

UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL - UFRGS
CENTRO ESTADUAL DE PESQUISA EM SENSORIAMENTO REMOTO E METEOROLOGIA - CEP SRM
PROGRAMA DE PÓS-GRADUAÇÃO EM SENSORIAMENTO REMOTO

Dissertação de Mestrado:

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

Por

Maciel ZORTEA

Sob orientação do

Prof. Vitor HAERTEL, PhD.

Porto Alegre, Abril de 2004.

UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL - UFRGS
CENTRO ESTADUAL DE PESQUISA EM SENSORIAMENTO REMOTO E METEOROLOGIA - CEPSRM
PROGRAMA DE PÓS-GRADUAÇÃO EM SENSORIAMENTO REMOTO

Dissertação de Mestrado:

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

Maciel ZORTEA

*Engenheiro Civil (2002 – UFRGS), Mestrando pelo Programa de
Pós-Graduação em Sensoriamento Remoto – PPGSR/UFRGS*

Sob orientação do

Prof. Vitor HAERTEL, PhD.

*Dissertação apresentada ao Programa de Pós-Graduação em
Sensoriamento Remoto da Universidade Federal do Rio Grande
do Sul como parte dos requisitos necessários para obtenção do
Grau de Mestre em Sensoriamento Remoto.*

*Porto Alegre
Abril de 2004.*

Folha de Homologação

É incluída no documento após a homologação pela Comissão de Pós-Graduação em Sensoriamento Remoto. Nela constam a identificação da Instituição, da Unidade do Curso, o título da dissertação, o nome e titulação do autor, seguido da sigla da Universidade onde se graduou, o requisito, a identificação da Instituição, da Unidade e a data de aprovação e de homologação e os nomes do professor orientador, dos componentes da Banca Examinadora, do Coordenador do CPGSR e do Diretor do Centro Estadual de Pesquisas em Sensoriamento Remoto e Meteorologia.

“Coragem, como toda virtude, é um equilíbrio”.

Aristóteles (384-322 a.C.)

AGRADECIMENTOS

A Deus, pela oportunidade divina.

Um agradecimento muito especial ao meu orientador, Prof. Victor Haertel, que sempre me apoiou e confiou na realização deste trabalho. Obrigado pela amizade, paciência, dedicação e disposição no auxílio para a conclusão desta dissertação de mestrado. Encontrei nele alguém com a visão que me fez olhar para mais longe.

Um agradecimento muito especial também à instituição Universidade Federal do Rio Grande do Sul, na qual com muito orgulho concluí minha graduação, me permitindo na seqüência continuar meus estudos de Pós-Graduação. Sou grato ao povo brasileiro que, mesmo sem saber, a custo de árduos impostos, financiaram a concretização de mais um sonho em minha vida. Sou igualmente grato àquelas pessoas que bravamente lutam por manter esta universidade viva, gratuita e de excelente qualidade. Que nunca lhes falte força e serenidade para que continuem sua missão. Agradeço também a CAPES pela bolsa de estudo.

Um agradecimento também especial a todos os demais colegas, professores e funcionários do Centro Estadual de Pesquisa em Sensoriamento Remoto e Meteorologia, pelos bons momentos que aí passei, pela amizade e ensinamentos adquiridos.

Agradeço também a banca examinadora, que certamente poderá colaborar com sugestões para o aperfeiçoamento deste trabalho.

Aos demais amigos e pessoas que sabem que são importantes para mim e souberam compreender meus momentos de ausência ou distância.

Dedico este trabalho, com muito carinho, a minha família.

INVESTIGAÇÃO SOBRE MÉTODOS PARA REDUÇÃO DA DIMENSIONALIDADE DOS DADOS EM IMAGENS HIPERESPECTRAIS *

Maciel ZORTEA

Orientador: Prof. PhD. Vitor HAERTEL

Os recentes avanços na tecnologia de sensores tem disponibilizado imagens em alta dimensionalidade para fins de sensoriamento Remoto. Análise e interpretação dos dados provenientes desta nova geração de sensores apresenta novas possibilidades e também novos desafios. Neste contexto, um dos maiores desafios consiste na estimação dos parâmetros em um classificador estatístico utilizando-se um número limitado de amostras de treinamento. Neste estudo, propõe-se uma nova metodologia de extração de feições para a redução da dimensionalidade dos dados em imagens hiperespectrais. Essa metodologia proposta é de fácil implementação e também eficiente do ponto de vista computacional. A hipótese básica consiste em assumir que a curva de resposta espectral do pixel, definida no espaço espectral, pelos contadores digitais (CD's) das bandas espectrais disponíveis, pode ser substituída por um número menor de estatísticas, descrevendo as principais características da resposta espectral dos pixels. Espera-se que este procedimento possa ser realizado sem uma perda significativa de informação. Os CD's em cada banda espectral são utilizados para o cálculo de um número reduzido de estatísticas que os substituirão no classificador. Propõe-se que toda a curva seja particionada em segmentos, cada segmento sendo então representado pela respectiva média e variância dos CD's. Propõem-se três algoritmos para segmentação da curva de resposta espectral dos pixels. O primeiro utiliza um procedimento muito simples. Utilizam-se segmentos de comprimento constante, isto é, não se faz nenhuma tentativa para ajustar o comprimento de cada segmento às características da curva espectral considerada. Os outros dois implementam um método que permite comprimentos variáveis para cada segmento, onde o comprimento dos segmentos ao longo da curva de resposta espectral é ajustado seqüencialmente. Um inconveniente neste procedimento está ligado ao fato de que uma vez selecionada uma partição, esta não pode ser alterada, tornando os algoritmos sub-ótimos. Realizam-se experimentos com um classificador paramétrico utilizando-se uma imagem do sensor AVIRIS. Obtiveram-se resultados animadores em termos de acurácia da classificação, sugerindo a eficácia dos algoritmos propostos.

* Dissertação de Mestrado em Sensoriamento Remoto, Centro Estadual de Pesquisas em Sensoriamento Remoto e Meteorologia, Curso de Pós-Graduação em Sensoriamento Remoto da Universidade Federal do Rio Grande do Sul. Porto Alegre/RS, (107 p.). Abril de 2004.

INVESTIGATION ON METHODS FOR DIMENSIONALITY REDUCTION ON HIPERESPECTRAL IMAGE DATA*

Maciel ZORTEA

Advisor: Prof. PhD. Vitor HAERTEL

Recent advances in sensor technology have made available high-dimensional image data for Remote Sensing applications. Analysis and interpretation of data provided by this new generation of sensors presents new possibilities and also new challenges. In this context, one major challenge consists in estimating the parameters in a statistically oriented classifier with a limited number of training samples. In the present study, we propose a new feature extraction approach to reduce the data dimensionality in hyperspectral image data. This proposed approach is easy to implement and also computationally efficient. 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 replaced by a smaller number of statistics, 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. We propose that the entire spectral curve be partitioned into segments, each segment being then represented by its respective DN's mean value and variance. We propose three algorithms to segment the pixel's spectral curve. The first one implements a very simple procedure. It uses constant length segments, i.e., no attempt is made to adjust the length of each segment to the characteristics of the spectral curves under consideration. The other ones implement a method allowing variable length for the segments, where the length of the segments along the spectral curves is adjusted sequentially. A drawback in this procedure lies on the fact that once selected, a partition cannot be changed, making the algorithms sub-optimal. Experiments using a parametric classifier are performed on an AVIRIS data set. Encouraging results have been obtained in terms of classification accuracy, suggesting the effectiveness of the proposed algorithms

*Master of Sciences Dissertation in Remote Sensing, Center for Remote Sensing and Meteorology, Graduate Program in Remote Sensing at Federal University of Rio Grande do Sul. Porto Alegre/RS, (107 p.). April 2004.

SUMÁRIO

1 - INTRODUÇÃO	1
1.1 - PERSPECTIVAS HISTÓRICAS: O INÍCIO DO SENSORIAMENTO REMOTO NA ERA ESPACIAL	1
1.2 - SENSORIAMENTO REMOTO: CONCEITOS BÁSICOS	3
1.3 - SENSORIAMENTO REMOTO: A ERA HIPERESPECTRAL	6
1.4 - APRESENTAÇÃO DO PROBLEMA	9
2 - REVISÃO BIBLIOGRÁFICA	12
2.1 - METODOLOGIAS PARA AUMENTO DA ACURÁCIA DA CLASSIFICAÇÃO DE DADOS EM ALTA DIMENSIONALIDADE: O ESTADO DA ARTE	12
2.1.1 - <i>Introdução de amostras não rotuladas</i>	12
2.1.2 - <i>Seleção de feições</i>	13
2.1.3 - <i>Extração de feições</i>	14
2.2 - EFEITO DO NÚMERO DE AMOSTRAS NA ESTIMAÇÃO DE PARÂMETROS	18
2.3 - CLASSIFICADOR PARAMÉTRICO: MÁXIMA VEROSSIMILHANÇA GAUSSIANA	19
2.3.1 - <i>Classificação de Bayes (bayesiana)</i>	19
2.3.2 - <i>A regra de decisão Máxima Verossimilhança</i>	20
2.3.3 - <i>Modelo para classes Normais Multivariadas</i>	21
2.4 - AVALIAÇÃO DA SEPARABILIDADE ENTRE CLASSES: DISTÂNCIA DE BHATTACHARYYA	23
2.5 - MÉTODOS DE CLASSIFICAÇÃO NÃO PARAMÉTRICOS	25
2.5.1 - <i>Método do paralelepípedo (Level-Slice Classifier)</i>	25
2.5.2 - <i>Classificador K-Vizinhos mais Próximos (K-Nearest Neighbor)</i>	26
2.5.3 - <i>Classificador Janela de Parzen (Parzen Window Classifier)</i>	26
2.5.4 - <i>Redes neurais</i>	27
3 - METODOLOGIA	28
3.1 - NOVAS ABORDAGENS PROPOSTAS PARA REDUÇÃO DA DIMENSIONALIDADE	32
3.2 - SEGMENTAÇÃO DA CURVA DE RESPOSTA ESPECTRAL EM SEGMENTOS DE COMPRIMENTO CONSTANTE – “SCC”	32
3.2.1 - <i>Algoritmo SCC – Segmentação com Comprimento Constante</i>	32
3.3 - SEGMENTAÇÃO DA CURVA DE RESPOSTA ESPECTRAL EM SEGMENTOS DE COMPRIMENTO VARIÁVEL - “SCV”	35
3.3.1 - <i>Algoritmo Otimização Global</i>	35
3.3.2 - <i>Algoritmo com Estrutura em Árvore (Tree)</i>	39
3.4 - CRITÉRIOS DE OTIMIZAÇÃO PARA AVALIAÇÃO DAS PARTIÇÕES GERADAS	41
3.4.1 - <i>CR1(UB): Critério 1 - Minimização do Limiar Superior do Erro Bayesiano</i>	43
3.4.2 - <i>CR2(DM): Critério 2 - Distância Mínima Entre Pares de Classes</i>	43
3.5 - CÁLCULO DA ACURÁCIA DA CLASSIFICAÇÃO	44
4 - ESTUDO DE CASO	45
4.1 - DADOS DISPONÍVEIS: IMAGEM HIPERESPECTRAL DO SENSOR AVIRIS	45
4.2 - SELEÇÃO DAS BANDAS ORIGINAIS	49
4.3 - REPRESENTAÇÃO DAS CURVAS DE RESPOSTA ESPECTRAL	51
4.4 - MATERIAIS NECESSÁRIOS	53
4.4.1 - <i>Equipamentos</i>	53
4.4.2 - <i>Aplicativos</i>	53
4.5 - REALIZAÇÃO DE EXPERIMENTOS	54
4.6 - EXPERIMENTO 01 - CLASSIFICAÇÃO UTILIZANDO A DIMENSIONALIDADE ORIGINAL	56
4.7 - EXPERIMENTO 02 - ALGORITMO SCC: SEGMENTAÇÃO COM COMPRIMENTO CONSTANTE	60
4.8 - EXPERIMENTO 03 - ALGORITMO OTIMIZAÇÃO GLOBAL	68
4.9 - EXPERIMENTO 04 - ALGORITMO COM ESTRUTURA EM ÁRVORE (TREE)	81
4.10 - COMPARAÇÃO DOS RESULTADOS DOS EXPERIMENTOS	91
5 - CONCLUSÕES E RECOMENDAÇÕES	99
6 - REFERÊNCIAS BIBLIOGRÁFICAS	104

LISTA DAS TABELAS

Tabela 1. Sensores Hiperespectrais atuais [SHI 03].	7
Tabela 2. Número total de pixels de treinamento e teste disponíveis para cada uma das classes em estudo.....	48
Tabela 3. Seleção das bandas ruidosas que foram excluídas (30 bandas).	49
Tabela 4. Relações obtidas utilizando o algoritmo SCC proposto: [1] Comprimento dos segmentos; [2] número total de segmentos; [3] número de bandas excedentes a serem acrescentadas ao último segmento e [4] número total de feições resultantes, utilizando-se duas feições por segmento. Os valores destacados representam as configurações que foram efetivamente utilizadas (para alternativas que geram a mesma quantidade de segmentos, são aquelas que minimizam o valor de bandas a serem acrescentadas ao último segmento), ou seja, a coluna [3].	61
Tabela 5. Acurácias máximas obtida (valores médios para 4 classes).....	66
Tabela 6. Resumo do desempenho de acurácia máxima (valor médio para 4 classes) obtida para o “algoritmo otimização global” frente ao número de amostras de treinamento e critério de segmentação utilizado. Apresenta-se também o correspondente número de feições para o qual obtiveram-se tais resultados.....	79
Tabela 7. Resumo do desempenho de acurácia máxima (valor médio para 4 classes) para o “algoritmo Tree” frente ao número de amostras de treinamento e critério de segmentação utilizado. Apresenta-se também o correspondente número de feições onde se obteve tais resultados.	90
Tabela 8. Comparação da acurácia de classificação (valores máximos) segundo o algoritmo e tamanho das amostras de treinamento e teste utilizadas.	91
Tabela 9. Tabela resumo com o número de feições para a qual ocorreu a acurácia de classificação máxima (média para 4 classes) segundo o algoritmo e tamanho das amostras de treinamento e teste utilizadas, apresentados na Tabela 8.	92
Tabela 10. Acréscimo de acurácia de classificação máxima (média para 4 classes), em relação à dimensionalidade original, segundo o algoritmo e tamanho das amostras de treinamento e teste utilizadas. Os resultados são calculados a partir da Tabela 8.	93

LISTA DAS FIGURAS

Figura 1. Espectro eletromagnético e espectro solar.....	4
Figura 2. Curvas de resposta espectral típicas do solo, vegetação e da água. Quanto maior a resolução espectral e radiométrica do sensor, melhor serão descritas as curvas de resposta espectral dos alvos imageados.	5
Figura 3. Representação da curva de resposta espectral adquirida por um sistema hiperespectral, neste caso o AVIRIS. O elevado número de bandas e sua pequena largura permitem uma descrição detalhada da resposta espectral de cada pixel presente na cena [JLP 03].....	6
Figura 4. Espectro de refletância de uma amostra contendo 3 minerais diferentes (Hematita, Kaolinita e Dolomita), obtida pelo sensor LANDSAT- TM [JLP 03].	8
Figura 5. Espectro de refletância de uma amostra contendo três minerais diferentes (Hematita, Kaolinita e Dolomita), obtida pelo sensor AVIRIS [JLP 03].....	8
Figura 6. Representação típica do fenômeno de Hughes. Cada um dos ramos da curva representa o desempenho típico de classificação frente a uma determinada quantidade m de amostras de treinamento.	10
Figura 7. Distribuição hipotética de duas classes num espaço de duas dimensões.....	15
Figura 8. Ilustração da importância de um número de pixels de treinamento suficiente para assegurar uma estimativa confiável da superfície de separação das classes. Quando um número de pixels muito pequeno é utilizado (a) pode-se obter uma boa separação dos pixels de treinamento, porém o desempenho do classificador para o conjunto de pixels de teste é pobre. Em (b) a situação persiste. Uma grande quantidade (posicionados aleatoriamente) de pixels de treinamento gera uma superfície que também é satisfatória para as amostras de teste (c) [RIC 99].	18
Figura 9. Fronteira de decisão entre duas classes	23
Figura 10. Limites teóricos inferior e superior do erro de classificação avaliado utilizando a distância de Bhattacharyya.	24
Figura 11. Exemplo da curva de resposta espectral de dois pixels de uma imagem hiperespectral obtida pelo sensor AVIRIS. No eixo das abscissas temos as bandas da imagem, nas ordenadas o respectivo contador digital (ou nível de cinza) para cada uma das bandas.	29
Figura 12. Exemplo de segmentação da curva de resposta espectral de cada pixel X , segundo três algoritmos distintos, gerando seis segmentos. Em (a) geram-se segmentos de comprimento constante. Em (b) e (c) segmentos de comprimento variável, sendo que a cada “nível hierárquico” adiciona-se uma partição adicional.	31
Figura 13. Algoritmo SCC: Exemplo de representação da segmentação da resposta original de cada pixel da imagem em 7 segmentos sucessivos, iguais e seqüenciais. Cada um dos segmentos contém $L = n$ bandas originais, que serão substituídas por duas feições. Para que ocorra redução da dimensionalidade em cada segmento, o número de feições extraídas deve ser menor do que o número de bandas em cada segmento.....	34

Figura 14. Algoritmo Otimização Global: Representação da segmentação da resposta original de cada pixel da imagem, num processo seqüencial ao final do quinto “nível hierárquico”. Note que uma vez que uma posição de divisão é gerada, esta permanece até o final do processo, sendo que cada nova posição é a ótima em relação às pré-existentes, para o critério aqui proposto. Para cada um dos segmentos são extraídas feições. A cada nível, a otimização é procurada utilizando-se o conjunto total de feições gerado, variando-se a posição da fronteira no novo segmento, mantendo inalterado os demais.	36
Figura 15. Algoritmo Tree: Representação da segmentação da resposta original de cada pixel da imagem, num processo seqüencial ao final do terceiro passo. Note que uma vez que uma posição de divisão é gerada, esta permanece até o final do processo, sendo que cada nova posição é a ótima em função das já existentes. Neste algoritmo, para a definição do próximo segmento, são testadas apenas as possibilidades de divisões sucessivas no centro de cada um dos segmentos já existentes.	39
Figura 16. Em (a) temos o recorte da área de interesse (88 x 70 pixels). Apresenta-se uma composição colorida no sistema RGB, bandas 50, 27, 17. Em (b) temos a respectiva verdade terrestre.	45
Figura 17. Vista de uma área contendo a cultivar <i>milho-cultivo direto (sorn-notill)</i> mostrando a baixa porcentagem que representa a planta frente à área emaciada no momento da coleta dos dados [LAN 2003].....	47
Figura 18. Vista de uma área da cultivar <i>soja com cultivo mínimo (soybeans-min)</i> mostrando a baixa porcentagem que representa a planta frente à área emaciada no momento da coleta dos dados [LAN 2003].....	47
Figura 19. Bandas originais: em (a) temos a banda 25, em (b) a 85, em (c) a 150 e em (d) temos a banda 200.	50
Figura 20. Envoltória das classes de treinamento. Em (a), (b), (c) e (d) são apresentados os valores da resposta mínima, média e máxima das classes de treinamento <i>Milho-cultivo direto, Gramíneas, Soja-cultivo direto e Soja-cultivo mínimo</i> , respectivamente, avaliada banda a banda, para todos os pixel de cada classe, conforme o total disponível em cada classe, anteriormente apresentado na Tabela 2.	51
Figura 21. Resposta média para as quatro classes de treinamento em estudo	52
Figura 22. . Experimento 1: Valor estimado para a acurácia do processo de classificação (MAXVER) obtidos para cada uma das quatro classes de interesse, frente a diferentes tamanhos de amostras de treinamento e teste. São utilizados dados em sua dimensionalidade original (190 bandas).	56
Figura 23. Experimento 1: Resultado médio da acurácia de classificação obtido frente a diferentes tamanhos de amostras de treinamento e teste utilizados para cada classe na dimensionalidade original dos dados (190 bandas).	58
Figura 24. Experimento 1: Tempo de processamento para o algoritmo desenvolvido, frente a diferentes tamanhos de amostras de treinamento e teste para cada classe, utilizando a dimensionalidade original dos dados (190 bandas).	59
Figura 25. Acurácia da classificação frente à dimensionalidade do problema utilizando-se 150 pixels de treinamento e teste para cada classe.	63

Figura 26. Acurácia da classificação frente à dimensionalidade do problema utilizando-se 200 pixels de treinamento e teste para cada classe.	63
Figura 27. Acurácia da classificação frente à dimensionalidade do problema utilizando-se 250 pixels de treinamento e teste para cada classe.	64
Figura 28. Acurácia da classificação frente à dimensionalidade do problema utilizando-se 300 pixels de treinamento e teste para cada classe.	64
Figura 29. Acurácia da classificação frente à dimensionalidade do problema utilizando-se 350 pixels de treinamento e teste para cada classe.	64
Figura 30. Acurácias médias obtidas frente ao número de pixels de treinamento e número de feições utilizadas.	65
Figura 31. Tempo de processamento total frente ao número de pixels de treinamento e número de feições utilizadas.	67
Figura 32. Experimento 3: Separabilidade entre classes avaliada utilizando a distância de Bhattacharyya (definida entre pares de classes), para todas as combinações de classes possíveis (4 classes: 6 combinações). As curvas apresentadas foram obtidas utilizando 150 pixels, na primeira iteração do algoritmo. Visualizam-se todas as possibilidades para o posicionamento da partição u_1 , computando-se o correspondente valor da distância de Bhattacharyya, avaliada com quatro feições, provenientes dos dois segmentos $[1:u_1] \cup [u_1+1:190]$	68
Figura 33. Experimento 3. Distância de Bhattacharyya média para as seis combinações de pares de classes; J utilizando o critério 1 e J utilizando o critério 2; respectivamente. As curvas apresentadas foram obtidas utilizando-se 150 pixels de treinamento, na primeira iteração do algoritmo, conforme resultados da Figura 32. Apresentam-se os pontos de máximo para os critérios 1 e 2.	70
Figura 34. Experimento 3: Alg. de Otimização Global - Acurácias de classificação obtidas utilizando-se 150 amostras de treinamento e teste, utilizando-se os critérios 1 e 2.	73
Figura 35. Experimento 3: Alg. de Otimização Global - Acurácias de classificação obtidas utilizando-se 200 amostras de treinamento e teste, utilizando-se os critérios 1 e 2.	74
Figura 36. Experimento 3: Alg. de Otimização Global - Acurácias de classificação obtidas utilizando-se 250 amostras de treinamento e teste, utilizando-se os critérios 1 e 2.	75
Figura 37. Experimento 3: Alg. de Otimização Global - Acurácias de classificação obtidas utilizando-se 300 amostras de treinamento e teste, utilizando-se os critérios 1 e 2.	76
Figura 38. Experimento 3: Alg. de Otimização Global - Acurácias de classificação obtidas utilizando-se 350 amostras de treinamento e teste, utilizando-se os critérios 1 e 2.	77
Figura 39. Experimento 3: Alg. de Otimização Global - Acurácia média de classificação (4 classes) obtida utilizando-se 150, 200, 250, 300 e 350 amostras de treinamento e teste, utilizando-se os critérios 1 e 2. Apresentam-se os resultados obtidos na extração de 4 a 94 feições.	78

Figura 40. Experimento 3: tempo de processamento necessário para o algoritmo desenvolvido, para diferentes tamanhos de amostras de treinamento e teste para cada classe, frente à dimensionalidade utilizada (Critério 1).....	80
Figura 41. Experimento 4: Alg. Tree - Resultados da classificação, utilizando-se 150 pixels de treinamento e teste, frente ao número de feições extraídas, segundo os critérios 1 e 2.	83
Figura 42. Experimento 4: Alg. Tree - Resultados da classificação, utilizando-se 200 pixels de treinamento e teste, frente ao número de feições extraídas, segundo os critérios 1 e 2.....	84
Figura 43. Experimento 4: Alg. Tree - Resultados da classificação, utilizando-se 250 pixels de treinamento e teste, frente ao número de feições extraídas, segundo os critérios 1 e 2.....	85
Figura 44. Experimento 4: Alg. Tree - Resultados da classificação, utilizando-se 300 pixels de treinamento e teste, frente ao número de feições utilizadas, segundo os critérios 1 e 2.....	86
Figura 45. Experimento 4: Alg. Tree - Resultados da classificação, utilizando-se 350 pixels de treinamento e teste, frente ao número de feições extraídas, segundo os critérios 1 e 2.....	87
Figura 46. Experimento 4: Alg. Tree - Acurácia média da classificação, variando-se o tamanho das amostras de treinamento e teste, frente ao número de feições extraídas, segundo os critérios 1 e 2.....	88
Figura 47. Experimento 4: tempo de processamento necessário para o algoritmo desenvolvido, para diferentes tamanhos de amostras de treinamento e teste para cada classe, frente à dimensionalidade utilizada (Critério 1).....	90
Figura 48. Comparação da acurácia de classificação máxima (média para 4 classes) segundo o algoritmo e tamanho das amostras de treinamento e teste utilizadas.....	92
Figura 49. Comparação da acurácia de classificação média (4 classes) segundo o algoritmo, critério utilizados e 150 pixels de treinamento e teste.....	94
Figura 50. Comparação da acurácia de classificação média (4 classes) segundo o algoritmo, critério utilizados e 200 pixels de treinamento e teste.....	95
Figura 51. Comparação da acurácia de classificação média (4 classes) segundo o algoritmo, critério utilizados e 250 pixels de treinamento e teste.....	96
Figura 52. Comparação da acurácia de classificação média (4 classes) segundo o algoritmo, critério utilizados e 300 pixels de treinamento e teste.....	97
Figura 53. Comparação da acurácia de classificação média (4 classes) segundo o algoritmo, critério utilizados e 350 pixels de treinamento e teste.....	98

LISTA DE ABREVIATURAS

AS	Algoritmo <i>Steepest Ascent</i> ;
AVIRIS	<i>Airborne Visible/Infrared Imaging Spectrometer</i> ;
CAPES	Coordenação de Aperfeiçoamento de Pessoal de Nível superior;
DBFE	<i>Decision Boundary Feature Extraction</i> ;
EM	Algoritmo <i>Expectation Maximization</i> ;
EUA	Estados Unidos da América
FCS	Algoritmo <i>Fast Constrained Search</i> ;
LANDSAT	<i>Land Remote Sensing Satellite</i> ;
MAXVER	Classificador Máxima Verossimilhança;
NASA	<i>National Aeronautics and Space Administration</i> ;
NC	Número de classes
PCA	Análise de componentes principais;
PP	Algoritmo <i>Projection Pursuit</i> ;
SAR	<i>Sinthetic Aperture Radar</i> ;
SBFS	Algoritmo <i>Sequential Backward Floating Selection</i> ;
SBS	Algoritmo <i>Sequential Backward Selection</i> ;
SFFS	Algoritmo <i>Sequential Forward Floating Selection</i> ;
SFS	Algoritmo <i>Sequential Forward Selection</i> ;
SLAR	<i>Side Looking Airbone Radar</i> ;
SPOT	<i>Systeme pour l'Observation de la Terre</i> ;

LISTA DE SÍMBOLOS

N_c	Número de combinações num processo de seleção de feições;
m	Número de feições;
n	Número de bandas presentes na imagem;
ϖ	Classe de interesse;
i	Índice de referência;
C	Número total de classes presentes na imagem;
x	Pixel da imagem;
$P(\omega_i)$	Probabilidade “a priori” de ocorrência da classe ϖ_i . É a probabilidade de ocorrência de um padrão da classe ϖ_i , independente de qualquer outra informação;
$p(x \omega_i)$	Função densidade de probabilidade para x condicional à classe ϖ_i . Representa a distribuição de probabilidades interna na classe ϖ_i . Esta função de probabilidade pode ser estimada através das amostras de treinamento;
$p(\omega_i x)$	Probabilidade “a posteriori” de ocorrência da classe ϖ_i condicionada ao padrão (pixel) x . Geralmente desconhecida e de difícil definição de sua função de probabilidade;
$p(x)$	Função densidade de probabilidade para x . Representa a probabilidade de encontrar um pixel de qualquer classe, na posição x . Pode ser definida por: $p(x) = \sum_{i=1}^k p(x \omega_i) p(\omega_i);$
CD's	Números ou contadores digitais. Valor numérico atribuído ao pixel em cada banda da imagem;
$g_i(x)$	Função discriminante da classe ϖ_i para fins de classificação;
N	Número de bandas ou feições;
m_i	Vetor média para a classe ϖ_i ;

Σ_i	Matriz de covariância para a classe ϖ_i ;
N_k	Número de amostras para a classe k ;
b_{ij}	Distância de Bhattacharyya entre as classes ϖ_i e ϖ_j ;
L	Comprimento dos segmentos de comprimento constante;
L'	Comprimento do segmento com as bandas “excedentes” aos de comprimento constante L ;
k	Número de segmentos;
NFS	Número de feições por segmento;
NF	Número de feições obtidas no processo de segmentação;
nps	Número máximo de posições de segmentação que devem ser analisadas para a geração de k segmentos;
σ^2	Variância dos dados.

1 - INTRODUÇÃO

1.1 - Perspectivas históricas: o início do sensoriamento remoto na era espacial

As técnicas de sensoriamento remoto vinham sendo empregadas já há algum tempo, utilizando dispositivos aerotransportados para fins de coleta de dados relativos à superfície da Terra. Foi entretanto, com o início da era espacial, que pode ser associada ao lançamento do satélite terrestre denominado de *Sputnik* pela antiga União Soviética em Novembro de 1957 e com a criação da NASA (*National Aeronautics and Space Administration*) pelos Estados Unidos, e com o lançamento, três anos mais tarde, do satélite denominado de *Tiros*, que se criaram as condições para o desenvolvimento de sistemas espaciais com fins específicos de sensoriamento remoto [LAN 03].

As operações relacionadas ao sensoriamento remoto da Terra são tidas como de alto impacto econômico, já que é na superfície do planeta que as pessoas vivem e que se desenvolvem a maior parte das atividades econômicas, associadas aos recursos naturais presentes. Desta forma, o potencial e a utilidade do uso do sensoriamento remoto foram tidos como de grande importância.

A principal diretriz que conduziu ao desenvolvimento de tais tecnologias foi portanto a econômica [LAN 03]. Coletar informações necessárias ao manejo dos recursos naturais a partir do solo é um processo lento e demorado, além de bastante oneroso. O uso de sensores aerotransportados aumenta a velocidade do processo e reduz seu custo. A partir do espaço, entretanto, o unitário é ainda mais reduzido, a área abrangida é superior e o tempo necessário diminui, especialmente quando se leva em consideração o caráter repetitivo, periódico na aquisição de dados, característica essencial em vários tipos de aplicações em sensoriamento remoto.

A obtenção de informação, em termos de concepção de sensores e sistemas de dados, foi discutida sob dois aspectos: (a) aquisição de informação sob forma de imagens de alta resolução espacial, isto é, dados priorizando informação de natureza espacial e (b) priorizando a informação de natureza espectral.

Sistemas de aquisição de imagens teriam um uso mais imediato se as imagens obtidas tivessem uma alta resolução espacial, ou seja, que os usuários pudessem “ver” detalhes de seu interesse na cena. Entretanto, a resolução espacial é um parâmetro de custo muito elevado no processo de coleta a partir do espaço. Por exemplo, para identificar uma espécie vegetal de milho, seria necessário uma resolução espacial de alguns centímetros, de modo que a forma da folha da cultivar milho pudesse ser convenientemente visualizada pelo usuário. Por outro lado, métodos espectrais, cujos estudos já haviam iniciado na época do início do sensoriamento remoto espacial, mostraram serem eficientes na solução do mesmo problema, sendo neste caso suficiente médios ou mesmo baixos níveis de resolução espacial. Desta forma, o custo para a coleta dos dados é menor e a redução no volume de dados por unidade de área é bastante significativa, reduzindo assim também o custo de processamento computacional. De uma maneira geral, esta tendência tem se mantido entre a comunidade que opera nesta área de sensoriamento remoto. Atualmente começam a ser disponibilizadas, para fins comerciais, imagens de uma nova geração de sensores ditos hiperespectrais. Estes sensores possibilitam uma resolução espectral muito superior aos tradicionais sensores que vinham sendo colocados em órbita terrestre, porém vale-se do conceito inicialmente definido: a identificação de alvos através de suas características espectrais.

1.2 - Sensoriamento Remoto: Conceitos Básicos

Sensoriamento remoto pode ser definido como o conjunto de técnicas e instrumentos destinados a estender e melhorar a capacidade perceptiva do olho humano, fornecendo informações qualitativas e quantitativas sob objetos colocados a próxima (*proximal sensing*) ou a grande (*remote sensing*) distância do ponto de observação [SWA 78].

O Sensoriamento Remoto representa um instrumento prático e sistemático para manter atualizadas informações geo-espaciais da superfície terrestre. Em particular, para múltiplos e diversificados setores aplicativos, por exemplo:

- (i) Agricultura: gestão dos processos produtivos, verificação de tipologia das culturas, previsão de colheitas, avaliação de danos pós-calamidades;
- (ii) Ciências florestais: atualizações cartográficas florestais, monitoramento de uso e cobertura do solo, monitoramento da poluição, controle e gestão de ecossistemas;
- (iii) gestão de riscos: monitoramento de catástrofes (enchentes, erosões e terremotos) e estimativa dos danos produzidos;
- (iv) defesa do território;
- (v) mar e áreas costeiras: estudos de fenômenos de erosão costeira, controle de rotas náuticas, etc.

A aquisição de informações sobre o objeto de interesse, que neste caso é uma determinada área da superfície terrestre, é possível graças ao uso de instrumentos ditos sensores, montados a bordo de satélites ou aviões, os quais fornecem medidas de grandezas eletromagnéticas (medidas normalmente em termos de radiância) geradas a partir da interação entre a radiação eletromagnética incidente proveniente de uma fonte, natural ou artificial, com objetos na superfície terrestre. A representação do espectro eletromagnético é apresentada na Figura 1, onde salienta-se a região correspondente a emissão de ondas eletromagnéticas pelo sol, a principal fonte de energia presentemente utilizada em sensoriamento remoto (Espectro Solar).

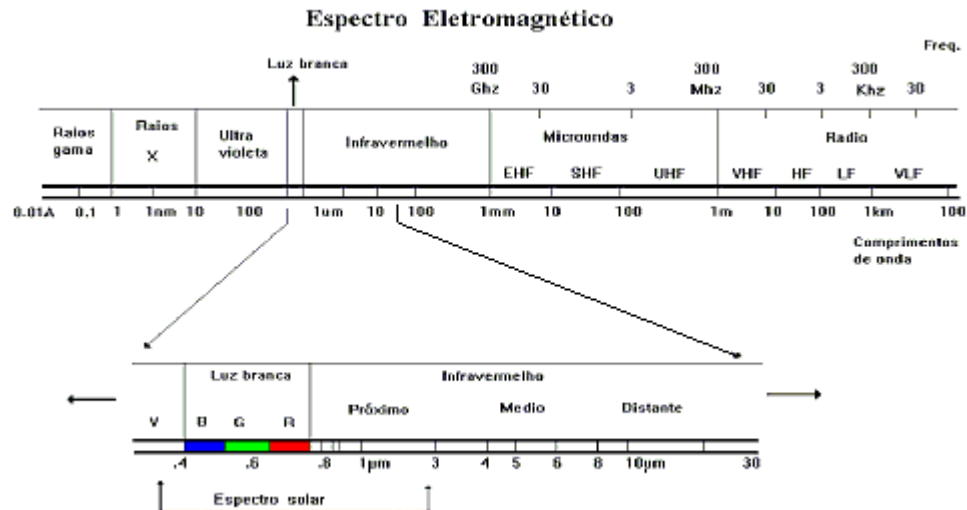


Figura 1. Espectro eletromagnético e espectro solar.

A informação é normalmente fornecida sob forma de medidas pontuais ou imagens digitais (arranjo matricial). Em particular, a respeito dos sensores imageadores, sua tipologia e a plataforma sobre a qual são instalados determinam o grau de detalhe com a qual os dados obtidos cobrem a área de interesse em termos de resolução espacial (área mínima visível representada pelo pixel), radiométrica (intervalo entre energia mínima e máxima mensurável), espectral (número e largura das bandas espectrais implementadas no sistema sensor) e temporal (intervalo de tempo que transcorre entre duas medidas consecutivas do mesmo local).

Supondo que seja de interesse medir a energia espontaneamente emitida da superfície terrestre (ligada a sua temperatura) ou aquela emitida pelo Sol e refletida pela Terra, se deverão considerar, respectivamente, intervalos de comprimento de ondas eletromagnéticas na região do infravermelho e no visível (0,4 a 0,7 micron). Este é o caso dos sensores óticos que, podendo obter instantaneamente várias vistas da mesma área de interesse (cena), cada uma delas associada um determinado intervalo de comprimento de onda (banda), tornam possível a obtenção da curva de resposta espectral. Essa curva de resposta espectral é definida como o valor da radiância em função do comprimento de onda do sinal captado. É uma grandeza característica das diversas tipologias de cobertura de solo e então útil para a sua discriminação na cena de interesse imagiada (Figura 2).

Como exemplo de sensores óticos tradicionais pode-se citar aqueles a bordo da série de satélites LANDSAT (*Land Remote Sensing Satellite*) e SPOT (*Système pour l'Observation de la Terre*). Tipicamente, estes sensores são ditos multiespectrais, com

um número de bandas inferior a uma dezena. Portanto, a cada pixel são atribuídas um conjunto de medidas quantitativas, cada uma correspondendo a uma das bandas, representando a resposta da interação da superfície com a radiação. Estas medidas passarão a quantificar estatisticamente cada pixel.

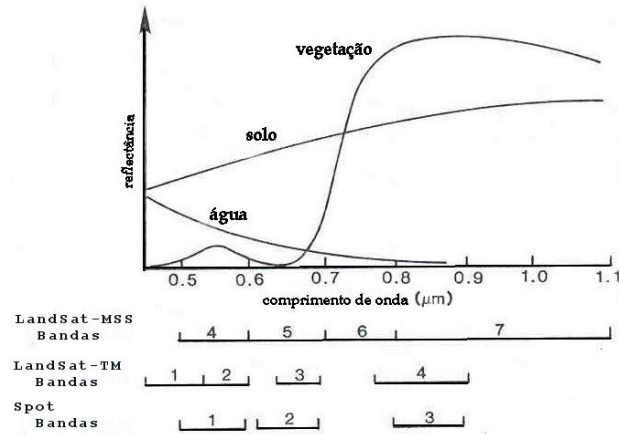


Figura 2. Curvas de resposta espectral típicas do solo, vegetação e da água. Quanto maior a resolução espectral e radiométrica do sensor, melhor serão descritas as curvas de resposta espectral dos alvos imageados.

A última fase do processo de sensoriamento remoto, do ponto de vista de elaboração dos dados, é a análise. Esta fase, destinada a extrair informações de interesse para uma dada aplicação específica, vem sendo executada utilizando-se técnicas de teoria do reconhecimento (*Pattern Recognition*), como por exemplo, técnicas de classificação (estatística, fuzzy, neural, etc). As técnicas de classificação podem ser genericamente divididas em duas famílias: técnicas supervisionadas e não supervisionadas. Na abordagem do tipo supervisionado, o analista deve fornecer informações ao classificador, como amostras de cada uma das classes espectrais na cena sendo analisada. No processo não supervisionado, o próprio algoritmo procura agrupar pixels com características semelhantes (segundo um determinado critério) em aglomerados que em princípio deveriam corresponder a classes espectrais existentes na cena. Neste caso a associação entre aglomerados e classes no terreno deverá ser feita a posteriori pelo analista.

Os parágrafos precedentes forneceram um resumo bastante sintético de alguns conceitos básicos em sensoriamento remoto. Não é o objetivo deste trabalho uma abordagem em maiores detalhes. Para tanto se sugerem as bibliografias [SWA 78] e [LAN 03].

1.3 - Sensoriamento Remoto: a Era Hiperespectral

Como consequência do avanço tecnológico no campo de sensores, as imagens em alta dimensionalidade têm se tornado cada vez mais disponíveis em Sensoriamento Remoto. A análise e interpretação destes dados provenientes desta nova geração de sensores como, por exemplo, o sensor AVIRIS (220 bandas que cobrem o espectro eletromagnético de 0.41 a 2.45 μm) apresenta novas possibilidades e também novos desafios. Imagens coletadas em um grande número de bandas espectrais contêm informações mais detalhadas com respeito às características espectrais das classes a serem identificadas, permitindo uma acurácia maior na classificação dos pixels do que aquela que seria possível com a utilização de imagens mais tradicionalmente utilizadas como aquelas fornecidas por sensores dos sistemas LANDSAT e SPOT, que fornecem dados espectrais em baixa dimensionalidade.

