

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

USO DA ANÁLISE DISCRIMINANTE REGULARIZADA (RDA) NO
RECONHECIMENTO DE PADRÕES EM IMAGENS DIGITAIS
HIPERESPECTRAL DE SENSORIAMENTO REMOTO

POR
MAURO ERBERT

PROF. PhD VITOR HAERTEL
ORIENTADOR

PORTO ALEGRE, RS
2001

SUMÁRIO

1	INTRODUÇÃO	11
1.1	CONCEITOS BÁSICOS	11
1.2	O SISTEMA SENSOR AVIRIS.....	21
1.3	RECONHECIMENTO DE PADRÕES.....	29
2	REVISÃO BIBLIOGRÁFICA	35
2.1	PROBLEMA DA DIMENSIONALIDADE	35
2.1.1	<i>Alta dimensionalidade</i>	35
2.1.2	<i>Fenômeno de Hughes, a “maldição da dimensionalidade”</i>	36
2.2	MÉTODOS PARA MITIGAR O FENÔMENO DE HUGHES	39
2.3	MÉTODOS DE REGULARIZAÇÃO DA MATRIZ COVARIÂNCIA.....	41
3	METODOLOGIA.....	43
3.1	TEORIA BAYESIANA DA DECISÃO	48
3.1.1	<i>Classificador de Bayes</i>	50
3.1.2	<i>Classificador de Máxima Verossimilhança Gaussiana</i>	52
3.2	DESCRIÇÃO DOS CLASSIFICADORES	54
3.2.1	<i>Classificador quadrático – QDA (Quadratic Discriminant Analysis)</i>	55
3.2.2	<i>Classificador linear – LDA (Linear Discriminant Analysis)</i>	56
3.3	MÉTODO REGULARIZADO – RDA (REGULARIZED DISCRIMINANT ANALYSIS).....	60
4	EXPERIMENTOS.....	67
4.1	MATERIAIS.....	67
4.2	CENA DE ESTUDO	68
4.2.1	<i>Seleção de Bandas Espectrais</i>	70
4.2.2	<i>Seleção de Classes</i>	71
4.2.3	<i>Amostras de treinamento e teste das classes selecionadas</i>	76
4.3	FERRAMENTA ADRIH (<i>ANÁLISE DISCRIMINANTE REGULARIZADA EM IMAGEM HIPERESPECTRAL</i>).....	77
4.4	TESTANDO O MÉTODO RDA COM DADOS DE SENSORIAMENTO REMOTO	77
4.4.1	<i>Experimento 01</i>	78
4.4.2	<i>Experimento 02</i>	86
4.4.3	<i>Experimento 03</i>	92
5	CONCLUSÕES/ SUGESTÕES.....	99
	ANEXO 1 – Gráfico demonstrativo do tempo de processamento com 200 amostras.....	103
	ANEXO 2 – Gráfico demonstrativo do tempo de processamento com 500 amostras.....	104
	ANEXO 3 – Gráfico demonstrativo do tempo de processamento com todas as amostras.....	105
	ANEXO 4 – Manual do usuário do software ADRIH.....	106
	BIBLIOGRAFIA.....	115

LISTA DE TABELAS

<i>Tabela 1 – Ficha técnica do sensor AVIRIS.....</i>	<i>24</i>
<i>Tabela 2 - Parâmetros necessários para o sensor LandSat-TM e o sensor AVIRIS</i>	<i>37</i>
<i>Tabela 3 – Seleção de bandas ruidosas que foram excluídas (35 bandas)</i>	<i>70</i>
<i>Tabela 4 – Bandas restantes que serão usadas no estudo (185 bandas).....</i>	<i>70</i>
<i>Tabela 5 - Total de pixels extraídos da imagem AVIRIS.....</i>	<i>72</i>
<i>Tabela 6 - Subconjuntos $N \times N$.....</i>	<i>72</i>
<i>Tabela 7 - Amostras de treinamento e teste das classes selecionadas</i>	<i>76</i>
<i>Tabela 8 - Quantidade de amostras das classes selecionadas, utilizadas no Experimento 01, para 200 amostras</i>	<i>79</i>
<i>Tabela 9 - Tabela de contingência do método QDA – 200 amostras</i>	<i>79</i>
<i>Tabela 10 - Tabela de contingência do método LDA – 200 amostras.....</i>	<i>80</i>
<i>Tabela 11 - Acurácia máxima dos classificadores com dimensionalidade total (185) para cada classe com 200 amostras de treinamento</i>	<i>84</i>
<i>Tabela 12 - Acurácia máxima alcançada, com dimensionalidade “d” com 200 amostras.....</i>	<i>84</i>
<i>Tabela 13 - Quantidade de amostras das classes selecionadas, utilizadas no Experimento 02, para 500 amostras</i>	<i>86</i>
<i>Tabela 14 - Tabela de contingência do método QDA – 500 amostras</i>	<i>87</i>
<i>Tabela 15 - Tabela de contingência do método LDA – 500 amostras.....</i>	<i>87</i>
<i>Tabela 16 - Acurácia máxima dos classificadores com dimensionalidade total (185) para cada classe com 500 amostras</i>	<i>90</i>
<i>Tabela 17 - Acurácia máxima alcançada, com dimensionalidade “d”, com 500 amostras.....</i>	<i>90</i>
<i>Tabela 18 - Quantidade de amostras das classes selecionadas, utilizadas no Experimento 03, com tamanhos variados.....</i>	<i>92</i>
<i>Tabela 19 - Relação entre o número de amostras de treinamento com a dimensionalidade</i>	<i>93</i>
<i>Tabela 20 - Tabela de contingência do método QDA – Amostras de tamanho diferentes</i>	<i>93</i>
<i>Tabela 21 - Tabela de contingência do método LDA – Amostras de tamanho diferentes.....</i>	<i>94</i>
<i>Tabela 22 - Acurácia máxima dos classificadores com dimensionalidade total (185) para cada classe.....</i>	<i>97</i>
<i>Tabela 23 - Acurácia máxima alcançada, com “d” dimensionalidade.....</i>	<i>97</i>

