número de amostras para cada classe, isto é, conjuntos de amostras de treinamento e teste igual a 500.

Figura 4 - Acurácia média dos classificadores (500 amostras por classe)



A análise destes resultados permite as seguintes conclusões: a metodologia da ODB aplicada em conjunto com o classificador árvore de decisão em estrutura binária (CAD+ODB) produz resultados acurados, desde que sejam utilizadas aproximadamente oito vezes mais amostras de treinamento do que a dimensionalidade dos dados (bandas), conforme ilustrado na Figura 4. Ou seja, partindo-se de um total de 64 bandas originais e treinando-se o classificador com 500 amostras ou mais para cada classe, são obtidos resultados compatíveis com os métodos tradicionais de classificação. Esse fato deve-se principalmente ao fato de que a ODB proposta é baseada preponderantemente na diferença entre as matrizes de covariância das classes, Σ_1 e Σ_2 . Este fato torna necessário um número maior de amostras de treinamento para uma estimativa razoavelmente acurada desses parâmetros.

Destaca-se que, para o experimento empregando um maior número de amostras de treinamento (Figura 4), o conjunto CAD+ODB obteve um percentual médio de acurácia igual a 80,1%, para 40 feições em cada nó. A curva que descreve o comportamento da acurácia média neste processo em função da dimensionalidade dos dados apresenta o comportamento esperado, com um incremento inicial na medida em que a dimensionalidade dos dados aumenta e um decréscimo a partir de

um certo valor da dimensionalidade (40 neste experimento), caracterizando assim o fenômeno de Hughes. Utilizando-se amostras de treinamento para cada classe em número de 70, 150 e 300, os resultados obtidos através da técnica CAD+ODB resultaram em média inferiores aos obtidos através das técnicas tradicionais, mas recuperando-se na medida em que o número de amostras de treinamento cresce.

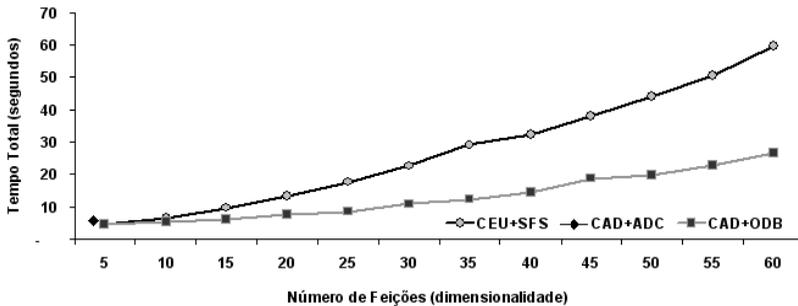
Em segundo lugar, o uso da metodologia conjunta da ADC aplicada com o classificador árvore de decisão (CAD+ADC) proporcionou uma acurácia média igual a 74,1%, utilizando um número decrescente de feições ao longo da árvore (4, 3, 2 e 1). Este valor é similar ao produzido pela metodologia CAD+ODB com dimensionalidade igual a 10, apresentando a vantagem de ser um pouco mais eficiente em termos de tempo de processamento que o primeiro (Figura 5). O uso da metodologia CAD+ADC mostra-se assim eficaz para casos quando se deseja que o número total de feições extraídas seja pequeno, ou seja, menor que o número de classes envolvidas.

O método de classificação em estágio único em conjunto com a técnica para redução de dimensionalidade *Sequential Forward Selection* (CAD+SFS) foi aplicado aos mesmos dados. Nesse processo, o critério de separação das classes foi a distância de *Jeffries-Matusita* (SERPICO *et al*, 2002), a qual é uma generalização da distância de Bhattacharyya para o caso multi-classe. Conforme ilustrado na Figura 4, utilizando-se 500 amostras de treinamento para cada classe, esse método é apresenta eficácia comparável ao método proposto (CAD+ODB) e superior ao método CAD+ADC em quase todo o espectro de feições examinados, exceto quando a dimensão utilizada para classificação é inferior a 10, onde o método CAD+ADC apresenta acurácia superior. O valor mais alto para a acurácia média neste classificador, foi obtido com o uso de 40 feições, sendo igual a 80,5%.

3.4 Tempo de processamento dos classificadores

Especificamente em relação ao tempo de processamento, o desempenho computacional das três técnicas é ilustrado na Figura 5. A metodologia que consumiu o maior tempo de processamento é aquela que implementa o conjunto CEU+SFS, causado pelo alto número de comparações requeridas. Também como salientado anteriormente, o método proposto CAD+ODB apresenta o melhor desempenho, obtendo um tempo total de processamento (treinamento e teste do classificador) inferior à metodologia tradicional em estágio único.