Nestas imagens, denominadas de hiperespectrais, a dimensionalidade dos dados (igual ao número de bandas espectrais) é muito alta, por exemplo, em torno de 200 variáveis por pixel. Este alto valor para a dimensionalidade dos dados, aliada ao número geralmente insuficiente de amostras de treinamento (na maioria das aplicações de sensoriamento remoto a aquisição de um grande número de amostras de treinamento é onerosa e de difícil aquisição) torna altamente problemática a estimação dos parâmetros necessários ao classificador, em especial, a matriz covariância.

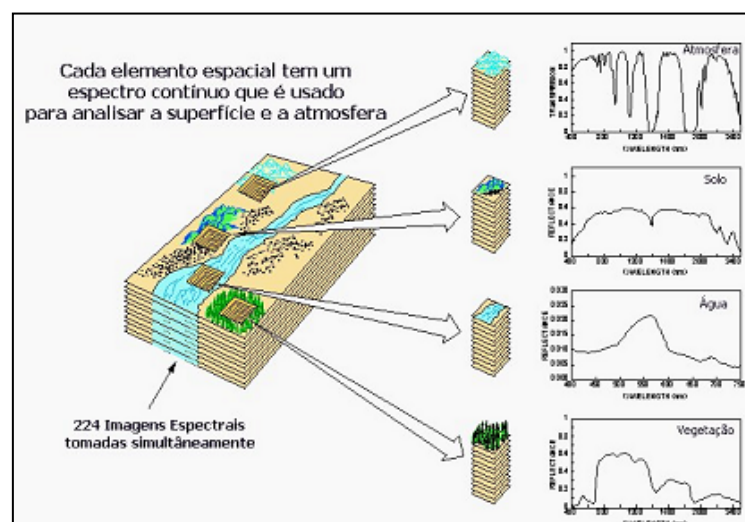


Figura 3. Representação da curva de resposta espectral adquirida por um sistema hiperespectral, neste caso o AVIRIS. O elevado número de bandas e sua pequena largura permitem uma descrição detalhada da resposta espectral de cada pixel presente na cena [JLP 03].

Neste tipo de imagem cada pixel pode ser representado por uma curva espectral praticamente contínua, como aquela ilustrada na Figura 3. A quantidade de energia refletida por cada pixel é medida sob a forma de radiância a intervalos de comprimento de onda, bastante estreitos (da ordem de 0.005 a 0.010 microns) e contínuos ao longo da porção do espectro eletromagnético denominada de região óptica do espectro, tipicamente entre 0.4 microns e 2.3 microns de comprimento de onda.

Tabela 1. Sensores Hiperespectrais atuais [SHI 03].

Sensores Satelitares	Construtor	Número de bandas	Comprimentos de onda (em microns)
FTHSI no MightySat II	Air Force Research Lab www.vs.af.mil/ TechProgs/MightySatII	256	0.35–1.05
Hyperion on EO-1	NASA Goddard Space Flight Center eo1.gsfc.nasa.gov	220	0.4–2.5
Sensores aero-transportados	Construtor	Número de bandas	Comprimentos de onda (em microns)
AVIRIS (<i>Airborne Visible Infrared Imaging Spectrometer</i>)	NASA Jet Propulsion Lab makalu.jpl.nasa.gov/	224	0.4–2.5
HYDICE (<i>Hyperspectral Digital Imagery Collection Experiment</i>)	Naval Research Lab	210	0.4–2.5
PROBE-1	Earth Search Sciences Inc. www.earthsearch.com	128	0.4–2.5
CASI (<i>Compact Airborne Spectrographic Imager</i>)	ITRES Research Limited www.itres.com	até 228	0.4–1.0
HyMap	Integrated Spectronics www.intspec.com	100 a 200	Visível ao infravermelho termal.
EPS-H (<i>Environmental Protection System</i>)	GER Corporation www.ger.com	VIS/NIR (76), SWIR1 (32), SWIR2 (32), TIR (12)	VIS/NIR (.43–1.05), SWIR1 (1.5–1.8), SWIR2 (2.0–2.5), e TIR (8–12.5)
DAIS 7915 (<i>Digital Airborne Imaging Spectrometer</i>)	GER Corporation	VIS/NIR (32), SWIR1 (8), SWIR2 (32), MIR (1), TIR (6)	VIS/NIR (0.43–1.05), SWIR1 (1.5–1.8), SWIR2 (2.0–2.5), MIR (3.0–5.0), e TIR (8.7–12.3)
DAIS 21115 (<i>Digital Airborne Imaging Spectrometer</i>)	GER Corporation	VIS/NIR (76), SWIR1 (64), SWIR2 (64), MIR (1), TIR (6)	VIS/NIR (0.40–1.0), SWIR1 (1.0–1.8), SWIR2 (2.0–2.5), MIR (3.0–5.0), e TIR (8.0–12.0)
AISA (<i>Airborne Imaging Spectrometer</i>)	Spectral Imaging	até 288	0.43–1.0

Nota: VIS = Visível; NIR = Infra-vermelho próximo; SWR = Infra-vermelho de ondas curtas; MIR = Infra-vermelho médio e TIR = Infra-vermelho termal.

Para melhor ilustrar a riqueza de detalhes fornecidas pelos sensores hiperespectrais (Tabela 1) se comparados aos tradicionais multiespectrais, apresentam-se as Figura 4 e 5, contendo duas curvas de resposta espectral obtidas da mesma cena de estudo. Trata-se de um exemplo onde é apresentada uma amostra em laboratório contendo uma mistura de três tipos de minerais. Cada um deles possui uma molécula distinta que absorve energia em regiões bem distintas do espectro eletromagnético.

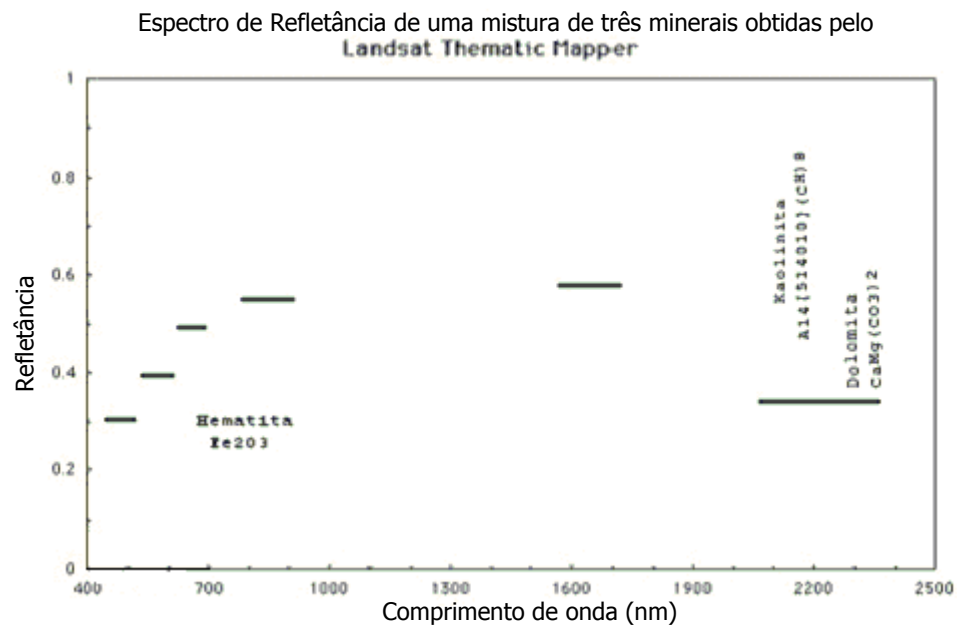


Figura 4. Espectro de refletância de uma amostra contendo 3 minerais diferentes (Hematita, Kaolinita e Dolomita), obtida pelo sensor LANDSAT- TM [JLP 03].

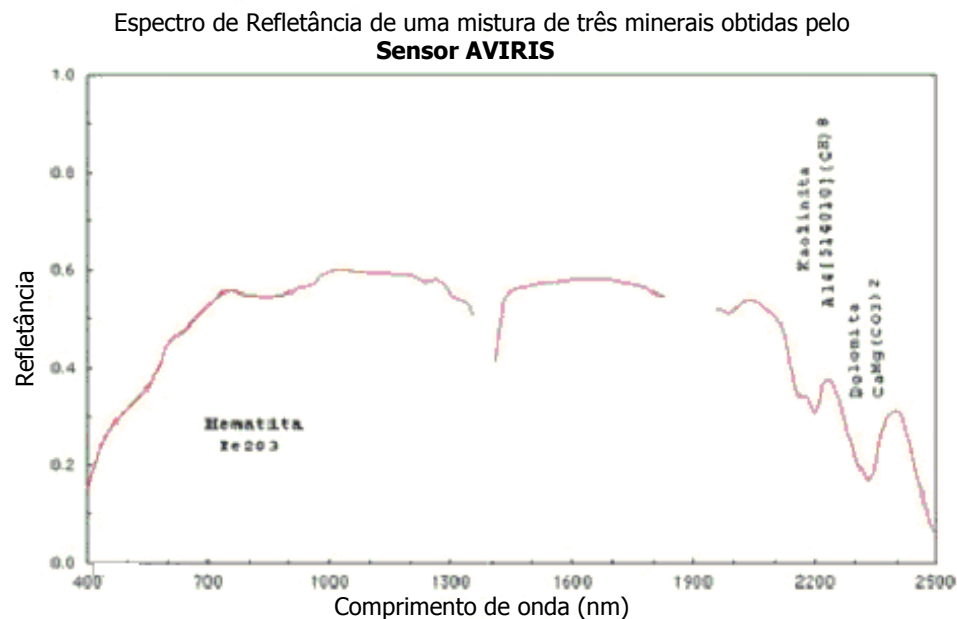


Figura 5. Espectro de refletância de uma amostra contendo três minerais diferentes (Hematita, Kaolinita e Dolomita), obtida pelo sensor AVIRIS [JLP 03].

Percebe-se que o sensor multiespectral TM a bordo do satélite LANDSAT descreve a curva de resposta espectral de maneira muito aproximada: o número de bandas é pequeno e elas são demasiado largas, inviabilizando a coleta de informações detalhadas, muitas vezes essenciais ao problema em estudo. Já na resposta espectral obtida pelo sensor AVIRIS (Figura 5), a resolução espectral é suficiente para que a presença dos minerais de interesse possa ser facilmente identificável, tornando possível, por exemplo, uma estimativa de suas concentrações relativas.

1.4 - Apresentação do Problema

A utilização prática de dados espectrais em alta dimensionalidade, implica entretanto, em alguns custos, como por exemplo custo computacional, geralmente elevado, requerido para fins de processamento e classificação destes dados. Mas, o problema maior que na prática acontece refere-se ao número quase sempre insuficiente de amostras de treinamento, necessárias à estimação dos parâmetros em classificadores paramétricos. Quando o número de amostras de treinamento é pequeno comparado com a dimensionalidade dos dados, as estimativas dos parâmetros do classificador tornam-se pouco confiáveis resultando num desempenho insatisfatório no processo de classificação. Normalmente, em um problema de classificação, é esperado que a acurácia dos resultados cresça na medida em que o número de bandas utilizadas (e, portanto, a quantidade de informação disponível) aumenta. A experiência mostra, entretanto, que este crescimento apresenta limites. Uma tendência comumente verificada consiste em um aumento inicial no valor da acurácia dos resultados a medida que o número de bandas utilizadas, isto é, a dimensionalidade dos dados, começa a aumentar. Em um determinado momento a acurácia atinge um valor máximo e após, começa a decrescer na medida em que mais informação, isto é, novas bandas espectrais é adicionada ao processo. Este fenômeno bem conhecido recebe o nome de “fenômeno de Hughes” (Figura 6) e deve-se à degradação na estimação dos parâmetros do classificador causado pelo número insuficiente de amostras de treinamento.

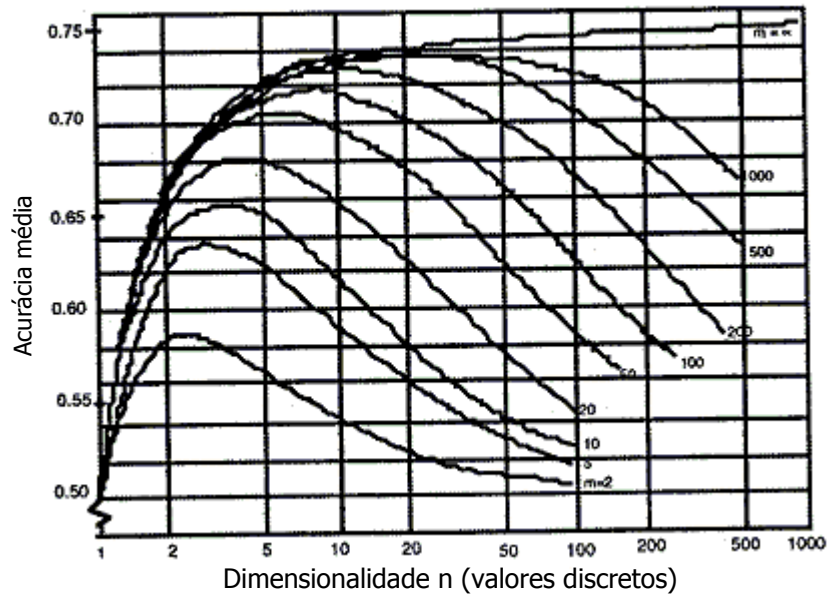


Figura 6. Representação típica do fenômeno de Hughes. Cada um dos ramos da curva representa o desempenho típico de classificação frente a uma determinada quantidade m de amostras de treinamento.

A solução para este problema está, atualmente, centrada em duas abordagens: (a) introdução de amostras não rotuladas (isto é, pixels cuja classe a que pertencem não é conhecida) no processo de estimação dos parâmetros através de utilização de técnicas como, por exemplo, o algoritmo *Expectation-Maximization* (EM), além de outras possíveis e (b) redução do número de feições fornecidas como entrada no classificador utilizado. Como um sensor hiperespectral adquire dados em bandas muito próximas e, portanto, correlacionadas, é possível reduzir o número de feições sem que haja perda significativa de informação [FUK 90]. Reduz-se desta forma os efeitos do fenômeno de Hughes e conseqüentemente melhoram-se os resultados fornecidos pelo classificador em relação àqueles obtidos utilizando diretamente número total de bandas originais.

O problema de redução do número de feições (redução na dimensionalidade dos dados) pode ser resolvido por duas abordagens, respectivamente baseadas na (a) seleção de um subconjunto de feições com maior poder discriminante, entre o total de bandas disponíveis, ou (b) extração de um número limitado de feições após uma transformação num novo espaço de feições [SER 03]. Desta forma, o termo “feição” tanto pode ser utilizado para designar bandas originais (seleção de feições) como também uma combinação destas (extração de feições).

Em imagens multiespectrais e para fins de classificação, cada pixel é normalmente representado por um vetor. Os elementos que compõem este vetor são as

respostas individuais fornecidas por cada uma das bandas espectrais disponíveis. Esta informação fornecida pelo conjunto das bandas espectrais utilizadas, pode também ser entendida como sendo a chamada curva de resposta espectral. Esta curva serve para caracterizar cada pixel individual. Desta forma, pixels que pertencem a uma mesma classe apresentam curvas de resposta espectral semelhantes. O objetivo deste estudo é investigar metodologias simples para redução da dimensionalidade dos dados por meio de extração de feições que ofereçam resultados confiáveis, sejam fáceis de implementar e eficientes do ponto de vista computacional. Ao invés de implementar um procedimento de procura objetivando encontrar um subconjunto ótimo ou sub-ótimo dentro do conjunto total de bandas espectrais disponíveis de acordo com algum dado critério, examina-se a curva de resposta espectral de cada pixel individualmente, isto é, o conjunto de todas as bandas disponíveis. Primeiramente captura-se, para cada pixel individualmente, a informação contida pelo conjunto total de bandas espectrais disponíveis, isto é, a curva espectral de cada pixel, substituindo as bandas espectrais por um pequeno número de estatísticas simples que objetivam descrever a posição e forma da resposta espectral do pixel.

A suposição básica é que a curva da resposta espectral de cada pixel pode ser particionada em segmentos e que cada segmento pode ser representado por um número menor de estatísticas (feições) sem muita perda de informação. Em outras palavras, esta operação consiste num processo de redução da dimensionalidade dos dados originais. Pelo que se constatou, a abordagem desenvolvida neste trabalho é original.

Este trabalho está assim organizado: no Capítulo 2 - Revisão Bibliográfica, procura-se situar o leitor no estado da arte, apresentando as principais técnicas que vem sendo utilizadas em sensoriamento remoto para redução da dimensionalidade em dados hiperespectrais. No Capítulo 3 - Metodologia, apresentam-se as novas idéias que são fruto deste trabalho. Posteriormente estas idéias são validadas através de experimentação, cuja descrição e resultados são apresentados e analisados no Capítulo 4 - Estudo de Caso. O Capítulo 5 - Conclusões e Recomendações - finaliza o trabalho confirmando os bons resultados obtidos com a utilização das novas metodologias para redução da dimensionalidade aqui propostas.

2 - REVISÃO BIBLIOGRÁFICA

2.1 - Metodologias para aumento da acurácia da classificação de dados em alta dimensionalidade: O estado da arte

2.1.1 - Introdução de amostras não rotuladas

A inclusão de amostras não rotuladas junto as rotuladas no conjunto de treinamento (*training samples*) contribui significativamente para a melhoria das estimativas da matriz de covariância das classes. Shahshahani e Landgrebe [SHA 94] apresentam um estudo detalhado de como os inconvenientes causados por amostras de treinamento em número insuficiente podem ser minorados pela introdução de amostras não rotuladas.

Utilizando o algoritmo “*Expectation Maximization*” (EM) eles conseguiram melhoria na acurácia da classificação. Nesta abordagem, o número de amostras disponíveis para a estimação dos parâmetros em cada classe é incrementado através da adição de um determinado número de amostras ditas “não rotuladas”, isto é, amostras (pixels) selecionados aleatoriamente ao longo da imagem, e cuja classe não é conhecida. O algoritmo EM introduz estas novas amostras, de uma forma ponderada, na estimação dos parâmetros necessários ao classificador (probabilidade a priori, vetor média e matriz covariância). O processo é repetido de forma iterativa até que um determinado critério de convergência tenha sido atingido. Os resultados, assim obtidos, demonstram que é possível um aumento da acurácia de classificação.

Uma outra abordagem, proposta por Jackson e Landgrebe [JAK 01], apresenta alguma similaridade com a abordagem anterior. Estes autores propõem também a adoção de um processo iterativo na estimação dos parâmetros do classificador. Em um passo inicial, os parâmetros estimados a partir das amostras

disponíveis são implementados no classificador e a imagem é classificada. Dado a pobreza nesta estimativa inicial, os resultados do processo de classificação são pouco confiáveis e necessitam serem refinados. Neste processo, estima-se, inicialmente, o grau de pertinência de cada pixel à classe na qual ele foi alocado. Este valor, é estimado pela razão entre o valor da função decisão assumida pela classe vencedora e o somatório dos valores de todas as funções decisão das classes em consideração. Note que esta estimativa varia no intervalo [0,1]. Em um segundo passo, adiciona-se um número de pixels às amostras originais de cada classe, ponderados pelo grau de pertinência que foi estimado conforme descrito acima. Um valor atualizado para os parâmetros do classificador pode, então, ser obtido. Este processo é continuado de uma forma iterativa, até que um determinado critério de convergência tenha sido atingido. Desta forma, pode-se também melhorar a acurácia de classificação, sendo os resultados finais similares aos obtidos utilizando-se a técnica anteriormente apresentada.

2.1.2 - Seleção de feições

Talvez a maneira mais direta de reduzir a dimensionalidade do problema seja simplesmente selecionar um subconjunto de bandas espectrais. Infelizmente, mesmo para um número pequeno de bandas originais, uma procura exaustiva torna-se computacionalmente proibitiva [JAI 97], visto que o número de combinações N_c possíveis para a seleção de um subconjunto contendo m feições entre um total de n bandas é dada pelo coeficiente binomial:

$$N_c = \binom{n}{m} = \frac{n!}{(n-m)!m!} \quad (1)$$

Desta forma, as técnicas de seleção de feições geralmente envolvem um algoritmo de procura associado a uma função critério [SWA 78] e [JAI 97]. O algoritmo de procura gera e compara possíveis “soluções” para o problema de seleção de feições (isto é, subconjunto de feições), aplicando uma função critério como medida da eficácia de cada solução.

Diversas abordagens sub-ótimas para seleção de feições têm sido propostas [BRU 95], [BRU 00] e [KIT 78]. As estratégias mais simples de procura sub-ótimas são os algoritmos “*Sequential Forward Selection*” (SFS) e o “*Sequential Backward*

Selection” (SBS), [JAI 97] e [PUD 94]. Estes algoritmos identificam iterativamente o melhor subconjunto de feições que pode ser obtido pela adição ou remoção, respectivamente, de uma feição por vez ao subconjunto selecionado. Infelizmente, ambos algoritmos têm um sério inconveniente: uma vez que a feição foi selecionada ela não pode ser descartada (SFS) ou, uma vez que a feição foi descartada ela não pode ser re-selecionada (SBS), [SER 03].

Outros algoritmos, como o “*Sequential Forward Floating Selection*” (SFFS) e o “*Sequential Backward Floating Selection*” (SBFS) trouxeram uma melhora aos algoritmos SFS e SBS, permitindo reconsiderar as feições incluídas ou removidas numa iteração previa.

Os dois algoritmos também de procura dita sub-ótima, apresentados por Serpico et al [SER 01] chamados de “*Steepest Ascent*” (AS) e “*Fast Constrained Search*” (FCS), permitem diferentes compromissos entre a eficiência das feições selecionadas e o tempo computacional requerido para encontrar uma solução, proporcionando melhores soluções do que o obtido com o algoritmo SFFS. Entretanto, apenas um máximo local de acurácia de classificação é atingido.

Em geral, com a utilização de métodos de seleção de feições, obtém-se um bom resultado, mas não é necessariamente o melhor que poderia ser obtido para o conjunto de bandas originais disponíveis.

Diversos outros métodos baseados em conceitos atrativos como “*graph-searching*” [ICH 84], “*genetic algorithms*” [SIE 89], “*simulated annealing*” [SIE 88], “*spectral mixture analysis*” [ADA 86], [HU 99] e [CHA 00], “*combined classification*” [BEN 99] e [JIM 99] e outros [STE 76] e [BAC 20] também já foram explorados na literatura [SER 03].

2.1.3 - Extração de feições

O principal objetivo das técnicas de extração de feições é reduzir a dimensionalidade dos dados, que minimize a perda de informações, através de uma transformação conveniente das feições.

2.1.3.1 Transformada de componentes principais

A transformada de componentes principais busca uma projeção dos dados num novo sistema de eixos tal que suas componentes não apresentem correlação entre

si. Ao fazer esta transformação, produz-se um novo espaço onde a maior variância dos dados está representada na primeira componente, a segunda maior variância na segunda componente, e assim sucessivamente. Espera-se, em geral, uma pequena variância ao longo das últimas componentes, podendo ser consideradas pouco contribuintes para a separabilidade dos dados. Desta forma, pode-se pensar em eliminá-las, reduzindo portanto a dimensionalidade dos dados.

A transformada de componentes principais é baseada na matriz de covariância global do conjunto de classes, e desta forma não é sensível à estrutura de cada classe individualmente. Geralmente esta abordagem funciona bem como ferramenta de redução da dimensionalidade porque, em dados provenientes de sensoriamento remoto, as classes são geralmente distribuídas na direção do máximo espalhamento dos dados (*maximum data scatter*).

2.1.3.2 Análise canônica

A técnica de componentes principais é útil apenas se os dados estiverem distribuídos substancialmente ao longo das primeiras componentes. Se os dados não apresentarem esta característica, possivelmente não se tem um ganho ao se utilizar esta transformada ao invés dos dados originais. Em tais condições, é possível que a técnica de análise canônica apresente melhores resultados.

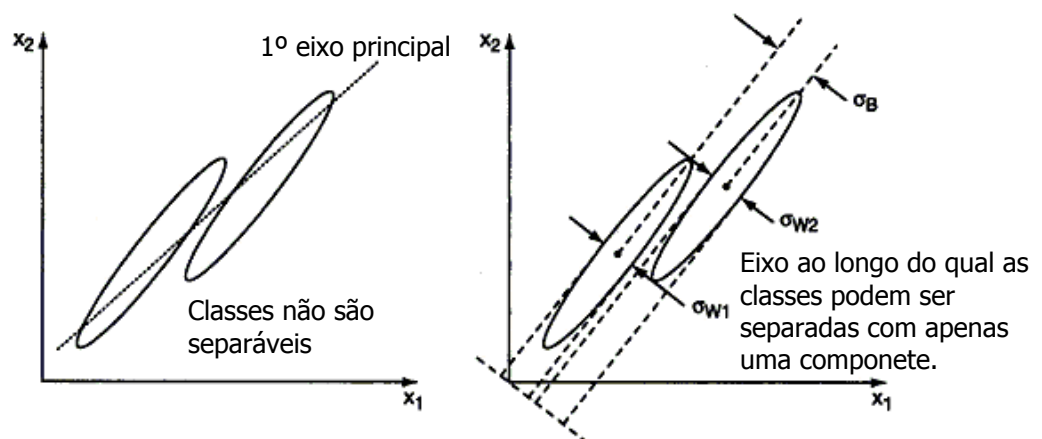


Figura 7. Distribuição hipotética de duas classes num espaço de duas dimensões.

- Na Figura 7 (à esquerda), percebe-se que as classes não são separáveis utilizando-se apenas uma feição, seja ela no espaço original ou utilizando componentes principais;

- A direita da Figura 7 percebe-se que as classes poderiam ser separadas utilizando-se apenas uma feição se o eixo apresentado pudesse ser encontrado. O problema está, então, em encontrar este eixo.

Uma inspeção na Figura 7 revela que o eixo desta nova transformada deveria ser orientado de tal forma que as classes tenham a máxima separabilidade entre suas médias, em relação ao eixo procurado, ao mesmo tempo minimizando ao máximo possível a dispersão individual das classes. Se a primeira for caracterizada por σ_B^2 e a segunda por σ_{w1} e σ_{w2} , então é necessário encontrar o novo eixo tal que

$$\frac{\sigma_B^2}{\left(\frac{\sigma_{w1} + \sigma_{w2}}{2}\right)^2} = \frac{\text{var entre classes}}{\text{var média dentro classes}} \quad \text{seja maximizado.} \quad (2)$$

Os novos vetores para definição das feições são escolhidos de modo a maximizar este critério.

Em forma matricial: seja Σ_w a matriz de dispersão dentro das classes (*within-class scatter matrix*) definida como sendo a média entre σ_{w1} e σ_{w2} . Seja Σ_B a matriz de dispersão entre classes (*between-class scatter matrix*). Em forma matricial, Σ_w e Σ_B são dadas por [LAN 03]:

$$\begin{aligned} \Sigma_w &= \sum_i P(\omega_i) \Sigma_i \\ \Sigma_B &= \sum_i P(\omega_i) (\mu_i - \mu_o)(\mu_i - \mu_o)^t \\ \mu_o &= \sum_i P(\omega_i) \mu_i \end{aligned} \quad (3)$$

Onde μ_i , Σ_i e $P(\omega_i)$ são o vetor média, a matriz de covariância, e a probabilidade a priori da classe ω_i , respectivamente. O critério de otimização para a obtenção das componentes pode ser definido como sendo:

$$J = \text{tr}(\Sigma_w^{-1} \Sigma_B) \quad (4)$$

2.1.3.3 *DBFE: Decision Boundary Feature Extraction*

A “*Decision Boundary Feature Extraction*” (DBFE) [LEE 93] se interessa por procurar o melhor conjunto de feições para uma dada dimensionalidade e também pela dimensionalidade ótima a ser utilizada. Utilizam-se, diretamente, as amostras de treinamento para localizar a fronteira de decisão entre classes, e não as estatísticas provenientes destas amostras. Define-se uma base de um possível novo espaço de feições. Representa-se cada pixel, omitindo uma das componentes desta base, e verifica-se se a classificação é alterada. Se permanecer a mesma para todos os pixels diz-se que esta componente é redundante, já que a classificação não muda. Emprega-se um conceito de fronteira “efetiva” de decisão, utilizando-se a regra de decisão de Bayes. Assim uma dimensionalidade intrínseca é computada e define-se uma transformação que possibilita o cálculo de feições ditas ótimas. A eficiência deste método, computacionalmente bastante oneroso, depende fortemente do tamanho da amostra de treinamento. Para lidar com este problema, Jimenez *et al.* [JIM 99] acoplaram o algoritmo DBFE com a técnica denominada “*Projection Pursuit*” (PP), que realiza os cálculos num espaço de menor dimensionalidade, otimizando uma função que os autores denominaram de “índice de projeção”. Entretanto, somente um máximo local do índice de projeção pode ser encontrado e, conseqüentemente, obtém-se uma solução que ao final é um conjunto sub-ótimo de feições [JIM 99].

2.1.3.4 *Agrupamento de feições*

Outra abordagem que a literatura apresenta é agrupar as feições originais em subgrupos de bandas altamente correlacionadas, transformando separadamente as feições de cada grupo. Desta forma, permite-se trabalhar em espaços de menor dimensionalidade. Agrupamento de feições combinadas com análise de componentes principais (PCA) foi a abordagem proposta por Jia *et al.* [JIA 99], e agrupamento de feições combinadas com a transformação de Fisher foi proposta por Kumar *et al.* [KUM 01].

Em geral, a utilização de feições obtidas a partir de métodos de extração de feições (assim como nos de seleção de feições) conduz a um bom resultado em termos de acurácia de classificação. Entretanto, este não é necessariamente o melhor que poderia ser obtido para o conjunto de bandas originais disponíveis.

2.2 - Efeito do número de amostras na estimação de parâmetros

Mesmo que conhecido desde o início da época do processamento digital de imagens provenientes de sensoriamento remoto, o fenômeno de Hughes não causou maiores problemas até o advento dos sensores hiperespectrais. Resumidamente, uma relação mínima entre o número de amostras de treinamento e número de bandas espectrais é necessária para assegurar que sejam obtidas estimativas confiáveis para o treinamento do classificador (num processo supervisionado). Na medida que a dimensionalidade dos dados aumenta, o número de pixels de treinamento necessários por classe também deve ser aumentado para que sejam preservadas as estimativas estatísticas. Assim, a adição de mais bandas espectrais, como no caso do sensor AVIRIS, não é útil a menos que mais amostras de treinamento (pixels) por classe estejam disponíveis [RIC 99]. Isto demonstra uma das maiores limitações da aplicação de procedimentos usuais de classificação em imagens hiperespectrais.

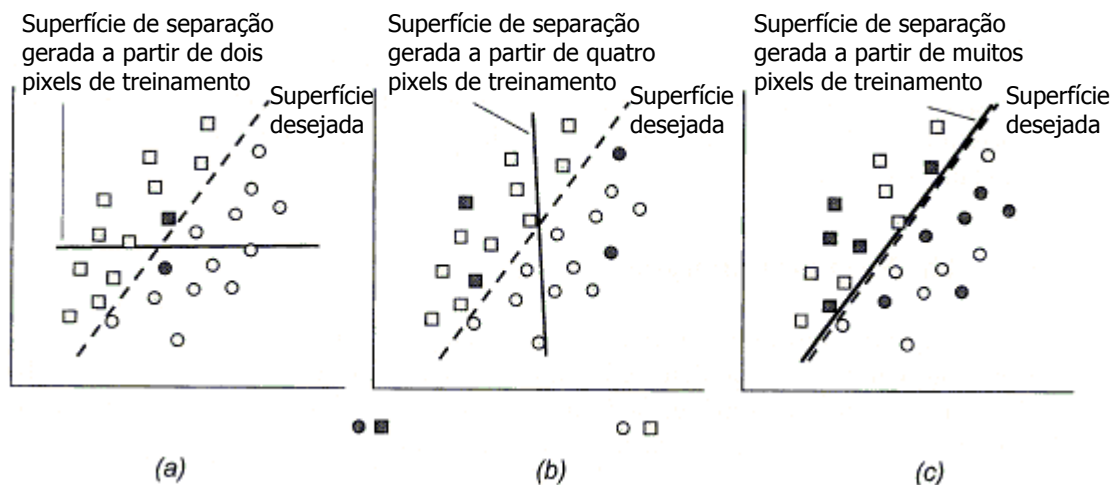


Figura 8. Ilustração da importância de um número de pixels de treinamento suficiente para assegurar uma estimativa confiável da superfície de separação das classes. Quando um número de pixels muito pequeno é utilizado (a) pode-se obter uma boa separação dos pixels de treinamento, porém o desempenho do classificador para o conjunto de pixels de teste é pobre. Em (b) a situação persiste. Uma grande quantidade (posicionados aleatoriamente) de pixels de treinamento gera uma superfície que também é satisfatória para as amostras de teste (c) [RIC 99].

Um exemplo simples, baseado apenas na tentativa de terminação da superfície de separação de duas classes, pode ser utilizado para ilustrar o problema. A Figura 8 mostra três diferentes conjuntos de amostras de treinamento para o mesmo conjunto de dados bidimensional. O primeiro tem apenas dois pixels por classe. Conforme pode ser observado, mesmo que a superfície de separação possa ser

encontrada, ela pode não ser acurada. Tendo-se dois pixel por classe (Figura 8(b)), tem-se uma melhor estimativa da superfície de separação das classes. Mas é apenas quando o número de pixel por classe é elevado, se comparado ao número de bandas da imagem (neste caso duas) que obtém-se uma boa estimativa para o classificador supervisionado (Figura 8 (c)).

2.3 - Classificador Paramétrico: Máxima Verossimilhança Gaussiana

Algoritmos de reconhecimento de padrões são freqüentemente classificados como paramétricos ou não-paramétricos. Uma técnica paramétrica assume uma distribuição probabilística para os dados dentro de cada classe. Técnicas não paramétricas não fazem este tipo de suposição.

No âmbito do sensoriamento remoto, destaca-se o classificador Máxima Verossimilhança (MAXVER). Este classificador pode ser visto como um caso particular do classificador Bayesiano.

2.3.1 - Classificação de Bayes (bayesiana)

O classificador Bayesiano é possivelmente o mais utilizado entre os vários classificadores disponíveis. Isto se deve ao fato de ser um classificador dito “ótimo”, no sentido de que minimiza a probabilidade de perda total média condicional, ou risco médio. O conceito de perda está ligado à perda ou penalidade que se incorre ao classificar erroneamente pixels em uma imagem.

Considerando que as classes espectrais para uma imagem sejam representadas por:

$$\omega_i, i = 1, 2, 3, \dots, C$$

onde C é o número total de classes. Para determinar a qual classe ou categoria pertence um dado pixel X , é necessário determinar sua probabilidade condicional:

$$p(\omega_i | x), i = 1, 2, 3, \dots, C$$

O vetor de posição x é uma coluna de valores de brilho para o pixel. Os elementos que formam este vetor são as respostas espectrais em cada uma das bandas espectrais utilizadas. Estas respostas espectrais, que representam a radiância recebida

em cada banda espectral, em uma imagem digital são quantificadas em valores inteiros denominados de números ou contadores digitais (CD). A probabilidade $p(\omega_i | x)$ fornece a verossimilhança associada a classe ω_i para um pixel na posição x . A classificação é executada de acordo com a seguinte regra:

$$x \in \omega_i \text{ se } p(\omega_i | x) > p(\omega_j | x) \text{ para todo } j \neq i \quad (5)$$

Esta regra de decisão intuitiva é um caso específico de uma regra mais geral, na qual a decisão pode ser ponderada dependendo da importância dada a cada tipo de classificação incorreta.

2.3.2 - A regra de decisão Máxima Verossimilhança

Apesar de sua simplicidade, as $p(\omega_i | x)$ definidas anteriormente são desconhecidas. Se os dados de treinamento disponíveis para cada classe forem em número suficiente, eles podem ser utilizados para estimar uma distribuição de probabilidades para cada classe, o que permitirá descrever a chance de encontrar um pixel x pertencente a classe ω_i . Definidos os dados de treinamento, a forma da função de distribuição será também especificada; porém, no momento, será utilizado apenas o termo geral representado pelo símbolo $p(x | \omega_i)$. Existirão tantas $p(x | \omega_i)$ quantas foram as classes, ou seja, para um dado pixel x no espaço multiespectral, serão calculadas um conjunto de probabilidades, fornecendo a relativa verossimilhança para este pixel pertencer a uma das classes disponíveis.

A probabilidade “a posteriori” $p(\omega_i | x)$ que se deseja calcular e a função densidade de probabilidade $p(x | \omega_i)$ disponível, estimada a partir dos dados de treinamento, estão relacionados pelo teorema de Bayes:

$$p(\omega_i | x) = \frac{p(x | \omega_i) p(\omega_i)}{p(x)} \quad (6)$$

onde $p(\omega_i)$ é a probabilidade da classe ω_i ocorrer na imagem. Se, por exemplo, 20% dos pixels da imagem pertencerem à classe ω_i , então $p(\omega_i) = 0.20$; $p(x)$ é a probabilidade de encontrar um pixel em qualquer classe na posição x . É interessante notar que:

$$p(x) = \sum_{i=1}^c p(x | \omega_i) p(\omega_i) \quad (7)$$

embora o próprio $p(x)$ não seja importante no desenvolvimento posterior.

As $p(\omega_i)$ são chamadas de probabilidade “a priori”, uma vez que elas representam as probabilidades com as quais as classes, associadas a um pixel, são definidas antes da classificação. Por analogia, as $p(\omega_i | x)$ são as probabilidades “a posteriori”. Usando (6) pode-se escrever a regra de decisão (7) como sendo:

$$x \in \omega_i \text{ se } p(x | \omega_i) p(\omega_i) > p(x | \omega_j) p(\omega_j) \text{ para todo } j \neq i \quad (8)$$

onde $p(x)$ foi removido por ser um fator comum aos dois membros. A regra (8) é mais adequada que a (5), uma vez que as $p(x | \omega_i)$ podem ser estimadas a partir dos dados de treinamento. Pressupõe-se que as $p(\omega_i)$ também sejam conhecidas ou possam ser estimadas a partir do conhecimento que o analista tem da imagem. Por conveniência matemática, a equação (8) pode ser escrita da seguinte forma:

$$x \in \omega_i \text{ se } g_i(x) > g_j(x) \text{ para todo } j \neq i \quad (9)$$

com

$$g_i(x) = \ln \{ p(x | \omega_i) p(\omega_i) \} \quad \text{ou} \\ g_i(x) = \ln p(x | \omega_i) + \ln p(\omega_i) \quad (10)$$

Onde \ln é o logaritmo neperiano.

Deste modo, o critério de classificação, que pode ser resumido por (9), é a regra de decisão generalizada usada em classificação máxima verossimilhança. As funções $g_i(x)$ e $g_j(x)$ são chamadas funções discriminantes.

2.3.3 - Modelo para classes Normais Multivariadas

A experiência prática utilizando dados multiespectrais em sensoriamento remoto tem mostrado que a função densidade de probabilidade multivariada Gaussiana descreve, bastante bem, o comportamento destes dados. Este fato pode ser explicado pelo teorema Central do Limite, pois a radiância refletida por alvos terrestres consistem do somatório de um número grande de outras variáveis aleatórias. Esta é mais uma suposição do que uma propriedade que possa ser demonstrável, sobre a natureza espectral das classes de informação. Entretanto, ela conduz a consideráveis

simplificações matemáticas e, além disso, é uma distribuição para a qual as propriedades de forma multivariada são bem conhecidas.

Assumindo-se agora, que serão utilizadas N bandas espectrais (ou feições), a equação (5) ficará:

$$p(x | \omega_i) = (2\pi)^{-N/2} |\Sigma_i|^{-1/2} \exp \left\{ -\frac{1}{2} (x - m_i)^T \Sigma_i^{-1} (x - m_i) \right\} \quad (11)$$

onde m_i representa o vetor média e Σ_i representa a matriz covariância da classe ω_i . Em situações práticas estes parâmetros são estimados a partir do conjunto das amostras de treinamento disponíveis para cada classe.

$$\hat{m}_k = \frac{1}{N_k} \sum_{i=1}^{N_k} x_i \quad \text{e} \quad \hat{\Sigma}_k = \frac{1}{N_k - 1} \sum_{i=1}^{N_k} (x_i - \hat{m}_k)(x_i - \hat{m}_k)^t \quad (12)$$

Onde N_k é o número de amostras de treinamento da classe k .