LISTA DE FIGURAS

Fig.1 - Espectro Eletromagnético - Espectro Solar	12
Fig. 2 - Bandas espectrais sistema LandSat-TM - (μm)	13
Fig. 3 - Reflectância do alvo "árvore" (Blue, Green, Red, InfraRed)	15
Fig. 4 - Curvas de resposta espectral padrão do solo, da vegetação e da água	15
Fig. 5 - Comparação entre baixa e alta resolução espectral.....	19
Fig. 6 - Satélite Orbview-4 e Orbview-3.....	21
Fig. 7 - Avião ER-2 que transporta o sensor AVIRIS.....	22
Fig. 8 - Sensor AVIRIS	23
Fig. 9 - Método de varredura "Whisk Broom"	24
Fig. 10 - Conceito AVIRIS.....	25
Fig. 11 - Espectro de reflectância de 3 minerais diferentes(Hematita, Kaolinita e Domilinta), segundo o sensor LandSat- TM	26
Fig. 12 - Espectro de reflectância de 3 minerais diferentes(Hematita, Kaolinita e Domilinta), segundo sensor AVIRIS.....	27
Fig. 13 - Imagem Cubo (Moffett Field)	28
Fig 14 - Esquema básico para reconhecimento de padrões.....	29
Fig. 15 - Número de parâmetros a serem estimados na matriz covariância	37
Fig. 16 - Efeito de Hughes. [HSI 98]	38
Fig. 17 - Conversão de sinal analógico para contador digital	43
Fig. 18 - Espaço multiespectral para $n=2$	44
Fig. 19 - Fronteiras de decisão para $n=2$	45
Figura 20 - Interpenetração de classes.....	46
Fig. 21 - Fronteira de decisão entre duas classes.....	47
Fig. 22 - Classes com $\mu_k \neq \mu_{k+1}$ e $\Sigma_k \neq \Sigma_{k+1}$	
Fig. 23 - Classes com $\mu_k = \mu_{k+1}$ e $\Sigma_k \neq \Sigma_{k+1}$	55
Fig. 24 - Classes com $\mu_k \neq \mu_{k+1}$ e $\Sigma_k = \Sigma_{k+1}$	
Fig. 25 - Classes com $\mu_k = \mu_{k+1}$ e $\Sigma_k = \Sigma_{k+1}$	56
Fig. 26 - Cubo de resultados da combinação de lambda (λ) e gama (γ)	65
Fig. 27 - Localização do estado de Indiana - EUA (região em verde)	68
Fig. 28 - AVIRIS: Composição RGB com as bandas 50, 27 e 17, da área experimental de teste, (NW Indiana's Indian Pine) - 220 bandas espectrais.....	69
Fig. 29 - Mapa temático da verdade terrestre (ground true) da figura 28.....	69
Fig. 30 - Classes selecionadas para o estudo e suas respectivas cores temáticas, segundo figura 29	71

Fig. 31 - As linhas dos gráficos (Selection Graph) da assinatura espectral (b), mostram os valores <i>máximo</i> , <i>mínimo</i> , <i>desvio padrão</i> e o valor <i>médio</i> dos pixels pertencentes a classe Corn.....	73
Fig. 32 - As linhas dos gráficos (Selection Graph) da assinatura espectral (b), mostram os valores <i>máximo</i> , <i>mínimo</i> , <i>desvio padrão</i> e o valor <i>médio</i> dos pixels pertencentes a classe Corn-min.....	73
Fig. 33 - As linhas dos gráficos (Selection Graph) da assinatura espectral (b), mostram os valores <i>máximo</i> , <i>mínimo</i> , <i>desvio padrão</i> e o valor <i>médio</i> dos pixels pertencentes a classe Soy-clean	74
Fig. 34 - As linhas dos gráficos (Selection Graph) da assinatura espectral (b), mostram os valores <i>máximo</i> , <i>mínimo</i> , <i>desvio padrão</i> e o valor <i>médio</i> dos pixels pertencentes a classe Soy-notill.....	74
Fig. 35 - As linhas dos gráficos (Selection Graph) da assinatura espectral (b), mostram os valores <i>máximo</i> , <i>mínimo</i> , <i>desvio padrão</i> e o valor <i>médio</i> dos pixels pertencentes a classe Woods	74
Fig. 36 - Sobreposição das assinaturas espectrais das classes selecionadas	75
Fig. 37 - Gráfico de acurácia das amostras teste da classe Woods, com estimação de parâmetros usando 200 amostras de treinamento	81
Fig. 38 - Gráfico de acurácia das amostras teste da classe Corn, com estimação de parâmetros usando 200 amostras de treinamento	82
Fig. 39 - Gráfico de acurácia das amostras teste da classe Corn-min, com estimação de parâmetros usando 200 amostras de treinamento	82
Fig. 40 - Gráfico de acurácia das amostras teste da classe Soy-clean, com estimação de parâmetros usando 200 amostras de treinamento	83
Fig. 41 - Gráfico de acurácia das amostras teste da classe Soy-notill, com estimação de parâmetros usando 200 amostras de treinamento	83
Fig. 42 - Gráfico de acurácia das amostras teste da classe Corn, com estimação de parâmetros usando 500 amostras de treinamento	88
Fig. 43 - Gráfico de acurácia das amostras teste da classe Corn-min, com estimação de parâmetros usando 500 amostras de treinamento	88
Fig. 44 - Gráfico de acurácia das amostras teste da classe Soy-clean, com estimação de parâmetros usando 500 amostras de treinamento	89
Fig. 45 - Gráfico de acurácia das amostras teste da classe Soy-notill com estimação de parâmetros usando 500 amostras de treinamento	89
Fig. 46 - Gráfico de acurácia das amostras teste da classe Corn, com estimação de parâmetros usando 1821 amostras de treinamento	95
Fig. 47 - Gráfico de acurácia das amostras teste da classe Corn-min, com estimação de parâmetros usando 1221 amostras de treinamento	95
Fig. 48 - Gráfico de acurácia das amostras teste da classe Soy-clean, com estimação de parâmetros usando 2327 amostras de treinamento	96
Fig. 49 - Gráfico de acurácia das amostras teste da classe Soy-notill, com estimação de parâmetros usando 1132 amostras de treinamento	96

AGRADECIMENTOS

Dedico este trabalho a minha família Gerhardt, Erna e Cynthia, por me incentivarem sempre a continuar estudando. A minha namorada Gabriele, por estar sempre comigo, me apoiando e tendo muita paciência, pois ♪...nossa vida vai melhorar...♪, um agradecimento especial para Elaine, Maria e o Zumba.

Agradecer a todos os professores, funcionários, demais amigos e colegas do Centro de Estadual de Pesquisas em Sensoriamento Remoto e Meteorologia - CEPSRM, do pessoal da Comunidade Evangélica Luterana Cristo – CELC - POA, em especial aos integrantes da banda.

Agradecer a CAPES pelo apoio com a bolsa de estudos. Ao professor Landgrebe da Purdue University pela imagem cedida para a realização deste trabalho.

Um agradecimento especial ao meu orientador PhD Vitor Haertel, pela paciência, dedicação e disposição no auxílio para a conclusão desta dissertação de mestrado.

Sou grato a Jesus Cristo, meu Salvador, por estar sempre ao meu lado...

Deus te abençoe...

SINOPSE

Em cenas naturais, ocorrem com certa frequência classes espectralmente muito similares, isto é, os vetores média são muito próximos. Em situações como esta, dados de baixa dimensionalidade (LandSat-TM, Spot) não permitem uma classificação acurada da cena. Por outro lado, sabe-se que dados em alta dimensionalidade [FUK 90] tornam possível a separação destas classes, desde que as matrizes covariância sejam suficientemente distintas. Neste caso, o problema de natureza prática que surge é o da estimação dos parâmetros que caracterizam a distribuição de cada classe. Na medida em que a dimensionalidade dos dados cresce, aumenta o número de parâmetros a serem estimados, especialmente na matriz covariância.

