

CLASSIFICADOR ADAPTATIVO PARA DADOS IMAGEM EM ALTA DIMENSIONALIDADE COM UM CONJUNTO PEQUENO DE AMOSTRAS DE TREINAMENTO

An adaptive classifier for high dimensional image data and a small training sample set

Delnir Monteiro de Lemos^{1,2}
Vitor Haertel²

¹ **FURG – Fundação Universidade Federal do Rio Grande**
CTI – Colégio Técnico Industrial Prof. Mário Alquatti
Rua Engenheiro Alfredo Huch, 475 – Centro – Rio Grande – RS – 96201 900 – Brasil
dmlemos@cti.furg.br

² **UFRGS - Universidade Federal do Rio Grande do Sul**
Centro Estadual de Pesquisas em Sensoriamento Remoto e Meteorologia
Cx. Postal 15044 – Porto Alegre – RS – 91501 970 – Brasil
victor.haertel@ufrgs.br

RESUMO

Neste trabalho é testado um classificador adaptativo que visa suavizar os efeitos causados por um número insuficiente de amostras de treinamento, fato este que pode degradar severamente a acurácia dos resultados obtidos por um classificador paramétrico utilizando dados com dimensão alta. O classificador adaptativo adiciona amostras semi-rotuladas ao conjunto das amostras de treinamento com o objetivo de reduzir os efeitos causados pelo pequeno número de amostras. O efeito das amostras semi-rotuladas é controlado por meio de um peso menor do que o peso atribuído as amostras originais. Os experimentos desenvolvidos mostram que este procedimento é eficiente na redução dos efeitos do fenômeno de Hughes contribuindo para aumentar a acurácia da imagem temática produzida.

Palavras chaves: dados em alta dimensionalidade, classificador adaptativo, reconhecimento de padrões.

ABSTRACT

In this paper, we test a self-learning and self-improving adaptive classifier to mitigate the problem of small training sample size that can severely affect the accuracy of the results produced by a parametric classifier employing high dimensional image data. The adaptive classifier mitigates the small training sample size by adding semi-labeled samples to the training set. In order to control the influence of semi-labeled samples, the proposed method assigns full weight to the training samples and reduced weight to semi-labeled samples. Experiments show that this procedure is effective in mitigating the Hughes phenomenon and increasing therefore the accuracy of the resulting thematic map.

Keywords: high dimensional image data, adaptable classifier, pattern recognition

1. INTRODUÇÃO

Em sensoriamento remoto aplicado, o incremento da resolução espectral fornecido pela tecnologia atual apresenta novas possibilidades e também novos desafios. Sistemas sensores capazes de coletar dados em um número elevado de bandas espectrais, como por exemplo o sistema AVIRIS,

passaram a ser disponibilizados. Esta disponibilidade de dados em alta dimensão, torna possível a separação de classes espectralmente semelhantes entre si, e, portanto, indistinguíveis quando se utiliza dados tradicionais em baixa dimensionalidade como por exemplo dados Landsat-TM (FUKUNAGA, 1990). Por outro lado, na medida em que a dimensão dos dados cresce,

aumenta também o número de parâmetros a serem estimados em um classificador paramétrico, o que por sua vez requer um número também crescente de amostras de treinamento, isto é, de pixels cujas classes foram previamente determinadas. Estas amostras são frequentemente obtidas por verificação diretamente no local (verdade terrestre). No contexto deste estudo, estas amostras são referidas como *amostras rotuladas*.

Na prática, dispõe-se geralmente apenas de um número limitado de amostras de treinamento (amostras rotuladas), fato este que resulta em uma estimativa pouco confiável para os parâmetros do classificador. Quando a razão entre o número de amostras de treinamento e o número de bandas espectrais é pequena, o valor estimado para os parâmetros se torna altamente variável provocando a deterioração no desempenho do classificador. Este fato fica ilustrado pelo comportamento do valor da acurácia no processo de classificação na medida em que um número crescente de bandas espectrais é utilizado e o número de amostras de treinamento é mantido constante. Inicialmente, a acurácia do processo de classificação vai aumentando com a introdução de informações adicionais, isto é, de novas bandas espectrais. Em um determinado ponto a acurácia atinge um valor máximo para em seguida passar a declinar, na medida em que bandas adicionais continuam a ser incluídas no processo de classificação. Este comportamento da acurácia versus dimensionalidade dos dados é conhecido como fenômeno de Hughes (*Hughes phenomenon*) ou a maldição da dimensionalidade (*the curse of dimensionality*), (LANDGREBE, 2003, TADJUDIN *et al.*, 1998).