O termo resultante $-N/2 \ln(2\pi)$ é comum a todos $g_i(x)$ e não auxilia na discriminação. Conseqüentemente, ele é ignorado na forma final da função discriminante para a classificação por máxima verossimilhança, de modo que $g_i(x)$ ficará:

$$g_i(x) = \ln p(\omega_i) - \frac{1}{2} \ln |\Sigma_i| - \frac{1}{2} (x - m_i)^T \Sigma_i^{-1} (x - m_i) \quad (13)$$

Freqüentemente o analista não dispõe de informações confiáveis sobre $p(\omega_i)$, e então, assume-se a situação de igualdade de probabilidades “a priori” para as classes. Como resultado, $\ln p(\omega_i)$ pode ser removido da eq. (13), desde que a $p(\omega_i)$ seja a mesma para todas as classes i . Pode-se também, remover da função, o termo comum $1/2$, resultando na seguinte função discriminante:

$$g_i(x) = -\ln |\Sigma_i| - (x - m_i)^t \Sigma_i^{-1} (x - m_i) \quad (14)$$

A implementação da regra de decisão de máxima verossimilhança envolve a utilização das equações (13) ou (14) em (9). A figura 9 ilustra um exemplo de fronteira de decisão entre duas classes.

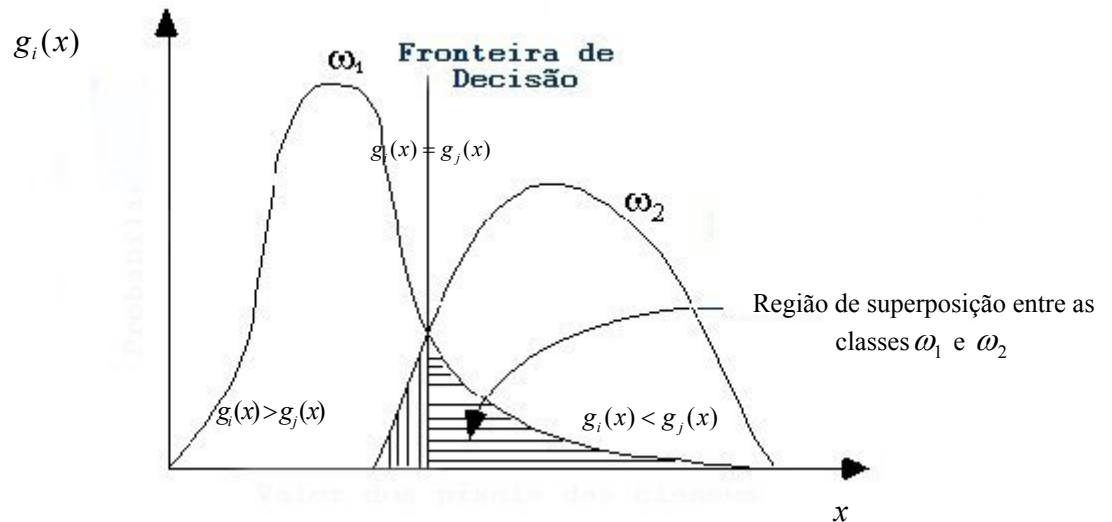


Figura 9. Fronteira de decisão entre duas classes

2.4 - Avaliação da separabilidade entre classes: distância de Bhattacharyya

Entre as várias alternativas para estimar a separabilidade entre pares de classes, a distância de Bhattacharyya é mencionada com frequência por vários autores, pelas características que apresenta [FUK 90], [LAN 03], [DUD 01] e [THE 01]. A forma geral da distância de Bhattacharyya é definida por:

$$b_{ij} = -\ln \left[\int_{-\alpha}^{\alpha} \sqrt{p(x|\omega_i)p(x|\omega_j)} dx \right]$$

Assumindo a distribuição Normal multivariada para os dados, esta distância estatística assume a seguinte forma:

$$b_{ij} = \frac{1}{8} (m_i - m_j)^t \left(\frac{\Sigma_i + \Sigma_j}{2} \right)^{-1} (m_i - m_j) + \frac{1}{2} \ln \left(\frac{\left| \frac{\Sigma_i + \Sigma_j}{2} \right|}{\sqrt{|\Sigma_i| + |\Sigma_j|}} \right) \quad (15)$$

onde b_{ij} é a distância de Bhattacharyya entre as classes i e j , m_i e m_j os vetores média e Σ_i e Σ_j as matrizes de covariância das classes (note que b_{ij} é um número real, variando no intervalo $[0, \infty]$).

Nesta medida de distância estatística, que é bastante utilizada pela comunidade de pesquisadores na área de reconhecimento de padrões, quanto maior o

valor da distância, mais separadas estão as funções de densidade de probabilidade das duas classes e teoricamente menor o erro de classificação obtido.

Se o erro de classificação puder ser estimado a partir de estatísticas das classes, esta informação pode ser utilizada para extração de feições e também para decidir quantas são necessárias para dado problema. Entretanto, a previsão do erro a partir das estatísticas das classes não é uma tarefa fácil, tendo sido tema de pesquisa de diversos autores [LEE 00].

Uma vantagem da utilização desta medida de separabilidade é que ela está diretamente relacionada à probabilidade de erro no processo de classificação. A distância de Bhattacharyya é uma medida teórica da distância entre duas distribuições gaussianas que é equivalente a um limiar superior (*upper bound*) do erro ótimo que pode ser obtido utilizando um classificador bayesiano. Ela apresenta a propriedade de ser computacionalmente simples.

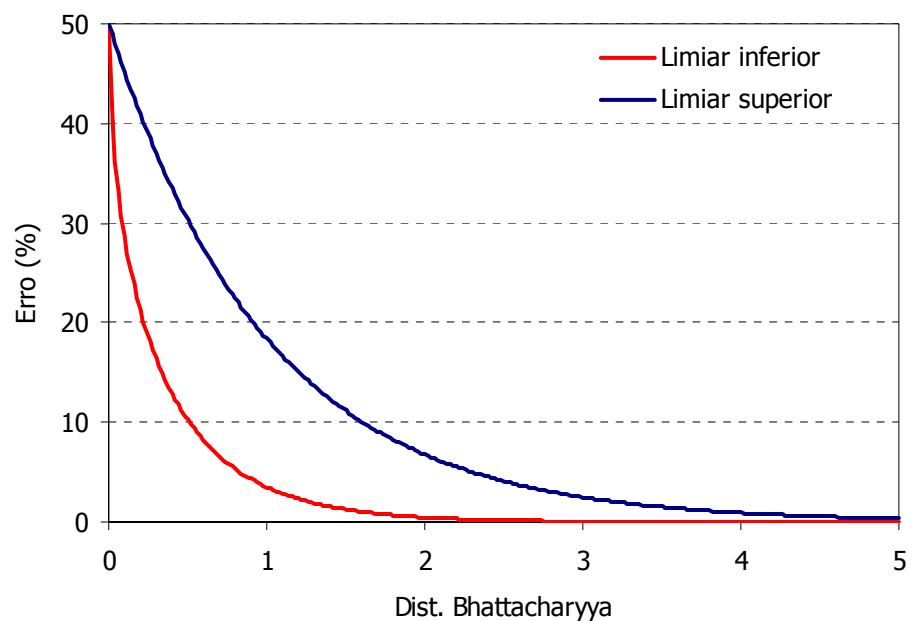


Figura 10. Limites teóricos inferior e superior do erro de classificação avaliado utilizando a distância de Bhattacharyya.

Uma vantagem da utilização desta medida de separabilidade é que ela está diretamente relacionada à precisão que pode ser obtida na classificação. Em termos práticos a literatura [FUK 90] e [LEE 00] apresentam os limiares inferior e superior do erro bayesiano ϵ_B que pode ser obtido numa classificação. Este limiar é dado como

função da distância de Bhattacharyya entre classes. Assumindo-se probabilidades de ocorrência a priori das classes idênticas:

$$\frac{1}{2}\left(1 - \sqrt{1 - e^{-2b_{ij}}}\right) \leq \epsilon_B \leq \frac{1}{2}e^{-b_{ij}} \quad (16)$$

Na Figura 10 apresentam-se os limites teóricos inferior e superior do erro de classificação estimado (16) utilizando-se a distância de Bhattacharyya (15).

2.5 - Métodos de Classificação Não Paramétricos

Classificadores não paramétricos são aqueles que não estão baseados em nenhuma forma particular de distribuição dos dados associadas às classes de interesse. Classificadores não paramétricos podem assumir diversas formas. A característica que os tornam atrativos é a sua generalidade.

2.5.1 - Método do paralelepípedo (Level-Slice Classifier)

Possivelmente este seja o mais simples de todos os classificadores. Neste algoritmo estima o vetor média de cada classe. Utilizando este valor como referência para cada classe, o espaço multiespectral é então particionado por meio de planos ou hiper-planos perpendiculares aos eixos de coordenadas gerando partições em forma de prismas retangulares (ou hiper-prismas). Normalmente o analista define estas dimensões em termos de unidades de desvio padrão. O pixel é classificado como pertencente à classe que corresponde a caixa onde ele se encontra. Os pixels que não estão contidos em nenhuma caixa não são classificados. Por outro lado, se um pixel se encontra numa região de superposição entre caixas, outro algoritmo deve ser utilizado para classificar o pixel, por exemplo avaliando-se a distância até a média mais próxima (introduzindo-se portanto uma modificação do tipo paramétrica). Este método implementa uma abordagem extremamente rudimentar que atualmente não vem mais sendo utilizada em sensoriamento remoto.

2.5.2 - Classificador *K*-Vizinhos mais Próximos (*K-Nearest Neighbor*).

Neste algoritmo, um vetor (pixel) desconhecido é classificado como pertencente à classe da maioria dos seus K vizinhos mais próximos. Desta forma, faz-se necessário avaliar a proximidade do pixel aos K pixels mais próximos. Para tanto, utiliza-se a distância euclidiana. É um método de classificação relativamente simples mas que em geral apresenta bons resultados em termos de acurácia de classificação.

2.5.3 - Classificador Janela de Parzen (*Parzen Window Classifier*)

Neste algoritmo, procura-se aproximar as funções densidade de probabilidade das classes em termos de uma série de funções para janelas onde se encontram os pixels de treinamento. Estas funções são também chamadas kernels. Uma função bastante utilizada é a exponencial, que é um kernel gaussiano. Neste caso, as funções de densidade de probabilidade (não paramétricas) são modeladas por uma combinação linear de funções gaussianas. Mesmo que uma classe seja claramente não gaussiana, percebe-se que ela pode ser satisfatoriamente representada em termos de uma combinação linear de “subclasses” gaussianas. Desta forma, tem-se uma abordagem bastante robusta e prática, já que faz uso das vantagens de uma abordagem paramétrica e também tem-se a possibilidade de generalização.

Ocorre que para a definição das funções kernel também são necessários parâmetros. Assim, a característica de elevada capacidade de generalização, para este tipo de classificadores, está geralmente associada a um também elevado número de parâmetros a serem determinados. Cada detalhe da forma das funções densidades de probabilidades precisa ser determinado neste processo. Isto pode rapidamente fugir do controle do analista. Num primeiro momento, pareceria óbvio que, ao utilizar-se uma grande quantidade de informação, a partir das amostras de treinamento, obter-se-iam melhores resultados. Entretanto, este raciocínio pode ser incorreto, visto que, especialmente em sensoriamento remoto, o conjunto de pixels de treinamento é em geral de tamanho limitado.

2.5.4 - Redes neurais

A classificação utilizando redes neurais é uma abordagem não paramétrica de utilização mais recente. É uma abordagem que difere substancialmente de outros métodos no sentido que as fronteiras de decisão não são definidas por uma regra determinística, obtida a partir das amostras de treinamento. As fronteiras de decisão são obtidas de maneira iterativa, buscando-se minimizar um determinado critério que é baseado no erro de classificação, avaliado utilizando-se os pixels de treinamento. Existem atualmente várias arquiteturas de redes neurais. Uma configuração clássica apresenta três layers: um de entrada dos dados, outro intermediário e o de saída. A exceção do primeiro, que contém nós de entrada dos dados, os demais layers apresentam elementos de processamento. Os parâmetros de entrada são as feições utilizadas para classificação, uma em cada nó do layer de entrada dos dados. As redes neurais podem ser interpretadas como sendo uma representação gráfica de um problema de expansão em séries, no qual o objetivo é aproximar a função discriminante que permite a separabilidade das classes. Os elementos contidos nos segundo e terceiro layers (elementos onde ocorrem processamentos) representam os coeficientes desta expansão em séries: tipicamente tem-se um layer com constantes “pesos”, outro com funções bases. As funções bases comumente citadas pela literatura são de dois tipos: transformações lineares (ridge construction), representando hiperplanos ou do tipo radial (radial construction). Estando a rede construída, a determinação dos parâmetros de cada elemento é efetuada utilizando-se métodos de otimização, em geral baseados no método do gradiente, por exemplo, o back propagation. Um problema para o analista é a determinação do número de elementos dos layers onde ocorre processamento. Para tal, geralmente utilizam-se técnicas de estimativa da probabilidade de erro de classificação dos pixels de treinamento (por exemplo, a cross-validation). As vantagens da utilização das redes neurais estão associadas a sua elevada capacidade de generalização, impondo hipóteses não muito severas sobre o comportamento do sistema. São uma boa solução para muitos casos aplicativos. As desvantagens estão associadas a problemas de convergência e elevado custo computacional.

3 - METODOLOGIA

Conforme mencionado anteriormente, os sistemas sensores presentemente utilizados são projetados para atender as características de um amplo conjunto de alvos de natureza as mais diversas. Em situações individuais, a utilização no processo de classificação de todas as feições originais ou de um número excessivo destas, além de desnecessário pode vir em prejuízo da eficiência do processo devido, principalmente, as dificuldades em se obter estimativas acuradas dos parâmetros empregados pelo classificador sempre que dados em alta dimensionalidade são utilizados [LAN 03]. É geralmente mais eficiente selecionar um sub-conjunto composto pelas feições mais adequadas a cada caso em particular. Desta forma uma etapa importante que deve preceder o processo de classificação refere-se à escolha de um sub-conjunto ótimo ou pelo menos sub-ótimo de feições para cada situação em particular. Este problema é conhecido como redução da dimensionalidade dos dados. Em termos gerais, existem duas abordagens gerais a este problema: (1) *seleção de feições* e (2) *extração de feições*. No primeiro caso, busca-se selecionar um sub-conjunto das feições originais que melhor atenda as características do problema em consideração. No segundo caso, o objetivo consiste em reduzir a dimensionalidade dos dados originais por meio da geração de um número menor de feições derivadas através de combinações lineares ou não, das feições originais. O objetivo desta dissertação consiste em investigar uma metodologia confiável, simples de implementar e computacionalmente eficiente, para fins de extração de feições em problemas envolvendo dados em alta dimensionalidade. Embora a metodologia proposta neste estudo tenha aplicações potenciais em várias outras áreas, a área de interesse mais especificamente tratada neste estudo refere-se a dados de imagens adquiridas por uma nova geração de sensores remotos, comumente referidas na literatura como dados “hiperespectrais”.

Em imagens hiperespectrais e para fins de classificação, cada pixel é normalmente representado por um vetor $X = [x_1, x_2, x_3, \dots, x_N]$. Os elementos x_i ($i=1, \dots, N$) que compõem este vetor são as respostas individuais fornecidas por cada uma das N bandas espectrais i disponíveis. Esta resposta é uma função da quantidade de energia refletida ou emitida pela superfície em determinado intervalo de comprimento de onda. Esta informação, fornecida pelo conjunto das bandas espectrais, pode também ser visualizada pela chamada curva de resposta (Figura 11). Esta curva descreve o comportamento espectral de cada pixel individualmente. Desta forma, pixels que pertencem a uma mesma classe apresentam curvas de resposta espectral semelhantes e vice-versa. Uma característica de dados multiespectrais em sensoriamento remoto é a alta correlação apresentada entre as bandas individuais, sugerindo a possibilidade de redução da dimensionalidade sem perda significativa de informação.

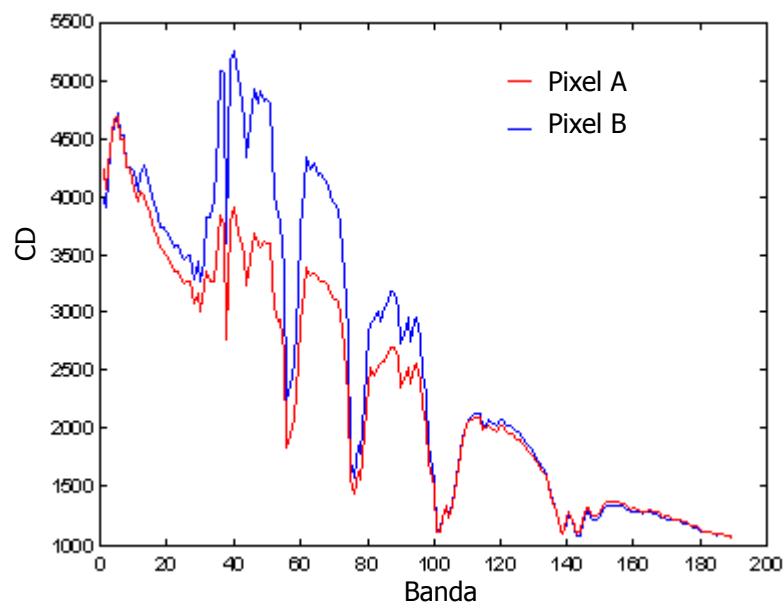


Figura 11. Exemplo da curva de resposta espectral de dois pixels de uma imagem hiperespectral obtida pelo sensor AVIRIS. No eixo das abscissas temos as bandas da imagem, nas ordenadas o respectivo contador digital (ou nível de cinza) para cada uma das bandas.

A hipótese básica que se faz neste trabalho é de que a curva de resposta espectral associada a cada pixel pode ser descrita por um número menor de variáveis (estatísticas) sem perda significativa de exatidão no processo de classificação. Em outras palavras, esta operação consiste num processo de redução da dimensionalidade dos dados originais através da extração de feições. Em presença de uma

dimensionalidade reduzida, espera-se minimizar ou mesmo eliminar os efeitos do fenômeno de Hughes no processo de classificação.

Na revisão bibliográfica desenvolvida ao longo deste estudo, não foi encontrada nenhuma proposta similar o que leva a crer que a metodologia aqui proposta apresenta caráter de originalidade. A metodologia proposta consiste em tentar capturar a informação contida no conjunto total de bandas espectrais disponíveis (curva espectral de cada pixel), por um pequeno número de estatísticas simples que objetivam descrever a *posição* e a *forma* da curva de resposta espectral do pixel.

Quanto às feições a serem extraídas, optou-se neste estudo por particionar a curva de resposta espectral em um certo número de segmentos e descrever a informação contida em cada segmento por duas estatísticas: média aritmética e variância. A escolha de tais parâmetros pode ser justificada pela expectativa de que a média represente adequadamente a posição do segmento enquanto que a variância descreva a variabilidade da curva dentro do segmento. Outras possibilidades que poderiam ser testadas seriam, por exemplo: utilizar as estatísticas dos momentos L (L moments), utilizar apenas um parâmetro de posição (média aritmética), utilizar a 1ª, 2ª, 3ª ... primeiras componentes da transformada discreta do co-seno (TDC), além de outras possibilidades originalmente consideradas neste estudo. Embora não sejam apresentados neste trabalho, alguns experimentos iniciais realizados indicaram que estas outras estatísticas também apresentariam bons resultados, embora levemente inferiores aos obtidos utilizando as estatísticas média e variância.

A metodologia aqui proposta consiste na partição da curva de resposta espectral de cada pixel (independente da classe a que pertençam), em segmentos s_j ($j=1, \dots, k$) de bandas adjacentes, conforme a Figura 12, tal que $s_1 \cup s_2 \cup s_3 \cup \dots \cup s_k = X$. A seguir, cada um dos segmentos s_j , contendo as bandas de um intervalo $x_{l,u} = [x_l, \dots, x_u]$, sendo $l \leq l \leq l+2$, $l+2 \leq u \leq N$, é substituído pela sua média aritmética e variância, ou seja:

$$\bar{x}_{l,u}(x) = \frac{1}{u-l+1} \sum_{i=l}^u x_i \quad (17)$$

$$\sigma^2_{l,u}(x) = \frac{1}{u-l+1} \sum_{i=l}^u (x_i - \bar{x}_{l,u}(x))^2 \quad (18)$$

Designando $f_j = [\bar{x}_{l,u}(x), \sigma^2_{l,u}(x)]$ as duas feições provenientes de cada segmento s_j , cada pixel passará a ser representado por um vetor, de dimensionalidade reduzida, com componentes $Y = [f_1, f_2, f_3, \dots, f_k]$, sendo $1 \leq k \leq (N/3)$. Desta forma, a transformação de X para Y equivale a uma transformação no espaço de $\mathfrak{R}^N \rightarrow \mathfrak{R}^{2k}$.

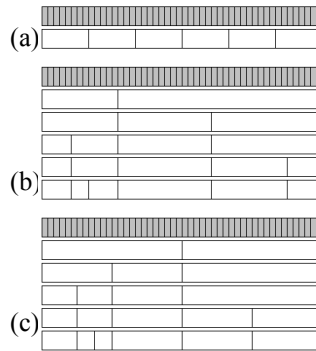


Figura 12. Exemplo de segmentação da curva de resposta espectral de cada pixel X , segundo três algoritmos distintos, gerando seis segmentos. Em (a) geram-se segmentos de comprimento constante. Em (b) e (c) segmentos de comprimento variável, sendo que a cada “nível hierárquico” adiciona-se uma partição adicional.

Neste processo deve-se inicialmente definir uma metodologia adequada para segmentação da curva de resposta espectral. Dentro da proposta de representar cada segmento pelo valores da média e da variância, cada segmento deve ser formado por no mínimo três bandas originais, ou seja, $(u - l + 1) \geq 3$. Para implementar esta metodologia, gerando partições conforme o exemplo indicado na Figura 12, são propostos três algoritmos. Um que gera partições que resulta em segmentos de comprimento constante. Os outros dois, com estrutura do tipo “*top-down*”, inicia com um único segmento, idêntico à totalidade da curva de resposta espectral, que vai sendo segmentado a cada etapa em um número crescente de segmentos de comprimentos menores até que as condições impostas por um determinado critério sejam atingidas. Este processo resulta em segmentos de comprimento variável.

Estas novas feições são então implementadas em um classificador paramétrico. Com a redução da dimensionalidade assim obtida ($Y = \mathfrak{R}^{2k}$), as estimativas dos parâmetros do classificador (vetor média e matriz covariância) tornam-se mais confiáveis, os efeitos do “fenômeno de Hughes” também referido na literatura como a “maldição da dimensionalidade” (*curse of dimensionality*) é reduzido ou mesmo eliminado proporcionando desta forma resultados mais acurados no processo de classificação.

A seguir, são descritos os algoritmos que foram propostos e desenvolvidos neste trabalho. Posteriormente, o desempenho de cada uma das três abordagens propostas é testado através de experimentos utilizando dados reais (dados hiperespectrais). Os resultados obtidos com a utilização das feições extraídas são comparados com aqueles obtidos utilizando-se as feições originais.

3.1 - Novas abordagens propostas para redução da dimensionalidade

Basicamente foram investigadas duas abordagens distintas para a representação da informação espectral em cada pixel, conforme a resposta fornecida pelo conjunto de bandas espectrais, substituindo-as por um número menor de estatísticas (feições), de modo a descrever de uma maneira aceitável a informação contida na curva de resposta espectral. São elas: (a) segmentação da curva de resposta espectral em segmentos de comprimento constante e (b) em segmentos com comprimentos variáveis.

3.2 - Segmentação da curva de resposta espectral em segmentos de comprimento constante – “SCC”

3.2.1 - Algoritmo SCC – Segmentação com Comprimento Constante

O primeiro algoritmo proposto consiste na segmentação da curva de resposta espectral de cada pixel em segmentos de comprimento $L = (u - l + 1)$ constantes e sucessivos ao longo de toda a curva de resposta espectral (Figura 12 (a)). Por ser uma solução de concepção simples, torna-se uma alternativa atraente do ponto de vista computacional, pois além da facilidade para implementação em ambiente de computador, o tempo de processamento requerido é pequeno. Em outras palavras, a execução do algoritmo é veloz. A partição não depende da natureza da resposta espectral nem do número de classes presentes na imagem. Nesta abordagem, não são levados em consideração itens importantes como questões de separabilidade estatística entre classes. Desta forma, este primeiro método não é ótimo, no sentido de que não procura adaptar o comprimento de cada segmento às características locais da curva de resposta espectral das classes envolvidas, visando otimizar a separabilidade entre as

classes. Esta preocupação é implementada apenas nas duas outras metodologias propostas neste estudo. A estratégia adotada nesta primeira abordagem na avaliação do melhor comprimento comum para os segmentos é baseada num processo iterativo, onde comprimento dos segmentos, e por consequência o número total de feições extraídas, é avaliado em termos da acurácia obtida no processo de classificação. Visto que o número de possibilidades para valores de comprimento comum a todos os segmentos é limitado, o comprimento ótimo segundo este critério pode ser facilmente encontrado examinando-se exaustivamente todas as possibilidades.

Em situações gerais, o comprimento dos segmentos de tamanho constantes L não necessariamente é um múltiplo do número total de bandas disponíveis N . Neste caso, no processo de segmentação, restarão bandas que, por facilidade de implementação computacional, sugerem-se que sejam agrupadas em um único segmento, por exemplo, o último. Desta forma, considerando-se a presença de k segmentos, têm-se $(k-1)$ segmentos de comprimento L e 1 segmento de comprimento L' , sendo $L \leq L' < 2L$. A relação entre comprimentos dos segmentos e número de total de bandas é estabelecida por:

$$(k-1)L + L' = N \quad (19)$$

Sendo NFS o número de feições obtidas por segmento, o número total de feições NF obtidas pelo processo de segmentação fica expresso por:

$$NF = k NFS \quad (20)$$

onde o valor de k (número de segmentos) é o valor inteiro da divisão do número total de bandas originais N pelo comprimento L de cada segmento, ou seja:

$$k = \text{int}\left(\frac{N}{L}\right) \quad (21)$$

Substituindo (21) em (19), e rearranjando os termos, chega-se ao número de bandas máximo de um dos segmentos, por exemplo o último:

$$L' = L + \left[N - \text{int}\left(\frac{N}{L}\right) L \right] \quad (22)$$

Percebe-se facilmente pelas equações (21) e (20) que os comprimentos L dos segmentos determinarão o número de feições totais NF que serão extraídas ao final

do processo. Porém, também estabelecem que não é possível obter um número arbitrário de feições, mas sim valores discretos e não necessariamente espaçados de maneira uniforme. Por exemplo, para $N = 190$ bandas e $NFS = 2$, ou seja, duas feições por segmento e comprimento dos segmentos $L=3$, têm-se $NF = 126$ feições, para $L=4$, temos 94, para $L=5$, temos 76, $L=6$, têm-se 62 feições, etc. Também estabelecem que vários comprimentos L distintos podem gerar o mesmo número de feições, por exemplo: para $28 \leq L \leq 31$, têm-se $NF = 12$. Neste caso, aconselha-se utilizar o valor de L que minimiza o comprimento do maior segmento L' (22), ou seja, que mantém o último segmento com comprimento o mais próximo aos demais.

Outra possível solução para lidar com o problema das possíveis bandas excedentes seria fazer uma redistribuição dos segmentos.

A Figura 13 ilustra a segmentação da resposta espectral que é obtida com a adoção deste algoritmo.

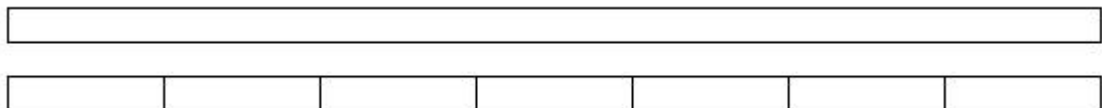


Figura 13. Algoritmo SCC: Exemplo de representação da segmentação da resposta original de cada pixel da imagem em 7 segmentos sucessivos, iguais e seqüenciais. Cada um dos segmentos contém $L = n$ bandas originais, que serão substituídas por duas feições. Para que ocorra redução da dimensionalidade em cada segmento, o número de feições extraídas deve ser menor do que o número de bandas em cada segmento.

Conforme mencionado anteriormente, neste algoritmo o comprimento L ótimo dos segmentos, medido em número de bandas espectrais, não é conhecido a priori. Porém, visto que o número de possibilidades para valores de comprimento comum a todos os segmentos é limitado, o comprimento ótimo dentro deste critério pode ser facilmente encontrado examinando exaustivamente todas as possibilidades.

3.3 - Segmentação da curva de resposta espectral em segmentos de comprimento variável - “SCV”

A segunda alternativa proposta consiste na segmentação da curva de resposta espectral de cada pixel em segmentos de comprimento não necessariamente iguais. Desta forma, espera-se que uma melhor adaptabilidade à curva de resposta espectral do conjunto de classes presentes na cena de estudo. Na abordagem aqui proposta, a otimização dos segmentos é feita seqüencialmente. Ao final de cada “nível hierárquico” (Figura 12 (b)(c)), gera-se uma partição adicional à(s) já existente(s), em uma posição m da curva de resposta espectral original, definida como sendo o ponto que maximiza uma determinada função de otimização $J(m)$ que avalia a qualidade das partições geradas. Esta avaliação é feita em termos de separabilidade estatística entre classes, estando a partição posicionada em m . Note que cada posição m testada altera o conjunto de segmentos $s_1 \dots s_k$. obtidos. Cada segmento é substituído pelas correspondentes feições extraídas, sendo $J(m)$ avaliado neste conjunto com dimensionalidade reduzida, ou seja, em Y . Para a aplicação desta alternativa foram desenvolvidos dois algoritmos, descritos a seguir:

3.3.1 - Algoritmo Otimização Global

Trata-se de um algoritmo com estrutura em árvore elaborado no sentido de buscar uma segmentação otimizada da curva de resposta espectral dos pixels.

Dado um conjunto de segmentos iniciais s_1, \dots, s_k , testam-se todas as possibilidades para geração de um novo segmento, testando a partição de cada um dos segmentos s_j ($j=1, \dots, k$), definido no intervalo $x_{l,u} = [x_l, \dots, x_u]$ em dois novos segmentos $[x_l, \dots, x_m] \cup [x_{m+1}, \dots, x_u]$, variando-se m no intervalo $l+2 \leq m < u-2$. Em outras palavras, esta procura é exaustiva, ou seja, testam-se todas as possíveis posições para posicionar m . Este processo é realizado para cada um dos segmentos já existentes, utilizando-se um critério adequado $J(m)$ como função de otimização para a avaliação das partições geradas, seja para o caso envolvendo duas classes ou também para o caso multiclassés.

Definida a melhor posição para a segmentação inicial (aquela que gerou os dois primeiros segmentos), mantém-se este particionamento e o processo continua, procurando, no interior de cada um destes dois segmentos, todas as possíveis posições para gerar um novo segmento. Ao final deste segundo passo (ou “nível hierárquico”), um dos segmentos gerados na iteração anterior fica particionado em dois e a curva de resposta espectral passa a contar portanto com três segmentos. Note que na aplicação da função de otimização $J(m)$ são sempre utilizadas as feições extraídas a partir dos segmentos existentes, ou seja, em dimensionalidade reduzida, porém crescente a cada “nível hierárquico”. Desta forma o processo continua. O número de segmentos e conseqüentemente o número de feições obtidas é portanto crescente.

A cada “nível hierárquico” o número de possibilidades a serem analisadas para segmentação decresce, visto que já existem os segmentos obtidos anteriormente. Eventualmente alguns destes segmentos ficam com um comprimento muito reduzido, inviabilizando uma segmentação adicional. Neste caso, o processo continua testando apenas os demais segmentos. A Figura 14 ilustra um exemplo de resposta do algoritmo em termos de partição da curva de resposta espectral original.

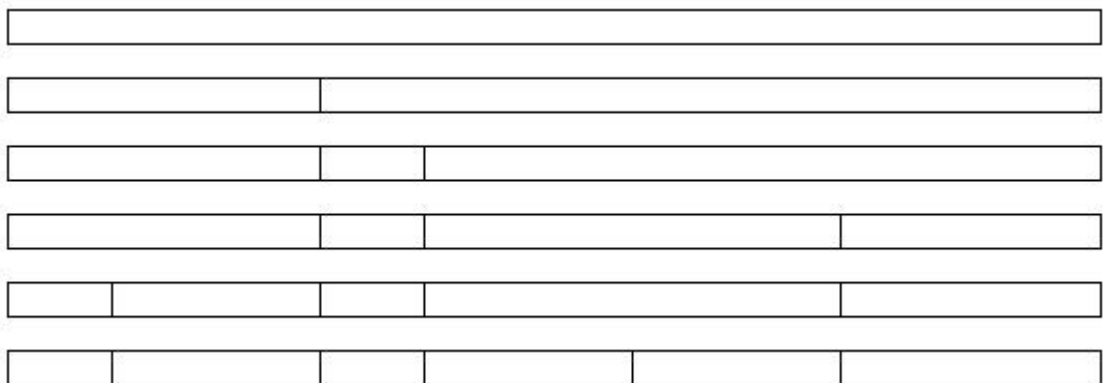


Figura 14. Algoritmo Otimização Global: Representação da segmentação da resposta original de cada pixel da imagem, num processo seqüencial ao final do quinto “nível hierárquico”. Note que uma vez que uma posição de divisão é gerada, esta permanece até o final do processo, sendo que cada nova posição é a ótima em relação às pré-existentes, para o critério aqui proposto. Para cada um dos segmentos são extraídas feições. A cada nível, a otimização é procurada utilizando-se o conjunto total de feições gerado, variando-se a posição da fronteira no novo segmento, mantendo inalterado os demais.

Um aspecto a ser mencionado é que uma vez definida a posição de uma segmentação, esta não é mais descartada, permanecendo inalterada ao longo do processo de subdivisões conseqüentes. Isto é um inconveniente, pois não existe garantia de que, em um determinado “nível hierárquico” não seria alcançado um particionamento

ainda melhor do que aquele obtido por este processo, caso fossem permitidas alterações nos particionamentos previamente realizados. Porém, um algoritmo de procura exaustiva sem impor restrições quanto às segmentações anteriores seria computacionalmente inviável, visto a gigantesca quantidade de combinações que deveriam ser testadas, mesmo para o caso de algumas poucas dezenas de bandas originais. Outro aspecto importante a ser mencionado é que a solução obtida, mesmo para a geração dos dois primeiros segmentos, não necessariamente é a ótima, mas sim é ótima para o critério $J(m)$ que está sendo utilizado.

Tipicamente, na medida em que vão sendo adicionados segmentos, a acurácia de classificação deste processo cresce, até um certo ponto, a seguir mantém-se constante ou mesmo decresce (devido ao fenômeno de Hughes). O comportamento do processo pode ser monitorado, em termos da acurácia que vai sendo obtida, de maneira que possa ser definido o melhor momento de parada e conseqüentemente, a melhor configuração para segmentação da curva de resposta espectral.

Cabe ainda ressaltar que este algoritmo é computacionalmente bem mais exigente do que a proposição do algoritmo SCC (segmentação comprimento constante), dado o número de possibilidades que são testadas, computando-se a separabilidade entre cada par de classes para o cálculo de $J(m)$.

A metodologia é utilizada repetidamente no processo de geração dos sucessivos segmentos. Mantida fixa a configuração da segmentação já obtida, cada um dos segmentos já gerados é investigado buscando a melhor posição para uma nova segmentação, segundo os mesmos critérios anteriormente descritos. Neste processo uma nova configuração é definida e o número total de segmentos fica incrementado de uma unidade.

A título exemplificativo, suponha um problema envolvendo uma imagem hiperespectral onde cada pixel é representado por um vetor de $N = 190$ componentes (cada componente representa a resposta espectral deste pixel em cada banda) e NFS (número de feições por segmento) igual a duas, por exemplo, média e variância. O número máximo de possibilidades a serem examinadas, a cada “nível hierárquico”, para que haja efetivamente redução da dimensionalidade em cada segmento é:

- $190 - 1 \times (2 + 3) = 185$, para a primeira partição (2 segmentos);
- $190 - 2 \times (2 + 3) = 180$, para a segunda partição (3 segmentos);

- $190 - 3 \times (2 + 3) = 175$, para a terceira partição (4 segmentos);
- \vdots
- $190 - n \times (NFS + (NFS+1))$, para o *n-ésimo* segmento e assim sucessivamente.

Genericamente, o número máximo de possíveis posições de segmentação *nps* totais que devem ser analisadas para a geração de *k* segmentos é dado pela equação:

$$nps = (k - 1) \times N - (2NFS + 1) \times \sum_{j=1}^{k-1} j \quad (23)$$

onde *N* é a dimensionalidade original do problema e *NFS* o número de feições geradas por segmento.

3.3.2 - Algoritmo com Estrutura em Árvore (Tree)

Este algoritmo é uma simplificação do anterior. Ao invés de testar-se todas as possíveis alternativas para definição de uma nova posição de partição m , testa-se apenas a *posição central* de cada um dos segmentos s_j ($j=1\dots k$) já existentes. 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 anterior. Também se mantém ainda alguma certa possibilidade de adaptação à resposta espectral do conjunto de pixels da cena de estudo.

Da mesma forma que no algoritmo Otimização Global, a posição escolhida para partição é aquela que maximiza $J(m)$. A Figura 15 ilustra um exemplo do processo de segmentação.

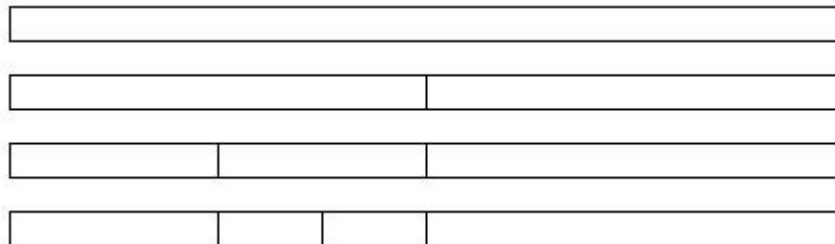
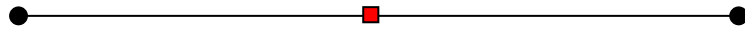


Figura 15. Algoritmo Tree: Representação da segmentação da resposta original de cada pixel da imagem, num processo seqüencial ao final do terceiro passo. Note que uma vez que uma posição de divisão é gerada, esta permanece até o final do processo, sendo que cada nova posição é a ótima em função das já existentes. Neste algoritmo, para a definição do próximo segmento, são testadas apenas as possibilidades de divisões sucessivas no centro de cada um dos segmentos já existentes.

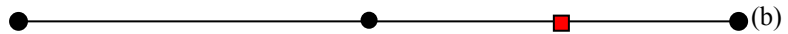
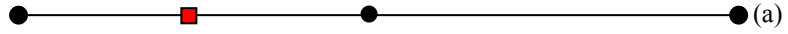
Este algoritmo em árvore tem a característica de ser mais flexível do que o algoritmo SCC, e é computacionalmente menos oneroso do que o algoritmo Otimização Global, visto que testa um número muito inferior de possibilidades para segmentação.

O número de posições a serem testadas a cada “nível hierárquico”, no processo de tomada de decisão para a geração dos segmentos, pode ser visualizado no exemplo abaixo:

- Uma, para a primeira partição:

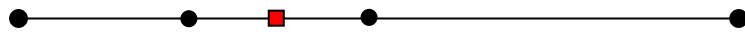
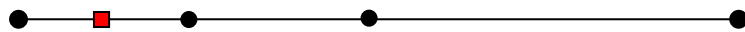


- Duas, para a segunda partição:

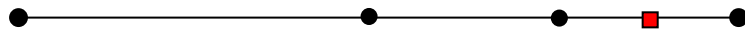
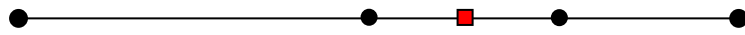
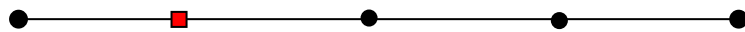


- Três, para a terceira partição:

Se a vencedora da segunda partição foi a primeira opção (a):



Se a vencedora da segunda partição foi a segunda opção (b):



- “n” possibilidades para a n-ésima partição. Note que o símbolo em círculo representa uma posição invariante e o quadrado a posição de teste para determinado “nível hierárquico”, determinando a configuração a ser utilizada para o cálculo da função de otimização $J(m)$. Definida qual a melhor opção, tal posição mantém-se até o final e o processo continua.