Contudo, é sabido que, no mundo real, a quantidade de amostras de treinamento disponíveis, é frequentemente muito limitada, ocasionando problemas na estimação dos parâmetros necessários ao classificador, degradando portanto a acurácia do processo de classificação, na medida em que a dimensionalidade dos dados aumenta. O Efeito de Hughes, como é chamado este fenômeno, já é bem conhecido no meio científico, e estudos vêm sendo realizados com o objetivo de mitigar este efeito. Entre as alternativas propostas com a finalidade de mitigar o Efeito de Hughes, encontram-se as técnicas de regularização da matriz covariância. Deste modo, técnicas de regularização para a estimação da matriz covariância das classes, tornam-se um tópico interessante de estudo, bem como o comportamento destas técnicas em ambientes de dados de imagens digitais de alta dimensionalidade em sensoriamento remoto, como por exemplo, os dados fornecidos pelo sensor AVIRIS.

Neste estudo, é feita uma contextualização em sensoriamento remoto, descrito o sistema sensor AVIRIS, os princípios da análise discriminante linear (LDA), quadrática (QDA) e regularizada (RDA) são apresentados, bem como os experimentos práticos dos métodos, usando dados reais do sensor. Os resultados mostram que, com um número limitado de amostras de treinamento, as técnicas de regularização da matriz covariância foram eficientes em reduzir o Efeito de Hughes. Quanto à acurácia, em alguns casos o modelo quadrático continua sendo o melhor, apesar do Efeito de Hughes, e em outros casos o método de regularização é superior, além de suavizar este efeito.

Esta dissertação está organizada da seguinte maneira:

No primeiro capítulo é feita uma introdução aos temas: sensoriamento remoto (radiação eletromagnética, espectro eletromagnético, bandas espectrais, assinatura espectral), são também descritos os conceitos, funcionamento do sensor hiperespectral AVIRIS, e os conceitos básicos de reconhecimento de padrões e da abordagem estatística.

No segundo capítulo, é feita uma revisão bibliográfica sobre os problemas associados à dimensionalidade dos dados, à descrição das técnicas paramétricas citadas anteriormente, aos métodos de QDA, LDA e RDA, e testes realizados com outros tipos de dados e seus resultados.

O terceiro capítulo versa sobre a metodologia que será utilizada nos dados hiperespectrais disponíveis.

O quarto capítulo apresenta os testes e experimentos da Análise Discriminante Regularizada (RDA) em imagens hiperespectrais obtidos pelo sensor AVIRIS.

No quinto capítulo são apresentados as conclusões e análise final.

A contribuição científica deste estudo, relaciona-se à utilização de métodos de regularização da matriz covariância, originalmente propostos por Friedman [FRI 89] para classificação de dados em alta dimensionalidade (dados sintéticos, dados de enologia), para o caso específico de dados de sensoriamento remoto em alta dimensionalidade (imagens hiperespectrais).

A conclusão principal desta dissertação é que o método RDA é útil no processo de classificação de imagens com dados em alta dimensionalidade e classes com características espectrais muito próximas.

ABSTRACT

The remote sensing analysis of natural scenes has been relying primarily on data collected by sensors that provide a relatively small number of spectral bands. In most of the cases, this low dimensional image data has proved capable of performing image data classification in an acceptable way. There are some cases, however, in which some or all of the classes involved are spectrally very similar, i.e., their mean vectors are nearly identical. In these cases, the low dimensional image data yield a very low classification accuracy. This problem may be solved by using high dimensional image data. It is well known that high dimensional image data allows for the separation of classes that are spectrally very similar, provided that their covariance matrices differ significantly. One problem with high dimensional data, however, is related to the estimation of the required parameters, specially the class covariance matrices. As the data dimension increases, so does the number of parameters to be estimated. In real world conditions, however, the sample size normally available for parameter estimation is limited, resulting in poor estimates for the parameters. This problem becomes apparent when one compares the classification accuracy against the data dimensionality. Initially the accuracy of the classifier tends to increase as the number of spectral bands increase, i.e., as the data dimensionality becomes larger. Eventually the accuracy peaks and as the data dimensionality continues to increase the classification accuracy tends to decrease. This pattern, known as the "Hughes Phenomenon" is caused by the gradual deterioration of the parameters estimation as the data dimensionality keeps increasing while the number of available training samples remains constant. One possible way to mitigate this problem consists in the procedure known as regularization of the covariance matrix. This procedure can be performed in two steps: (1) replacing the class covariance matrix by a linear combination of the class covariance matrix and the common covariance matrix for all classes involved in the process and (2) scaling the class covariance matrices to counter the bias that tend to estimate the small eigenvalues too small and the large eigenvalues too large. This bias occur whenever the number of training samples is too small compared with the data dimensionality. This study is concerned with the applications of the regularization techniques to remote sensing high dimensional image data such as the data provided by the AVIRIS sensor system. Methods using the conventional quadratic classifier (QDA) and the classifier implementing the regularized covariance matrices (RDA) are tested and compared. The experiments have shown that the regularization techniques were very efficient in mitigating the Hughes phenomenon. In some experiments, the maximum accuracy was obtained using the conventional quadratic classifier (QDA), with a smaller number of spectral bands. In other cases the regularized procedure (RDA) produced more accurate results at higher dimensionality having mitigated the Hughes phenomenon. The investigation of the behavior of the regularization procedure (RDA) in remote sensing image data constitutes the main contribution of this study. The regularization procedure, originally proposed by Friedman is shown to perform well in remote sensing digital image classification. The regularization procedure (RDA) seems to be capable of improving the accuracy of the classification procedure when high dimensional image data is used.

1 INTRODUÇÃO

1.1 Conceitos básicos

Sensoriamento remoto tem por objetivo estudar objetos ou regiões na superfície da Terra sem entrar diretamente em contato com os mesmos. Estes objetos de interesse, vegetação, culturas agrícolas, solos, formações rochosas, corpos de água, além de outros, são em seu conjunto genericamente denominados de *alvos*. Estas informações de interesse são obtidas basicamente através do estudo da interação da radiação eletromagnética emitida por fontes naturais (Sol) ou artificiais (Radar) com estes alvos.

A principal função de um sensor remoto consiste em captar e medir a quantidade de energia refletida e/ou emitida por alvos, e desta forma obter informações sobre a natureza e/ou condições destes alvos (culturas agrícolas, florestas, rochas, solos, corpos d'água, etc). Da radiação eletromagnética que é refletida pela superfície da Terra, apenas uma porção do espectro eletromagnético é captada pelo sensores.

Segundo a literatura, a palavra *espectro* (do latim "spectrum", que significa fantasma ou aparição) foi usada por Isaac Newton, no século XVII, para descrever a faixa de cores que apareceu, quando numa experiência a luz do Sol atravessou um prisma de vidro em sua trajetória. Atualmente chama-se *espectro eletromagnético* à faixa de frequências e respectivos comprimentos de ondas que caracterizam os diversos tipos de ondas eletromagnéticas. [PRÄ 01]

As principais faixas do espectro eletromagnético (visível, infravermelho próximo, infravermelho médio, infravermelho emissivo, etc) estão descritas e representados na figura 1.