Métodos para diminuir os efeitos do fenômeno de Hughes vem sendo objeto de estudos por parte da comunidade internacional de sensoriamento remoto. Neste contexto, as propostas feitas podem ser agrupadas em três categorias: (1) métodos para regularização da matriz de covariância (FRIEDMAN, 1989, AEBERHARD *et al.*, 1994), (2) métodos para redução na dimensionalidade dos dados, via seleção ou extração de feições, e (3) métodos que buscam incrementar o número de amostras de treinamento por meio da inclusão das chamadas amostras semi-rotuladas (SHAHSHAHANI *et al.*, 1994, LANDGREBE, 2003). A metodologia avaliada neste estudo insere-se nesta categoria. Este estudo testa um classificador adaptativo que implementa um processo de classificação iterativo, visando a cada iteração aumentar a confiabilidade nos valores estimados para os parâmetros do classificador e conseqüentemente melhorar a acurácia da classificação. Neste processo iterativo, uma primeira imagem temática é produzida pelo classificador, utilizando estimativas para os parâmetros baseadas somente nas amostras

rotuladas disponíveis. Nas iterações seguintes, novos pixels, cujas classes passam a ser estimadas pelo classificador (amostras semi-rotuladas), são adicionados, de uma forma ponderada, às amostras rotuladas na atualização das estimativas dos parâmetros do classificador. Este processo iterativo é repetido até que um determinado critério de convergência seja alcançado. O classificador utilizado neste processo é adaptativo no sentido que a cada nova iteração são utilizadas informações extraídas da iteração anterior, melhorando a cada etapa o valor estimado para os parâmetros e conseqüentemente a acurácia da imagem temática produzida.

Os experimentos desenvolvidos neste estudo mostraram ser possível iniciar o processo de classificação com um número pequeno de amostras de treinamento (contribuição mínima do analista). No decorrer do processo, o classificador vai extraindo estimativas mais confiáveis para os parâmetros, e gerando gradualmente resultados mais acurados a cada nova iteração.

2. METODOLOGIA

Um dos classificadores mais utilizados no processo de classificação de imagens digitais em sensoriamento remoto é o da Máxima Verossimilhança (MV). Neste caso, a forma geral da função de decisão é dada por:

$$G_i(\mathbf{X}) = p(\mathbf{X} | \omega_i)P(\omega_i) \quad (1)$$

sendo a regra de classificação dada por:

$$\mathbf{X} \in \omega_i \quad \text{se } G_i(\mathbf{X}) > G_j(\mathbf{X}) \quad \forall j \neq i \quad (2)$$

onde, \mathbf{X} é o vetor que representa um *pixel* individual a ser classificado, $p(\mathbf{X} | \omega_i)$ representa a função densidade de probabilidade condicional à classe ω_i e $P(\omega_i)$ representa a probabilidade *a priori* associada à mesma classe.

A utilização prática do classificador MV requer a especificação da função densidade de probabilidade $p(\mathbf{X} | \omega_i)$. A experiência prática indica a função densidade de probabilidade Gaussiana como a mais adequada aos dados de sensoriamento remoto. Neste caso, o classificador passa a denominar-se de Máxima Verossimilhança Gaussiana (MVG) e a função de decisão (1), após algumas simplificações bem conhecidas adquire a forma:

$$G_i = -\ln|\Sigma_i| - (\mathbf{X} - \mu_i)' \Sigma_i^{-1} (\mathbf{X} - \mu_i) + 2\ln(P(\omega_i)) \quad (3)$$

$$i = 1 \dots m$$

sendo μ_i o vetor de médias e Σ_i a matriz de covariância associados à classe ω_i respectivamente e m o número de classes. Os parâmetros μ_i e Σ_i são normalmente estimados a partir das amostras de

treinamento disponíveis, (amostras rotuladas), utilizando os estimadores:

$$\hat{\mu}_i = \frac{1}{N_i} \sum_{k=1}^{N_i} \mathbf{X}_k \quad (4)$$

$$\hat{\Sigma}_i = \frac{1}{N_i - 1} \sum_{k=1}^{N_i} (\mathbf{X}_k - \hat{\mu}_i)(\mathbf{X}_k - \hat{\mu}_i)'$$

sendo N_i o número de amostras de treinamento disponíveis (amostras rotuladas) para a classe ω_i . Note-se que a função de decisão (3) é quadrática. Este classificador se enquadra portanto na categoria dos *Classificadores Quadráticos (CQ)*.