O conjunto CAD+ADC exigiu um tempo de processamento similar ao demandado pelas demais técnicas na dimensionalidade mínima testada (cinco feições), sendo que este desempenho está diretamente relacionado ao menor número de feições extraídas por esse método (quatro feições).

Figura 5 - Tempo de processamento dos classificadores

4. CONCLUSÕES

Os experimentos mostram que um classificador árvore de decisão como o testado neste estudo pode oferecer um desempenho superior a um classificador em estágio único, pois reduz um problema global multi-classe em problemas localmente mais simples, comparando apenas duas classes em cada estágio. Em relação à acurácia de classificação, a otimização da distância de Bhattacharyya aplicada em um classificador árvore de decisão apresenta resultados compatíveis com a metodologia tradicional, desde que seja utilizado um número de amostras suficientemente grande para estimar as matrizes de covariância das classes com acuidade, isto é, desde que seja utilizado um conjunto de amostras de treinamento superior a oito vezes a dimensionalidade dos dados.

Esse fato evidencia que o aumento no número de comparações feitas entre pares de classes com o uso da técnica CAD+ODB pode tornar possível resultados na acurácia final de classificação em dados de imagens hiperespectrais similares, ou até mesmo superiores, aos obtidos através de técnicas tradicionais, com um melhor desempenho computacional. Como esperado, o método de classificação em estágio único é superior ao método proposto sempre que o número de amostras de treinamento disponíveis seja próximo à dimensionalidade dos dados.

Estudos no sentido de otimizar o processo de seleção das árvores são recomendados. Os experimentos mostraram que valores para LV inferiores a 100%, resultam em árvores menores, mas tendem a apresentar resultados instáveis. Assim, não é possível afirmar que a estrutura do CAD utilizado seja ótima, isto é, que não exista uma estrutura menos complexa do que a utilizada neste estudo, produzindo localmente resultados superiores a acurácia média geral.

5. REFERÊNCIAS

- AEBERHARD, S., COOMANS, D. and 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. 358p.
- FRIEDMAN, J. H. Regularized Discriminant Analysis, *Journal of the American Statistical Association*, v. 84, n. 405, p. 165-175, 1989.
- FUKUNAGA, K. *Introduction to Statistical Pattern Recognition*. 2nd. Ed. Boston: Academic Press, 1990. 591p.
- HUGHES, G. F. On the mean accuracy of statistical pattern recognizers, *IEEE Trans. Inform. Theory*, vol. IT-14, p. 55-63, 1968.
- JACKSON, Q. and 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.
- LANDGREBE, D. *Signal Theory Methods in Multispectral Remote Sensing*, Hoboken, NJ, Wiley Interscience, 2003. 508 p.
- MORAES, D. A. O. e HAERTEL, V. Extração de Feições em Dados Imagem com Alta Dimensão por Otimização da Distância de Bhattacharyya em um Classificador de Decisão em Árvore. *Dissertação de Mestrado em Sensoriamento Remoto - PPGSR*, Universidade Federal do Rio Grande do Sul, CNPq, 99 p., 2005.
- RICHARDS, J. A. and JIA, X. *Remote Sensing Digital Image Analysis: an introduction*, 3rd. ed. Springer, 1998. 362p.
- SAFAVIAN, S. R. and LANDGREBE, D., A Survey of Decision Tree Methodology, *IEEE Trans. Systems, Man and Cybernetics*, vol. 21, no. 3, May/June 1991. p. 660-674.
- SERPICO, S. B., D'INCA, M., MELGANI, F. and MOSER, G. A comparison of feature reduction techniques for classification of hyperspectral remote-sensing data. In: *Proceedings of SPIE - The International Society for Optical Engineering*, 2002. Agia Pelagia, Greece. v 4885, p 347-358.
- SHASHAHANI, B. and 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.
- THEODORIDIS, S. and KOUTROUMBAS, K. *Pattern Recognition*, 2nd ed., Elsevier, Amsterdam, 689 p., 2003.
- THERRIEN, C. W. *Decision Estimation and Classification: An Introduction to Pattern Recognition and Related Topics*, New York, John Wiley & Sons, 1989, 251 p.
- TADJUDIN, S. and LANDGREBE, D. Classification of High Dimensional Data with Limited Training Samples, *Ph.D. thesis, School of Electrical and Computer Engineering*, Purdue University, West Lafayette, IN, 1998.
- WEBB, A. *Statistical Pattern Recognition*, 2nd ed., John Wiley, Hoboken, 496 p., 2002.

(Recebido em junho de 2006. Aceito em novembro de 2006.)