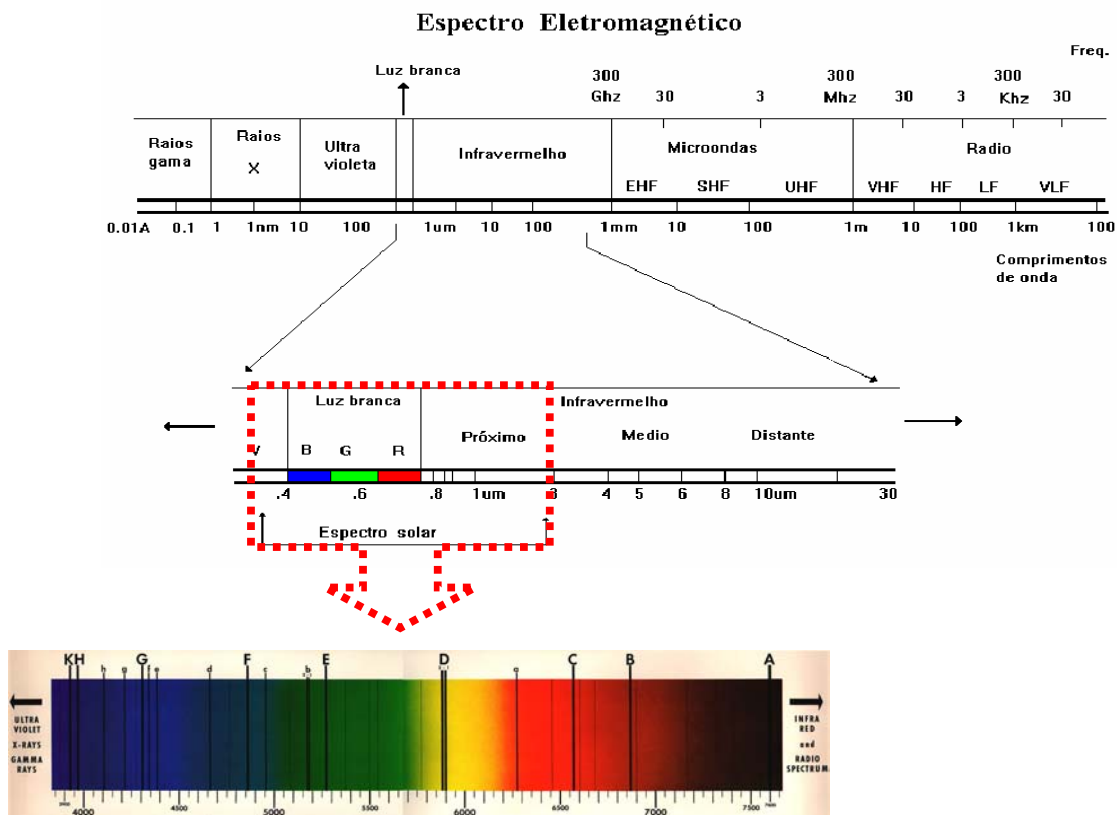


Fig.1 - Espectro Eletromagnético - Espectro Solar

Com a tecnologia atualmente disponível, pode-se gerar ou detectar a radiação eletromagnética numa extensa faixa de frequência, que se estende de 1 a 1024 Hz, ou comprimentos de onda na faixa de 108 metros a 0.01A. [INP 01]

Em sensoriamento remoto é costume definir faixas ou porções do espectro eletromagnético, compreendendo o conjunto de frequências (ou comprimento de ondas) nos quais cada sensor individual opera. Estas faixas, em Sensoriamento Remoto são denominadas de *bandas espectrais*.

0.45 μm	0.52 μm	Azul	banda 1
0.52 μm	0.60 μm	Verde	banda 2
0.63 μm	0.69 μm	Vermelho	banda 3
0.76 μm	0.90 μm	Infra-Vermelho-Próximo	banda 4
1.55 μm	1.75 μm	Infra-Vermelho-Médio	banda 5
2.08 μm	2.35 μm	Infra-Vermelho-Médio	banda 7
10.4 μm	12.5 μm	Infra-Vermelho-Termal	banda 6

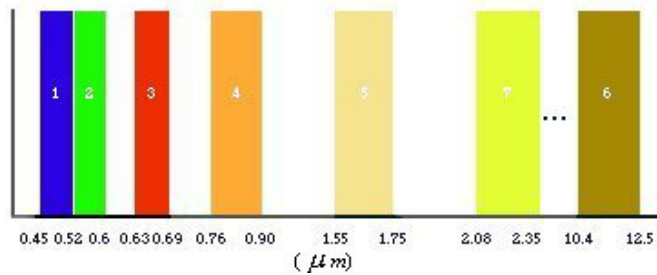


Fig. 2 - Bandas espectrais sistema LandSat-TM - (μm)

A figura 2, ilustra as diferentes bandas espectrais associadas ao sistema sensor LandSat-TM. Este sistema sensor possui um total de sete bandas, com numeração de 1 a 7, sendo que cada banda representa uma faixa do espectro eletromagnético captada pelo sensor a bordo do satélite.

Quanto maior o número de bandas, e consequentemente mais estreitas elas forem individualmente, mais informações se tem sobre a distribuição da energia refletida pela cena e melhor será a resolução espectral, isto é, a *curva de resposta espectral* de cada alvo pode ser conhecida em maiores detalhes.

O conceito de *resolução espectral* de um determinado sistema sensor, está ligado ao número de bandas espectrais e a largura individual de cada banda espectral. Um sistema com alta resolução espectral (sensor AVIRIS, por exemplo), tipicamente contém um número elevado de bandas espectrais, as quais individualmente abrangem um intervalo estreito de comprimentos de onda.

A informação espectral de uma cena pode ser representada em cada banda individual por uma imagem em formato raster, na qual a posição espacial de cada pixel é definida por duas coordenadas “x” e “y” ((linha,coluna) respectivamente) e uma terceira “z” que representa a radiância emitida ou refletida na região abrangida pelo pixel. No caso de imagens multiespectrais, a variável “z” passa a ser multidimensional, assumindo a forma de um vetor, no qual cada componente está associado a uma banda em particular. A representação gráfica deste vetor, pode então, ser entendida como uma visualização aproximada da curva de resposta espectral do pixel em questão. Cada alvo em particular pode ser caracterizado por uma curva de resposta espectral própria.

Todos os alvos *absorvem, transmitem, refletem e emitem* seletivamente radiação eletromagnética.[STE 93] Através destas características, é possível discriminar espectralmente alvos distintos, segundo o comportamento destes fenômenos na interação do alvo com a radiação eletromagnética. Por exemplo, para medir a radiância refletida por uma vegetação verde, nas faixas (ou bandas) espectrais correspondentes ao azul, verde, vermelho (visível) e infravermelho próximo, gera-se um gráfico como o da figura 3, onde estão representados em colunas, os valores médios da reflectância da vegetação nessas 4 bandas (ou faixas).