O número de parâmetros a serem estimados em (3), especialmente na matriz de covariância, cresce rapidamente com a dimensionalidade dos dados. Assim, por exemplo, no caso de dados Landsat-TM (7 bandas) o número de parâmetros a serem estimados é igual a 28. No caso do sensor AVIRIS (210 bandas) este número cresce para 22155, resultando numa rápida degradação no processo de classificação se o número de amostras de treinamento disponíveis (amostras rotuladas) não crescer na mesma proporção. Este problema, bem como possíveis alternativas para mitigá-lo, vem sendo objeto de estudo por parte de vários pesquisadores (SHASHAHANI *et al.*, 1994, HSIEH *et al.*, 1998, FRIEDMAN, 1989, AEBERHARD *et al.*, 1993, CORTIJO, 1999, JACKSON *et al.*, 2001, LANDGREBE, 2003).

3. O CLASSIFICADOR ADAPTATIVO

A metodologia investigada neste estudo consiste em um procedimento iterativo, utilizando o classificador quadrático (3), gerando a cada iteração um certo número de amostras denominadas de semi-rotuladas, que são adicionadas de uma forma ponderada, as amostras de treinamento originais (amostras rotuladas). Desta forma, o número total de amostras para treinamento do classificador é incrementado, permitindo assim estimativas mais confiáveis para os parâmetros do classificador.

Esta metodologia está sucintamente descrita no fluxograma ilustrado na figura 1. O procedimento consiste essencialmente da adição de uma etapa adicional ao classificador Máxima Verossimilhança Gaussiana (3). O processo inicia por uma classificação prévia onde somente as amostras rotuladas são utilizadas na estimação dos parâmetros. Como resultado desta classificação inicial, pixels na imagem (amostras não-rotuladas) passam a ser associados a uma das classes existentes de acordo com a regra de decisão em (3). Desta forma amostras não-rotuladas passam a ser denominadas de amostras semi-rotuladas. Esta designação significa que alguma informação com relação à classe à qual o pixel pertence (rótulo do pixel) passa a ser disponível, embora com um grau de incerteza. A partir da segunda iteração, as

amostras semi-rotuladas passam a ser utilizadas conjuntamente com as amostras rotuladas originais para re-estimar os parâmetros do classificador (*retro-alimentação*) e posterior classificação dos dados na imagem.

As amostras na categoria de semi-rotuladas apresentam portanto um grau de incerteza que varia com cada amostra individual e a cada iteração, podendo eventualmente ocorrer inclusive a troca de rótulo no decorrer do processo iterativo. Desta forma, nesta metodologia cada amostra semi-rotulada deve receber um fator de peso, expressando o grau de confiança em seu rótulo, e a estimação dos parâmetros do classificador é feita levando em consideração este fator.

As amostras rotuladas recebem peso igual à unidade. As amostras semi-rotuladas recebem uma ponderação proporcional ao grau de pertencimento à classe vencedora na iteração imediatamente anterior, expresso no intervalo [0,1]. A estimação dos pesos para as amostras semi-rotuladas é realizado por meio da expressão (5) e a estimação ponderada dos parâmetros (vetor de médias e matriz de covariância) pelas expressões (6) e (7).

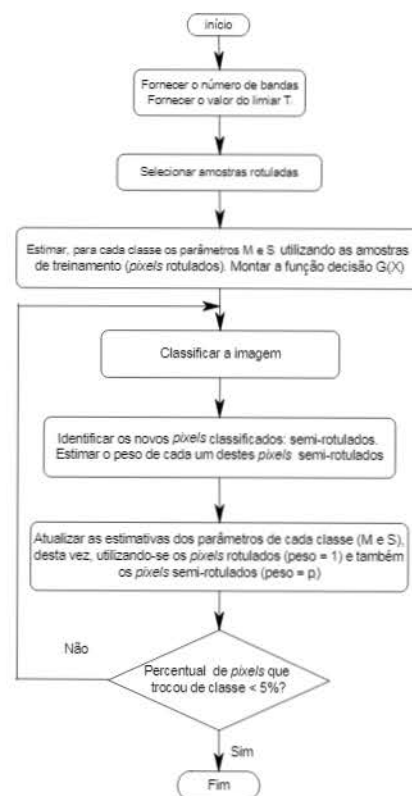


Fig. 1 – Um procedimento adaptativo de classificação.