Genericamente, o número máximo de posições de segmentação que devem ser analisadas para a geração de k segmentos é dado por :

$$nps = \sum_{j=1}^k (j-1) \quad (24)$$

3.4 - Critérios de otimização para avaliação das partições geradas

Definidos os algoritmos que possibilitam a geração de segmentos de comprimento variável, resta agora definir o critério $J(m)$ de avaliação das partições geradas. Para tanto, se faz necessária a adoção de uma medida de separabilidade estatística entre classes.

Conforme apresentado anteriormente, entre as várias alternativas para estimar a separabilidade entre pares de classes, a distância de Bhattacharyya, devido as suas características, é mencionada com frequência por vários autores [FUK 90], [LAN 03], [DUD 01] e [THE 01]. Assumindo a distribuição Normal multivariada para os dados, esta distância estatística assume a seguinte forma:

$$b_{ij} = \frac{1}{8} (m_i - m_j)^t \left(\frac{\Sigma_i + \Sigma_j}{2} \right)^{-1} (m_i - m_j) + \frac{1}{2} \ln \left(\frac{\left| \frac{\Sigma_i + \Sigma_j}{2} \right|}{\sqrt{|\Sigma_i| |\Sigma_j|}} \right)$$

onde b_{ij} é a distância de Bhattacharyya entre as classes i e j , m_i e m_j os vetores média e Σ_i e Σ_j as matrizes de covariância das classes, obtidas a partir de Y , estando a partição definida na posição m .

Conforme já foi mencionado anteriormente, quanto maior o valor da distância de Bhattacharyya, mais separadas estão as funções de densidade de probabilidade das duas classes e teoricamente menor o erro de classificação obtido. Uma vantagem da utilização desta medida é que ela está diretamente relacionada à probabilidade de erro no processo de classificação. O valor mínimo para probabilidade de erro é conhecido como erro de Bayes. Como é sabido, o erro de Bayes é causado pela superposição que ocorre entre as funções densidades de probabilidade $-p(x|\omega_i)-$ das classes envolvidas no processo de classificação. Este erro é uma propriedade inerente ao problema de classificação que não pode ser eliminado [DUD 01]. O valor do erro de Bayes, constitui-se portanto em um outro critério para avaliação da separabilidade entre pares de classes. Dada, entretanto, a dificuldade em estimar diretamente o valor do erro de Bayes para cada caso em particular, é prática comum a utilização de limiares (*upper and lower bounds*) que limitam este erro e que sejam computacionalmente mais

tratáveis. Fukunaga [FUK 90] sugere a seguinte expressão para limiar superior ao erro de Bayes, fazendo uso da distância de Bhattacharyya:

$$\varepsilon_B \leq \sqrt{p(\omega_i) p(\omega_j)} e^{-b_{ij}} \leq \frac{1}{2} e^{-b_{ij}} \quad (25)$$

onde $p(\omega_i)$ e $p(\omega_j)$ são as probabilidades de ocorrência a priori das classes.

Uma extensão deste limiar para o caso multiclases foi sugerido por Bruzzone [BRU 95]:

$$\varepsilon_B \leq \sum_{i=1}^C \sum_{j>i}^C \sqrt{p(\omega_i)p(\omega_j)} e^{-b_{ij}} \leq \frac{1}{C} \sum_{i=1}^C \sum_{j>i}^C e^{-b_{ij}} \quad (26)$$

onde C é o número de classes presentes e b_{ij} a distância de Bhattacharyya (15) entre cada par de classes.

Definido o critério para avaliação da separabilidade entre classes, resta agora estabelecer um critério de otimização $J(m)$ propriamente dito. Para um problema envolvendo apenas duas classes, um critério aceitável consistiria em definir $J(m)$ em termos de um limiar superior para o erro bayesiano:

$$J(m) = -\varepsilon_B \geq -\sqrt{p(\omega_i)p(\omega_j)} e^{-b_{ij}} \quad (27)$$

sendo que o sinal negativo em (27) valida a procura por um valor máximo. O par de classes que apresenta o valor mais alto para o erro de Bayes pode, portanto, ser considerado como o par que apresenta a maior dificuldade de separação constituindo-se em um critério adequado ao processo de extração das feições.

Uma outra possibilidade seria definir $J(m)$ diretamente pela distância de Bhattacharyya dentro da suposição de normalidade multivariada dos dados. Neste caso, o par de classes apresentando o menor valor para a distância de Bhattacharyya é aquele de mais difícil separabilidade.

$$J(m) = b_{ij} \quad (28)$$

ou seja, a distância de Bhattacharyya entre classes, já que o produto $\sqrt{p(\omega_i)p(\omega_j)}$ neste caso é uma constante.

Para o problema genérico, envolvendo um número qualquer de classes (problema multiclasse) a tarefa de otimização não é tão evidente, na medida que possivelmente a melhor otimização, em termos de segmentação, que poderia ser obtida para um determinado par de classes não é necessariamente também a melhor para os demais pares. Para tanto, propõe-se dois critérios para lidar com este tipo de situação:

3.4.1 - CR1(UB): Critério 1 - Minimização do Limiar Superior do Erro Bayesiano

Assumindo que as classes apresentem distribuição de probabilidades do tipo normal, o primeiro critério $J(m)$ explorado neste trabalho para avaliação das partições geradas é baseado na minimização do limiar superior (*upper bound*) do erro bayesiano ε_B (26) que obtem-se num processo de classificação. Minimizando-se esse limiar, espera-se também minimizar o valor do erro de Bayes obtido posteriormente na classificação. Neste critério, a otimização envolve todas as classes presentes. Define-se:

$$J(m) = - \sum_{i=1}^C \sum_{j>i}^C \sqrt{p(\omega_i)p(\omega_j)} e^{-b_{ij}} \quad (29)$$

Neste critério, a situação genérica, envolvendo um número qualquer de classes, é avaliada sob a ótica de pares de classes. Estando a partição posicionada em m , para o cálculo de $J(m)$ faz-se necessária a avaliação da separabilidade estatística entre todos os possíveis pares de classes, para tanto se utiliza a distância de Bhattacharyya b_{ij} (15) (que é definida entre pares de classes).

3.4.2 - CR2(DM): Critério 2 - Distância Mínima Entre Pares de Classes

Neste critério, propõe-se uma segmentação definida de modo a otimizar a separabilidade do par de classes que apresenta o menor valor para separabilidade, ou seja, de maior dificuldade para classificação, e aplicar esta segmentação para os demais pares de classes. Para tanto, a cada posição de partição m testada computam-se os valores de b_{ij} para todos os pares de classes. Depois de analisadas as posições m , segundo o algoritmo utilizado, e calculados os valores de b_{ij} para todas as possíveis combinações de pares de classes, verifica-se qual o par de classes obteve a menor média aritmética de distâncias b_{ij} (a média é computada para cada um dos pares de classes,

utilizando-se os valores de b_{ij} em todas as posições de m que foram testadas). O par cuja média das distâncias b_{ij} é a menor, é teoricamente o par de maior dificuldade para separação. Define-se $J(m)$ como sendo o valor b_{ij} obtido na posição m para esse par:

$$J(m) = \text{valor}(\text{Classe} \min\{\text{media}(b_{ij})\}) \quad , \quad (30)$$

sendo $1 \leq i < C$ e $i < j \leq C$.

Em ambos os critérios 1 e 2, o problema de otimização acaba sendo avaliado dentro de uma ótica de pares de classes. Nota-se que o valor negativo em (29) serve apenas para indicar o procedimento usual de procura por um máximo. Vale lembrar que a procura é pelo menor valor em módulo e que este valor está associado ao do limiar mais baixo estimado para o erro de Bayes.

Ao maximizar-se $J(m)$, procura-se o ponto que minimize o limiar superior para o erro bayesiano de classificação, que se espera estar relacionado ao real erro de classificação que será obtido posteriormente. Na realidade, garante-se estar no ponto de mínimo do limiar superior do erro de Bayes. Isto se deve ao fato de que a função de otimização, seja qual for o critério utilizado, estar ligada ao limiar ou *upper bound* do erro bayesiano e não ao valor exato do erro de classificação. Embora o limiar seja um indicador do limite máximo do erro, não se conhece de antemão exatamente qual o valor que se obterá posteriormente na classificação. Não é possível afirmar categoricamente que é exatamente o ponto de mínimo erro na classificação. Porém, como o cálculo da probabilidade de erro e em geral uma tarefa bastante difícil, a utilização de limiares é descrita pela literatura como uma boa alternativa.

3.5 - Cálculo da acurácia da classificação

A acurácia de classificação, para cada classe, é obtida pelo cociente entre o número de pixels corretamente classificados na classe pelo número total de pixels utilizados na classe em questão.

4 - ESTUDO DE CASO

4.1 - Dados disponíveis: imagem hiperespectral do sensor AVIRIS

Para a implementação e validação das metodologias propostas neste trabalho, utilizou-se um segmento obtido a partir de dados coletados pelo sistema sensor AVIRIS, disponibilizada gratuitamente para fins científicos pelo Prof. David Landgrebe, da Purdue University. Este segmento é composto por 88 linhas e 70 colunas, totalizando 6160 pixels com um total de 220 bandas, cobrindo uma faixa do espectro eletromagnético que vai desde 400 até 2500 nanômetros imageando, portanto, a região do visível até o infravermelho. Uma composição colorida da imagem pode ser visualizada na Figura 16(a).

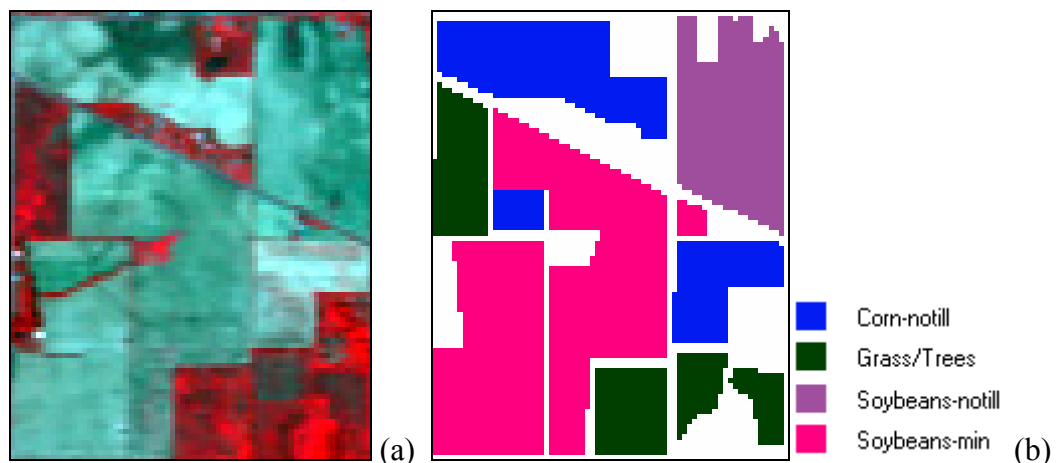


Figura 16. Em (a) temos o recorte da área de interesse (88 x 70 pixels). Apresenta-se uma composição colorida no sistema RGB, bandas 50, 27, 17. Em (b) temos a respectiva verdade terrestre.

Esta imagem, obtida em Junho de 1992, com resolução espacial de 20 metros, cobre uma região agrícola (região experimental de testes da Universidade Purdue) denominada de *Indian Pines*. Esta área está localizada no noroeste do estado de

Indiana, nos EUA. Nela apresentam-se cultivos de soja, milho entre outros. A imagem foi obtida no início da época de crescimento, ocorrendo por isso grande quantidade de solo exposto. Foram ainda utilizadas técnicas de cultivo distintas, com variações desde o cultivo tradicional até o cultivo direto. Esta situação resulta em classes espectrais de resposta muito semelhante, sendo portanto de difícil separação.

Esta imagem é particularmente atraente para fins de pesquisa pois, sendo um local controlado para fins de experimentos, os cultivos estão convenientemente demarcados e sua localização, ou verdade terrestre (Figura 16(b)), é disponível com boa exatidão permitindo, desta forma, testar acuradamente os resultados obtidos de processos de classificação.

Tendo a imagem sido obtida no início do crescimento dos cultivares, a parcela de resposta espectral de cada pixel é função não apenas da resposta da própria planta mas, também, do solo. Se ao longo da imagem o solo apresentar, por exemplo, diferentes concentrações de nutrientes, a resposta obtida para diferentes pixels, mesmo que representando o mesmo cultivar na sua forma de plantio, pode ser bastante diferente. Em termos estatísticos, esta falta de homogeneidade faz com que a distribuição de valores de resposta espectral dos pixel de algumas classes de informação (classes temáticas) não apresente uma distribuição normal gaussiana mas, em alguns casos, mostre características multimodais, caracterizando a existência de duas ou mais classes espectrais dentro da mesma classe de informação. Isto é uma condição bastante indesejada ao utilizar-se, por exemplo, um classificador paramétrico que supõe uma distribuição normal de probabilidades. Por esta razão, a fim de minimizar uma possível falta de homogeneidade, optou-se por trabalhar com apenas um segmento (Figura 16) da imagem total disponível.

No recorte, convenientemente efetuado para realização dos experimentos visando à validação das soluções aqui propostas, encontram-se as classes temáticas milho com cultivo direto (*corn-notill*), gramíneas (*grass*), soja com cultivo direto (*soybeans-notill*) e soja com cultivo mínimo (*soybeans-min*).



Figura 17. Vista de uma área contendo a cultivar *milho-cultivo direto (sorn-notill)* mostrando a baixa porcentagem que representa a planta frente à área emaciada no momento da coleta dos dados [LAN 2003]



Figura 18. Vista de uma área da cultivar *soja com cultivo mínimo (soybeans-min)* mostrando a baixa porcentagem que representa a planta frente à área emaciada no momento da coleta dos dados [LAN 2003].

Os dados utilizados nestes experimentos são muito desafiadores. A imagem foi tomada no início da época de crescimento dos cultivos milho e soja, onde apenas aproximadamente 5% da área está efetivamente coberta pela vegetação, sendo os restantes 95% composto por solo exposto e resíduos de colheitas anteriores. Além disto, diferentes práticas de cultivo foram aplicadas na região. Desta forma, a superfície apresenta diferentes quantidades de resíduo do ano precedente. As diferenças no processo de preparação do solo podem ser assim descritas:

1. Cultivo direto (no till): é uma prática que está se tornando cada vez mais comum, na qual o novo plantio é feito sem cobrir o resíduo da cultura do ano anterior. Este resíduo permanece na superfície, protegendo o solo da erosão;
2. Cultivo mínimo (minimum till): é feita apenas uma preparação mínima do terreno antes do plantio da nova cultura;

3. Cultivo tradicional (clean till) : é o processo no qual o terreno é todo removido e preparado antecipadamente para o plantio, de tal forma que a superfície do solo permanece exposta;

Desta forma, os solos podem ser bastante heterogêneos e os resíduos na superfície provenientes do ano anterior podem ainda estar presentes. Nestas circunstâncias, a discriminação destes cultivos torna-se um problema bastante desafiador.

Utilizando os dados disponíveis de verdade terrestre, foram construídos, para cada uma das classes de informação, dois conjuntos de pixels: o primeiro para fins de treinamento e o segundo para fins de teste (training set e test set). Como as áreas de plantio ocupadas por estas quatro classes de informação não são iguais, o tamanho das amostras é variável de uma classe para outra. A Tabela 2 abaixo ilustra o número total de pixels disponíveis.

Tabela 2. Número total de pixels de treinamento e teste disponíveis para cada uma das classes em estudo.

Classe:	Treinamento	Teste	Total
Milho-cultivo direto	504	504	1008
Gramíneas	373	373	746
Soja-cultivo direto	368	368	736
Soja-cultivo mínimo	973	973	1946

Para a realização dos experimentos, de forma a melhor poder comparar os resultados obtidos, dada a quantidade diferente de pixels em cada uma das classes de estudo, optou-se em uniformizar o tamanho das amostras das classes em estudo, ao longo dos vários experimentos. Foram assim selecionando subconjuntos menores de pixels de treinamento e teste, de tamanhos previamente definidos (150, 200, 250, 300 e 350 pixels) do total disponível para cada classe. Este processo foi aplicado tanto aos pixels de treinamento como aos de teste. Os subconjuntos (de treinamento e de teste) foram selecionados de forma a cobrirem de uma maneira o mais uniforme possível a totalidade dos conjuntos de treinamento e de teste. Outra possível alternativa seria uma seleção aleatória de pixels nos dois conjuntos, de maneira a formar os subconjuntos desejados. Esta uniformização com relação ao tamanho das amostras torna-se necessária

para que os resultados obtidos das várias classes de informação sejam comparáveis entre si. Uniformidade quanto ao número de amostras de treinamento para todas as classes, torna a confiabilidade nas estimativas dos parâmetros do classificador (especialmente com relação à matriz covariância) mais uniforme para as classes em estudo. A uniformidade com relação ao número de amostras de teste oferece melhores condições de comparar os vários resultados entre si.

4.2 - Seleção das bandas originais

A energia eletromagnética proveniente do sol sofre atenuações ao atravessar a atmosfera da Terra. Para determinados comprimentos de onda, coincidentes com a energia de vibração das moléculas de alguns gases que compõe a atmosfera (CO_2 , H_2O , O_3 e outros) esta energia é completamente absorvida. Como o sensor AVIRIS opera de uma forma quase que continua no intervalo de comprimentos de onda de 400 até 2500 nanômetros, algumas das 220 bandas disponíveis coincidem com as bandas de absorção acima mencionadas, resultando em uma resposta que é basicamente ruído (Figura 19(c)). Assim sendo, estas bandas não agregam informação alguma, e sua exclusão contribui ainda para a redução na dimensionalidade dos dados. Analisando o total de bandas disponíveis neste experimento, concluiu-se que do total de 220 bandas, apenas 190 bandas realmente apresentam poder discriminante, podendo as demais 30 bandas serem descartadas (Tabela 3). Desta forma, cada pixel pode ser entendido como sendo um vetor de dimensão $[1 \times 190]$, onde cada componente representa a quantidade de energia (radiância) refletida ou emitida em uma determinada faixa de comprimentos de onda de radiação eletromagnética incidente.

Tabela 3. Seleção das bandas ruidosas que foram excluídas (30 bandas).

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40
41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60
61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80
81	82	83	84	85	86	87	88	89	90	91	92	93	94	95	96	97	98	99	100
101	102	103	104	105	106	107	108	109	110	111	112	113	114	115	116	117	118	119	120
121	122	123	124	125	126	127	128	129	130	131	132	133	134	135	136	137	138	139	140
141	142	143	144	145	146	147	148	149	150	151	152	153	154	155	156	157	158	159	160
161	162	163	164	165	166	167	168	169	170	171	172	173	174	175	176	177	178	179	180
181	182	183	184	185	186	187	188	189	190	191	192	193	194	195	196	197	198	199	200
201	202	203	204	205	206	207	208	209	210	211	212	213	214	215	216	217	218	219	220

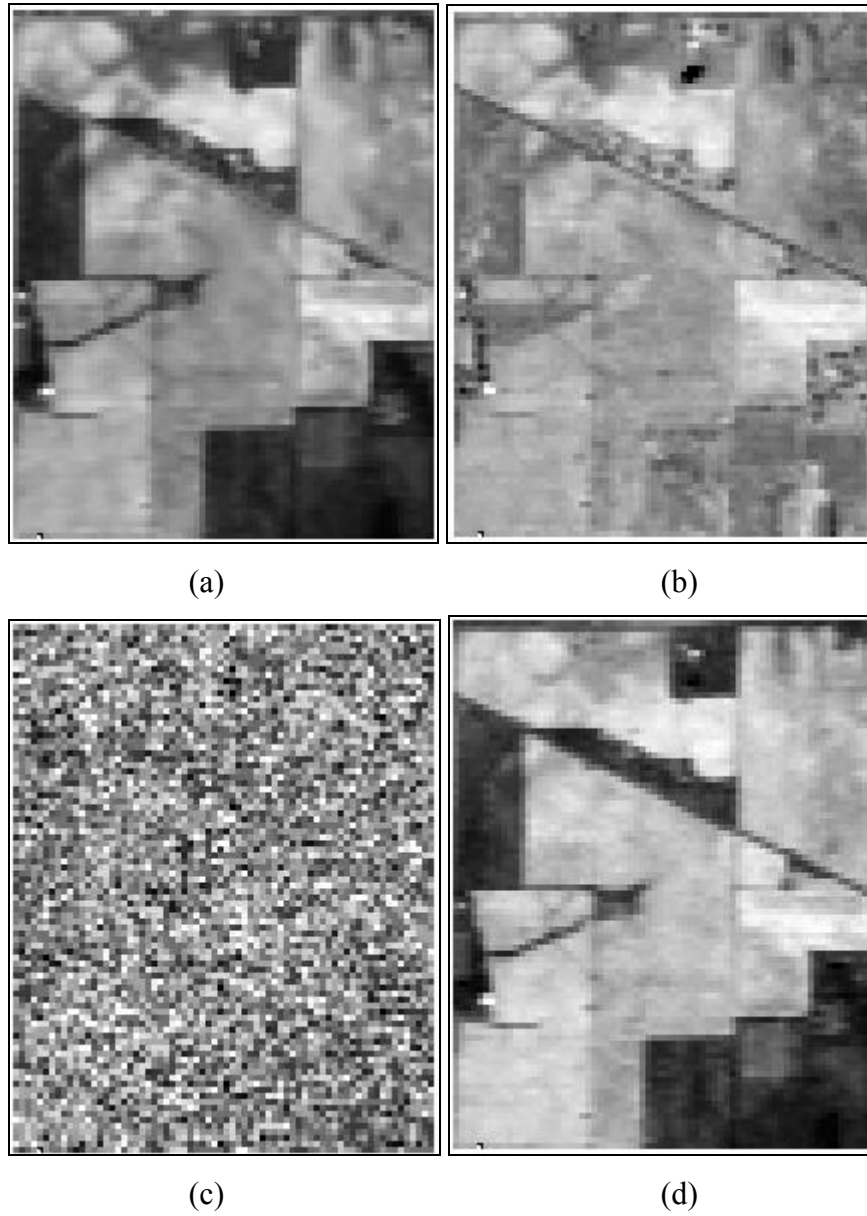


Figura 19. Bandas originais: em (a) temos a banda 25, em (b) a 85, em (c) a 150 e em (d) temos a banda 200.

4.3 - Representação das curvas de resposta espectral

A fim de poder melhor visualizar o comportamento de cada uma das quatro classes de interesse examinadas neste trabalho, foram elaborados gráficos com a resposta espectral obtida (assinatura espectral) onde se apresenta a envoltória dos valores de contador digital mínimo e máximo e o valor médio, banda a banda, para cada uma das quatro classes de treinamento. A Figura 20 ilustra os resultados obtidos.

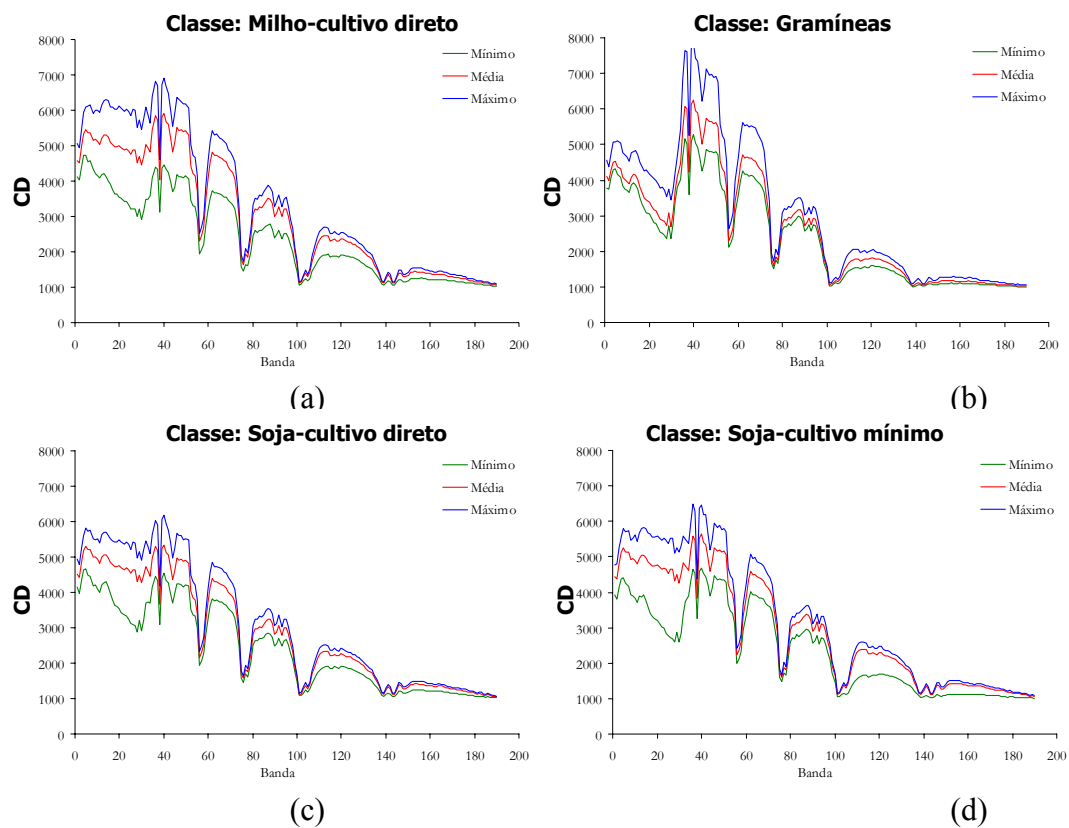


Figura 20. Envoltória das classes de treinamento. Em (a), (b), (c) e (d) são apresentados os valores da resposta mínima, média e máxima das classes de treinamento *Milho-cultivo direto*, *Gramíneas*, *Soja-cultivo direto* e *Soja-cultivo mínimo*, respectivamente, avaliada banda a banda, para todos os pixel de cada classe, conforme o total disponível em cada classe, anteriormente apresentado na Tabela 2.

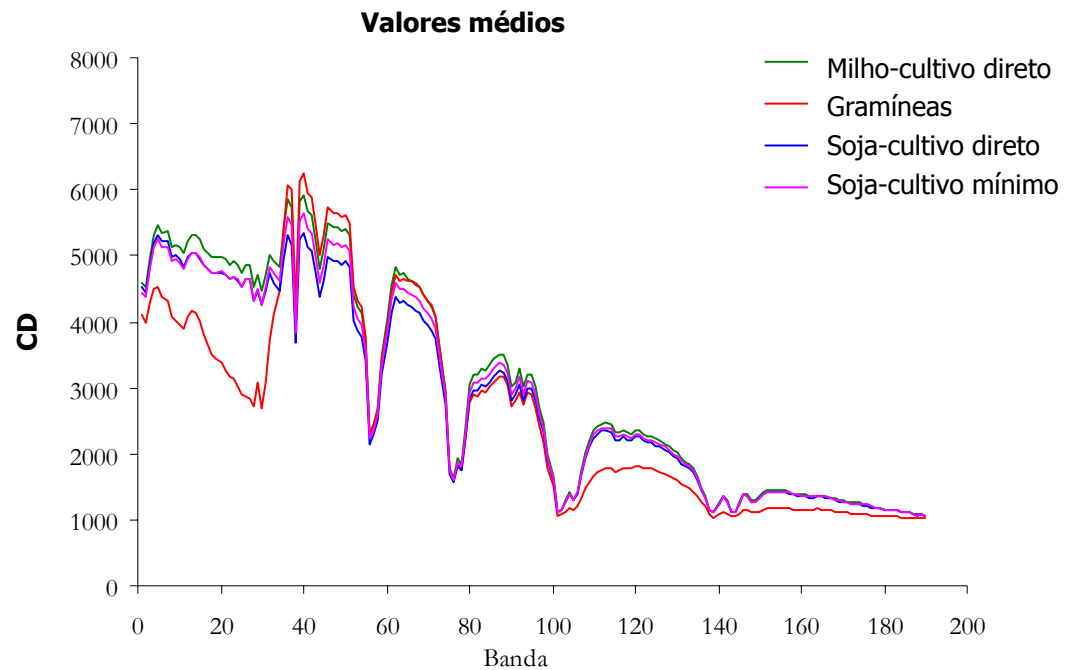


Figura 21. Resposta média para as quatro classes de treinamento em estudo

Visualmente pode-se perceber que, à exceção da classe gramíneas (*grass*), o comportamento das demais classes são muito semelhantes. Fica assim evidenciada a dificuldade de obtenção de uma elevada acurácia na classificação. Entretanto, era de se esperar tal comportamento, visto que se tratam de cultivares espectralmente muito semelhantes. No caso das duas classes de soja (*soybeans*), elas diferem entre si apenas pela forma de cultivo.

Cabe perceber ainda que para cada classe em estudo, as curvas de resposta espectral dos vários pixels que compõem as amostras apresentam uma dispersão em torno de um valor médio, sendo os limites estabelecidos pelas envoltórias de valores mínimos e máximos. Na Figura 20 está ilustrado o quão dispersos estão os valores da resposta espectral dos dados coletados pelo sensor AVIRIS no caso do experimento sob análise. Percebe-se que a amplitude desta dispersão não é constante ao longo da curva de resposta espectral.

Deve-se atentar para o fato que as envoltórias de valores máximos e mínimos servem apenas para ilustrar a dispersão de valores em torno da curva média. Para fins de classificação, cada pixel das amostras de teste a serem classificadas será comparado com a curva de resposta média de cada uma das classes. Além do vetor

média, cada classe é também caracterizada pela sua matriz de covariância, que pode ser entendida como uma representação da orientação espacial dos dados da classe no espaço das feições. Pode-se demonstrar [FUK 90] que, utilizando-se dados em alta dimensionalidade, duas classes que compartilham do mesmo vetor média, podem, não obstante, serem separadas com alta acurácia desde que suas matrizes de covariância difiram substancialmente.

4.4 - Materiais necessários

Para a realização deste trabalho de pesquisa, foram utilizados os seguintes materiais:

4.4.1 - Equipamentos

Para a realização dos experimentos utilizou-se computador PC com a seguinte configuração: processador AMD Athlon Xp 1800+, Placa mãe ASUS A7V266-C, memória RAM Samsung 512 Mb (2 x 256) DDR PC 266, HD MAXTOR 40GB 7200 rpm, placa de vídeo Prolink Gforce4 MX420, Gravador CD LG 32X10X40. Para a redação do texto, utilizou-se este e outros computadores com processador de texto *Microsoft Word*, planilha *Microsoft Excel* e *Paint*.

Impressora Jato de tinta HP 860 e materiais de escritório convencionais (folhas de papel A4, mídias em disquete e CDs, etc).

4.4.2 - Aplicativos

Os programas necessários a este estudo foram desenvolvidos em linguagem MATLAB. Foi ainda utilizado o software Multispec for Windows (produzido pela Purdue University e de acesso livre). A maior parte do trabalho foi desenvolvida em linguagem MATLAB, deste a extração dos dados da imagem, tratamento de informações e elaboração de rotinas e programas específicos para as técnicas que foram desenvolvidas neste trabalho.

4.5 - Realização de experimentos

O objetivo deste estudo consiste no desenvolvimento de novos métodos de extração de feições em dados de imagem em alta dimensionalidade (dados hiper-espectrais) com a finalidade de reduzir a dimensionalidade dos dados e desta forma minimizar os efeitos do fenômeno de Hughes. Estes experimentos visam quantificar numericamente os resultados de desempenho das metodologias propostas, especialmente no que tange o comportamento da acurácia no processo de classificação de imagens digitais utilizando feições extraídas. A eficiência computacional dos métodos de extração de feições propostos neste estudo é também avaliada.

Inicialmente procedeu-se a uma classificação do tipo supervisionada, utilizando-se o classificador máxima Verossimilhança Gaussiana (sob hipótese de que os dados apresentam distribuição gaussiana multivariada) onde se utilizaram os dados das classes na sua dimensionalidade original (190 bandas), variando o número de amostras utilizadas para treinamento e teste. O objetivo aqui consiste em estabelecer um patamar de referência para comparação dos resultados obtidos utilizando feições extraídas, em lugar das feições (bandas) originais.

Neste processamento inicial, foram portanto utilizadas todas as 190 bandas AVIRIS não ruidosas, resultando em pixels aqui representados por vetores de dimensão 190. Com a finalidade de tornar os resultados obtidos para as várias classes comparáveis entre si, foram utilizados, para todas as classes, sub-conjuntos de treinamento e de teste de mesmo tamanho para todas as classes em estudo. Estes sub-conjuntos foram extraídos do conjunto das amostras disponíveis (Tabela 2).

A partir dos dados originais de cada classe (190 bandas espectrais), foram extraídas feições. Para tanto, foram aplicados os algoritmos propostos no Capítulo 3, particionando as curvas de resposta espectral em um número de segmentos, de forma idêntica para todas as classes envolvidas. Em cada segmento, os contadores digitais associados as bandas espectrais foram então substituídos por um número menor de feições (média e variância). O número de feições extraídas para cada pixel torna-se então igual ao dobro do número de segmentos obtidos no particionamento da curva de resposta espectral. Os parâmetros do classificador foram então estimados a partir das amostras de treinamento, utilizando as feições extraídas. As acurácias assim obtidas no processo de classificação foram então estimadas por meio das amostras de teste. A

eficiência de cada um dos métodos propostos no Capítulo 3 for então avaliada, em termos da acurácia obtida, conforme o valor estimado a partir dos sub-conjuntos de amostras de teste. A eficiência computacional de cada algoritmo foi também medida.

Foram desenvolvidos um total de quatro experimentos. No experimento 1, foram utilizadas as 190 bandas espectrais originais. Nos demais experimentos foram utilizadas as feições extraídas. Em todos os experimentos, foi empregado o classificador Máxima Verossimilhança Gaussiana, assumindo-se ainda iguais valores para as probabilidades “a priori” de ocorrência das classes.

4.6 - Experimento 01 - classificação utilizando a dimensionalidade original

Este experimento reporta os resultados de classificação obtidos utilizando as 190 bandas originais disponíveis. O principal objetivo deste experimento consiste em fornecer um patamar de referência aos experimentos envolvendo feições extraídas. São utilizadas as 190 bandas espectrais significativas, e a acurácia obtida é estimada para diferentes valores do tamanho da amostra de treinamento, i.e., diferentes graus de confiabilidade na estimativa amostral dos parâmetros do classificador, em especial a matriz de covariância. Os tamanhos utilizados para as amostras de treinamento foram: 200, 250, 300, 350 (pixels). Note-se que o valor mínimo admissível para o tamanho das amostras de treinamento é igual à dimensionalidade dos dados mais um. Um valor inferior resultará em que a estimativa amostral da matriz de covariância será singular e portanto não utilizável no classificador Máxima Verossimilhança Gaussiana [LAN 03].

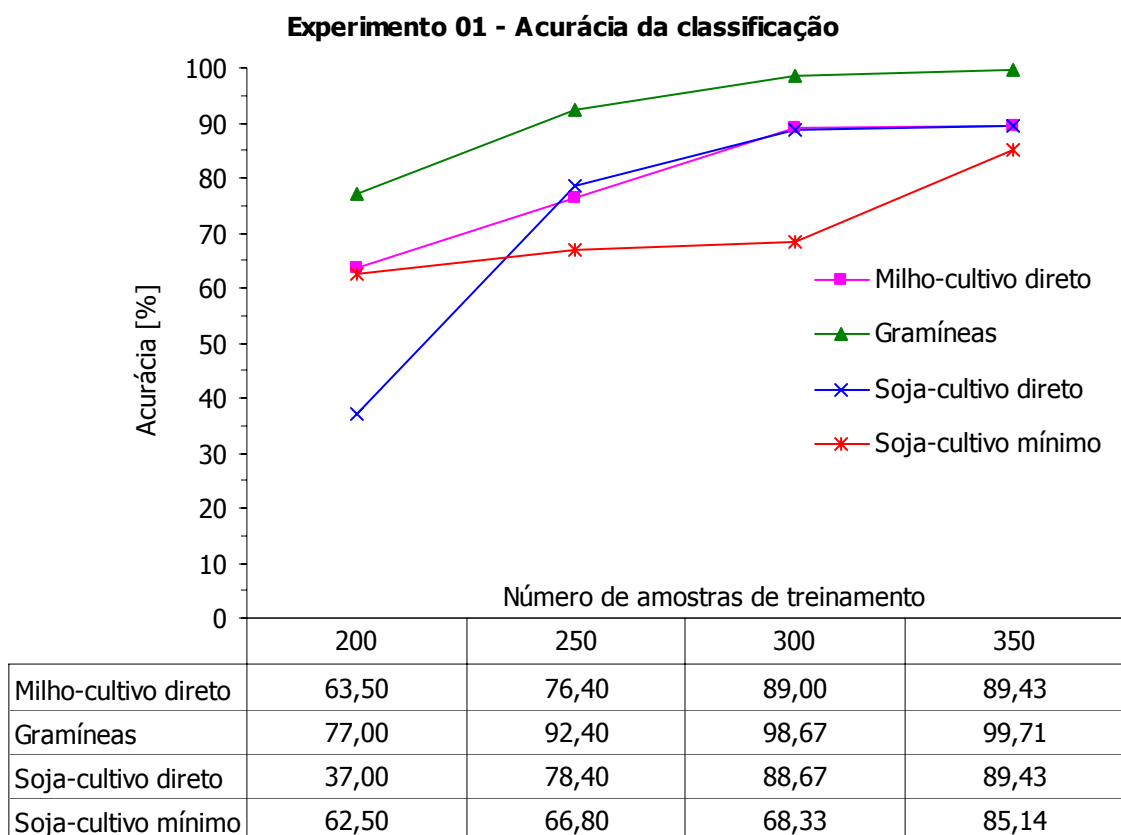


Figura 22. . Experimento 1: Valor estimado para a acurácia do processo de classificação (MAXVER) obtidos para cada uma das quatro classes de interesse, frente a diferentes tamanhos de amostras de treinamento e teste. São utilizados dados em sua dimensionalidade original (190 bandas).

A Figura 22 ilustra os resultados obtidos, em termos de valores estimados para a acurácia de classificação, frente ao número de amostras de treinamento de teste que foram utilizadas para cada classe. Os resultados apresentam uma tendência geral já esperada: o valor estimado para a acurácia obtida no processo de classificação tende a aumentar na medida em que o tamanho das amostras de treinamento também aumenta. Esta tendência era esperada, pois na medida em que aumenta o número das amostras disponíveis aumenta também a confiabilidade na estimativa amostral dos parâmetros do classificador, em especial, da matriz de covariância. Este fato pode também ser verificado em termos dos valores estimados para a média da classificação, apresentada na Figura 23.

Utilizando-se apenas 200 pixels de treinamento, a razão entre o tamanho da amostra e dimensionalidade dos dados é muito baixa ($200/190=1.05$). Na literatura é geralmente recomendado 10 como valor mínimo adequado para esta relação, sendo bastante desejáveis valores bem superiores (da ordem de 100). Nestas condições, os efeitos do fenômeno de Hughes tende a se apresentar de maneira mais intensa. Isto explica a acurácia de classificação de apenas 37% obtida para a classe *soja-cultivo direto* (Figura 22). A acurácia média para as quatro classes nestas condições foi de apenas 60% (Figura 23). Na medida em que foram adicionados mais pixels às amostras de treinamento, a acurácia média sofreu um considerável aumento, passando para 78.5% (+18.5%) com a adição de apenas 50 pixels, chegando a 86.17% (+26.87%) ao se adicionar 100 pixels e atingindo 90.93% (+30.93%) ao utilizar-se 350 pixels de treinamento (Figura 23).

O fato da classe *gramíneas* apresentar os maiores valores de acurácia (Figura 22) também era previsível, visto esta ser a classe espectralmente mais distinta das demais.