Conclui-se que:

- a) o alvo reflete com mais intensidade a componente de radiação **verde e infravermelho;**
- b) e de forma menos intensa as componentes **azul e vermelha** (absorvidas no processo de fotossíntese pela clorofila).

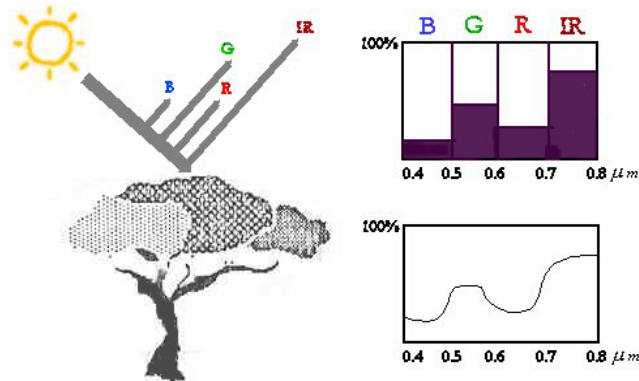


Fig. 3 - Reflectância do alvo “árvore” (Blue, Green, Red, InfraRed)

A representação destas faixas uma ao lado da outra, resulta na, já citada, *curva de resposta espectral* aproximada do alvo, que representa a interação da radiação eletromagnética com a vegetação, e neste exemplo, se caracteriza pela maior absorção na região do visível e pela maior reflectância da região do infravermelho. A figura 4 ilustra, as curvas de resposta espectral padrão para os três principais tipos de alvos naturais: solo, água e vegetação, bem como a localização das bandas espectrais disponíveis nos sistemas: LandSat-MSS, LandSat-TM e Spot.

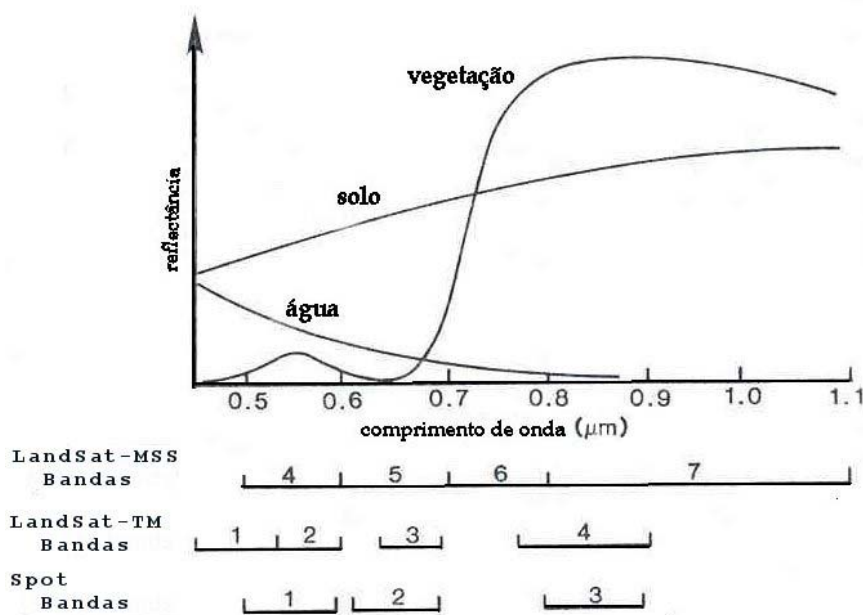


Fig. 4 – Curvas de resposta espectral padrão do solo, da vegetação e da água

Neste contexto, nota-se que as curvas são totalmente distintas, o que facilita na separação das classes. No mundo real, entretanto, ocorrem com frequência alvos cujas curvas de resposta espectral são muito semelhantes (classes diferentes de culturas agrícolas, por exemplo). Neste caso, a separação destas classes requer um detalhamento maior na curva de resposta espectral, isto é, o uso de um sensor que ofereça uma resolução espectral mais alta.

Portanto, a disponibilidade de obtenção de informações mais detalhadas a respeito de grandes áreas da superfície da Terra, pelos sistemas sensores com resolução espectral mais alta, oferecem, uma alternativa rápida, eficiente e de baixo custo, quando comparada a processos tradicionais.

Um das plataformas mais utilizadas até há algum tempo atrás para transportar um sensor, e ainda importante nos dias de hoje, é o avião. Aeronaves, contudo, apresentam limitações e desvantagens, pois a área de estudo é relativamente pequena, o processo é demorado e o custo muito grande, apesar da grande vantagem de permitir uma resolução espacial alta nos dados assim coletados.

A partir da década de 1970, o desenvolvimento dos dispositivos sensores passaram a usar, principalmente, plataformas orbitais, isto é, satélites ou estações espaciais, que apresentam uma série de vantagens, tais como o custo de uma imagem, que torna-se relativamente acessível. Apesar do investimento inicial ser grande, a coleta de dados em uma forma contínua e sistemática tende a aumentar o número de usuários, o que resulta numa diminuição de custos. Ainda, a área abrangida é muito maior e também a rapidez na obtenção de dados sobre grandes áreas da superfície da Terra. A maioria destes sistemas

sensores ainda possuem resolução espacial inferior à da fotografia aérea (30 x 30 m – LandSat, por exemplo). Sensores atuais colocados em órbita, nos últimos anos, já possuem uma resolução espacial mais alta (chegando a 1 x 1 m – Ikonos, por exemplo), igualando-se praticamente à resolução espacial de uma fotografia aérea de alta altitude. Desta forma, ficam disponíveis ao usuário várias opções de resoluções espaciais, cada uma adequada à uma dada finalidade.

Nas décadas compreendidas entre os anos de 1970, 1980 e 1990, os sistemas mais usados e citados na literatura eram o LandSat-TM, Spot, NOAA, SkyLab, SIR, ATM. Na época, eram considerados sistemas avançados, pois dispunham de sensores que captavam energia refletida ou emitida, em vários intervalos de comprimentos de onda (bandas espectrais). Possuíam 3, 5, 7, 11 e até 12 bandas espectrais e uma resolução espacial (dimensão do pixel) entre 10m e 1km, o que permitia obter uma quantidade considerável de informações, que foram e ainda são úteis, tanto para fins de pesquisa quanto para uso comercial. Estes sistemas são chamados de *sensores multiespectrais*.

O produto de um sistema multiespectral, é um conjunto de imagens do mesmo objeto ou cena. Cada imagem é obtida em uma banda espectral distinta, sendo todas as bandas registradas, ou seja, abrangem exatamente a mesma região. Os dados espectrais obtidos pelos sensores multiespectrais possuem baixa dimensionalidade, e as técnicas atuais para fins de classificação podem ser consideradas satisfatórias, neste contexto.