O classificador adaptativo visa portanto maximizar o número de amostras de treinamento, pela integração das amostras de treinamento disponíveis (amostras rotuladas) com as amostras semi-rotuladas obtidas a cada etapa do processo,

através da aplicação de pesos no decorrer do processo iterativo. Trata-se portanto de um processo iterativo, de aproximações sucessivas, iniciando-se com uma estimativa inicial baseada somente nas amostras rotuladas disponíveis e repetindo os passos abaixo relacionados, a cada iteração usando as amostras rotuladas e as amostras semi-rotuladas:

A- Estimação dos pesos, para cada pixel na imagem, exceto aqueles que pertencem ao conjunto das amostras rotuladas:

$$p(\mathbf{X}) = \frac{G_i(\mathbf{X})}{\sum_{j=1}^m G_j(\mathbf{X})} \quad (5)$$

sendo $G_i(\mathbf{X})$ ao valor da função de decisão (3) correspondente à classe vencedora (ω_i) na presente iteração para o pixel \mathbf{X} , m o número de classes sob consideração.

B- Estimação ponderada do vetor de médias μ_k e da matriz de covariância Σ_k para a classe ω_k ($k=1, \dots, m$):

$$\mu_k = \frac{1}{\sum_{j=1}^{N_k} p(\mathbf{X}_j)} \sum_{j=1}^{N_k} p(\mathbf{X}_j) \mathbf{X}_j \quad (6)$$

$$\Sigma_k = \frac{1}{\sum_{j=1}^{N_k} p(\mathbf{X}_j) - 1} \sum_{j=1}^{N_k} p(\mathbf{X}_j) (\mathbf{X}_j - \mu_k) (\mathbf{X}_j - \mu_k)' \quad (7)$$

sendo N_k o número total de amostras (rotuladas e semi-rotuladas) disponíveis para a classe ω_k . Note-se que no caso de amostras rotuladas é atribuído peso igual a 1, isto é, $p(\mathbf{X}) = 1$.

C- Classificação da imagem por meio da função de decisão (3), na suposição de valores iguais para as m probabilidades *a priori* $P(\omega_j)$:

$$G_i(\mathbf{X}) = \ln |\Sigma_i| - (\mathbf{X} - \mu_i)' \Sigma_i^{-1} (\mathbf{X} - \mu_i)$$

Note-se que as estimativas para vetores de médias (6) e matrizes de covariância (7), são misturas ponderadas de estimativas de máxima verossimilhança, envolvendo amostras rotuladas e amostras semi-rotuladas.

Um procedimento nesta linha foi proposto recentemente por JACKSON *et al.* (2002). Aqueles autores investigaram a utilização de amostras semi-rotuladas no treinamento de um classificador quadrático relatando resultados positivos com relação ao aumento da acurácia no processo de classificação, sempre que o número de amostras de treinamento (amostras rotuladas) disponíveis é pequeno quando comparado com a dimensionalidade dos dados.

Diferentemente do trabalho desenvolvido por aqueles autores, neste estudo não foi utilizado

nenhum método de estimação regularizada dos parâmetros estatísticos.

4. EXPERIMENTOS

Com a finalidade de testar a metodologia proposta, foram desenvolvidos experimentos utilizando dados hiperespectrais obtidos pelo sensor AVIRIS. Os dados coletados por este sensor consistem de 210 bandas espectrais, cobrindo a região espectral no intervalo 0.4 μ m–2.4 μ m. Trata-se de um sistema em fase experimental, que utiliza uma plataforma aero-transportada. A cena utilizada cobre uma área experimental, localizada no Estado de Indiana, Estados Unidos. Dados de campo para fins de verdade terrestre foram obtidos pelo grupo da Universidade Purdue (West Lafayette, Indiana) e são utilizados neste estudo. Estes dados foram divididos em dois grupos. Um primeiro grupo (amostras de treinamento rotuladas) é empregado para fins de estimação dos parâmetros do classificador e um segundo grupo (amostras de teste), para fins de estimação da acurácia obtida.