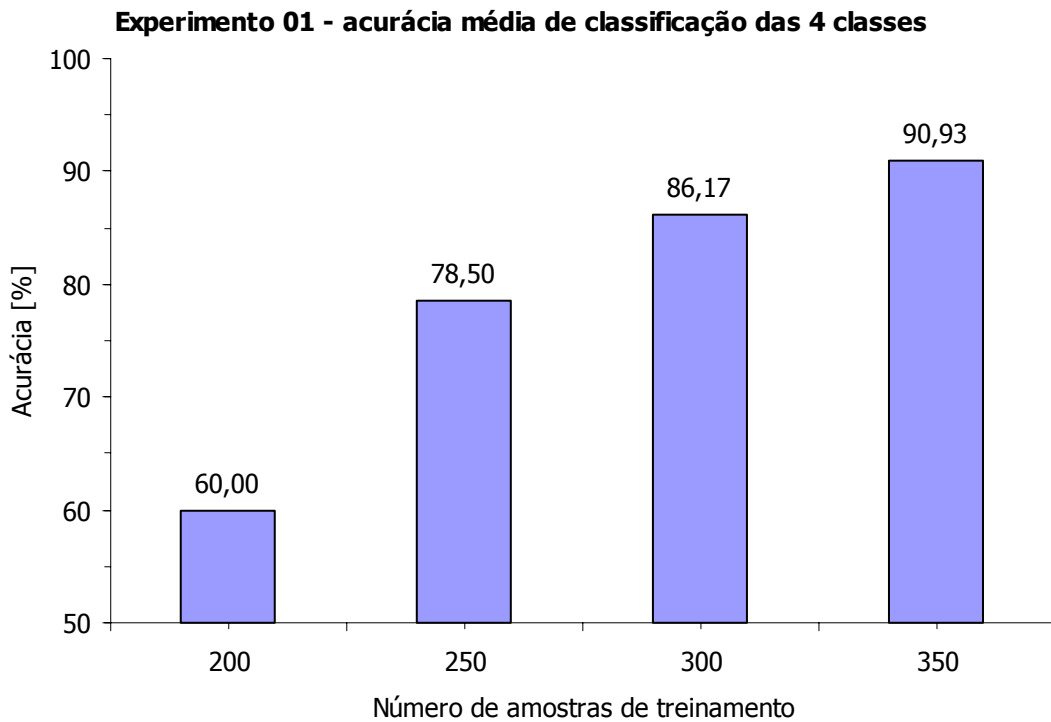


Figura 23. Experimento 1: Resultado médio da acurácia de classificação obtido frente a diferentes tamanhos de amostras de treinamento e teste utilizados para cada classe na dimensionalidade original dos dados (190 bandas).

A Figura 24 apresenta o tempo de processamento total necessário pelo algoritmo desenvolvido na maneira como foi implementado. Este tempo pode ser entendido como a soma do tempo necessário para o treinamento do classificador e o tempo necessário para classificar os pixels nas amostras de teste. Verifica-se que a relação obtida entre tempo de processamento e número de amostras utilizadas é muito aproximadamente linear.

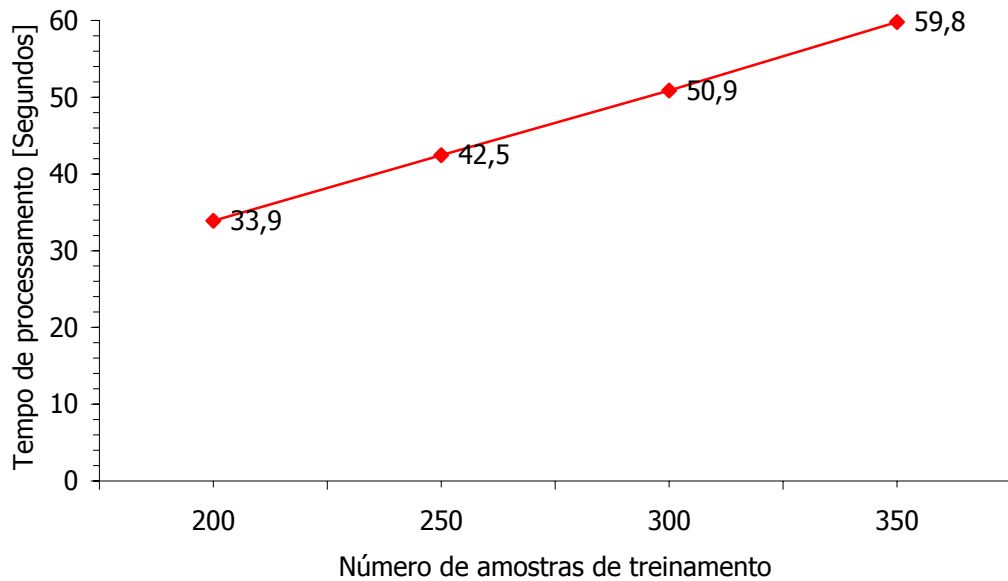
Experimento 01 - Tempo de processamento total

Figura 24. Experimento 1: Tempo de processamento para o algoritmo desenvolvido, frente a diferentes tamanhos de amostras de treinamento e teste para cada classe, utilizando a dimensionalidade original dos dados (190 bandas).

Considerando que a dimensionalidade do problema se manteve inalterada, o tempo para a estimação dos parâmetros do classificador (vetor média de dimensão $[1 \times 190]$ e matriz covariância $[190 \times 190]$) é proporcional ao número de amostras de treinamento. Imagina-se que o tempo de classificação seja linearmente proporcional ao número de amostras a serem classificadas. Desta forma, estaria explicado o comportamento linear obtido no gráfico que apresenta o número de amostras de treinamento versus o tempo de processamento total (Figura 24). Para um número de amostras de treinamento e teste de 200 pixels, o tempo gasto foi de 33.9 segundos. Utilizando-se 350 pixels, o tempo necessário ficou em 59.8 segundos.

A título de esclarecimento com respeito à terminologia utilizada neste estudo com relação aos dados de entrada, neste experimento, não existe distinção entre feição e banda. Vale lembrar que uma feição tanto pode ser entendida como uma banda original ou o resultado de uma combinação destas bandas.

4.7 - Experimento 02 - Algoritmo SCC: Segmentação com Comprimento Constante

Neste experimento, implementa-se o algoritmo para extração de feições por segmentação da curva de resposta espectral de cada classe em segmentos de comprimento “n” constante. As feições extraídas consistem do valor médio dos contadores digitais a cada segmento e do correspondente valor para a variância. Estão aqui reportados os valores estimados para a acurácia da classificação e o tempo de processamento. O interesse principal deste experimento é poder avaliar o desempenho do classificador frente à dimensionalidade dos dados (número de feições utilizadas), variando-se também o número de amostras de treinamento utilizado para estimação dos parâmetros.

Para o algoritmo SCC, são testados diferentes comprimentos de segmentação (três, quatro, cinco, seis bandas, etc) conforme a Tabela 4. Para segmentos de tamanho não múltiplo do total de bandas originais disponíveis (190), as bandas restantes foram agrupadas no último segmento (o último segmento finaliza-se na banda 190). Para situações onde diferentes comprimentos de segmento, por exemplo, comprimento igual a 16 e comprimento igual a 17 bandas, resulta no mesmo número de segmentos, 11, (Tabela 4) utilizou-se aquele que o torna o mais próximo possível dos demais segmentos de tamanho constante. No exemplo apresentado, foi utilizado o comprimento $n = 17$. Este valor gerou 10 segmentos contendo 17 bandas cada um e mais um último segmento 20 bandas ($10 \times 17 + 20 = 190$). Esta distribuição de segmentos encontra-se ilustrada na Tabela 3.

Deve-se observar que a metodologia proposta para partição da curva de resposta espectral em segmentos de comprimento constante (com a possível exceção do último segmento), não permite a ocorrência de segmentos com comprimentos arbitrários quaisquer.

Tabela 4. Relações obtidas utilizando o algoritmo SCC proposto: [1] Comprimento dos segmentos; [2] número total de segmentos; [3] número de bandas excedentes a serem acrescentadas ao último segmento e [4] número total de feições resultantes, utilizando-se duas feições por segmento. Os valores destacados representam as configurações que foram efetivamente utilizadas (para alternativas que geram a mesma quantidade de segmentos, são aquelas que minimizam o valor de bandas a serem acrescentadas ao último segmento), ou seja, a coluna [3].

[1] Compr. Segmento	[2] Núm. Segmentos	[3] Band. Último. Se	Número de feições		[1] Compr. Segmento	[2] Núm. Segmentos	[3] Band. Último. Se	Número de feições	
				[4] NFS = 2					[4] NFS = 2
3	63	1		126	50	3	40	6	
4	47	2		94	51	3	37	6	
5	38	0		76	52	3	34	6	
6	31	4		62	53	3	31	6	
7	27	1		54	54	3	28	6	
8	23	6		46	55	3	25	6	
9	21	1		42	56	3	22	6	
10	19	0		38	57	3	19	6	
11	17	3		34	58	3	16	6	
12	15	10		30	59	3	13	6	
13	14	8		28	60	3	10	6	
14	13	8		26	61	3	7	6	
15	12	10		24	62	3	4	6	
16	11	14		22	63	3	1	6	
17	11	3		22	64	2	62	4	
18	10	10		20	65	2	60	4	
19	10	0		20	66	2	58	4	
20	9	10		18	67	2	56	4	
21	9	1		18	68	2	54	4	
22	8	14		16	69	2	52	4	
23	8	6		16	70	2	50	4	
24	7	22		14	71	2	48	4	
25	7	15		14	72	2	46	4	
26	7	8		14	73	2	44	4	
27	7	1		14	74	2	42	4	
28	6	22		12	75	2	40	4	
29	6	16		12	76	2	38	4	
30	6	10		12	77	2	36	4	
31	6	4		12	78	2	34	4	
32	5	30		10	79	2	32	4	
33	5	25		10	80	2	30	4	
34	5	20		10	81	2	28	4	
35	5	15		10	82	2	26	4	
36	5	10		10	83	2	24	4	
37	5	5		10	84	2	22	4	
38	5	0		10	85	2	20	4	
39	4	34		8	86	2	18	4	
40	4	30		8	87	2	16	4	
41	4	26		8	88	2	14	4	
42	4	22		8	89	2	12	4	
43	4	18		8	90	2	10	4	
44	4	14		8	91	2	8	4	
45	4	10		8	92	2	6	4	
46	4	6		8	93	2	4	4	
47	4	2		8	94	2	2	4	
48	3	46		6	95	2	0	4	
49	3	43		6	130	1	0	2	

O passo seguinte consiste no processo de extração de feições propriamente dito. Conforme descrito no Capítulo 3, o conjunto de valores dos contadores digitais em cada segmento é utilizado no cálculo de duas estatísticas, a média e variância, as quais se espera que capturem informações com respeito, respectivamente, à posição e à forma das curvas de resposta espectral no segmento. Experimentos foram realizados com a finalidade de avaliar esta abordagem no processo de classificação de imagens multi-dimensionais. Os parâmetros testados nestes experimentos foram (i) dimensionalidade dos dados em termos de feições extraídas e (ii) número de amostras utilizadas no treinamento do classificador. O primeiro parâmetro é obviamente uma função do número de segmentos resultantes do particionamento da curva de resposta espectral e encontra-se ilustrado na Tabela 3. A dimensionalidade mínima em termos das feições extraídas é obviamente igual a 2 (dois), quando um único segmento, idêntico à curva de resposta espectral é adotado. A dimensionalidade máxima neste experimento corresponde a 126 feições, resultantes de um particionamento da curva de resposta espectral em 63 segmentos. Valores intermediários estão ilustrados na Tabela 3.

Os experimentos desenvolvidos mostram que a acurácia obtida no processo de classificação apresenta um comportamento típico esperado. Com exceção da classe *gramíneas*, que se manteve sempre com elevada acurácia de classificação (valores superiores a 98%), as demais apresentaram uma acurácia crescente com o aumento da dimensionalidade, atingindo um máximo para em seguida decrescer, na medida em que a dimensionalidade dos dados continuava a aumentar, numa clara ilustração do fenômeno de Hughes (Figuras 25 a 29)

Os efeitos do segundo parâmetro (tamanho das amostras de treinamento) com relação à acurácia do processo de classificação, foram também testados. Os resultados apresentados pelos experimentos, mostraram também um comportamento esperado. O valor estimado para a acurácia global mostrou uma relação positiva com o tamanho das amostras de treinamento. Também conforme o esperado, o pico na curva de acurácia, causado pelo fenômeno de Hughes tende a ocorrer para valores mais altos da dimensionalidade dos dados na medida em que o tamanho das amostras de treinamento aumenta. Os efeitos do tamanho das amostras na acurácia do processo de classificação é bem conhecido. Uma amostra maior permite uma estimativa mais confiável dos parâmetros do classificador, em especial da matriz de covariância, resultando em valores mais altos para a acurácia dos resultados. A contribuição nova, a

ser destacada neste estudo, consiste na observação de que a metodologia proposta para extração de feições mostrou-se capaz de capturar as informações contidas na curva de resposta espectral constituindo-se portanto em um possível método para redução da dimensionalidade de dados, como os encontrados em imagens digitais hiperespectrais.

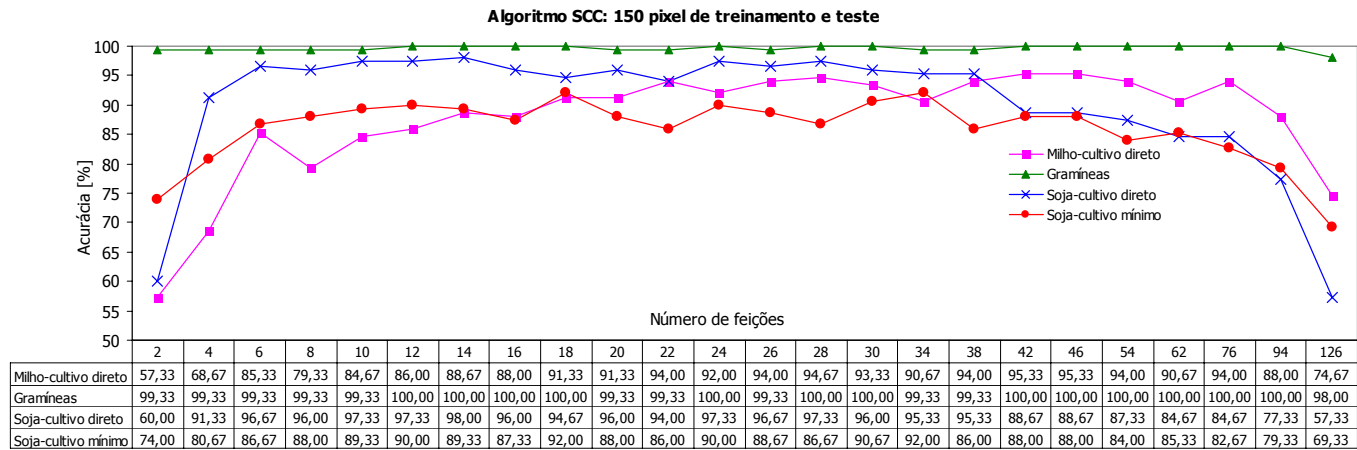


Figura 25. Acurácia da classificação frente à dimensionalidade do problema utilizando-se 150 pixels de treinamento e teste para cada classe.

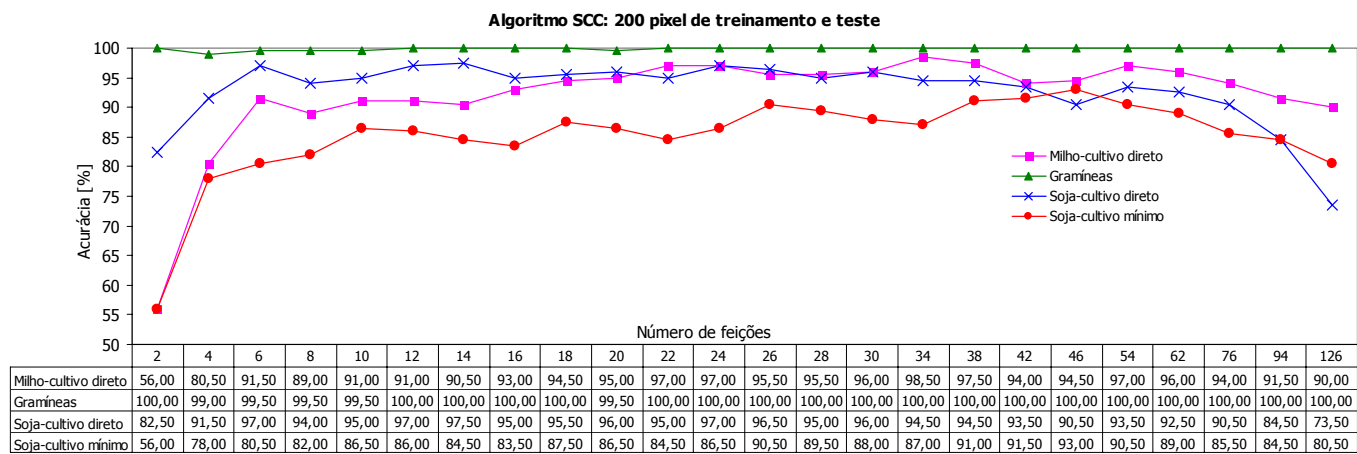


Figura 26. Acurácia da classificação frente à dimensionalidade do problema utilizando-se 200 pixels de treinamento e teste para cada classe.

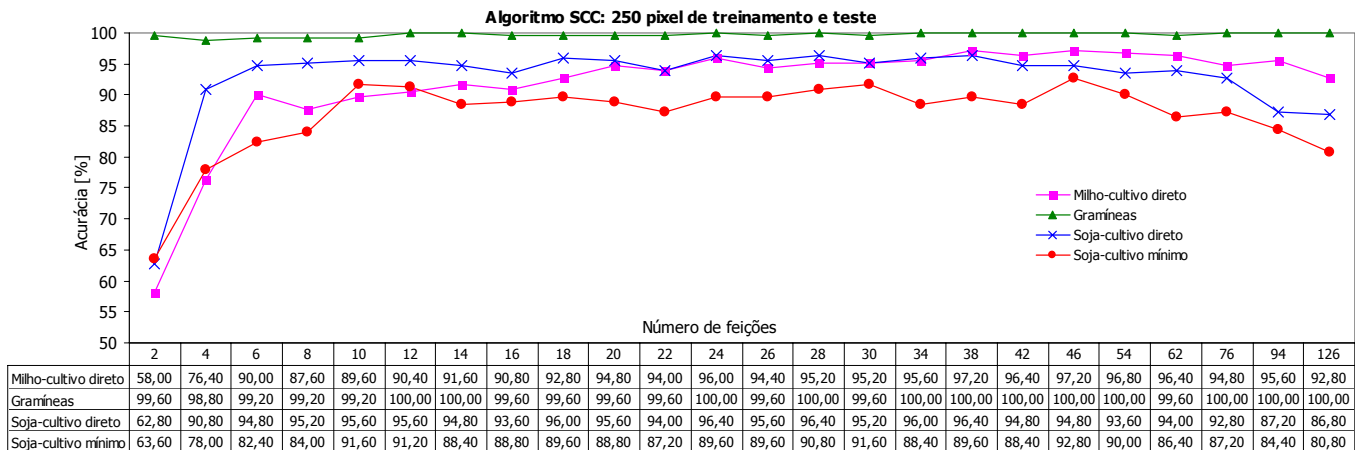


Figura 27. Acurácia da classificação frente à dimensionalidade do problema utilizando-se 250 pixels de treinamento e teste para cada classe.

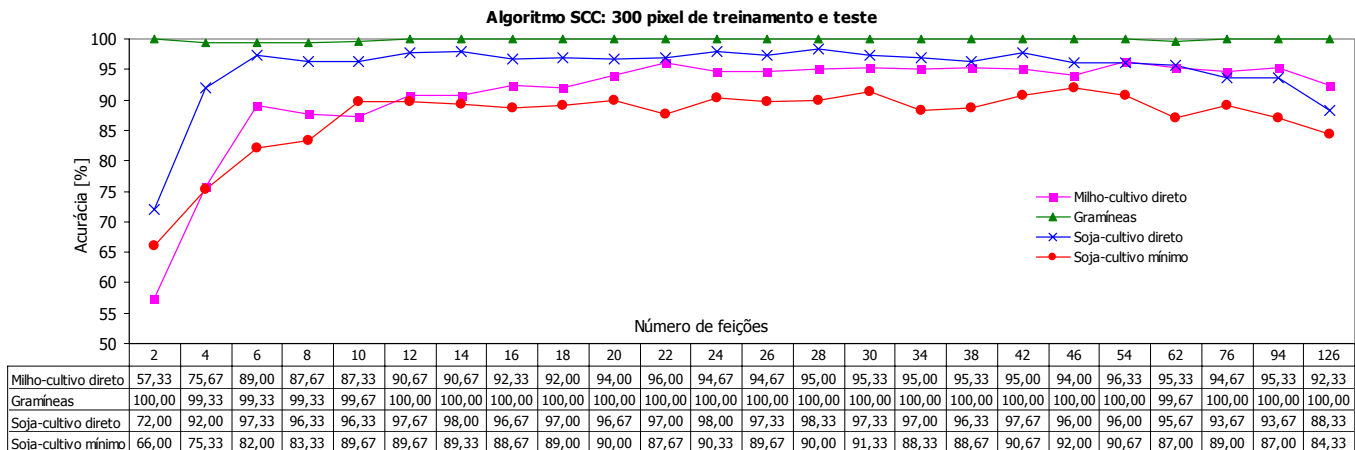


Figura 28. Acurácia da classificação frente à dimensionalidade do problema utilizando-se 300 pixels de treinamento e teste para cada classe.

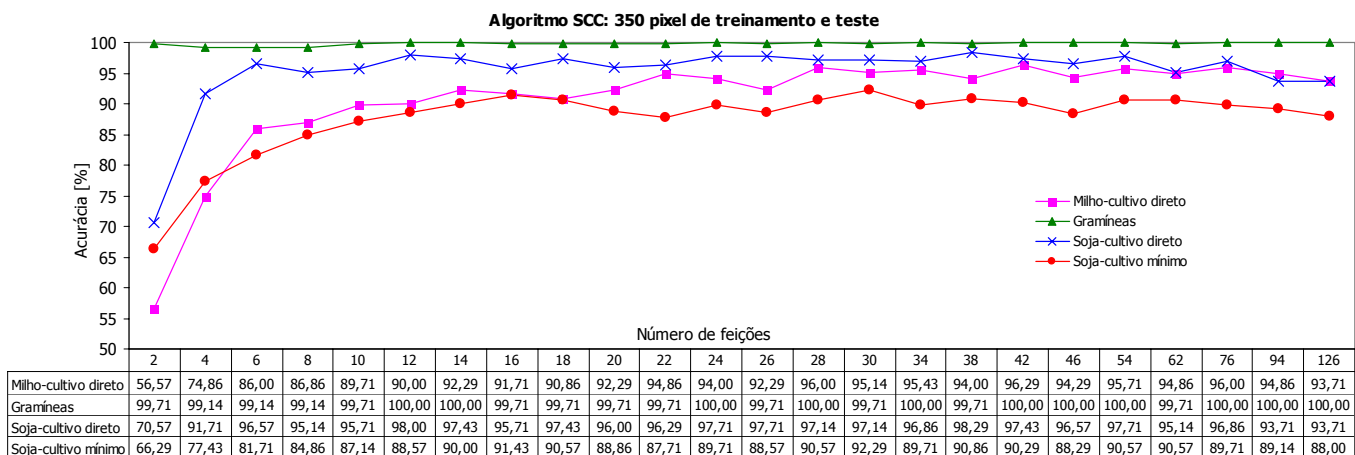


Figura 29. Acurácia da classificação frente à dimensionalidade do problema utilizando-se 350 pixels de treinamento e teste para cada classe.

A Figura 30 apresenta um resumo dos resultados obtidos em termos da acurácia média de classificação (média aritmética dos valores obtidos para cada uma das classes individualmente) frente ao número de amostras de treinamento e teste que foram utilizadas. Percebe-se claramente que, na medida em que o tamanho das amostras decresce, a acurácia no processo de classificação também decresce significativamente. Este efeito é especialmente visível no caso de espaços em alta dimensionalidade (fenômeno de Hughes). Este fato evidencia a necessidade de dispor-se de métodos que permitam a redução de dimensionalidade dos dados utilizados no processo de classificação.

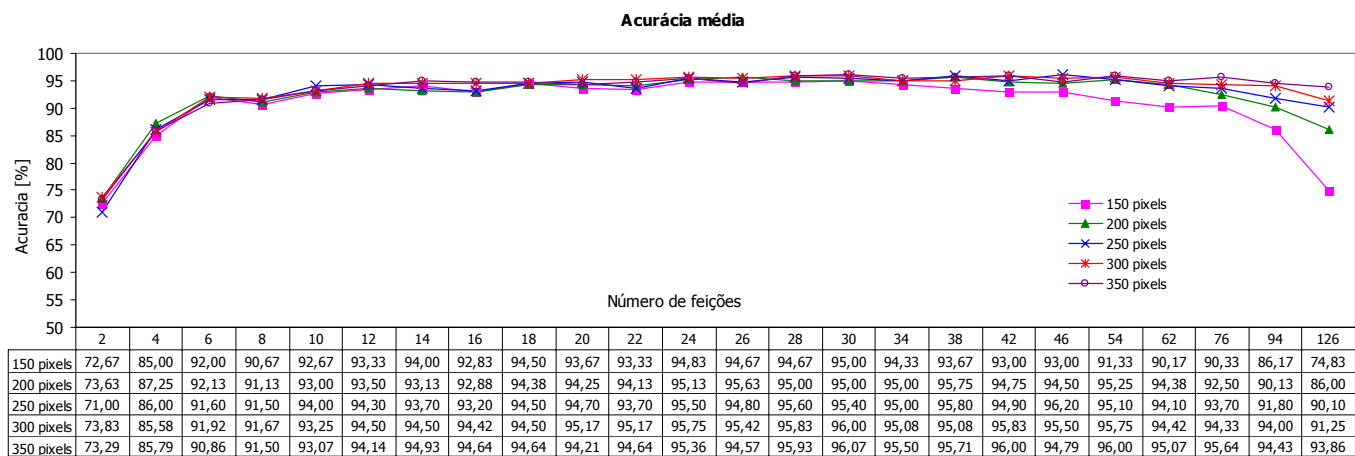


Figura 30. Acurácias médias obtidas frente ao número de pixels de treinamento e número de feições utilizadas.

Para o caso de apenas um segmento (duas feições), a acurácia média obtida apresenta um valor baixo, próximo a 73% (Figura 30). A interpretação óbvia para este baixo valor reside na incapacidade de um tão baixo número de feições extraídas capturar todas as informações relevantes contidas na curva de resposta espectral (190 bandas espectrais) resultando desta forma em uma significativa perda de informação. Adicionando-se mais duas feições (utilizando dois segmentos com um comprimento igual a 95 bandas espectrais, a acurácia mostra um incremento para aproximadamente 86%, ou seja, duas feições adicionais, nestas condições, acrescentaram aproximadamente 13% ao valor estimado para a acurácia da classificação. Aumentando o número de feições para 6, o que corresponde a um comprimento de 63 bandas para os segmentos, a acurácia estimada no experimento supera os 90%. O mesmo experimento mostra que o resultado da acurácia média fornecida pelo classificador atinge suavemente um máximo para apresentar a seguir uma queda, que é justamente mais

pronunciada quando o número de amostras de treinamento é mais reduzido. A Tabela 5 apresenta um resumo dos valores máximos de acurácia de classificação obtidos, com a correspondente comprimento dos segmentos obtidos e o número total de feições obtidas, com diferentes valores para o comprimento dos segmentos e o correspondente número de feições resultantes, frente ao número de amostras de treinamento e teste utilizadas.

Tabela 5. Acurácias máximas obtida (valores médios para 4 classes)

Número de pixel de treinamento e teste	150	200	250	300	350
Comprimento dos segmentos	12	10	8	12	12
Número de feições	30	38	46	30	30
Acurácia média [%]	95.00	95.75	96.20	96.00	96.07

Os resultados ilustrados na Tabela 5 (utilizando-se 150, 200 e 250 pixels de treinamento) estão coerentes com comportamento esperado, conforme relatado na literatura atual. Para as amostras contendo 300 e 350 pixels de treinamento e teste, esperava-se um desempenho um pouco superior, visto ser a relação entre número de amostras e dimensionalidade um pouco maior. De qualquer forma, os valores obtidos são bastante aceitáveis.

A Figura 31 apresenta o tempo de processamento total necessário para a classificação das amostras utilizando o algoritmo proposto e segundo a maneira como foi implementado em linguagem MATLAB. Salienta-se que tal tempo de processamento deve ser tomado como um indicativo de valor apenas relativo, já que o tempo de processamento está intimamente ligado à capacidade do equipamento utilizado. Os valores apresentados são, neste caso, os reais tempos de processamento total que foram necessários para a realização destes experimentos para elaboração desta dissertação, com o equipamento disponível e anteriormente citado. O tempo de processamento apresentado é o total, ou seja, corresponde ao tempo necessário para geração das amostras (segmentação dos dados originais das amostras de treinamento e teste, para a geração das feições) e sua posterior classificação.

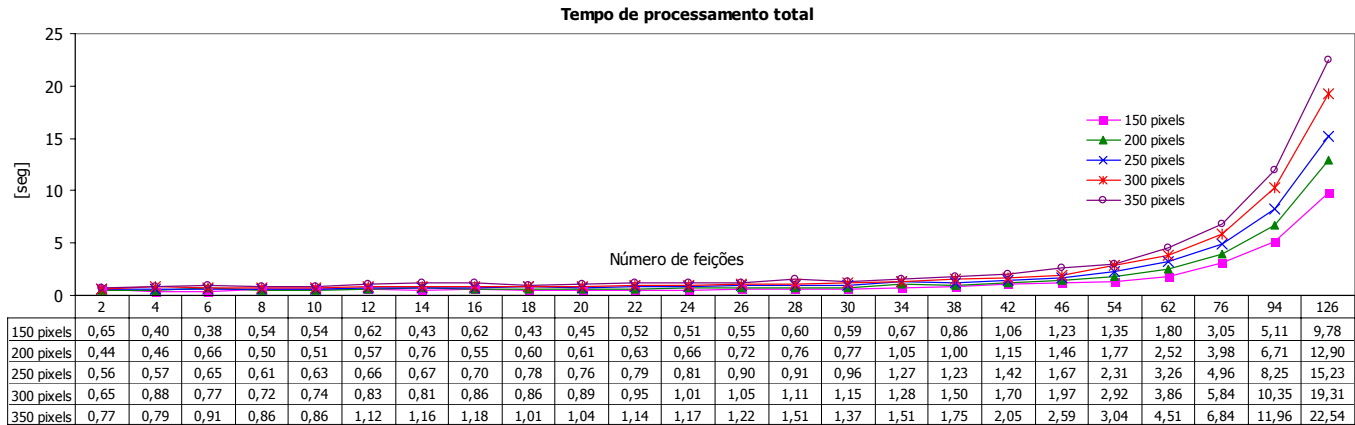


Figura 31. Tempo de processamento total frente ao número de pixels de treinamento e número de feições utilizadas.

Percebe-se que na medida que o número de feições utilizadas aumenta, o número de parâmetros a serem estimados pelo classificador aumenta de maneira quadrática. Por exemplo, utilizando-se duas feições, temos que estimar apenas um vetor média contendo duas componentes, e uma matriz de covariância (simétrica) que terá quatro componentes (três a serem estimados), utilizando quatro feições, devem ser estimado um vetor média de quatro componentes e uma matriz covariância de 16 componentes (10 a serem estimados), e assim sucessivamente. Para o caso de 126 feições, a matriz covariância tem 15876 parâmetros (sendo 8001 a serem estimados).

4.8 - Experimento 03 - Algoritmo Otimização Global

Neste experimento, implementa-se o algoritmo “*Otimização Global*” que corresponde à proposta de segmentação da curva de resposta espectral de cada classe em segmentos de comprimento variável. Estão aqui reportados os resultados da acurácia estimada para os resultados obtidos no processo de classificação e os tempos de processamento correspondentes. Os resultados são apresentados e discutidos.

O interesse principal deste experimento foi testar a adequabilidade do processo de geração dos segmentos com comprimentos variáveis, em termos do desempenho do classificador frente à dimensionalidade dos dados, i.e., do número de feições extraídas. Avaliou-se também os efeitos do tamanho das amostras de treinamento utilizadas na estimação dos parâmetros. A justificativa desta abordagem reside na expectativa de que uma maior flexibilidade no processo de segmentação da curva de resposta espectral, permita a formação de segmentos com uma maior capacidade de adaptação às variações locais daquela curva e portanto com melhores condições para extração de feições capazes de capturar de uma forma mais eficiente as características mais relevantes das curvas de resposta espectral. Os resultados, apresentados a seguir, confirmam tal expectativa.

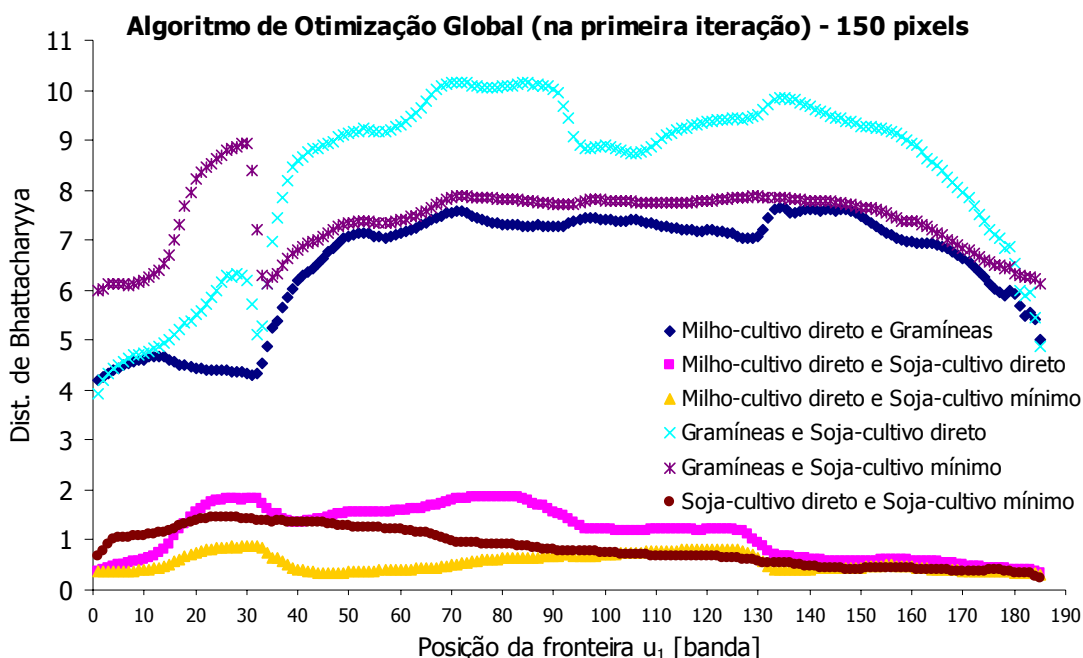


Figura 32. Experimento 3: Separabilidade entre classes avaliada utilizando a distância de Bhattacharyya (definida entre pares de classes), para todas as combinações de classes possíveis (4 classes: 6 combinações). As curvas apresentadas foram obtidas utilizando 150 pixels, na primeira iteração do algoritmo. Visualizam-se todas as possibilidades para o posicionamento da partição u_1 , computando-se o correspondente valor da distância de Bhattacharyya, avaliada com quatro feições, provenientes dos dois segmentos $[1:u_1] \cup [u_1+1:190]$.

A utilização deste algoritmo pressupõe a utilização de uma medida de separabilidade entre classes. Para tanto se utilizou a distância de Bhattacharyya, definida entre pares de classes (15). O algoritmo inicia percorrendo todas as possibilidades possíveis para a partição da curva de resposta espectral, gerando desta forma os dois segmentos iniciais. A cada posição testada são extraídas as correspondentes feições (para os segmentos gerados). Desta forma, na primeira iteração, obtêm-se um total de quatro feições extraídas, reduzindo a dimensionalidade dos dados originais de 190 para apenas 4. Realiza-se uma procura exaustiva para a tomada de decisão com respeito à melhor partição da curva. Os resultados podem ser observados na Figura 32.

Observa-se que as distâncias medidas para os pares de classes onde se apresenta a classe *Grass* são bastante superiores às demais. Isto é de certa forma um comportamento esperado, visto que esta classe apresenta uma curva de resposta espectral mais facilmente diferenciável entre as quatro utilizadas, em outras palavras, é a classe que mais facilmente pode ser separada das demais.

Foram avaliadas todas as possibilidades para a segmentação inicial, calculando-se a função J , (29) e (30) buscando-se saber qual a posição que otimiza a partição. Para tanto, utilizaram-se os dois critérios propostos anteriormente (Capítulo 3): no critério 1 procura-se a posição que minimiza o “*upper bound*” do erro bayesiano (29). No critério 2 avalia-se qual o par de classes que apresenta a menor distância (estatística) entre as suas duas componentes (e portanto com uma maior dificuldade para uma correta classificação) procurando-se nesta curva qual o ponto que maximiza a distância de Bhattacharyya entre classes (30). O ponto que maximiza o valor obtido é indicado, correspondendo à posição onde deve ser feita a primeira partição.

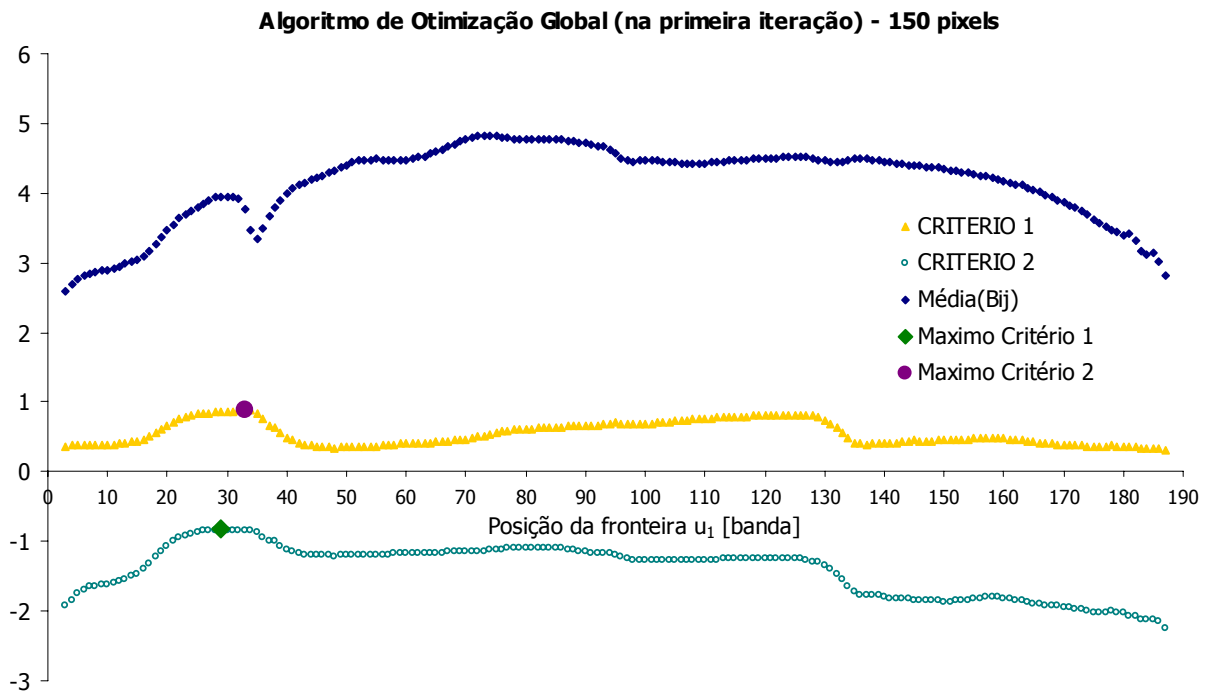


Figura 33. Experimento 3. Distância de Bhattacharyya média para as seis combinações de pares de classes; J utilizando o critério 1 e J utilizando o critério 2; respectivamente. As curvas apresentadas foram obtidas utilizando-se 150 pixels de treinamento, na primeira iteração do algoritmo, conforme resultados da Figura 32. Apresentam-se os pontos de máximo para os critérios 1 e 2.

Os resultados obtidos na primeira iteração, utilizando 150 pixels de treinamento e teste, podem ser visualizados na Figura 33. Nela apresenta-se o valor para os dois critérios a cada possível posição da partição u_1 . No valor apresentado para critério 1, não aplicou-se o fator $\sqrt{p(\omega_i)p(\omega_j)}$ (presente na equação (27)), visto que foi feita a suposição de igual probabilidade de ocorrência $p(\omega)$ “a priori” para cada classe.