Desde a década de 1960 até hoje, um volume grande de recursos vem sendo investido em pesquisa, visando o desenvolvimento de novos sistemas sensores (hardware)

bem como processos para a extração de informações, isto é, a utilização destes dados (desenvolvimento de novas metodologias e softwares).

As técnicas e os conhecimentos desenvolvidos na área da ciência da computação são de especial valia para sensoriamento remoto, pelo motivo de que os dados de tais sensores são melhor processados e analisados, utilizando os recursos da computação. O avanço dos processadores, memórias e condutores atuais, torna o trabalho de processamento destes dados e a implementação de novos algoritmos para análise das mesmas, mais eficiente e preciso do que nas décadas passadas, quando o processo era essencialmente manual, o que praticamente inviabilizava análises mais completas

Assim como ocorre com a área de Informática, o aperfeiçoamento na tecnologia (hardware) dos sensores remotos vem constantemente progredindo. Um dos avanços mais notáveis que ocorreram mais recentemente, refere-se a geração e utilização de dados com resolução espectral extremamente elevada. Estes dados, designados como *dados hiperespectrais* consistem de observações feitas, simultaneamente, em um número grande de bandas espectrais (224 bandas espectrais no caso do sensor AVIRIS, por exemplo). Estes dados são caracterizados, portanto, por uma alta dimensionalidade, exigindo em muitos casos, por esta razão, uma abordagem distinta daquelas tradicionalmente empregadas em dados de baixa dimensionalidade (dados LandSat, Spot, por exemplo).

Os sensores que coletam dados em vários canais, contíguos e em estreitas bandas espectrais, nas regiões do visível, infravermelho próximo e infravermelho médio são denominados de *sensores hiperespectrais*. Com o surgimento destes novos sensores hiperespectrais, que possuem um número muito maior de bandas espectrais (46, 210, 224,

256, 288), torna-se necessário o desenvolvimento de novas metodologias e técnicas para fins de classificação.

Os sensores *ultraespectrais* compartilham da mesma definição, porém com intervalos de comprimento de onda extremamente estreitos, na ordem de 1-2nm no visível e infravermelho-próximo. Já os sensores *multiespectrais* são tipicamente intervalos de largura igual a 10nm e não necessariamente contíguos [GOE 85], deixando portanto descobertas, extensas regiões do espectro.

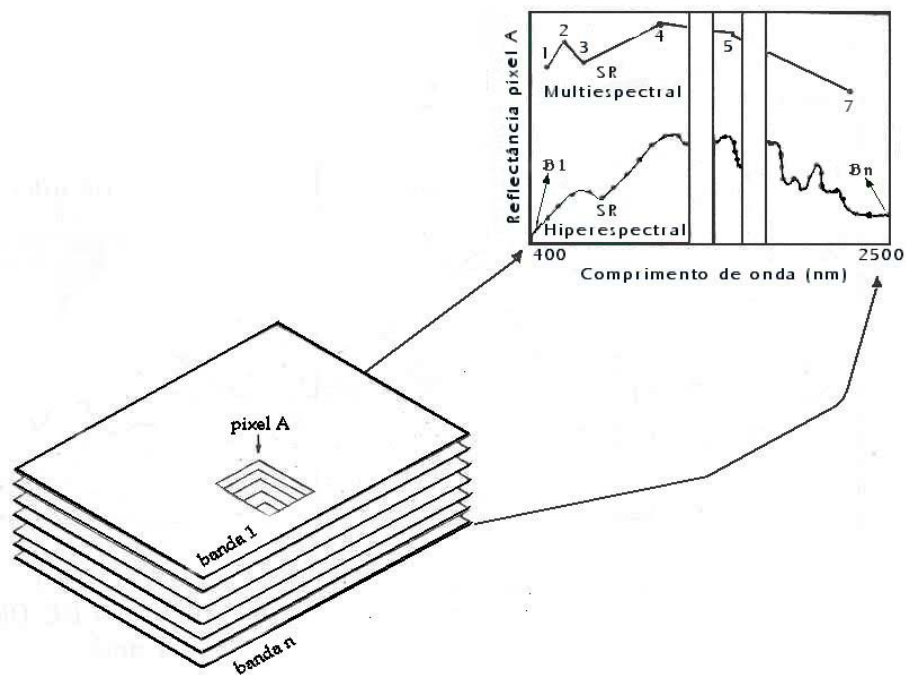


Fig. 5 – Comparação entre baixa e alta resolução espectral

Na figura 5, observa-se a curva de resposta espectral de um pixel em uma cena qualquer, geradas pelos sensores que recebem a denominação de *multiespectrais* e os denominados *hiperespectrais*. A curva de resposta espectral do pixel, é muito mais detalhada com as informações obtidas do sensor hiperespectral, do que a curva obtida de

um sensor multiespectral. Este atributo dos dados hiperespectrais permite a identificação de maiores detalhes nas características do comportamento espectral do alvo em questão.

A grande contribuição dos sensores hiperespectrais reside em sua capacidade de registrar o comportamento espectral de cada pixel em grande detalhe. Este detalhamento fino das características espectrais dos alvos terrestres, oferece ainda oportunidades interessantes em geologia, identificando picos de absorção que servem para caracterizar minerais.

Existem projetos em andamento para disponibilizar esta tecnologia hiperespectral em satélites. O que há de mais recente, neste campo de aquisição de imagens via sensores de alta resolução espectral e espacial, são os sensores hiperespectrais instalados nos satélites Orbview-4 e Orbview-3, mostrados na figura 6, os quais estão programados para serem colocados em órbita entre 2001* e 2002. Estes sensores hiperespectrais, possuem uma resolução espacial de 8m (para uso de agências governamentais norte-americanas e outros usuários habilitados) e de 20m (para os demais usuários), cobrindo um total de 200 bandas espectrais. Serão posicionados em órbita a uma altitude de 470km, e o tempo estimado de vida é de 5 anos. Além do sensor hiperespectral, os satélites possuem ainda um sensor multiespectral e um sensor pancromático com resolução de 4m e 1m, respectivamente.

* A primeira tentativa de colocar o satélite Orbview-4 em órbita fracassou (21/09/2001). Os recursos estão focados no Orbview-3 [ORB 01]

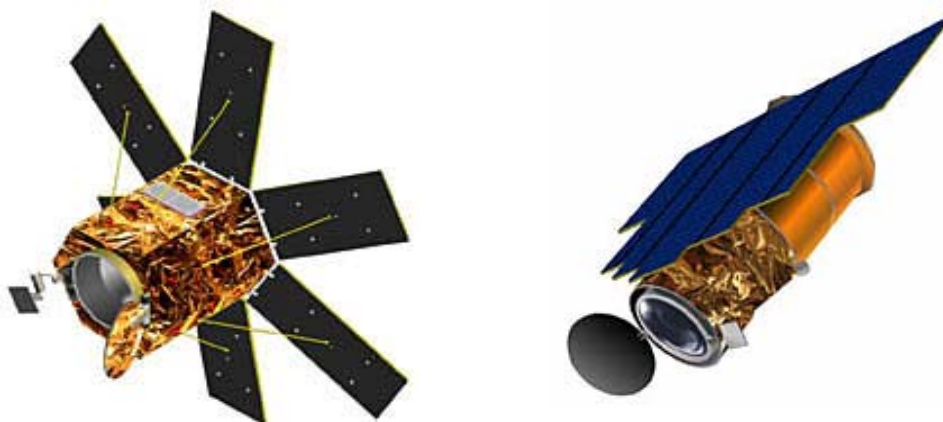


Fig. 6 - Satélite Orbview-4 e Orbview-3

1.2 O sistema sensor AVIRIS

Como ainda não existem dados em alta dimensionalidade gerados em sensores instalados em satélites para fins de pesquisa, e pela disponibilidade atual de dados hiperespectrais do sensor AVIRIS, optou-se pela utilização destes dados.