Esta área está ocupada por culturas agrícolas que apresentam respostas espectrais muito semelhantes entre si e, por esta razão, não são adequadamente separáveis utilizando-se dados tradicionais em baixa dimensionalidade como, por exemplo, dados Landsat-TM. Trata-se portanto de uma área adequada para testar a eficiência das técnicas investigadas neste estudo.

Com este objetivo, foi desenvolvido uma série de experimentos utilizando a técnica do classificador adaptativo investigado neste estudo, tomando-se a dimensionalidade dos dados como variável independente e a resultante acurácia na classificação como variável dependente.

Da totalidade das 210 bandas espectrais disponíveis, foram inicialmente removidas aquelas que apresentavam um nível de ruído muito elevado (em geral devido a presença de vapor de água na atmosfera), restando 185 bandas espectrais para os experimentos. Estes experimentos foram realizados utilizando-se desde 10 até 180 bandas espectrais, selecionadas em cada caso a intervalos iguais ao longo das 185 bandas disponíveis.

Para fins de ilustração dos resultados obtidos, são aqui apresentados dois destes experimentos: o primeiro utilizando 80 bandas espectrais e o segundo empregando 90 bandas espectrais. Nestes experimentos, foi empregado a função de decisão (3), com um limiar estimado com auxílio de uma estatística χ^2 a um nível de 95% (RICHARDS *et al.*, 1999).

Em ambos os experimentos foi utilizado um número igual de amostras rotuladas (234), para cada uma das classes envolvidas. Este número de amostras foi escolhido deliberadamente pequeno com relação à dimensionalidade dos dados para melhor evidenciar as conseqüências de um número pequeno de amostras de treinamento na acurácia da

classificação.

As tabelas 1 e 3 apresentam o número total de *pixels* que foram classificados em cada uma das classes definidas para o experimento, bem como a acurácia obtida, empregando-se 80 e 90 bandas espectrais, respectivamente e utilizando-se somente as amostras rotuladas na estimação dos parâmetros do classificador (234 amostras por classe).

TABELA 1 – DISTRIBUIÇÃO DE CLASSES PARA A ÁREA SELECIONADA (80 BANDAS), CLASSIFICAÇÃO INICIAL.

Nome da Classe	Número Total de <i>pixels</i> Classificados na Imagem	Acurácia (%)
Milho cultivado direto	1380	62.1
Milho cultivado mínimo	946	65.7
Milho cultivado tradicional	301	99.6
Soja cultivado direto	1130	64.5
Soja cultivado mínimo	2046	57.6
Soja cultivado tradicional	609	70.0
Floresta	1723	72.2

TABELA 2 – DISTRIBUIÇÃO DE CLASSES PARA A ÁREA SELECIONADA (80 BANDAS), CLASSIFICAÇÃO FINAL.

Nome da Classe	Número Total de <i>pixels</i> Classificados na Imagem	Acurácia (%)
Milho cultivado direto	1291	65.5
Milho cultivado mínimo	744	69.9
Milho cultivado tradicional	279	99.6
Soja cultivado direto	1707	64.9
Soja cultivado mínimo	3300	79.8
Soja cultivado tradicional	572	70.4
Floresta	3249	94.1

TABELA 3 – DISTRIBUIÇÃO DE CLASSES PARA A ÁREA SELECIONADA (90 BANDAS), CLASSIFICAÇÃO INICIAL.

Nome da Classe	Número Total de <i>pixels</i> Classificados na Imagem	Acurácia (%)
Milho cultivado direto	1240	56.1
Milho cultivado mínimo	866	66.4
Milho cultivado tradicional	259	99.6
Soja cultivado direto	953	55.5
Soja cultivado mínimo	1638	47.9
Soja cultivado tradicional	520	66.4
Floresta	1343	61.3

TABELA 4 – DISTRIBUIÇÃO DE CLASSES PARA A ÁREA SELECIONADA (90 BANDAS), CLASSIFICAÇÃO FINAL.