Conforme pode ser visualizado na Figura 33, os dois critérios apresentam resultados próximos, mas não idênticos. Isto se deve ao fato de que um dos critérios avalia a melhor alternativa considerando todas as classes presentes, ao passo que o segundo critério baseia-se nas duas classes com maior dificuldade para separação. A segmentação para as demais fica definida em função deste par (*milho-cultivo direto* e *soja-cultivo mínimo*). Genericamente, pode-se dizer que o primeiro critério faz uma ponderação do tipo $\exp(-b_{ij})$, sendo b_{ij} a distância de Bhattacharyya entre o par de classes i e j , resultando portanto em uma ponderação mais alta para pares de classes de mais difícil separação (menor valor de b_{ij}) e uma baixa ponderação às classes com elevada separabilidade. Entretanto, este processo leva também em consideração a presença de todas as demais classes presentes, já que é feito um somatório destes valores (29). De qualquer forma, a posição definida por ambos os critérios fica bastante próxima, devido justamente ao “peso” elevado que é atribuído ao par de classes

de menor b_{ij} . Pelo critério 1, a segmentação inicial da curva de resposta espectral fica: banda [1:29] \cup [30:190] agrupamento das bandas 1 a 29, gerando o primeiro segmento e bandas 30 a 190, gerando o segundo segmento. Nesta condição, o valor do critério 1 mostrou-se igual a -0.83. Utilizando-se o critério 2, geraram-se os segmentos de bandas [1:33] \cup [34:190]. Corresponde ao par de classes *corn-notill* e *soybeans-min*, com distância de Bhattacharyya média de 0.55 (média dos valores para todas as posições de segmentação testadas), sendo o valor máximo 0.89, definindo a posição da segmentação.

Uma vantagem apresentada pelo critério 1 é que ele permite estimar diretamente qual o “acerto mínimo” que se poderá obter na classificação, antes mesmo de realizá-la. Este valor pode ser facilmente estimado da seguinte forma: sendo o número de classes presentes C igual a 4, o limiar superior do erro bayesiano estará limitado em $\varepsilon_B \leq \frac{1}{4} \times 0.83 = 0.2075$ (Equação 27, onde o termo $\frac{1}{4}$ provém do produto da raiz das probabilidades a priori de cada classe). Pode-se afirmar, então, que a acurácia média obtida deverá ser superior a $(1 - 0.2075) = 0.7925$, ou seja, 79.25%. Constatou-se que o valor obtido posteriormente (Figura 39), quando efetuada de fato a classificação, foi de 89.83%.

Na Figura 33 também se apresenta a posição que seria obtida se o critério adotado fosse simplesmente a média aritmética das distâncias de Bhattacharyya obtidas para todos os pares de classes. Claramente, a posição resultante para a partição da curva de resposta espectral seria bem diferente daquela nos casos anteriores. Neste caso, o valor médio é bastante influenciado pela presença da classe *Grass*, que tem características que a tornam facilmente separável das demais. É por esta razão que num primeiro momento não se sugere um critério que seja simplesmente a média aritmética das combinações de distância de Bhattacharyya obtidas. Neste caso, o ponto de máximo da curva média dos pares de distâncias de Bhattacharyya seria obtido na posição [1:74] \cup [75:190], com valor 4.83.

Matematicamente, a diferença de resultados entre os critérios 1 e do valor médio para a distância de Bhattacharyya pode ser explicada da seguinte forma: dado uma série de vetores x com componentes positivas $x_k = [d_1, d_2, d_3, \dots, d_n]$ (neste caso d_i representa a separabilidade entre pares de classes na posição k onde $i=1\dots n$, e k varia de 1 ao número total de possíveis posições de partição), o vetor x_k que contém a maior média aritmética de suas componentes, ou seja,

$$\bar{x}_k = \frac{1}{n} \sum_{i=1}^n d_i \quad (31)$$

não é, necessariamente, aquele que minimiza a média dos vetores de componentes, dada por

$$\bar{z}_k = \frac{1}{n} \sum_{i=1}^n e^{-d_i} \quad (32)$$

Apesar de aparentemente simples, esta constatação não é tão evidente. O que acontece é que uma a posição de mínimo em (31) está associada a um máximo em (32). Porém, a posição de mínimo, que é realmente aquela que interessa, em (32) não está necessariamente associada à posição de máximo da equação (31). Este comportamento deve-se ao fato da ordem de grandeza dos valores das componentes d_i e da função exponencial não ser do tipo linear.

Definida a primeira segmentação o processo continua de maneira análoga, ou seja, mantendo a posição da primeira partição inalterada e procurando a melhor posição para a partição seguinte, através do cálculo ponto que maximiza a função de otimização J utilizando-se a medida da separabilidade entre pares de classes em dimensionalidade reduzida. Esta dimensionalidade aumenta progressivamente na medida em que são introduzidos segmentos adicionais. Cada novo segmento contribui individualmente com duas feições extraídas adicionais. Como exemplo, utilizando-se 150 pixels de treinamento e teste, o critério 1 gera inicialmente dois segmentos: um contendo as 29 bandas espectrais iniciais e o outro as restantes 161 bandas espectrais. Na segunda iteração, o segmento com 161 bandas é particionado, resultado em dois segmentos: um contendo 93 bandas e outro com 68 bandas [29+93+68]. Na iteração seguinte o segmento com 93 bandas é particionado em dois: um com 87 bandas e outro com apenas 6 bandas[29+68+87+6]. Este processo é repetido até que um determinado número de feições desejado seja atingido.

A título de brevidade, foram apresentados com detalhes apenas os resultados que correspondem à procura da primeira partição, utilizando-se 150 pixels de treinamento e teste (Figuras 32 e 33). Para os demais experimentos, envolvendo amostras com diferentes tamanhos, são apresentados apenas os resultados finais obtidos.

A acurácia obtida nos experimentos utilizando amostras de treinamento (e de teste) com tamanhos de 150, 200, 250, 300 e 350 pixels, envolvendo um número crescente de feições extraídas (4 a 94), estão apresentadas nas Figura 34 a Figura 39. Cada figura é composta por dois gráficos, um correspondendo aos resultados obtidos utilizando o critério de otimização 1 (conforme equação 25) e o outro se utilizando o critério 2 (conforme equação 26).

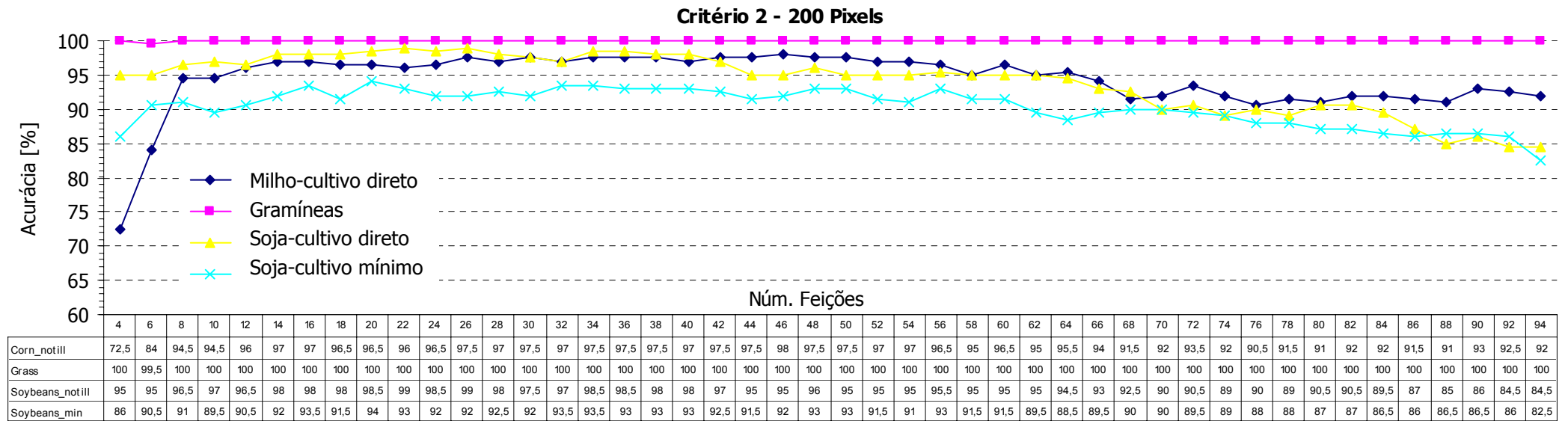
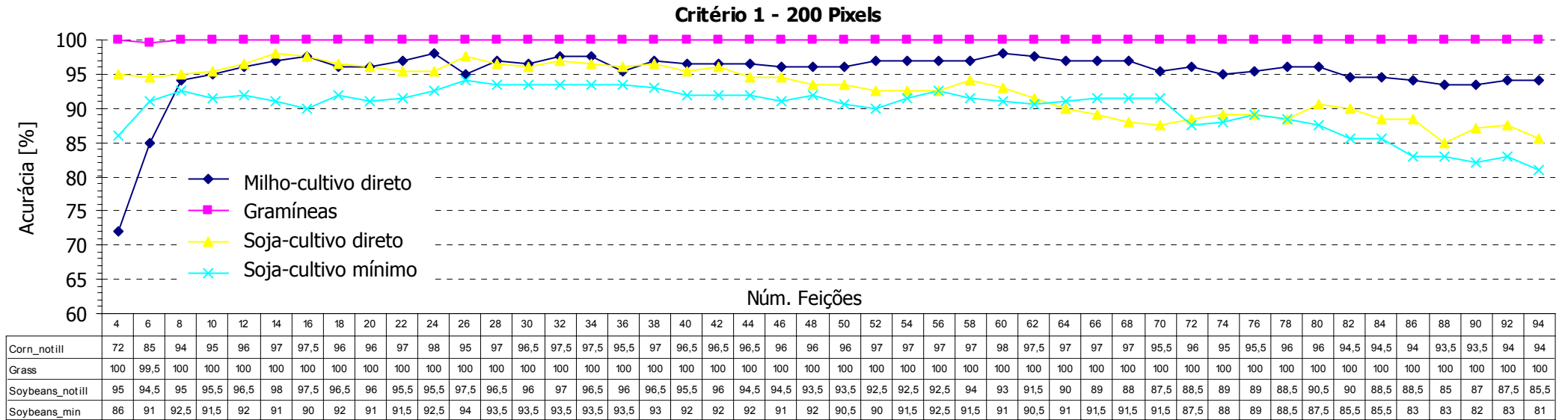


Figura 35. Experimento 3: Alg. de Otimização Global - Acurácias de classificação obtidas utilizando-se 200 amostras de treinamento e teste, utilizando-se os critérios 1 e 2.

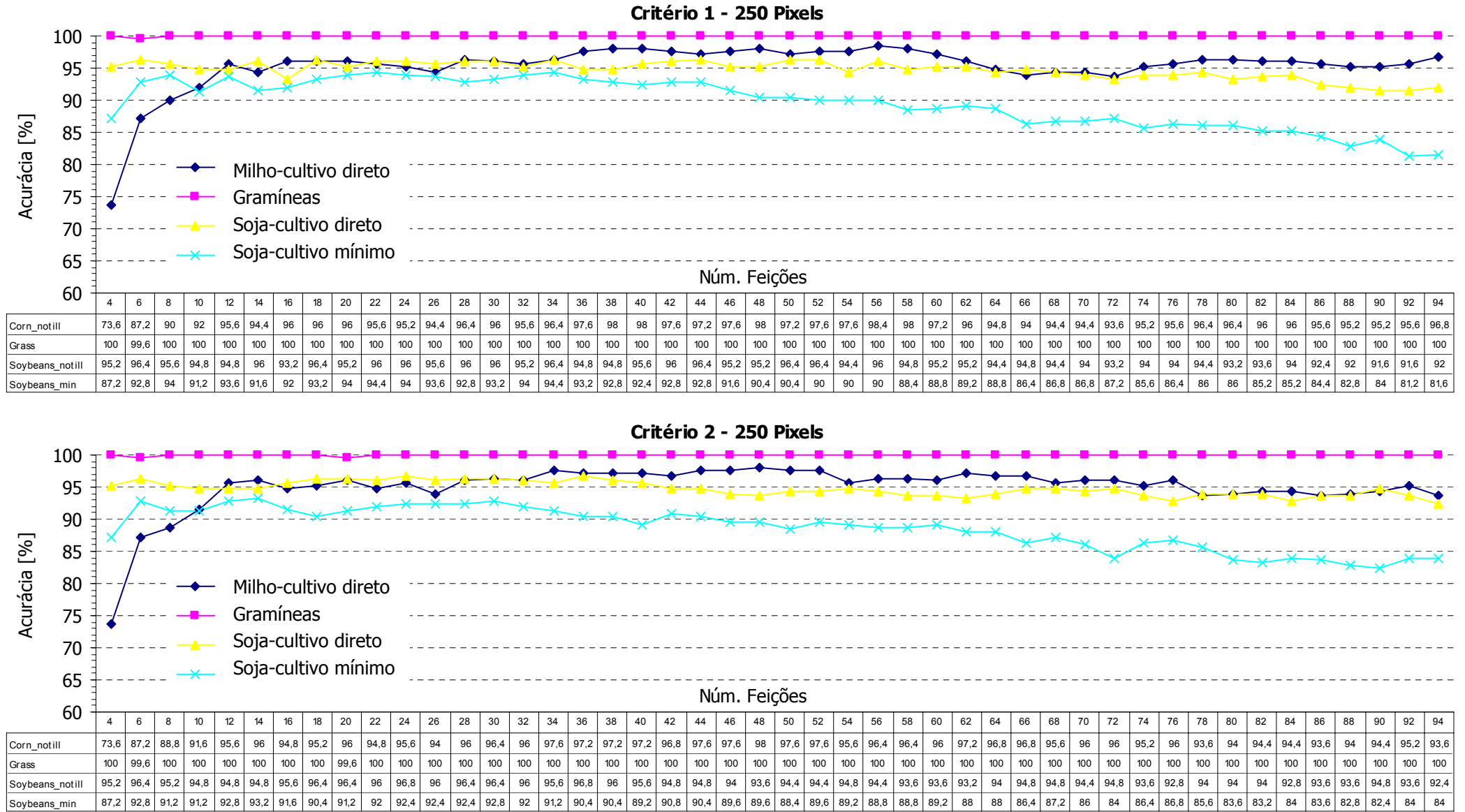


Figura 36. Experimento 3: Alg. de Otimização Global - Acurácias de classificação obtidas utilizando-se 250 amostras de treinamento e teste, utilizando-se os critérios 1 e 2.

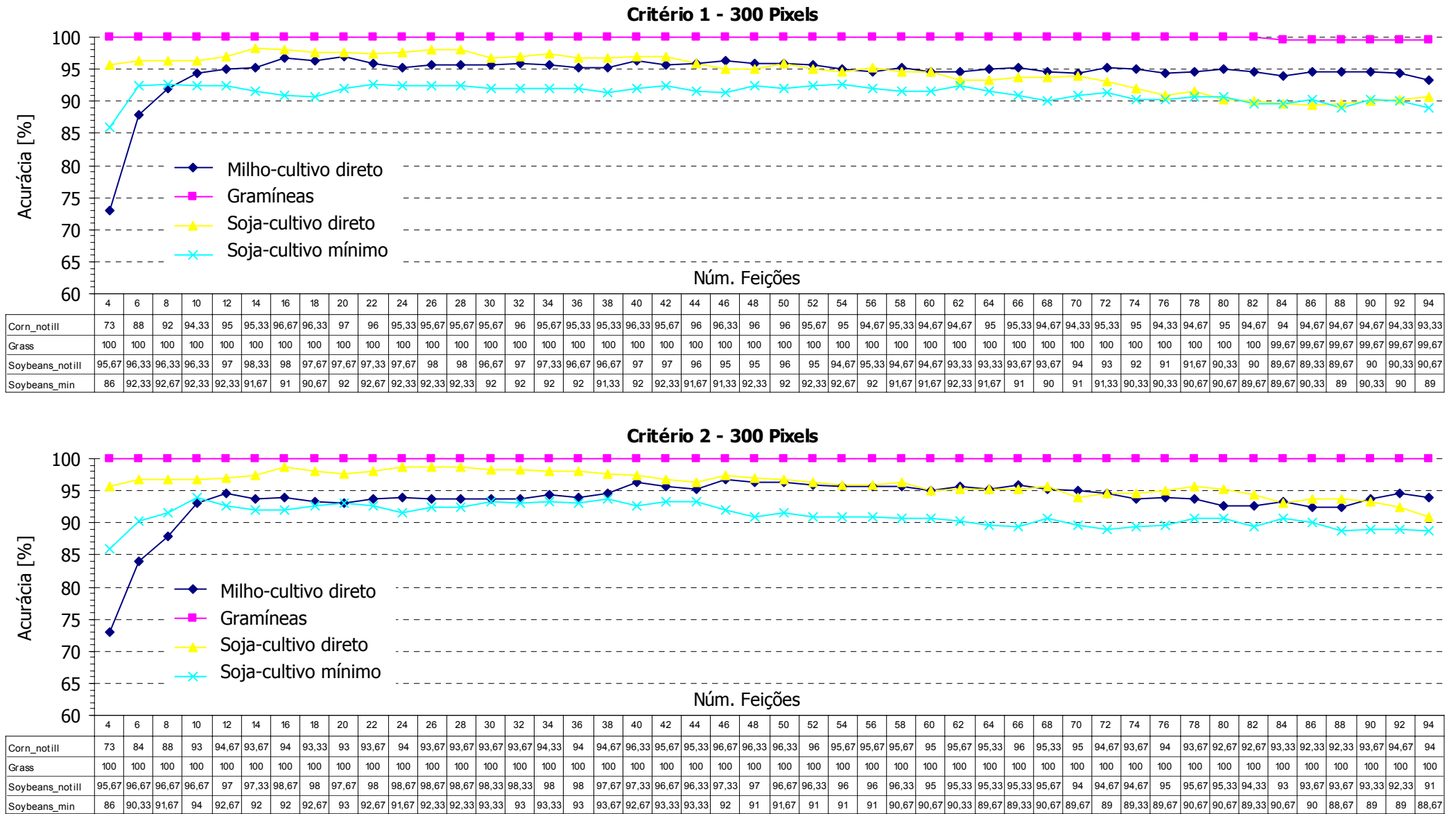


Figura 37. Experimento 3: Alg. de Otimização Global - Acurácias de classificação obtidas utilizando-se 300 amostras de treinamento e teste, utilizando-se os critérios 1 e 2.

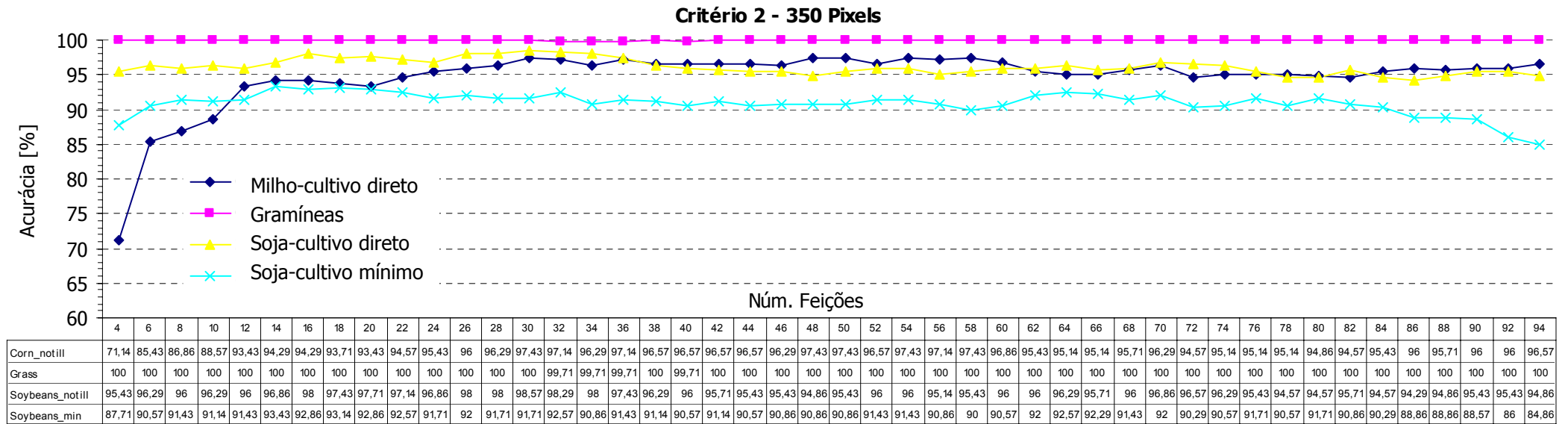
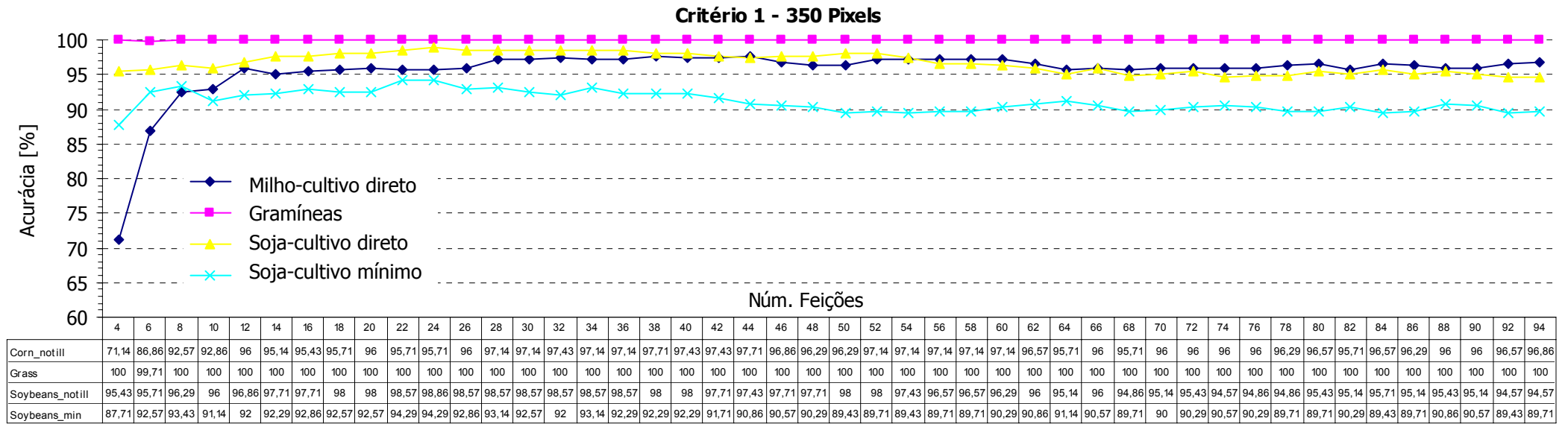
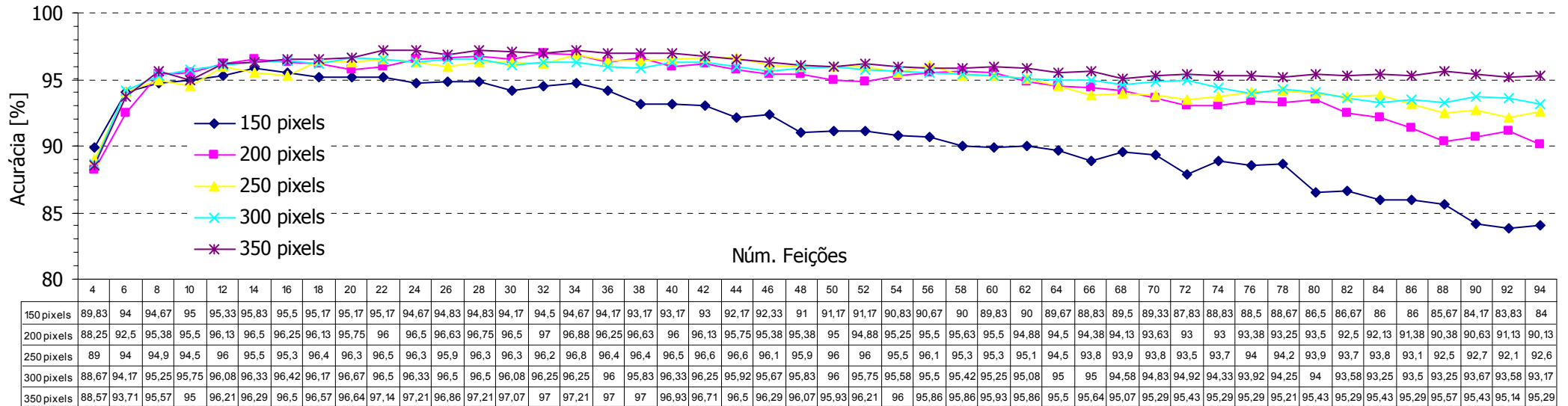


Figura 38. Experimento 3: Alg. de Otimização Global - Acurácias de classificação obtidas utilizando-se 350 amostras de treinamento e teste, utilizando-se os critérios 1 e 2.

Critério 1 - Acuracias Médias



Critério 2 - Acuracias Médias

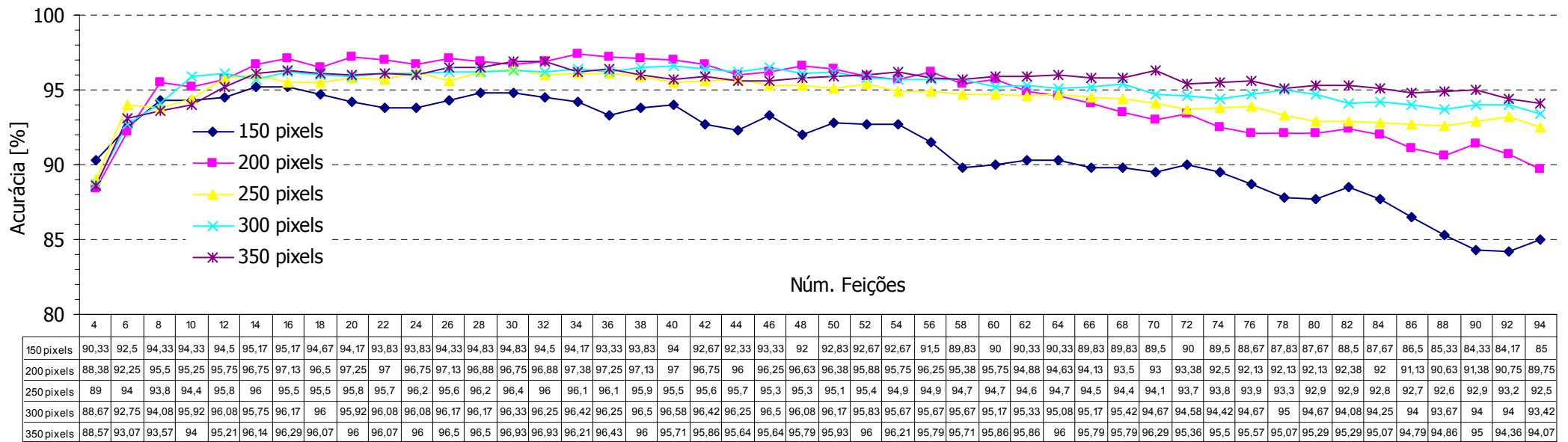


Figura 39. Experimento 3: Alg. de Otimização Global - Acurácia média de classificação (4 classes) obtida utilizando-se 150, 200, 250, 300 e 350 amostras de treinamento e teste, utilizando-se os critérios 1 e 2. Apresentam-se os resultados obtidos na extração de 4 a 94 feições.

Utilizando-se 150 pixels de treinamento e teste, o valor mais alto para a acurácia no processo de classificação obtido, utilizando-se o critério 1, foi de 95.83% (14 feições) e para o critério 2 obteve-se 95.17% (para 14 feições e também para 16 feições). Considerando-se os resultados obtidos nos experimentos, utilizando-se 4 feições verificam-se acurácias médias superiores a 88,25% (Figura 39). O valor máximo obtido ocorreu para um número de feições que variou conforme o critério de otimização e tamanho das amostras utilizado. Os resultados máximos considerando a média das quatro classes ficaram compreendidos num intervalo de 95.17 a 97.38%. A Tabela 6 resume os resultados. Percebe-se que ambos os critérios apresentam resultados que podem ser considerados satisfatórios.

Tabela 6. Resumo do desempenho de acurácia máxima (valor médio para 4 classes) obtida para o “algoritmo otimização global” frente ao número de amostras de treinamento e critério de segmentação utilizado. Apresenta-se também o correspondente número de feições para o qual obtiveram-se tais resultados.

Npixel	OT.GLOB: Critério 1	Num. Feiçoes	OT.GLOB: Critério 2	Num. Feiçoes
150	95,83	14	95,17	14
200	97,00	32	97,38	34
250	96,80	34	96,40	30
300	96,67	20	96,58	40
350	97,21	24	96,93	30
<i>Média</i>	<i>96,70</i>	<i>24,8</i>	<i>96,49</i>	<i>29,6</i>

Os resultados obtidos e apresentados na Tabela 6 sugerem uma pequena vantagem para o critério 1 em relação ao critério 2 visto que, em média e para os tamanhos de amostras utilizadas, os resultados gerados pelo critério 1 apresentam uma acurácia levemente superior àquela obtida via critério 2. Estes valores foram atingidos, no caso do critério 1, para um número de feições em média menor do que no caso do critério 2. Um maior número de experimentos seria entretanto necessário para confirmar esta suposição.

Percebe-se que a classe *grass* é facilmente separável das demais utilizando-se apenas 150 amostras de treinamento e apenas 4 feições. Os resultados obtidos empregando ambos os critérios indicam uma acurácia de 100% (Figura 34). Verifica-se que o comportamento geral das curvas obtidas para a acurácia no processo de classificação seguiu um comportamento esperado, ou seja, à medida que são adicionadas novas feições, extraídas pelo processo de segmentação, a acurácia cresce, inicialmente de maneira mais rápida, apresentando a seguir uma tendência a estabilizar-se. A partir de um certo valor para a dimensionalidade dos dados (feições extraídas), a acurácia tende a se deteriorar caracterizando o fenômeno de Hughes. Também conforme o esperado, este decréscimo no

valor da acurácia ocorre de maneira mais acentuada para amostras de treinamento de pequeno tamanho. O número máximo de feições obtidas pelo algoritmo, para todos os tamanhos de amostras de treinamento e teste utilizadas, ficou compreendido no intervalo 104 a 112 feições, para ambos os critérios. Nestas condições foram obtidos uma totalidade de segmentos, cujos comprimentos foram 3, 4 ou 5 bandas originais, ou seja, comprimentos que não podem mais ser divididos de forma que viessem a gerar mais dois novos segmentos e que nestes novos segmentos pudesse haver redução da dimensionalidade. O número total de feições é coerente com o esperado, ou seja, inferior a 126 (que seriam obtidas se todos os segmentos tivessem comprimento 3 bandas, conforme Tabela 2). Entretanto, para simplificar a apresentação dos resultados, nas Figura 34 a Figura 39 foram apresentados os resultados de classificação até um número máximo de 94 feições.

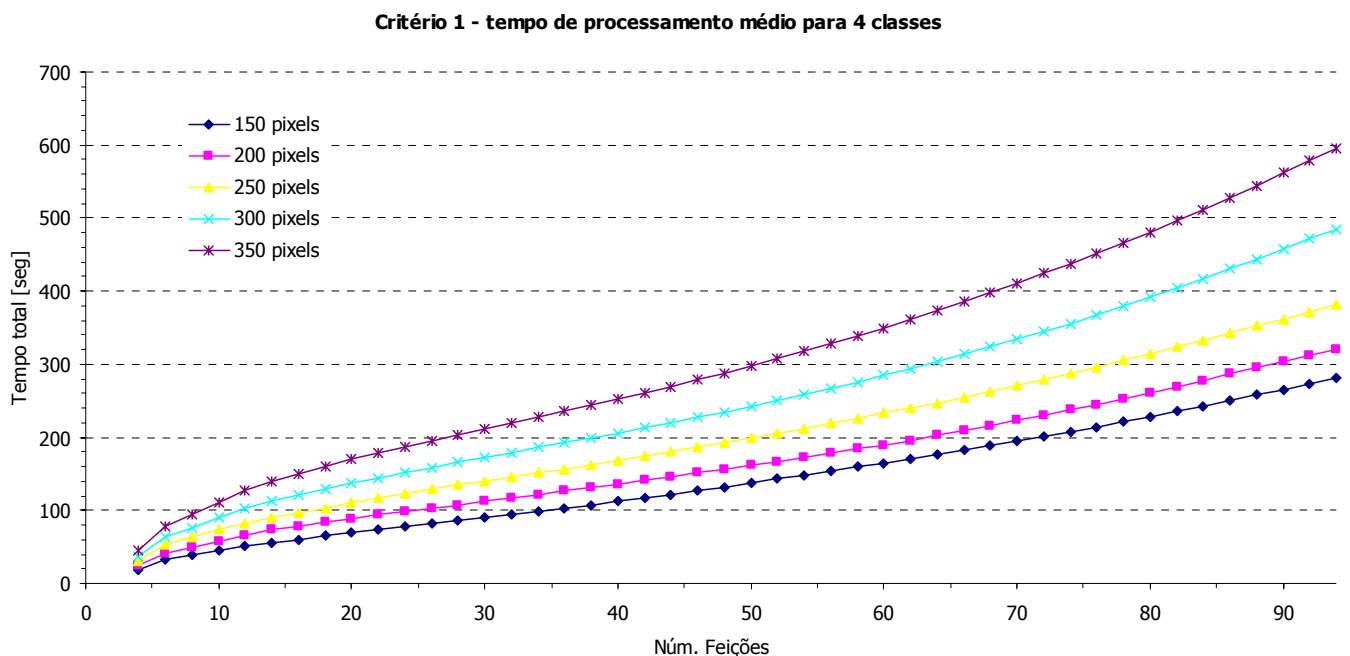


Figura 40. Experimento 3: tempo de processamento necessário para o algoritmo desenvolvido, para diferentes tamanhos de amostras de treinamento e teste para cada classe, frente à dimensionalidade utilizada (Critério 1).

O tempo de processamento segundo a forma como foi implementado o “Algoritmo Otimização Global” pode ser visualizado na Figura 40. Os resultados fazem referência à utilização do critério 1. Resultados bastante semelhantes foram obtidos utilizando-se o critério de otimização 2. Percebe-se que, para um determinado número de pixels utilizados, o acréscimo no tempo de processamento, para um número de feições crescente, é do tipo exponencial.

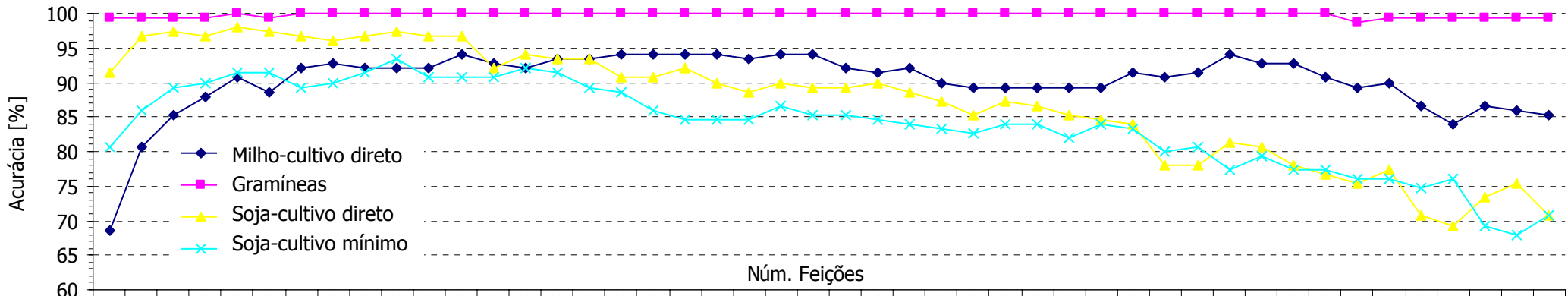
4.9 - Experimento 04 - Algoritmo com Estrutura em Árvore (Tree)

Neste experimento implementou-se a segunda proposta de partição da resposta espectral original de cada pixel do tipo segmentos de comprimento “n” variável. Esta abordagem está referenciada como “algoritmo Tree”. Os resultados são apresentados e discutidos. Conforme apresentado anteriormente no Capítulo 3, pode-se afirmar que o funcionamento deste algoritmo é uma simplificação do “algoritmo otimização global”. Enquanto naquele buscava-se a avaliação de todas as possibilidades de partição e com isto geração dos segmentos, neste testam-se apenas todos os centros dos segmentos que vão sendo gerados no “nível hierárquico” anterior. A idéia é a mesma, porém evidentemente a execução do algoritmo é mais rápida. O mesmo inconveniente anterior permanece, ou seja, uma vez gerada uma posição de partição, esta se mantém até o final do processo. A cada iteração, testadas as possíveis partições, extraem-se feições e avalia-se qual a posição que maximiza a separabilidade entre classes, novamente utilizando os critérios de otimização 1 e 2.

Uma particularidade deste algoritmo é que a partição inicial da curva de resposta espectral resulta idêntica àquela obtida no processo envolvendo o algoritmo SCC. Conseqüentemente, as quatro feições iniciais extraídas a partir dos dois algoritmos são idênticas. Somente nas iterações seguintes é que efetivamente as características que diferenciam os dois algoritmos evidenciam-se. No presente caso, o critério de otimização é utilizado na decisão de qual dos dois segmentos gerados na iteração anterior será particionado.

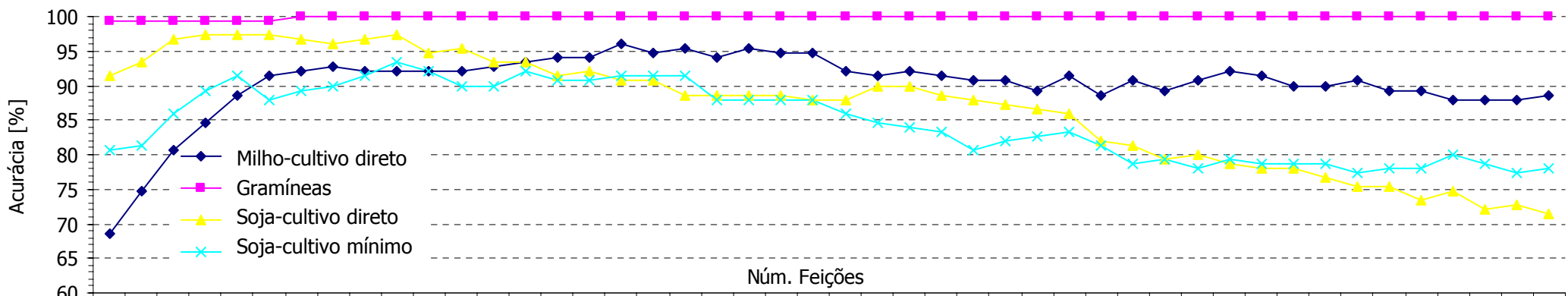
As Figuras 41 a 45 apresentam a acurácia da classificação para as quatro classes utilizadas neste trabalho, utilizando-se amostras de treinamento e de teste com 150, 200, 250, 300 e 350, para um número crescente de feições extraídas (4 a 94). Os valores médios correspondentes são apresentados na Figura 46. Cada figura exhibe dois gráficos, um correspondendo aos resultados obtidos utilizando-se o critério de otimização 1 e o outro utilizando-se o critério 2, conforme definidos no Capítulo 3.