Deste modo, sendo o sensor AVIRIS um sistema experimental, o seu uso inicial é realizado através do transporte aéreo. O primeiro voo do sensor AVIRIS foi realizado em 1987. A plataforma atual do sistema AVIRIS é um avião modelo U-2, como mostra a figura 7, tendo sido modificado com a finalidade de adaptação ao sensor. Voa a aproximadamente 20 km acima do nível do mar, e a uma velocidade aproximada de 730 km/h. O sensor AVIRIS tem sido usado na América do Norte, Europa e porções da América do Sul (mais recentemente na Argentina). [INP 00]

No Brasil, já foram feitos vôos (simulações) em algumas regiões em 1995, em Cuiabá, no pantanal Mato-grossense, em Brasília, e em Porto Nacional. Algumas imagens e composições estão disponíveis no site do Inpe.[INP 00]



Fig. 7 - Avião ER-2 que transporta o sensor AVIRIS

Tendo como base a figura 8, o sistema óptico do sensor AVIRIS, consiste de um scanner, 4 espectrômetros, e uma fonte de calibração acoplada por fibra óptica. Detetores de linha vetorial são usados, necessitando que a varredura seja feita mecanicamente no modo “whisk broom”(figura 9). Existe um par de detetores vetoriais (espectrômetro / linha) para cada porção do espectro eletromagnético no visível, infravermelho próximo e dois no infravermelho médio. As fibras ópticas conduzem o sinal do scanner/frente óptica para os espectrômetros, onde o sinal é dividido em bandas discretas registradas pelos detetores. O filtro de Silicon (Si) é usado para o visível, um de Indium-Antimonide (InSb) para o infravermelho próximo, e um filtro de Fluoride é usado para a região infravermelho médio. Os detetores de Silicon possuem 32 elementos de linha vetorial na porção do visível, e as outras 3 regiões do espectro são registradas por 64 elementos de linha vetorial por detetores de Indium Antimonide.

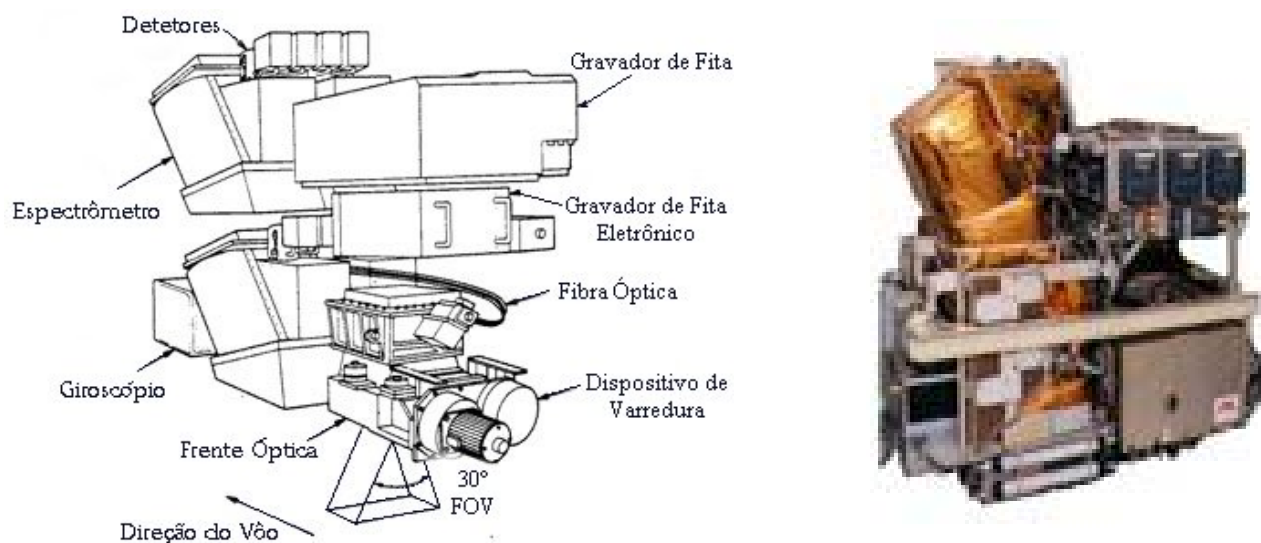


Fig. 8 - Sensor AVIRIS

Os instrumentos do sensor AVIRIS contêm 224 detetores diferentes, cada um cobrindo um intervalo de comprimento de onda (banda espectral), de aproximadamente 10nm, permitindo uma cobertura completa entre 400nm e 2500nm.

O conjunto das informações coletadas pelos detetores permite a construção de curvas de resposta espectral, que, comparadas com dados de campo ou de laboratório, permitem a identificação dos alvos.

O sensor AVIRIS usa um espelho “scanner” que varre no sentido longitudinal (modo “Whisk Broom”), produzindo 614 pixels para cada varredura dos 224 detetores. Cada pixel cobre, aproximadamente, 20x20m de área do terreno (com alguma sobreposição entre os pixels), permitindo uma cobertura total de aproximadamente 11 Km de extensão, no sentido da varredura.

Os dados obtidos pelo instrumento são gravados pelo sistema juntamente com os dados de navegação, engenharia e as leituras do calibrador do sensor. Estes dados, processados e gravados, ocupam 140 Mbytes para cada 512 varreduras (ou linhas) de dados. Cada 512 linhas de conjunto de dados é denominado de *cena*. A tabela 1 mostra as características técnicas do sensor AVIRIS.