Nome da Classe	Número Total de <i>pixels</i> Classificados na Imagem	Acurácia (%)
Milho cultivado direto	1256	61.9
Milho cultivado mínimo	746	68.7
Milho cultivado tradicional	252	99.6
Soja cultivado direto	1039	60.0
Soja cultivado mínimo	3211	76.8
Soja cultivado tradicional	485	66.9
Floresta	2482	86.1

As tabelas 2 e 4 apresentam os resultados obtidos após 06 iterações do classificador adaptativo. Em cada iteração foram utilizadas as mesmas 234 amostras rotuladas por classe, acrescidas de amostras semi-rotuladas, sempre obtidas da iteração imediatamente anterior. Na metodologia adotada, as amostras semi-rotuladas são formadas por 10% dos *pixels* originalmente não rotulados, e que passam a semi-rotulados pela classificação produzida na iteração imediatamente anterior. São selecionados aqueles 10% que apresentam o peso mais alto conforme definido em (5).

A comparação dos dados na Tabela 1 com

aqueles na Tabela 2 e também dos dados na Tabela 3 com aqueles na Tabela 4, evidencia os incrementos obtidos nos valores estimados para a acurácia e os correspondentes ajustes nos pixels atribuídos a cada classe, ilustrando desta forma a validade da metodologia proposta.

Nestes experimentos, a acurácia das imagens classificadas foi estimada com o auxílio das amostras de teste disponíveis, e representam a proporção destas amostras classificadas corretamente conforme os dados de referência (verdade terrestre). Esta é a estimativa denominada de *producer's accuracy* por Congalton (1991). Nesta abordagem, portanto, o valor da acurácia estima a probabilidade de que o classificador reconheça um pixel como pertencente à classe definida pela verdade terrestre.

A variação da acurácia média em função do número de iterações está ilustrada nas figuras 2 e 3 para os experimentos com 80 e 90 bandas espectrais respectivamente. Pode-se observar claramente em cada caso, um incremento na acurácia média com o número de iterações, ilustrando a crescente contribuição das amostras semi-rotuladas ao processo de classificação.

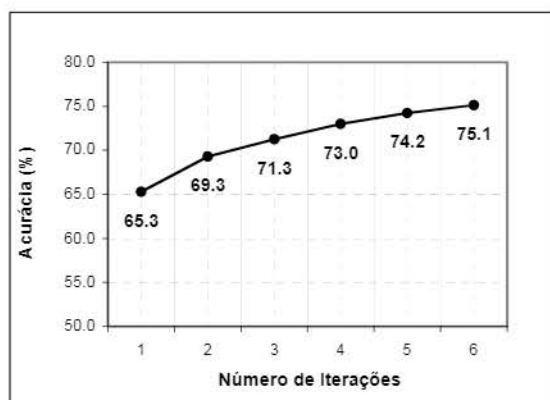


Fig. 2 – Acurácia média da classificação para 80 bandas espectrais.

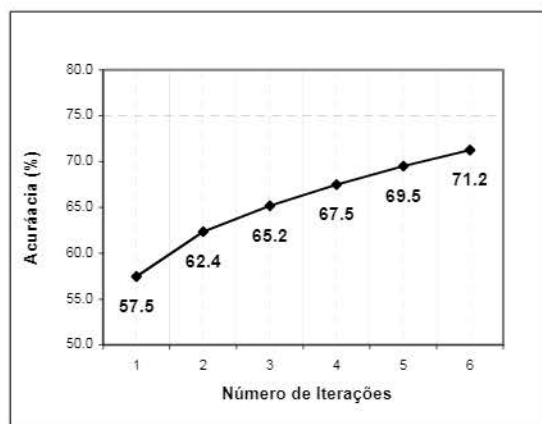


Fig. 3 – Acurácia média da classificação para 90 bandas espectrais.

Deve-se aqui observar que, pelo processo empregado na seleção do sub-conjunto de bandas espectrais utilizadas em cada experimento, as bandas individuais utilizadas em dois experimentos distintos não são necessariamente as mesmas, o que explica certas diferenças verificadas entre os valores estimados para a acurácia nestes dois experimentos.

Estes resultados servem para ilustrar que, para configurações que possuam um número pequeno de amostras de treinamento, insuficiente para uma estimativa confiável dos parâmetros do classificador, a metodologia apresentada neste estudo se constitui em uma alternativa capaz de minorar este problema.