Critério 1 - 150 Pixels



	4	6	8	10	12	14	16	18	20	22	24	26	28	30	32	34	36	38	40	42	44	46	48	50	52	54	56	58	60	62	64	66	68	70	72	74	76	78	80	82	84	86	88	90	92	94	
Corn_notill	68,67	80,67	85,33	88	90,67	88,67	92	92,67	92	92	92	94	92,67	92	93,33	93,33	94	94	94	94	93,33	94	94	92	91,33	92	90	89,33	89,33	89,33	89,33	91,33	90,67	91,33	90,67	91,33	94	92,67	92,67	90,67	89,33	90	86,67	84	86,67	86	85,33
Grass	99,33	99,33	99,33	99,33	100	99,33	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	98,67	99,33	99,33	99,33	99,33	99,33	
Soybeans_notill	91,33	96,67	97,33	96,67	98	97,33	96,67	96	96,67	97,33	96,67	96,67	92	94	93,33	93,33	90,67	90,67	92	90	88,67	90	89,33	89,33	90	88,67	87,33	85,33	87,33	86,67	85,33	84,67	84	78	78	81,33	80,67	78	76,67	75,33	77,33	70,67	69,33	73,33	75,33	70,67	
Soybeans_min	80,67	86	89,33	90	91,33	91,33	89,33	90	91,33	93,33	90,67	90,67	90,67	92	91,33	89,33	88,67	86	84,67	84,67	84,67	86,67	85,33	85,33	84,67	84	83,33	82,67	84	84	82	84	83,33	80	80,67	77,33	79,33	77,33	77,33	76	76	74,67	76	69,33	68	70,67	

Critério 2 - 150 Pixels



	4	6	8	10	12	14	16	18	20	22	24	26	28	30	32	34	36	38	40	42	44	46	48	50	52	54	56	58	60	62	64	66	68	70	72	74	76	78	80	82	84	86	88	90	92	94		
Corn_notill	68,67	74,67	80,67	84,67	88,67	91,33	92	92,67	92	92	92	92	92,67	93,33	94	94	96	94,67	95,33	94	95,33	94,67	94,67	92	91,33	92	91,33	90,67	90,67	89,33	91,33	88,67	90,67	89,33	89,33	92	91,33	90	90	90,67	89,33	89,33	88	88	88	88	88,67	
Grass	99,33	99,33	99,33	99,33	99,33	99,33	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
Soybeans_notill	91,33	93,33	96,67	97,33	97,33	97,33	96,67	96	96,67	97,33	94,67	95,33	93,33	93,33	91,33	92	90,67	90,67	88,67	88,67	88,67	88,67	88	88	90	90	88,67	88	87,33	86,67	86	82	81,33	79,33	80	78,67	78	78	76,67	75,33	75,33	73,33	74,67	72	72,67	71,33		
Soybeans_min	80,67	81,33	86	89,33	91,33	88	89,33	90	91,33	93,33	92	90	90	92	90,67	90,67	91,33	91,33	91,33	88	88	88	88	86	84,67	84	83,33	80,67	82	82,67	83,33	81,33	78,67	79,33	78	79,33	78,67	78,67	78,67	77,33	78	78	80	78,67	77,33	78		

Figura 41. Experimento 4: Alg. Tree - Resultados da classificação, utilizando-se 150 pixels de treinamento e teste, frente ao número de feições extraídas, segundo os critérios 1 e 2.

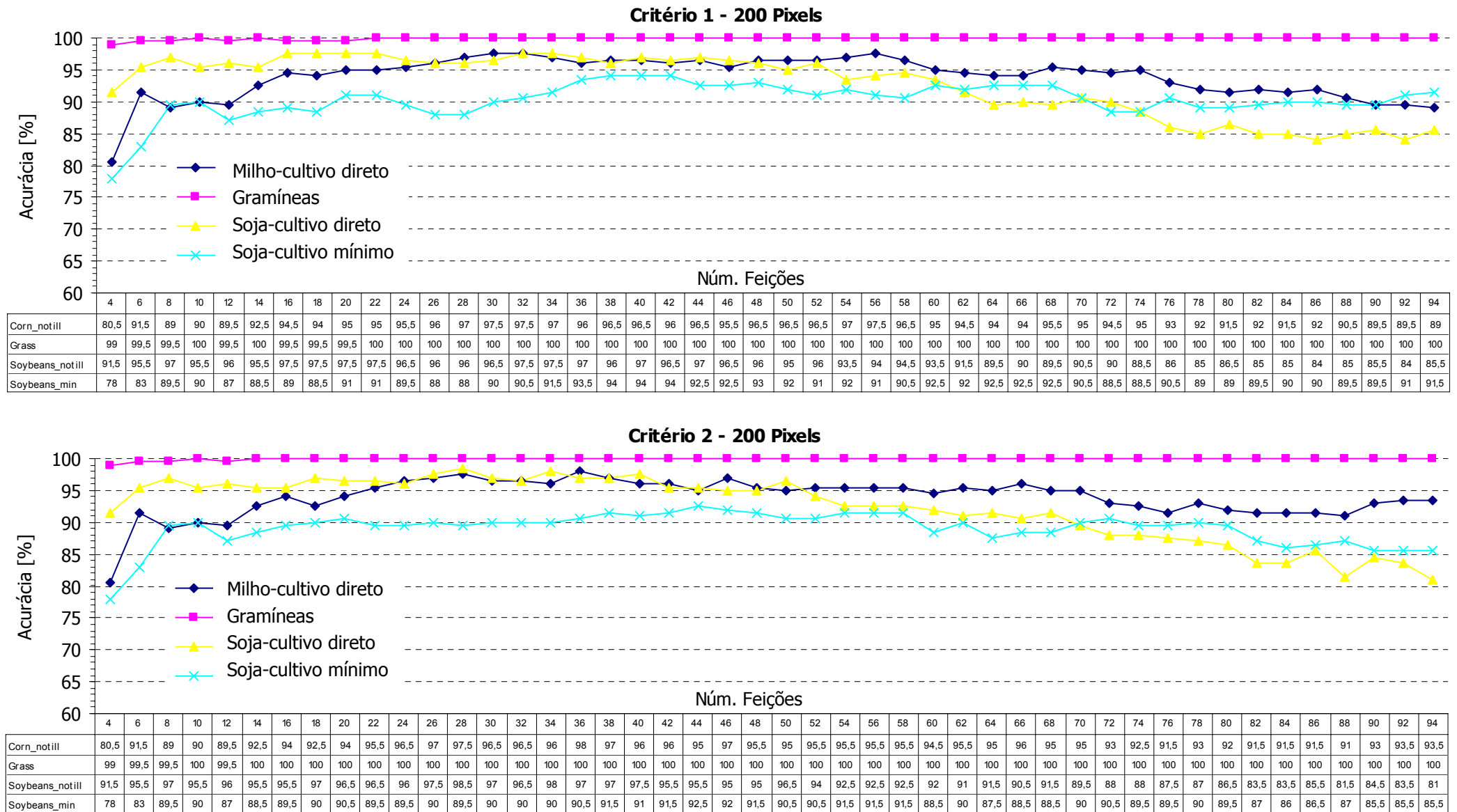


Figura 42. Experimento 4: Alg. Tree - Resultados da classificação, utilizando-se 200 pixels de treinamento e teste, frente ao número de feições extraídas, segundo os critérios 1 e 2.

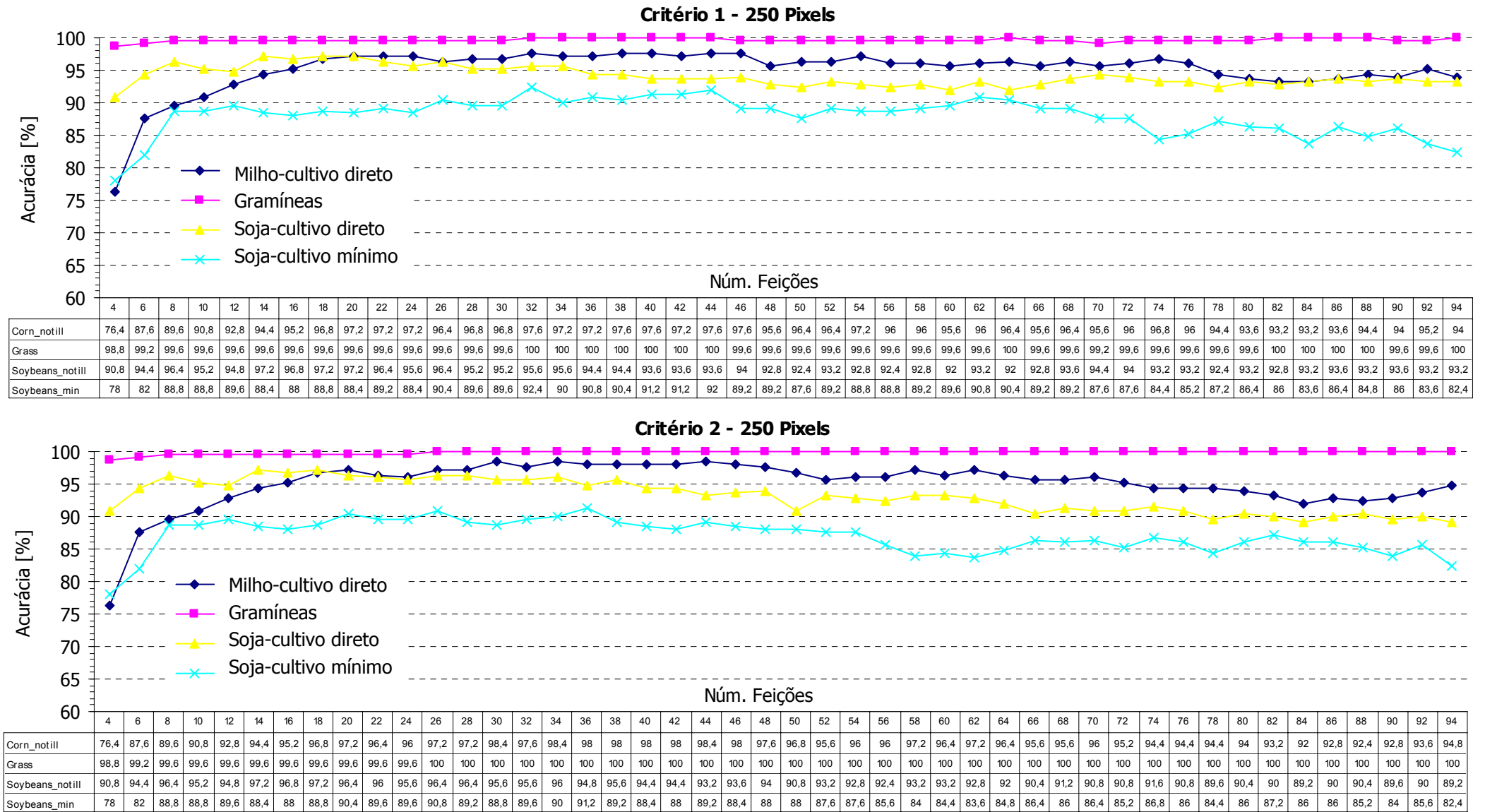


Figura 43. Experimento 4: Alg. Tree - Resultados da classificação, utilizando-se 250 pixels de treinamento e teste, frente ao número de feições extraídas, segundo os critérios 1 e 2.

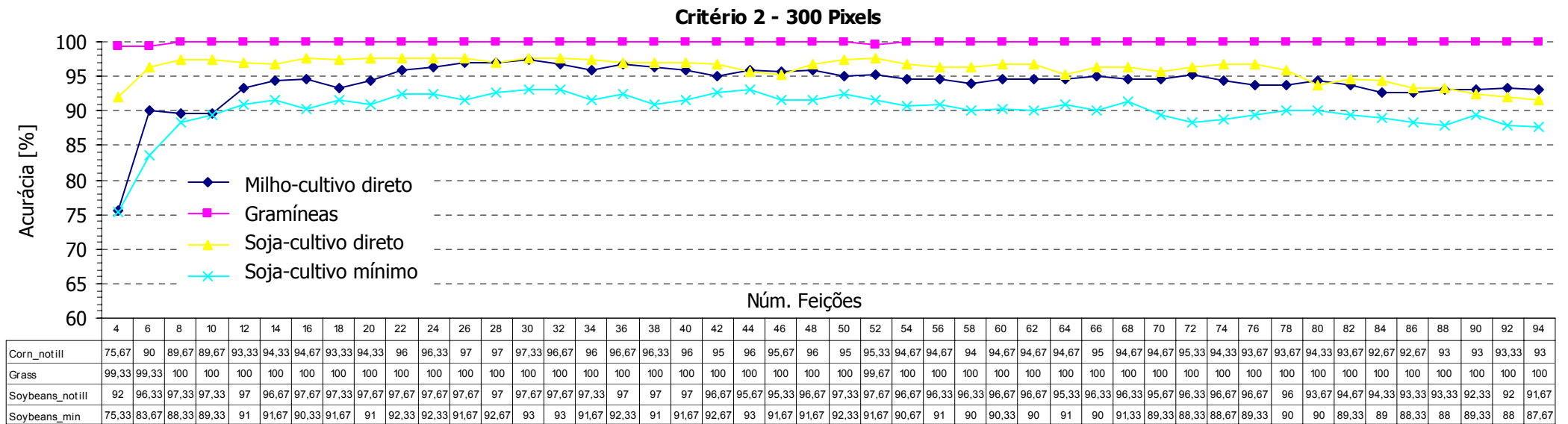
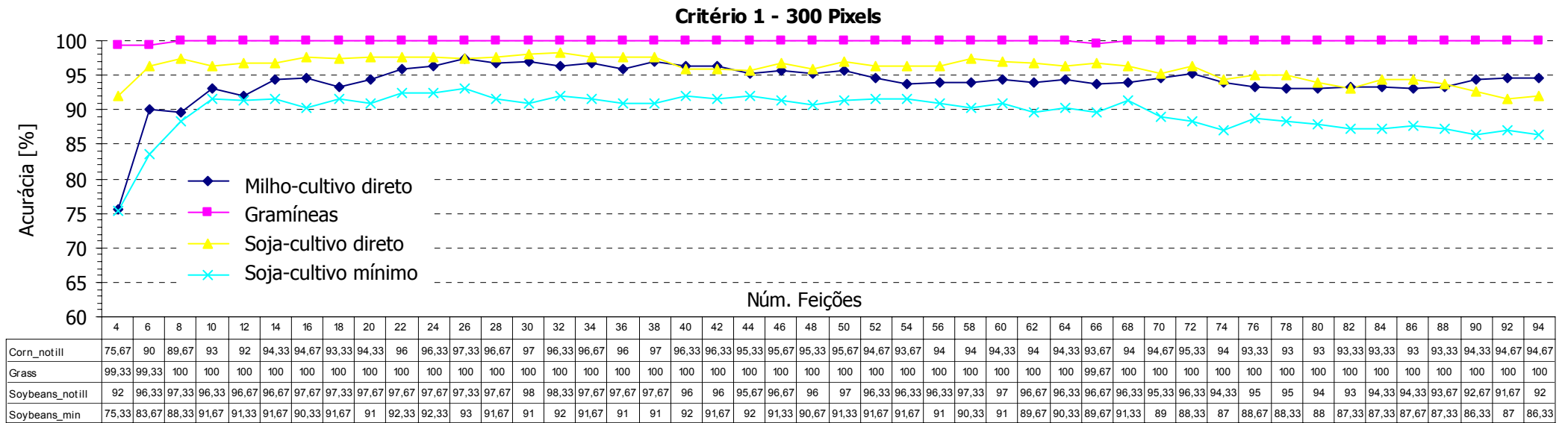
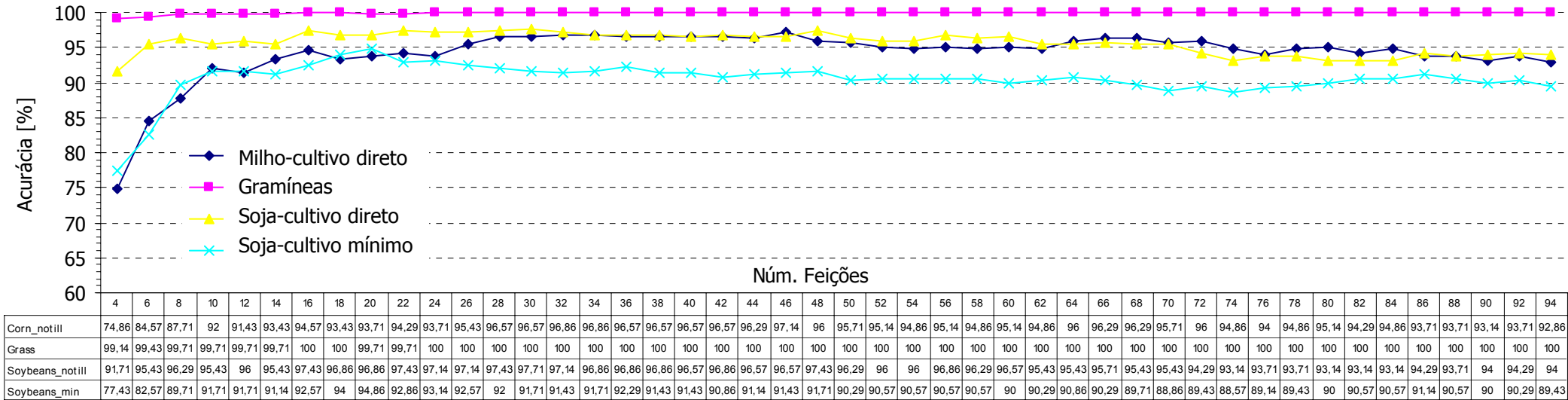


Figura 44. Experimento 4: Alg. Tree - Resultados da classificação, utilizando-se 300 pixels de treinamento e teste, frente ao número de feições utilizadas, segundo os critérios 1 e 2.

Critério 1 - 350 Pixels



Critério 2 - 350 Pixels

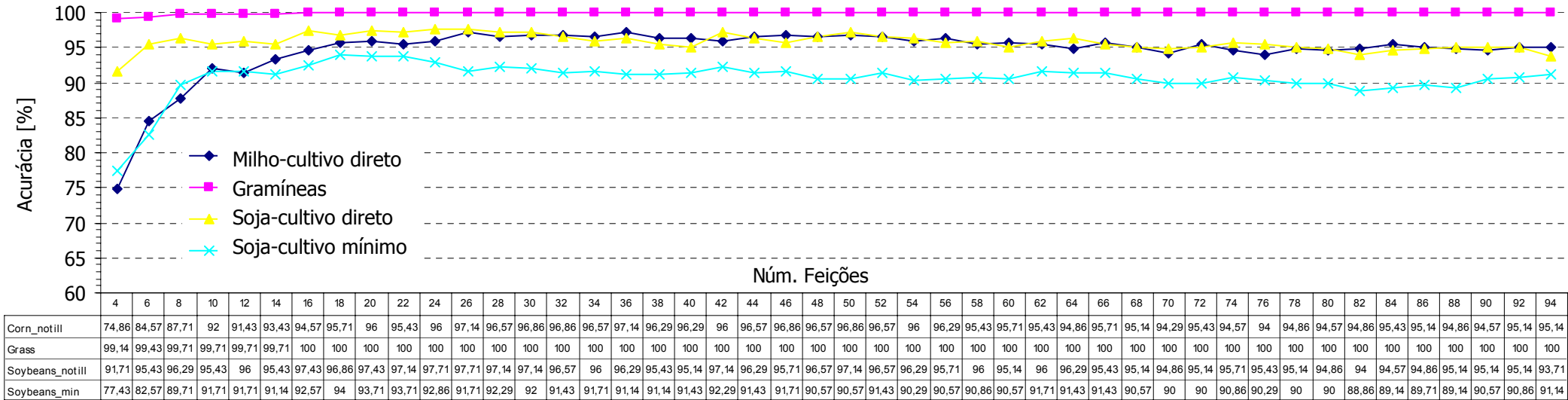


Figura 45. Experimento 4: Alg. Tree - Resultados da classificação, utilizando-se 350 pixels de treinamento e teste, frente ao número de feições extraídas, segundo os critérios 1 e 2.

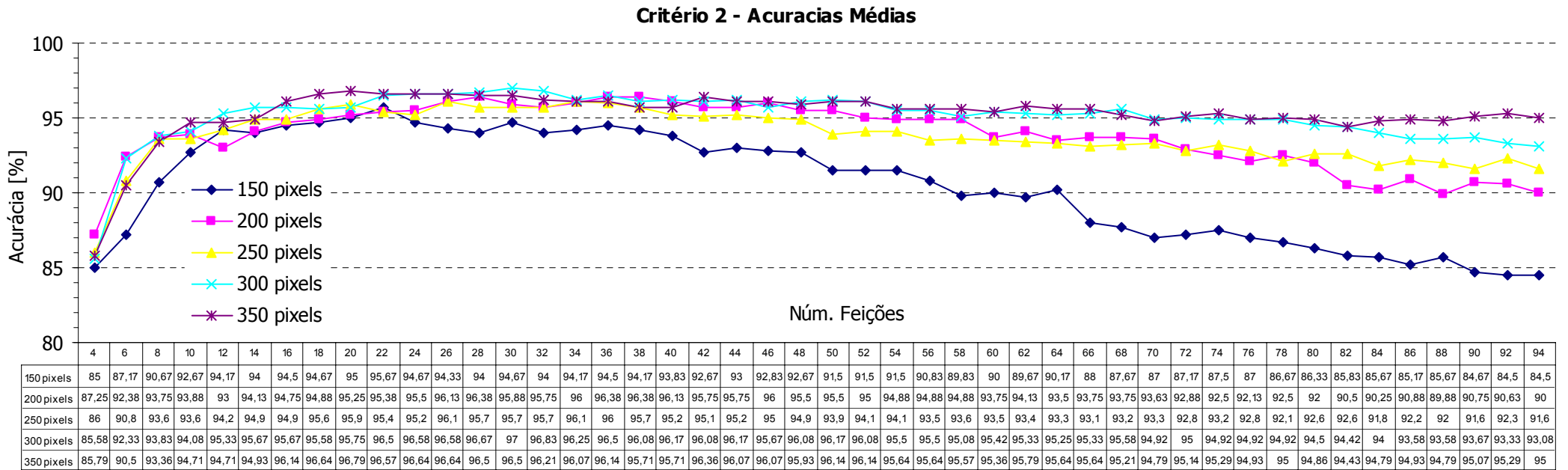
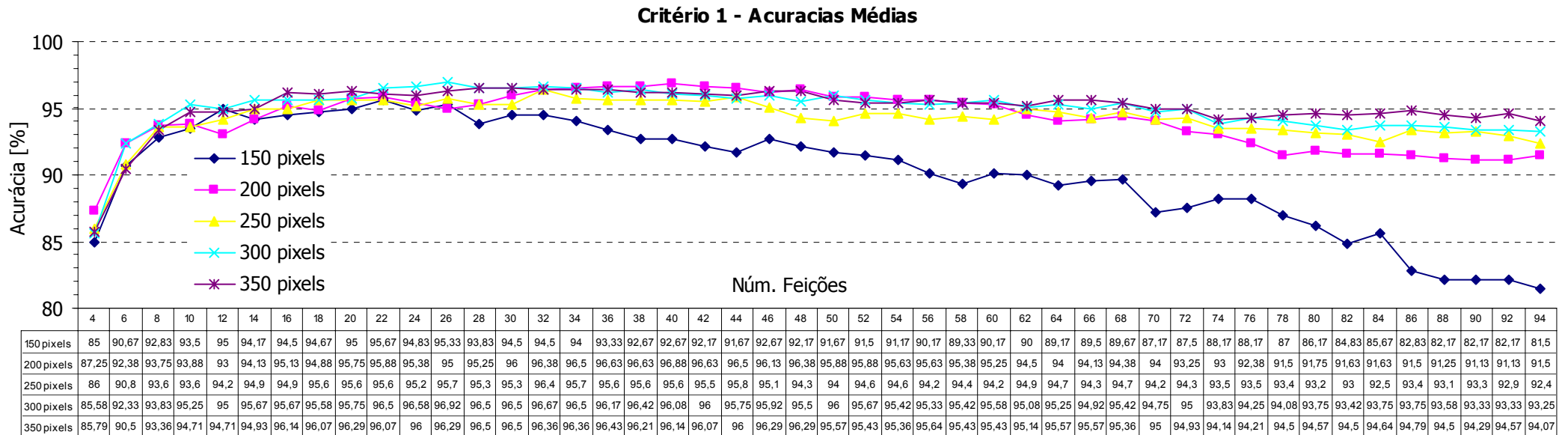


Figura 46. Experimento 4: Alg. Tree - Acurácia média da classificação, variando-se o tamanho das amostras de treinamento e teste, frente ao número de feições extraídas, segundo os critérios 1 e 2.

Os experimentos realizados, utilizando-se este algoritmo, mostraram resultados semelhantes, em sua forma geral, àqueles obtidos por outras abordagens: a acurácia alcançada no processo de classificação cresce inicialmente, na medida em que novas informações, isto é, novas feições são acrescentadas. Eventualmente um máximo é atingido para o valor da acurácia. A partir deste ponto, a introdução de novas feições passa a provocar um decréscimo no valor da acurácia, caracterizando o fenômeno de Hughes. A forma geral da curva que exprime o valor da acurácia obtida em função do número de feições utilizadas no processo de classificação segue aqui também os padrões gerais conhecidos: quanto menor o número de amostras de treinamento, maior a intensidade com que se apresenta o fenômeno de Hughes. Em particular, para um tamanho das amostras de 150 pixels, o resultado de acurácia média (Figura 46) utilizando-se 82 feições (84.83% para o critério 1) e 92 feições (84.67% para o critério 2) é inferior à acurácia obtida no início do processo: 85%, quando utilizaram-se apenas 4 feições.

O número máximo de feições geradas pelo algoritmo para todos os tamanhos de amostras de treinamento e teste utilizadas, para ambos os critérios, atingiu o valor de 124 feições, obtidas ao final da 61ª iteração. Nesta condição foram obtidos 62 segmentos, cujos comprimentos eram 3, 4 ou 5 bandas, ou seja, comprimentos que não permitem um particionamento adicional. O número total de feições é coerente com o esperado, ou seja, inferior a 126 (que seriam obtidas se todos os segmentos tivessem comprimento 3 bandas, conforme Tabela 2).

Considerando-se os resultados obtidos neste experimento, utilizando-se 4 feições foram obtidas acurácias médias superiores a 85%. O valor máximo obtido ocorreu para um número de feições que variou conforme o critério de otimização e tamanho das amostras de treinamento utilizadas. Os valores máximos considerando a média das quatro classes ficaram compreendidos num intervalo de 95.67% a 97%, ou seja, resultados normalmente considerados de bom nível. Maiores detalhes quanto ao valor médio máximo obtido, bem como o correspondente número de feições para cada caso podem ser visualizados na Tabela 7.

Tabela 7. Resumo do desempenho de acurácia máxima (valor médio para 4 classes) para o “algoritmo Tree” frente ao número de amostras de treinamento e critério de segmentação utilizado. Apresenta-se também o correspondente número de feições onde se obteve tais resultados.

Npixel	TREE: Critério 1	Num. Feiçoes	TREE: Critério 2	Num. Feiçoes
150	95,67	22	95,67	22
200	96,88	40	96,38	28
250	96,40	32	96,10	26
300	96,92	26	97,00	30
350	96,50	28	96,79	20
<i>Média</i>	<i>96,47</i>	<i>29,6</i>	<i>96,39</i>	<i>25,2</i>

Estes resultados sugerem que desempenho de ambos os critérios é bastante semelhante, visto que os valores máximos obtidos para as acurácias de classificação ficaram bastante próximos (Tabela 7).

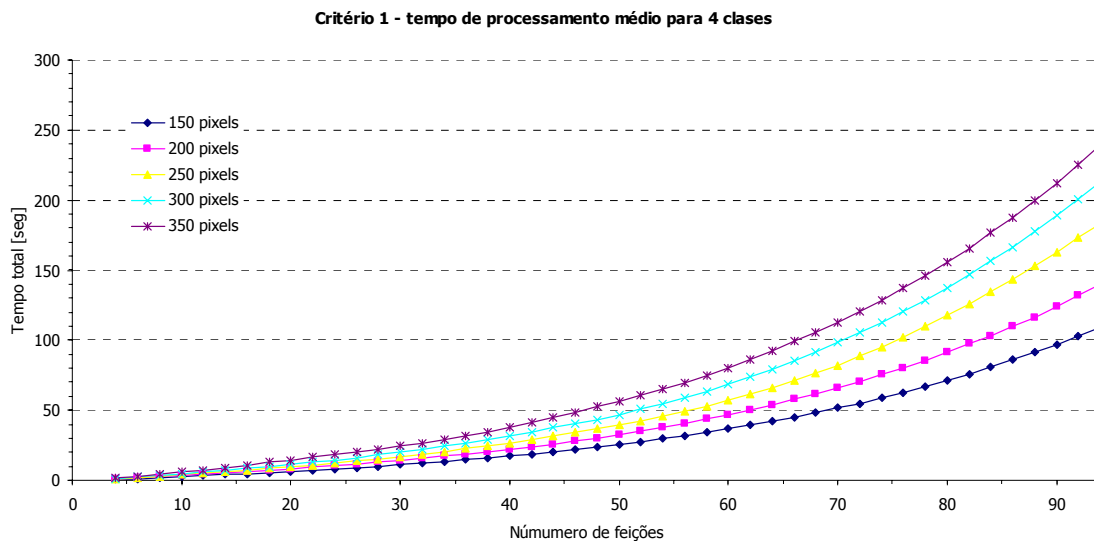


Figura 47. Experimento 4: tempo de processamento necessário para o algoritmo desenvolvido, para diferentes tamanhos de amostras de treinamento e teste para cada classe, frente à dimensionalidade utilizada (Critério 1).

O tempo de processamento segundo a forma como foi implementado o “algoritmo Tree” pode ser visualizado na Figura 47. Os resultados fazem referência à utilização do critério de otimização 1. Resultados muito semelhantes foram obtidos utilizando-se o critério de otimização 2. Percebe-se que, para os diferentes números de amostras (pixels) de treinamento e teste utilizados, o acréscimo no tempo de processamento, para um número de feições crescente, é do tipo exponencial.

4.10 - Comparação dos Resultados dos Experimentos

Feitas as considerações anteriores sobre o desempenho dos três algoritmos propostos, torna-se oportuno compará-los. Esta comparação pode ser feita a partir de diferentes pontos de vista. Em termos de uma análise mais conceitual pode-se dizer que, teoricamente, o algoritmo mais robusto é o “otimização global”, visto que neste caso é feita uma análise mais exaustiva no processo de partição dos segmentos. O algoritmo “Tree” pode ser visto neste contexto como um caso particular do algoritmo Otimização Global. Neste caso, um número menor de possíveis partições é analisado, resultando em um processo computacionalmente mais leve do que o de Otimização Global. Deve-se notar que neste algoritmo é mantida a possibilidade de segmentos de comprimento desigual. O algoritmo “SCC – Segmentação Comprimento Constante”, apesar de sua concepção ser extremamente simples, mostrou ao longo dos experimentos realizados, ser capaz de particionar a curva de resposta espectral, para fins de extração de feições, de uma forma que levou a resultados bastante aceitáveis, apresentando bons resultados de classificação.

Tabela 8. Comparação da acurácia de classificação (valores máximos) segundo o algoritmo e tamanho das amostras de treinamento e teste utilizadas.

Npixel	Classe	190 bandas	SCC	TREE: CRIT 1	TREE: CRIT 2	OTIM - CRIT 1	OTIM - CRIT 2
150	Milho-cultivo direto	-	93,33	92,00	92,00	95,33	93,33
	Gramíneas	-	100,00	100,00	100,00	100,00	100,00
	Soja-cultivo direto	-	96,00	97,33	97,33	97,33	96,00
	Soja-cultivo mínimo	-	90,67	93,33	93,33	90,67	91,33
	<i>Valor médio</i>	-	95,00	95,67	95,67	95,83	95,17
200	Milho-cultivo direto	63,50	97,50	96,50	97,50	97,50	97,50
	Gramíneas	77,00	100,00	100,00	100,00	100,00	100,00
	Soja-cultivo direto	37,00	94,50	97,00	98,50	97,00	98,50
	Soja-cultivo mínimo	62,50	91,00	94,00	89,50	93,50	93,50
	<i>Valor médio</i>	60,00	95,75	96,88	96,38	97,00	97,38
250	Milho-cultivo direto	76,40	97,20	97,60	97,20	96,40	96,40
	Gramíneas	92,40	100,00	100,00	100,00	100,00	100,00
	Soja-cultivo direto	78,40	94,80	95,60	96,40	96,40	96,40
	Soja-cultivo mínimo	66,80	92,80	92,40	90,80	94,40	92,80
	<i>Valor médio</i>	78,50	96,20	96,40	96,10	96,80	96,40
300	Milho-cultivo direto	89,00	95,33	97,33	97,33	97,00	96,33
	Gramíneas	98,67	100,00	100,00	100,00	100,00	100,00
	Soja-cultivo direto	88,67	97,33	97,33	97,67	97,67	97,33
	Soja-cultivo mínimo	68,33	91,33	93,00	93,00	92,00	92,67
	<i>Valor médio</i>	86,17	96,00	96,92	97,00	96,67	96,58
350	Milho-cultivo direto	89,43	95,14	96,57	96,00	95,71	97,43
	Gramíneas	99,71	99,71	100,00	100,00	100,00	100,00
	Soja-cultivo direto	89,43	97,14	97,43	97,43	98,86	98,57
	Soja-cultivo mínimo	85,14	92,29	92,00	93,71	94,29	91,71
	<i>Valor médio</i>	90,93	96,07	96,50	96,79	97,21	96,93

Tabela 9. Tabela resumo com o número de feições para a qual ocorreu a acurácia de classificação máxima (média para 4 classes) segundo o algoritmo e tamanho das amostras de treinamento e teste utilizadas, apresentados na Tabela 8.

Npixel	190 BANDAS	SCC	TREE: CRIT 1	TREE: CRIT 2	OTIM - CRIT 1	OTIM - CRIT 2
150	-	93	22	22	14	14
200	190	98	40	28	32	34
250	190	97	32	26	34	30
300	190	95	26	30	20	40
350	190	95	28	20	24	30
	<i>190</i>	<i>95,7</i>	<i>29,6</i>	<i>25,2</i>	<i>24,8</i>	<i>29,6</i>

Avaliando-se apenas o valor da acurácia máxima de classificação (valor médio para as 4 classes), conclui-se que todos os algoritmos apresentaram resultados bastante superiores se comparados ao resultado de classificação obtido utilizando-se todas as bandas espectrais originais. As diferenças, para os valores máximos de acurácia no processo de classificação obtidos nos experimentos, são pouco significativas entre os três algoritmos propostos.

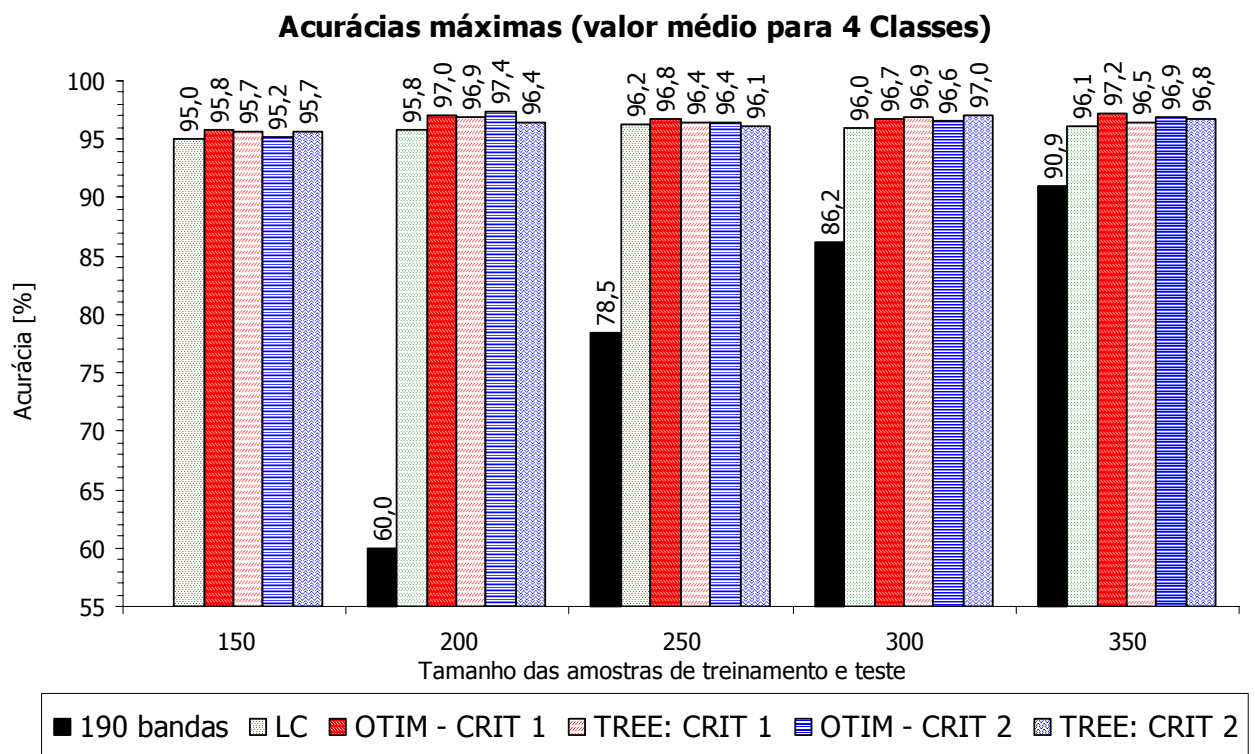


Figura 48. Comparação da acurácia de classificação máxima (média para 4 classes) segundo o algoritmo e tamanho das amostras de treinamento e teste utilizadas.

O acréscimo de acurácia de classificação em relação àquela obtida utilizando-se as bandas espectrais originais pode ser visualizada na Tabela 10. Observa-se que neste caso não se torna viável a utilização de amostras de treinamento com tamanho inferior ao da dimensionalidade dos dados originais mais 1 (191 no presente caso).

Tabela 10. Acréscimo de acurácia de classificação máxima (média para 4 classes), em relação à dimensionalidade original, segundo o algoritmo e tamanho das amostras de treinamento e teste utilizadas. Os resultados são calculados a partir da Tabela 8.

Npixel	SCC	TREE: CRIT 1	OTIM - CRIT 1	TREE: CRIT 2	OTIM - CRIT 2
200	35,75	36,88	37,00	36,38	37,38
250	17,70	17,90	18,30	17,60	17,90
300	9,83	10,75	10,50	10,83	10,42
350	5,14	5,57	6,29	5,86	6,00
<i>Média</i>	<i>17,11</i>	<i>17,77</i>	<i>18,02</i>	<i>17,67</i>	<i>17,92</i>

Se em termos de acurácias máximas a diferença entre os três algoritmos propostos, obtida na classificação, é pequena, o mesmo não ocorre quando selecionados um número menor de feições. Nestas condições, as performances dos três diferentes algoritmos ficam melhor evidenciadas. Estes resultados podem ser visualizados na seqüência de figuras Figura 49 a Figura 53. A necessidade de escolha de um número reduzido de feições pode ser relevante, por exemplo, no caso destas serem utilizados como dados de entrada (*input*) em outro processo de otimização.