5. CONCLUSÕES

No presente estudo foi apresentada e testada uma metodologia que faz uso conjunto de amostras rotuladas e não rotuladas para fins de estimação dos parâmetros (vetor de médias e matriz de covariância) em um classificador quadrático (Máxima Verossimilhança Gaussiana), quando se utiliza dados em alta dimensionalidade. Como é bem conhecido, o atrativo da utilização destes dados reside na possibilidade de separar com razoável acurácia classes com características espectrais muito semelhantes (FUKUNAGA, 1990). Esta vantagem fica entretanto prejudicada sempre que o número de amostras de treinamento disponíveis não é suficiente para uma estimativa confiável dos parâmetros do classificador, fato este que ocorre com frequência em situações reais. Os experimentos mostraram que este problema pode ser contornado com sucesso quando se incrementa o conjunto das amostras de treinamento disponíveis (amostras rotuladas) com amostras adicionais, não-rotuladas. A adição de amostras não-rotuladas mostrou que as estimativas dos parâmetros requeridos pelo classificador quadrático tornam-se mais confiáveis, resultando em um incremento na acurácia do classificador. Os experimentos realizados mostram, portanto, que a metodologia proposta é eficiente para mitigar os efeitos do fenômeno de Hughes, permitindo desta forma a utilização efetiva de um número maior de bandas espectrais no classificador, com um correspondente valor mais alto para a acurácia final.

Apesar dos resultados promissores, o classificador adaptativo proposto tem limitações. Em particular, para um problema onde a dimensionalidade dos dados, isto é, o número de bandas espectrais é muito maior que o número de amostras de treinamento (amostras rotuladas), a classificação inicial pode ser muito imprecisa. Como resultado o classificador adaptativo proposto pode não convergir.

Como sugestão para trabalhos futuros, pode-se citar: o estudo de um método que combine

o emprego de amostras semi-rotuladas com métodos de estimação regularizada dos parâmetros do classificador, e a otimização do software CADHI (Classificador Adaptativo para Imagens Hiperspectrais), desenvolvido neste estudo, pela adição de novas ferramentas que ampliem a sua utilização.

REFERÊNCIAS BIBLIOGRÁFICAS

- 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.
- CONGALTON, R. A Review of Assessing the Accuracy of Classifications of Remotely Sensed Data, **Remote Sensing of Environment**, v. 37, p. 35-46, 1991.
- CORTIJO F.J.; DE LA BLANCA N. P. The Performance of Regularized Discriminant Analysis Versus Non-parametric Classifiers Applied to High-Dimensional Image Classification. **International Journal of Remote Sensing**, v. 20, n. 17, p. 3345-3365, 1999.
- FRIEDMAN, J. H. Regularized Discriminant Analysis, **Journal of the American Statistical Association. Theory and Methods**, v. 84, n. 405, p. 165-175, 1989.
- FUKUNAGA, K. **Introduction to Statistical Pattern Recognition**, 2^o ed., School of Electrical Engineering, Purdue University. New York, Academic Press Inc. 1990. 590 p.
- HSIEH, P.; LANDGREBE, D.A. **Classification of High Dimensional Data**. TR-ECE 98-4, School of Electrical and Computer Engineering., Purdue University, May 1998. 121 p.
- JACKSON, Q.; LANDGREBE, D.A., An Adaptive Classifier Design for High-Dimensional Data Analysis with a Limited Training Data Set, **IEEE Transactions on Geoscience and Remote Sensing**, v. 39, n. 12, p. 2664-2679, 2001.
- JACKSON, Q.; LANDGREBE, D.A. An Adaptive Method for Combined Covariance Estimation and Classification, **IEEE Transactions on Geoscience and Remote Sensing**, v. 40, n. 5, p. 1082-1087, 2002.
- JIA, X.; RICHARDS, J. Segmented Principal Components Transformation for Efficient Hyperspectral Remote-Sensing Image Display and Classification. **IEEE Transactions on Geoscience and Remote Sensing**, v. 37, n. 1, p. 538-542, 1999.
- LANDGREBE, D.A. **Signal Theory Methods in Multispectral Remote Sensing**, John Wiley & Sons, Hoboken, New Jersey, 2003. 508 p.
- RICHARDS, J.A.; JIA, X. **Remote Sensing Digital Image Analysis, An Introduction**, 3rd ed, Springer, 1999. 365 p.
- SHAHSHAHANI, B. M.; LANDGREBE D. A. 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. e 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. 123 p.

Recebido em 04 de julho de 2005 – Aceito em 30 de dezembro de 2005.