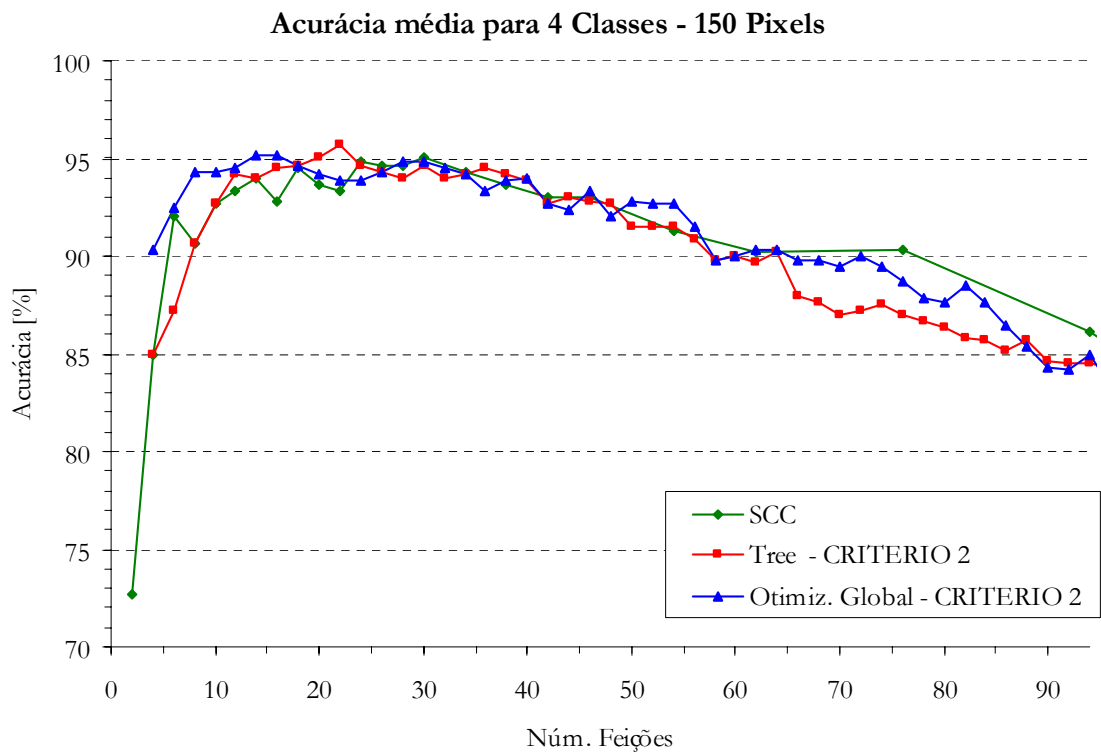
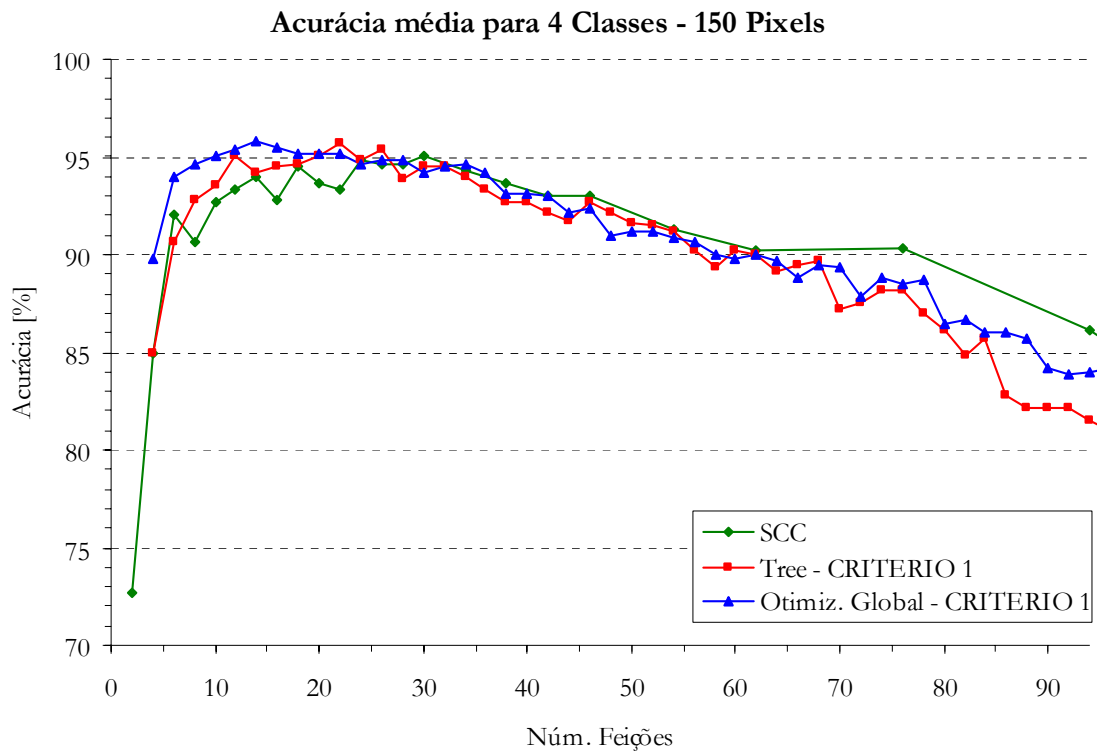


Figura 49. Comparação da acurácia de classificação média (4 classes) segundo o algoritmo, critério utilizados e 150 pixels de treinamento e teste.

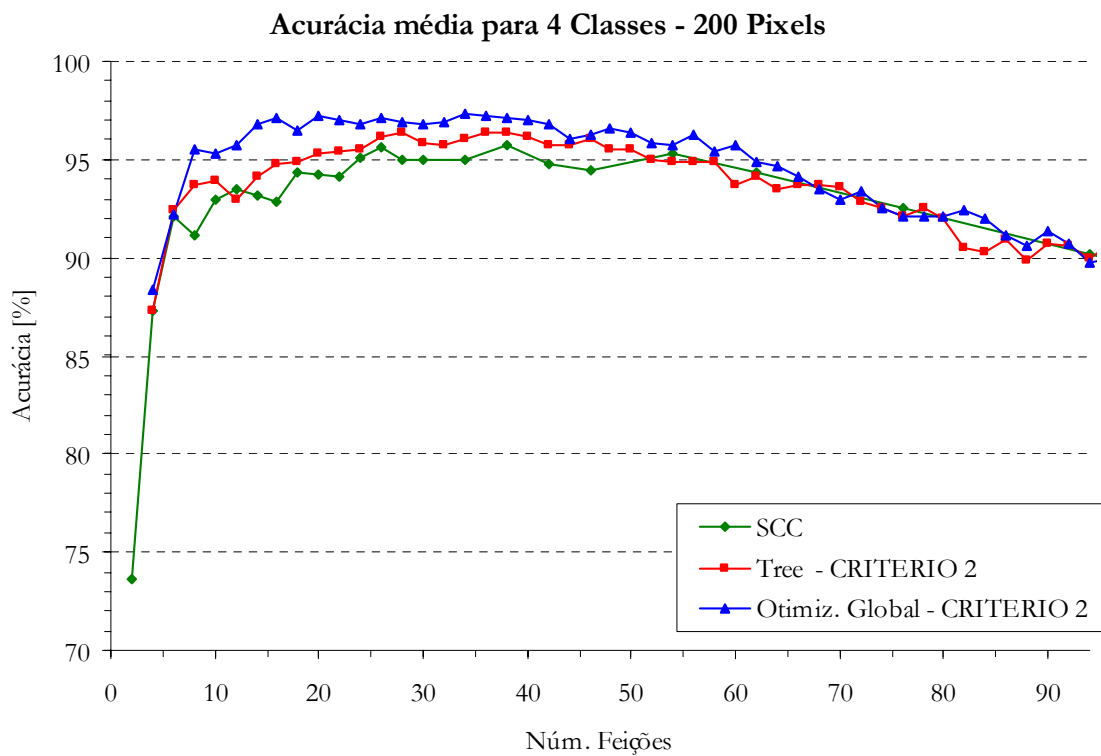
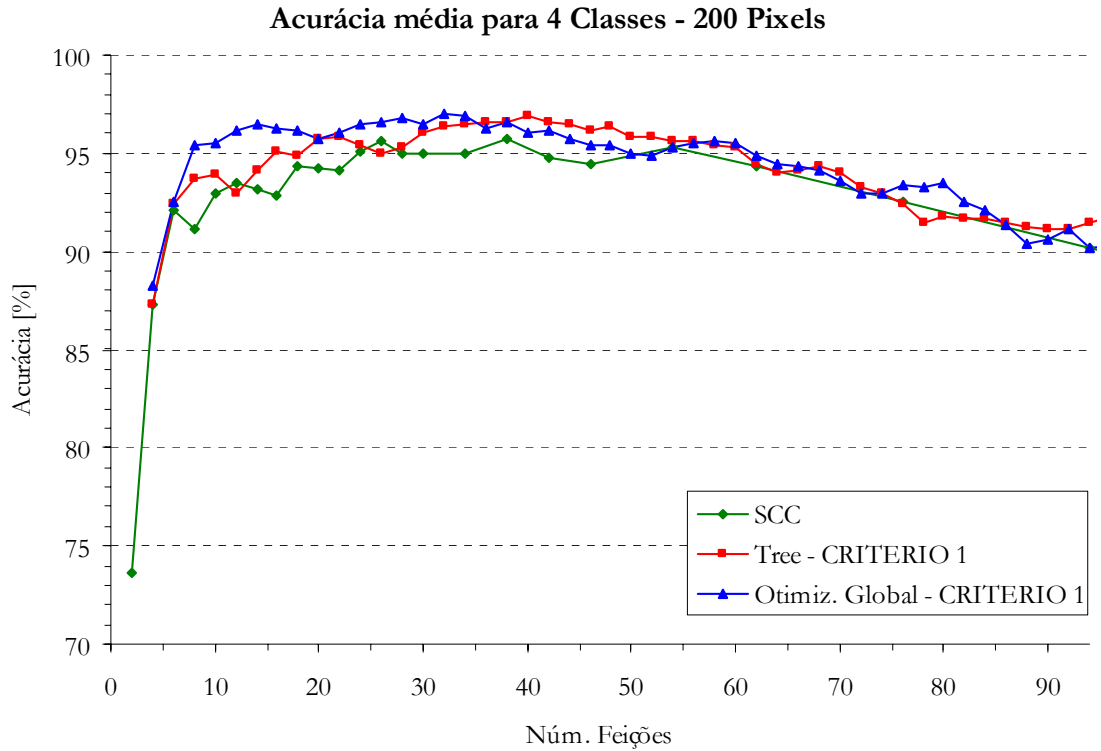


Figura 50. Comparação da acurácia de classificação média (4 classes) segundo o algoritmo, critério utilizados e 200 pixels de treinamento e teste.

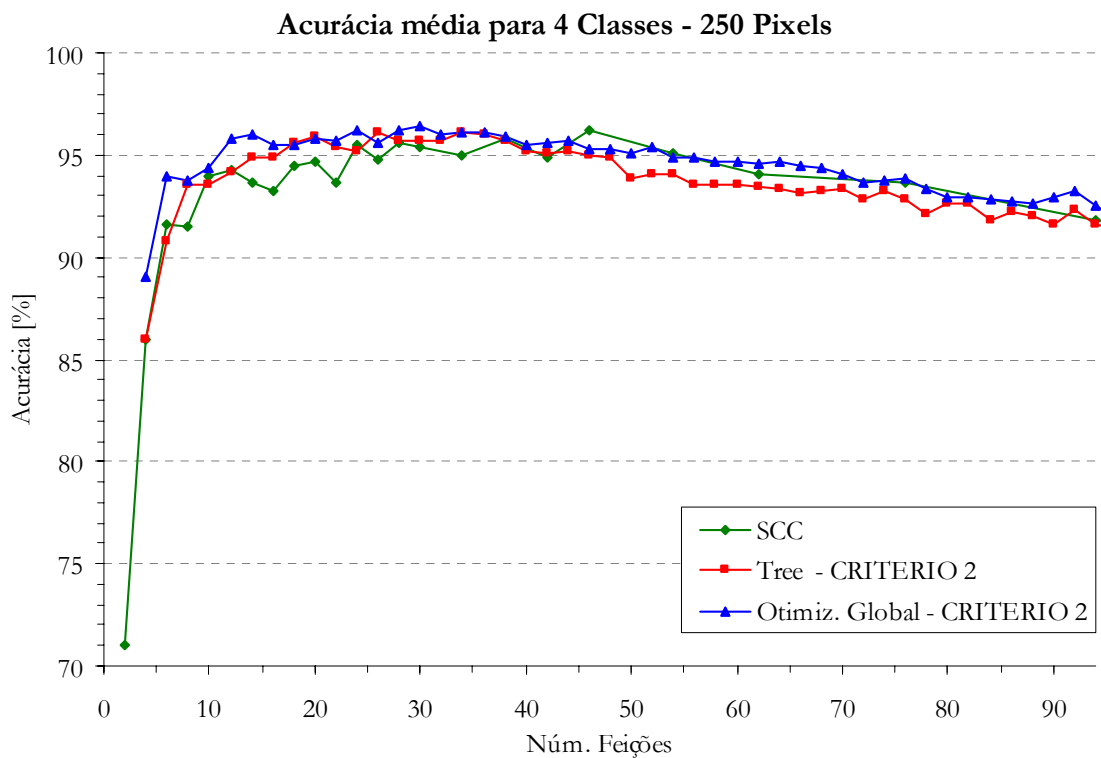
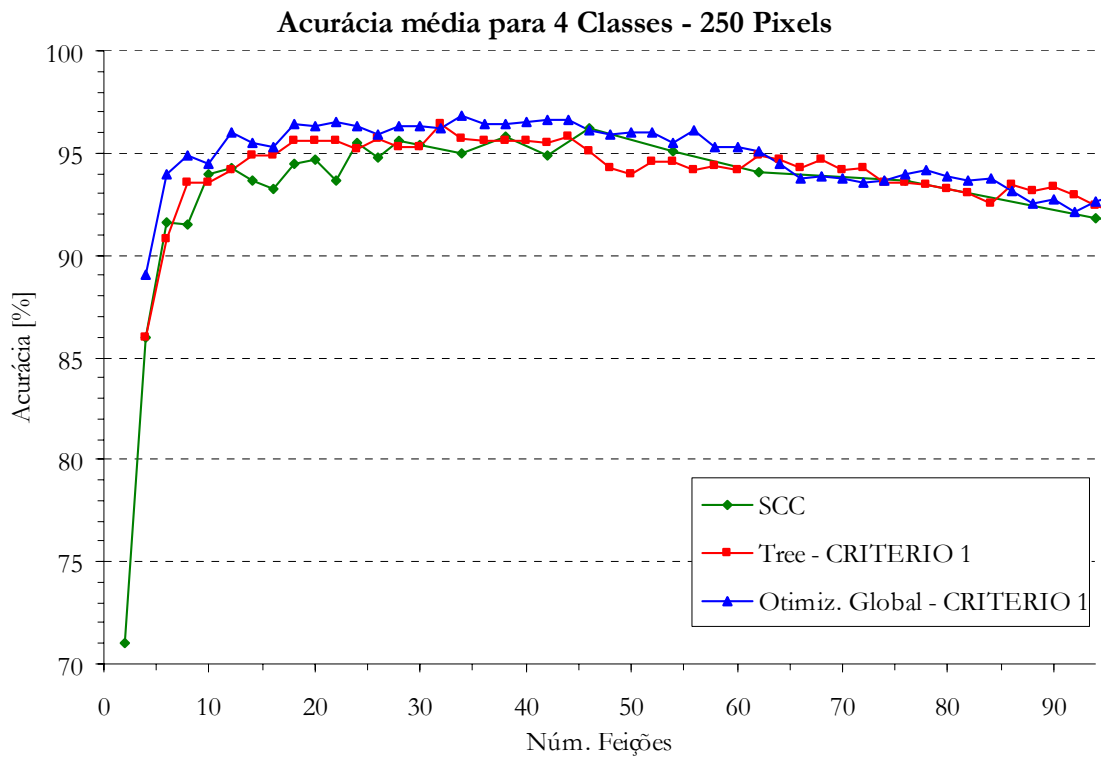


Figura 51. Comparação da acurácia de classificação média (4 classes) segundo o algoritmo, critério utilizados e 250 pixels de treinamento e teste.

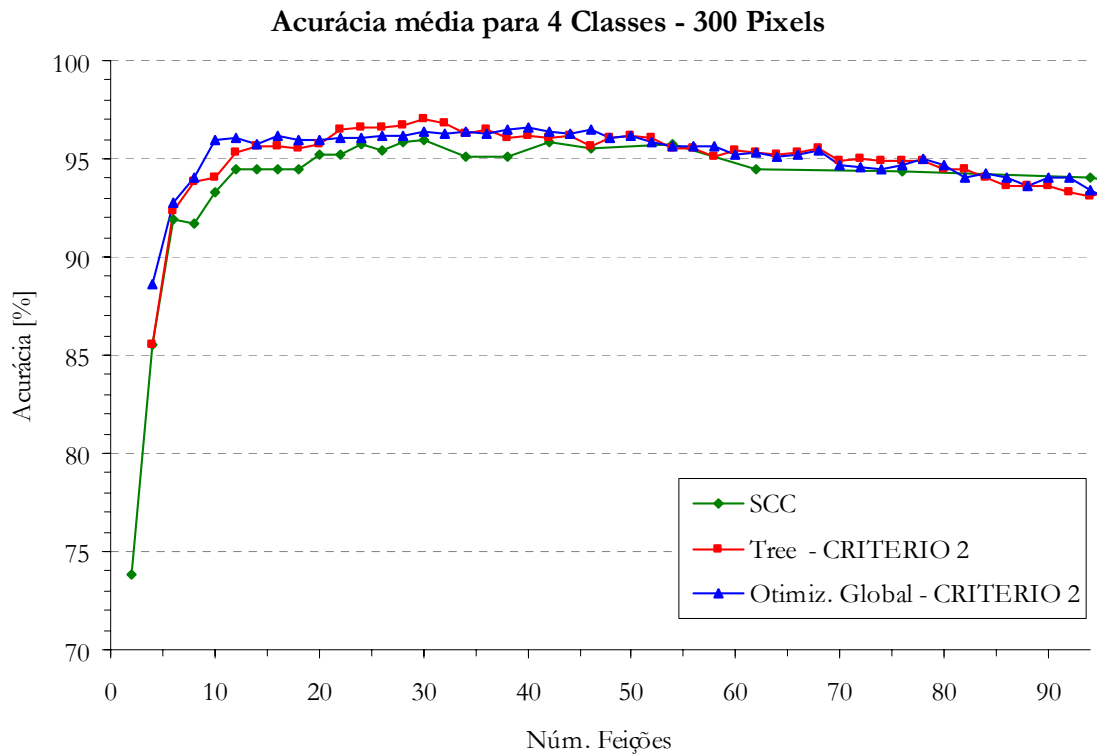
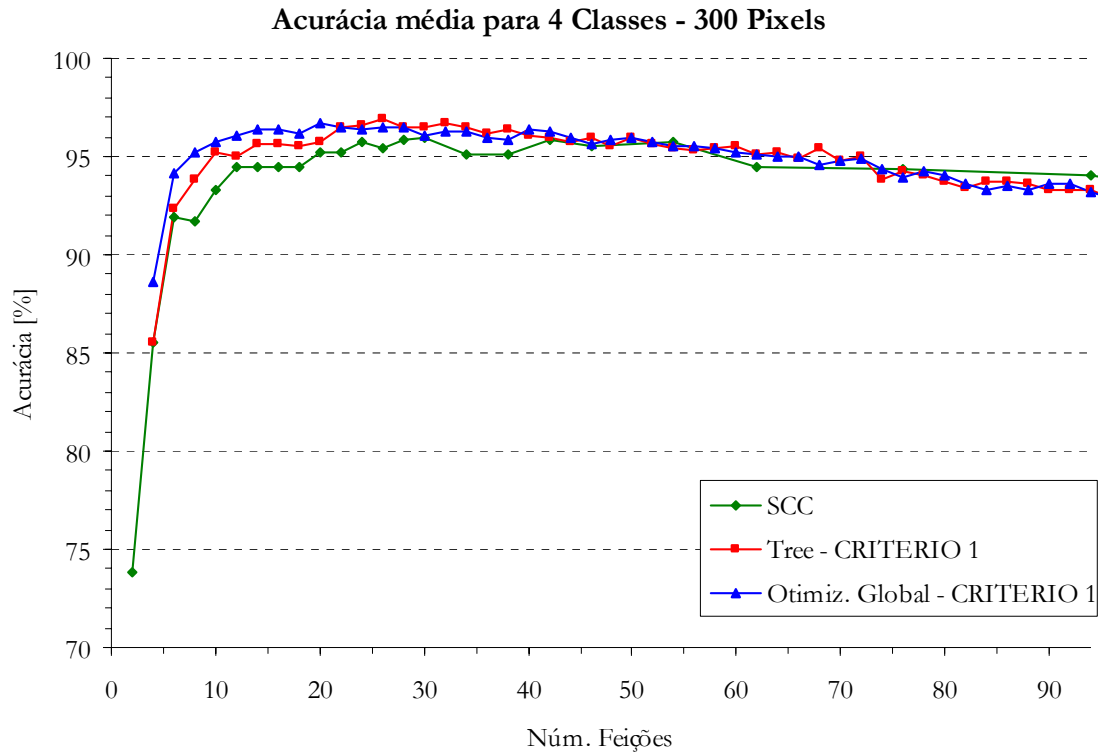


Figura 52. Comparação da acurácia de classificação média (4 classes) segundo o algoritmo, critério utilizados e 300 pixels de treinamento e teste.

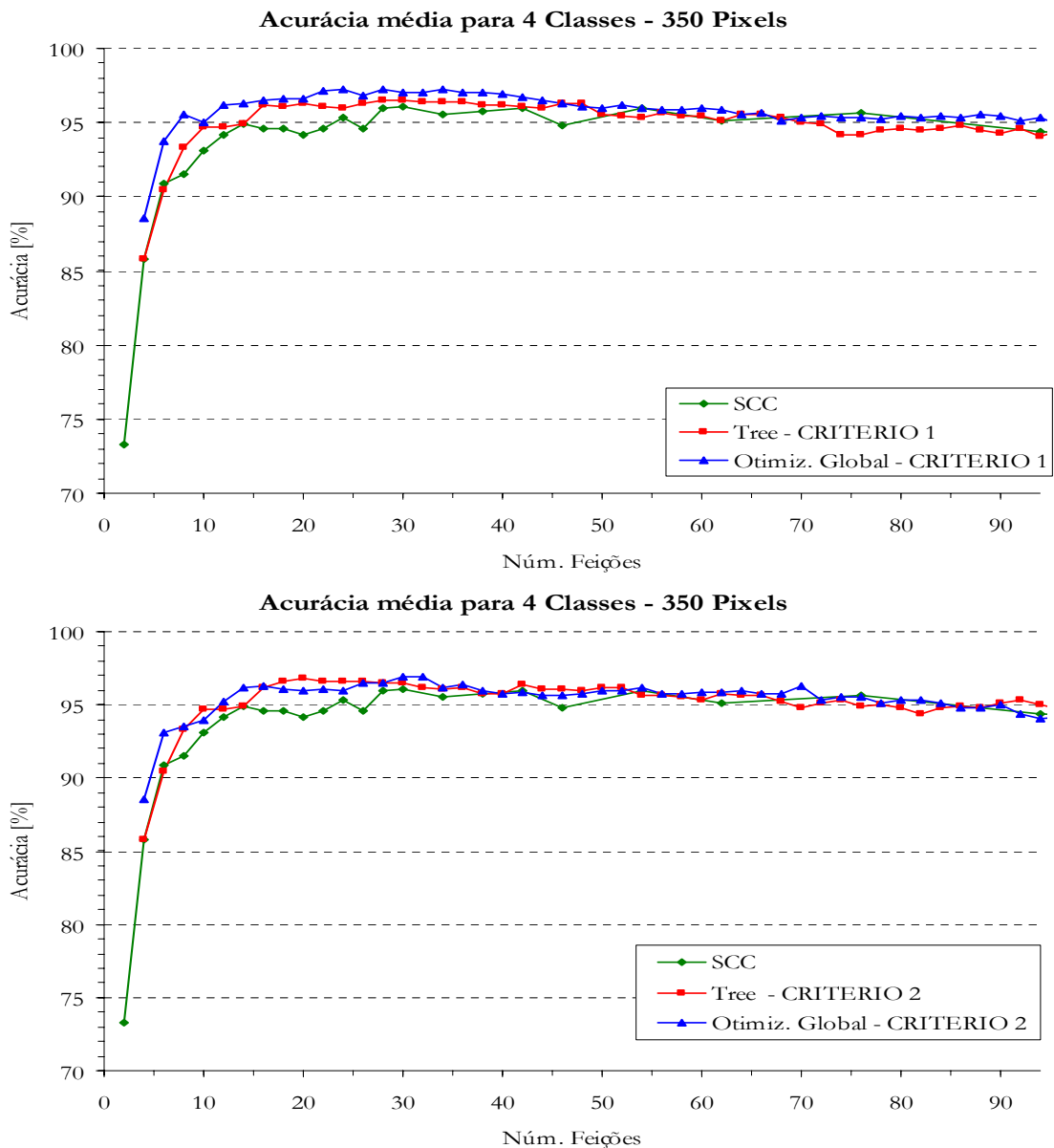


Figura 53. Comparação da acurácia de classificação média (4 classes) segundo o algoritmo, critério utilizados e 350 pixels de treinamento e teste.

Também é importante observar que o algoritmo “Otimização Global”, ao avaliar todas as possibilidades para a primeira partição, inicia com uma acurácia, a quatro feições, superior ao algoritmo “Tree”, que por sua vez neste ponto tem o mesmo valor que o algoritmo SCC, devido a partição da curva de resposta espectral original na posição central. Em geral, a medida em que o número de feições aumenta, os resultados obtidos para os algoritmos que implementam comprimentos variáveis para os segmentos tornam-se “menos ótimos” (em relação aos resultados obtidos utilizando-se o algoritmo mais simples que utiliza comprimentos constantes), já que admitem uma grande quantidade de posições de partição fixas, provenientes das iterações precedente. Entretanto, percebeu-se que o máximo ocorre para um número relativamente reduzido de feições (Tabela 9).

5 - CONCLUSÕES E RECOMENDAÇÕES

Dados em alta dimensionalidade como aqueles adquiridos por uma nova geração de sistemas sensores denominados de *hiperespectrais*, consistindo de um conjunto de dados em forma de imagens digitais coletadas em um número grande de regiões do espectro eletromagnético, contêm informações mais detalhadas com respeito às classes a serem identificadas. Estes dados permitem uma maior acurácia na classificação dos pixels do que aquela que seria possível com a utilização de imagens mais tradicionalmente utilizadas, como aquelas fornecidas por sensores dos sistemas LANDSAT e SPOT, que fornecem dados espectrais em baixa dimensionalidade.

A alta dimensionalidade dos dados aliada ao número geralmente limitado e insuficiente de amostras de treinamento (na maioria das aplicações de sensoriamento remoto a aquisição de um grande número de amostras de treinamento é e pouco prática) torna inviável uma estimação confiável dos parâmetros necessários ao classificador, em especial, a matriz covariância. Este fato resulta em valores baixos para a acurácia dos resultados obtidos no processo de classificação.

Por esta razão é, geralmente, mais eficiente reduzir a dimensionalidade dos dados originais, selecionando um sub-conjunto composto pelas feições mais adequadas a cada caso em particular. Assim sendo, uma etapa importante que deve preceder o processo de classificação refere-se à escolha de um sub-conjunto ótimo ou pelo menos sub-ótimo de feições para cada situação em particular. Este problema é conhecido com o de redução da dimensionalidade dos dados.

Em termos gerais, existem duas abordagens gerais a este problema: seleção de feições e extração de feições. No primeiro caso, busca-se selecionar um sub-conjunto das feições originais que melhor atenda às características do problema em consideração. No segundo caso, o objetivo consiste em diminuir a dimensionalidade dos dados

originais por meio da geração de um número menor de feições derivadas através de combinações lineares ou não, das feições originais.

Neste estudo optou-se por investigar uma metodologia confiável, simples de implementar e computacionalmente eficiente baseada no conceito geral de extração de feições. Ao invés de implementar um procedimento de procura objetivando encontrar um bom subconjunto ótimo ou sub-ótimo entre o total de bandas espectrais disponíveis de acordo com algum dado critério, optou-se por examinar a curva de resposta espectral de cada pixel individualmente, isto é, a sua resposta ao longo do conjunto de todas as bandas disponíveis. Para cada pixel individual, procurou-se capturar a informação disponível na totalidade curva de resposta espectral, por um pequeno número de estatísticas simples que objetivam descrever a posição e forma da resposta espectral do pixel. Em outras palavras, esta operação consiste num processo de redução da dimensionalidade dos dados originais.

Os resultados obtidos com a metodologia proposta para extração de feições em dados hiper-espectrais foram satisfatórios. A suposição básica de que a curva de resposta espectral em cada pixel pode ser segmentada e que a informação contida em cada segmento pode ser adequadamente representada por meio do valor médio dos contadores digitais e da variância destes, mostrou-se adequada nos experimentos realizados. Realizados os experimentos, percebeu-se, por exemplo, que o algoritmo “Otimização Global”, logo nas primeiras iterações, apresenta a tendência de gerar segmentos bastante diferenciados, alguns com pequeno e outros com valores maiores para o comprimento. Este fato parece confirmar uma outra suposição inicial, utilizada como premissa neste trabalho, a de que a partição da curva de resposta espectral original de cada pixel em segmentos de comprimento não necessariamente constante, mas sim otimizado de acordo com um determinado critério, consegue melhor adaptar-se a suposição intuitiva de que entre o total de bandas disponíveis, existem algumas que tem alto poder discriminante, e outras que contribuem muito pouco para separabilidade entre classes, podendo estas últimas serem agrupadas em segmentos maiores. Visto os bons resultados obtidos em termos de acurácia de classificação, conclui-se que os algoritmos propostos neste trabalho conseguem distinguir estas regiões de maneira bastante aceitável.

Com base nos dados que foram utilizados e nos experimentos realizados neste trabalho, pode-se afirmar que os três algoritmos propostos forneceram resultados

aceitáveis. O ganho de acurácia, apresentado na Tabela 10, em relação aos resultados que seriam obtidos na classificação, para o mesmo problema utilizando-se os dados originais, é bastante significativo: superior a 36.38% ao se utilizar uma amostra de treinamento e teste com apenas 200 pixels e ainda superior a 5.57% utilizando 350 pixels para fins de treinamento. Vale lembrar que a obtenção de amostras de treinamento para fins de estimação dos parâmetros do classificador é uma tarefa em geral bastante onerosa o que, em termos práticos, inviabiliza a aquisição de um número suficientemente grande de amostras, compatível com a dimensionalidade que caracteriza os dados hiper-espectrais. De qualquer forma, lembra-se que seria também desejável uma série de experimentos mais exaustiva, utilizando inclusive outros dados (imagens), para melhor testar a metodologia proposta.

Era prevista a ocorrência do fenômeno de Hughes, e ele de fato se manifestou, tanto mais intensamente quando menor a relação entre o tamanho da amostra de treinamento e a dimensionalidade dos dados utilizados no classificador. O método proposto para redução na dimensionalidade do dados via extração de feições serviu claramente para atenuar o fenômeno de Hughes.

Entende-se como sendo um fato positivo a decisão de se ter utilizado o mesmo número de amostras de treinamento para cada uma das quatro classes utilizadas. Desta forma, as estimativas do classificador foram obtidas nas mesmas condições para todas as classes, estando os resultados de classificação associados às características espectrais de cada classe e não a uma maior ou menor confiabilidade nos valores estimados para os parâmetros de cada classe.

Os experimentos inicialmente desenvolvidos ao longo deste estudo abordaram o problema da extração de feições de uma forma local, isto é, considerando cada segmento individualmente. Os resultados insatisfatórios obtidos evidenciaram a necessidade de uma abordagem mais global do problema, levando em consideração a totalidade dos segmentos existentes.

Um inconveniente teórico que existe nos algoritmos aqui propostos para a otimização dos segmentos é que uma vez definida uma posição de partição, esta não pode mais ser alterada nas iterações sucessivas. É um inconveniente que também está presente em alguns outros métodos propostos na literatura, sendo difícil uma solução eficaz, visto a enorme quantidade de combinações que podem ser geradas no processo de seleção de um subconjunto de feições, mesmo para o caso de algumas poucas

dezenas de bandas espectrais como dados de entrada. De qualquer forma, pode-se afirmar aqui, a existência de uma otimização parcial nesta abordagem, no sentido que o resultado que é apresentado a cada iteração é sempre o ótimo (dentro dos critérios propostos) em relação às partições previamente geradas. Este processo de otimização, embora parcial, mostrou-se suficiente para se atingir elevadas acurácias nos processos de classificação desenvolvidos nos experimentos.

Em termos de valor máximo para a acurácia de classificação (resultado médio para as quatro classes, Tabela 8), todos os três algoritmos apresentaram resultados muito semelhantes, com valores superiores a 95.17%. Estes resultados foram obtidos para um número de feições que variou conforme o algoritmo, o número de amostras de treinamento e os critérios utilizados. A diferença de performance entre os algoritmos propostos pode ser melhor observada através dos experimentos desenvolvidos nas condições mais desfavoráveis, isto é, nos casos em que o número de feições utilizadas são os mais reduzidos. Nestas condições, os algoritmos que implementam o processo de segmentação utilizando segmentos de comprimento variável apresentam resultados de acurácia de classificação superiores aos obtidos utilizando segmentos com comprimento constantes, ou seja, o “ganho” é mais rápido. Tais resultados puderam ser observados nas Figuras 49 a Figura 53. Embora na maioria das aplicações, utilizando dados de sensoriamento remoto proveniente desta nova geração de sensores hiperespectrais, o interesse seja pela procura de um subconjunto ótimo de feições que, para um determinado número de amostras de treinamento disponíveis, conduza a um máximo de acurácia de classificação, existem também aplicações onde se mostra necessário a escolha de um número reduzido de feições, que podem, por exemplo, serem utilizados como *dados de entrada* em algum outro processo de otimização ou até mesmo para fins de visualização em aplicações onde se trabalha com interpretação visual da imagem. Nestas condições, a utilização de um subconjunto bastante reduzido de feições, provenientes dos algoritmos que implementam as abordagens que permitem a segmentação da curva de resposta espectral em comprimentos não constantes ao invés de comprimentos constantes, mostra-se bastante atraente.

Quanto aos critérios que foram utilizados para tomada de decisão na partição de segmentos, foram sugeridos dois, ambos baseados na utilização da distância de Bahattacharyya, para estimar a da separabilidade entre classes. O chamado critério 1

está diretamente associado ao limiar do erro bayesiano implícito no processo de classificação, levando em consideração a presença de todas as classes, já o critério 2 toma como referência o par de classes com maior dificuldade de separação e define a segmentação para todas as classes em função deste par crítico. Os resultados obtidos utilizando-se ambos foram muito semelhantes. Note também que podem ser propostos outros critérios para a medida da separabilidade entre classes sem que seja necessário alterar os algoritmos.

Os resultados de acurácia de classificação obtidos sugerem que a suposição de normalidade dos dados utilizados neste experimento é bem razoável. Uma possibilidade de investigação para futuros trabalhos também poderia ser a alteração deste algoritmo, implementando abordagens não paramétricas, de tal forma que não seja necessário fazer suposição sobre a distribuição de probabilidade dos dados de entrada.

Quanto à complexidade computacional exigida pelos algoritmos propostos, pode-se dizer que estão coerentes com a capacidade atual dos processadores disponíveis em computadores pessoais normalmente utilizados. Portanto, parece adequado supor-se que a proposta desenvolvida nesta dissertação pode ser considerada operacional, sendo portanto viável sua implementação em *softwares* para fins de classificação de dados de imagens digitais com dimensionalidade elevada. Os experimentos realizados permitem supor que a metodologia proposta nesta dissertação se constitui em uma alternativa para o problema de extração de feições em imagens hiperespectrais.

Como sugestão para trabalhos futuros, poder-se-ia mencionar uma investigação sobre a utilização de outras estatísticas a serem empregadas para extração das feições. Uma outra possibilidade que poderia ser sugerida seria a utilização de comprimentos constantes, porém variando o número de feições utilizadas para representar cada segmento. Neste trabalho utiliza-se a média aritmética e a variância, parâmetros que buscam representar a posição e a forma de cada segmento. Talvez pudesse ser utilizada apenas a média, ou também um terceiro parâmetro estatístico. Também poderia ser investigada a utilização das primeiras componentes da DCT (Transformada Discreta do Cosseno) ou estatísticas baseadas nos “momentos L”. Em particular, embora não apresentados explicitamente neste estudo, experimentos preliminares utilizando estas duas últimas abordagens apresentadas sugeriram bons resultados.

6 - REFERÊNCIAS BIBLIOGRÁFICAS

- [ADA 86] J. B. ADANS, M. O. SMITH, A. R. GILLEPIE, “Spectral Mixture: A New Analysis of Rock and Sol Types at The Viking Lander 1 Suite”, in J. Geophysical Res., vol. 91. no. B8, July 10, pp. 8098-8112, 1986.
- [BEN 99] J. A. BENEDIKTSSON, I. KANELLOPOULOS, “Classification of Multisource and Hyperspectral Data Based on Decision Fusion” IEEE Transactions on Geoscience and Remote Sensing, vol. 37. no. 3 pp.1367-1377, 1999.
- [BRU 00] L. BRUZZONE, S. B. SERPICO, “A Technique for Feature Selection In Multiclass Cases”, International Journal of Remote Sensing, vol. 21, pp. 549-563, 2000.
- [BRU 95] 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.
- [CHA 00] C. I. CHANG, “Information-Theoretic Approach to Spectral Variability, Similarity, and Discrimination for Hyperspectral Image Analysis”, IEEE Transactions on Information Theory, vol. 46, no. 5 pp. 1927-1932, 2000.
- [DUD 01] R. O DUDA, P. E. HART, D. G. STORK, “Pattern Classification”. 2^a edição, John Wiley & Sons, Estados Unidos da América, 1991.
- [ERB 02] M. ERBERT, “Uso da Análise Discriminante Regularizada (RDA) no Reconhecimento de Padrões em Imagens Digitais Hiperespectral de Sensoriamento Remoto”, Dissertação de Mestrado, Programa de Pós-Graduação em Sensoriamento Remoto - UFRGS. Porto Alegre, 2002.
- [FUK 90] K. FUKUNAGA, “Introduction to Statistical Pattern Recognition”, 2^a edição, Academic Press, New York, 1990.

-
- [HU 99] Y. H. HU, H. B. LEE, F. L. SCARPACE, "Optimal Linear Spectral Unmixing", IEEE Trans. on Geoscience and Remote Sensing. Vol. 37, no. 1, pp.639-644, 1999.
- [ICH 84] M. ICHINO, J. SKLANSKY, "Optimum Feature Selection by Zero-One Integer Programming", IEEE Transactions on Systems, Man and Cybernetics, vol. 14, pp. 737-746, 1984.
- [JAI 97] A. JAIN, D. ZONGKER, "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.
- [JAK 01] Q. JACKSON, D. LANDGREBE, "Na 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.
- [JIA 99] 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.
- [JIM 99] L. O. JIMENEZ, A. MORALES-MORELL, A. CREUS, "Classification of Hyperdimensional Data Based on Feature and Decision Fusion Approaches Using Projection Pursuit, Majority Voting, and Neural Networks", IEEE Transactions on Geoscience and Remote Sensing, vol. 37, no. 3, pp. 1360-1366, 1999.
- [JIM 99] L. O. JIMENEZ, D. A. LANDGREBE, "Hyperspectral Data Analysis and Feature Reduction Via Projection Pursuit", IEEE Transactions on Geoscience and Remote Sensing vol. 37, pp. 2653-2667, 1999.
- [JLP 03] JET PROPULSION LABORATORY, Homepage do AVIRIS (Airborne Visible/Infrared Imaging Spectrometer), <http://aviris.jpl.nasa.gov/>, acessado em 12 de dezembro de 2003.
- [KIT 78] J. KITTLER, "Feature Set Search Algorithm", Pattern Recognition and Signal Processing, pp. 41-60, C.H.Chen, Sijthoff and Noordhoff, Alpena an den Rijn, Netherlands, 1978.
- [KUM 01] S. KUMAR, J. GHOSH, M.M. CRAWFORD, "Best-Bases Feature Extraction Algorithms for Classification of Hyperspectral Data", IEEE Trans. on Geoscience and Remote Sensing, vol. 39, no. 7, pp. 1368-1379, 2001.

-
- [LAN 03] D. LANDGREBE, "Signal Theory Methods in Multispectral Remote Sensing". Editora Wiley, Nova Jersey, 2003. ISBN 0-471-42028-X.
- [LEE 93] C. LEE, D. A. LANDGREBE, D. "Analyzing High-Dimensional Multispectral Data", IEEE. Transaction on Geoscience and Remote Sensing, vol. 31, pp. 792-800, 1993.
- [LEE 93] C. LEE, D. A. LANDGREBE, "Feature Extraction Based on Decision Boundaries", IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 15, pp 388-4000, 1993.
- [LEE 00] C. LEE, D. A. LANDGREBE, "Bayes Error Evaluation of the Gaussian ML Classifier", IEEE Trans. on Geoscience and Remote Sensing, vol. 38, no. 3, pp. 1471-1475, 2000.
- [PUD 94] P. PUDIL, J. NOVOVICOVA, e J. KITTLER, "Floating Search Methods in Feature Selection", Pattern Recognition Letters, vol. 15, pp.1119-1125, 1994.
- [RIC 99] J. A. RICHARDAS, X. JIA, "Remote Sensing Digital Image Analysis: An Introduction", Springer-Verlag, Berlim, 1999.
- [SER 01] S. B. SERPICO, 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
- [SER 03] 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, 2003.
- [SHA 94] B. SHAHSHANI, D. 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.
- [SHI 03] P. SHIPPERT, "Spotlight on Hyperspectral", <http://www.uni.edu/storm/rs/2002/aars/hyperspectral-wp.pdf>, acessado em 12 de dezembro de 2003.
- [SIE 89] W. SIEDLECKI, J. SKLANSKY, "On Automatic Feature Selection", Int. Journal of Pattern Recognition and Artificial Intelligence, vol. 2, pp. 197-220, 1988.

- [STE 76] S. D. STEARNS, "On Selection Features For Pattern Classifiers", Third Int. Conf. On Pattern Recognition, Coronado (CA). pp. 71-75, 1976.
- [SWA 78] P. H. SWAIN, S. M. DAVIS, Remote Sensing: The Quantitative Approach, McGraw-Hill, New York, 1978.
- [THE 89] C. W. THERRIEN, "Decision, Estimation and Classification. An Introduction to Pattern Recognition and Related Topics". Wiley, New York.