Tabela 1 – Ficha técnica do sensor AVIRIS

Taxa de dados em 1994, 17 Mbps, a partir de 1995 passou para 20.4 Mbps
Pacote de dados em 1994, de 10 bits, a partir de 1995, passou para 12 bits.
Detetores para os λ 's da faixa do visível da Silicon (Si), e detetores de Indium-Antimonide (InSb) para o infravermelho-próximo.
Varredura do tipo "Whisk Broom" (Fig. 9)
Taxa de varredura de 12 Hz
Detetores frios ("cooled") de Liquid Nitrogen (LN2)
Canais de bandas nominais de 10 nm, calibradas para 1 nm
Campo de visada total de 30° graus (completo 614 amostras)
Campo de visada instantâneo de 1 mrad (IFOV, 1 amostra), calibrados para 0.1 mrad
Fita cassete média para gravação S-VHS ST-120



Fig. 9 - Método de varredura "Whisk Broom"

Dada a sua alta resolução espectral, o sistema AVIRIS é capaz de reproduzir com muito boa aproximação as curvas de resposta espectral de vários tipos de alvos (Fig.10), permitindo uma boa caracterização tanto dos picos de absorção como dos de reflectância. Estes picos são essenciais para caracterizar certos tipos de alvo (minerais em geologia, por exemplo). O conceito do sistema AVIRIS pode ser observado na figura 10 .

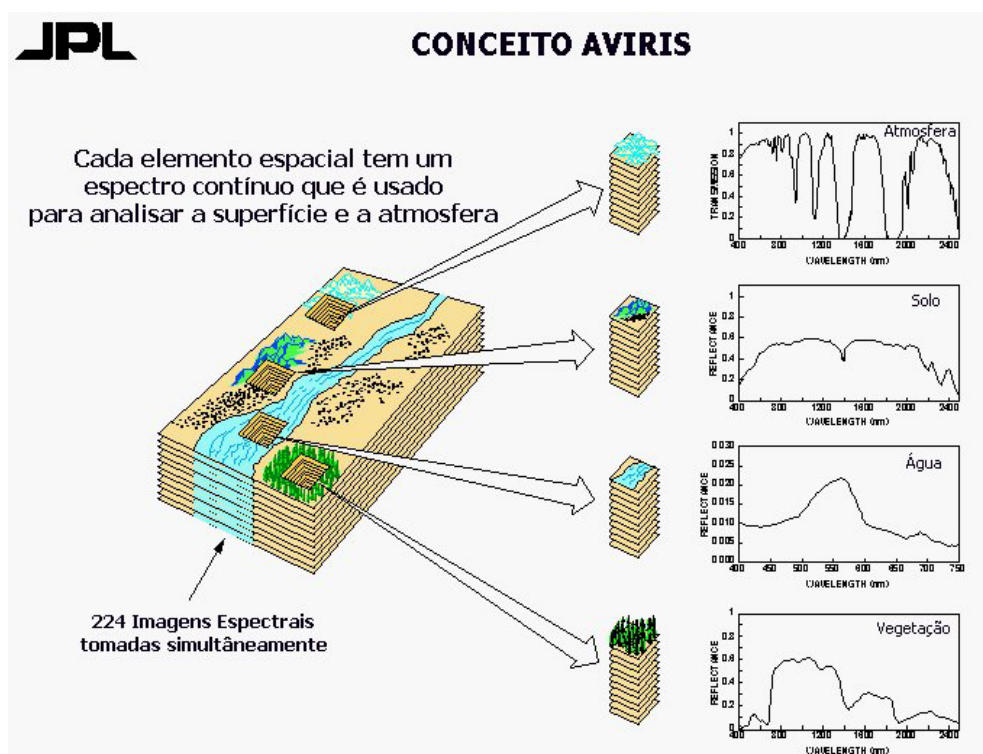


Fig. 10 - Conceito AVIRIS

A inadequacidade dos dados fornecidos por sensores convencionais, quando comparados com a informação espectral apresentada pelos sensores hiperespectrais, é ilustrada nas figuras 11 e 12, onde são apresentados três tipos de minerais (kaolinita, Dolomita, Hematita). Cada um destes minerais possui uma molécula distinta que absorve energia em regiões também distintas do espectro eletromagnético.

Comparando as curvas de resposta espectral das amostras num sensor multiespectral LANDSAT-TM (figura 11) e num sensor hiperespectral AVIRIS (figura 12), é possível detectar a presença destes minerais segundo seus picos de absorção apenas no sensor hiperespectral. Já no sensor multiespectral os três minerais não podem ser identificados, pois as informações se confundem e nenhuma concentração relativa pode ser determinada.

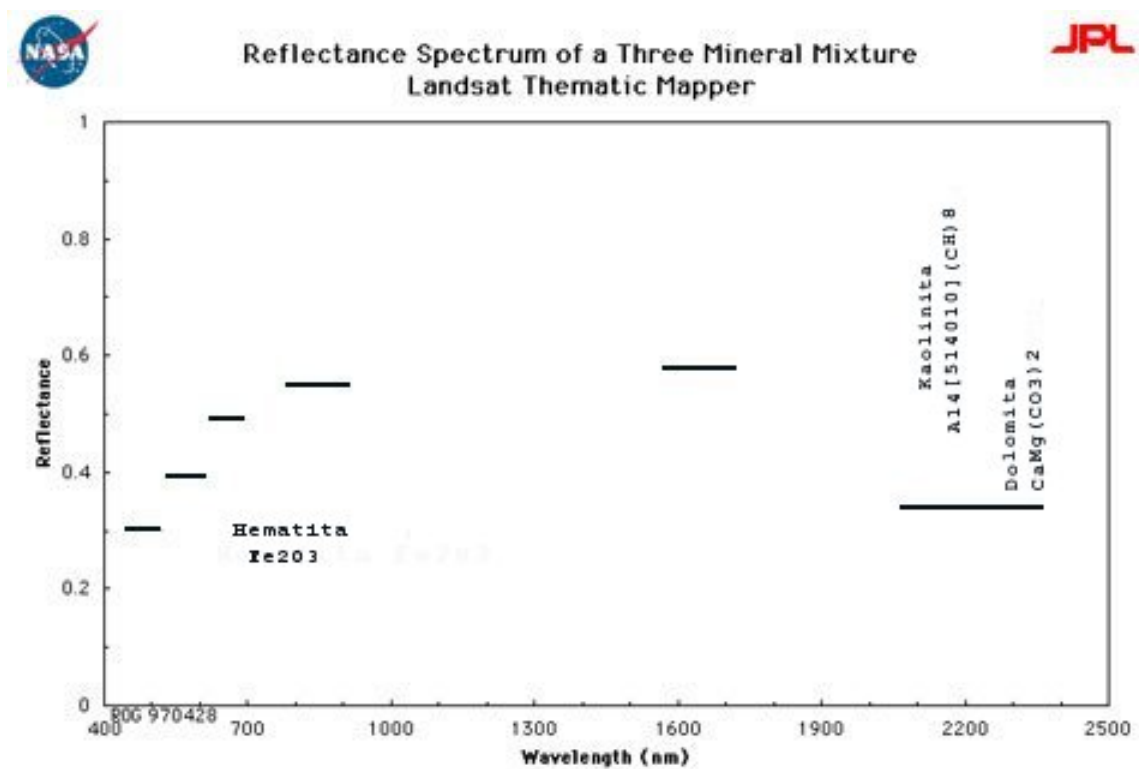


Fig. 11 – Espectro de reflectância de 3 minerais diferentes(Hematita, Kaolinita e Dolomita), segundo o sensor LandSat- TM

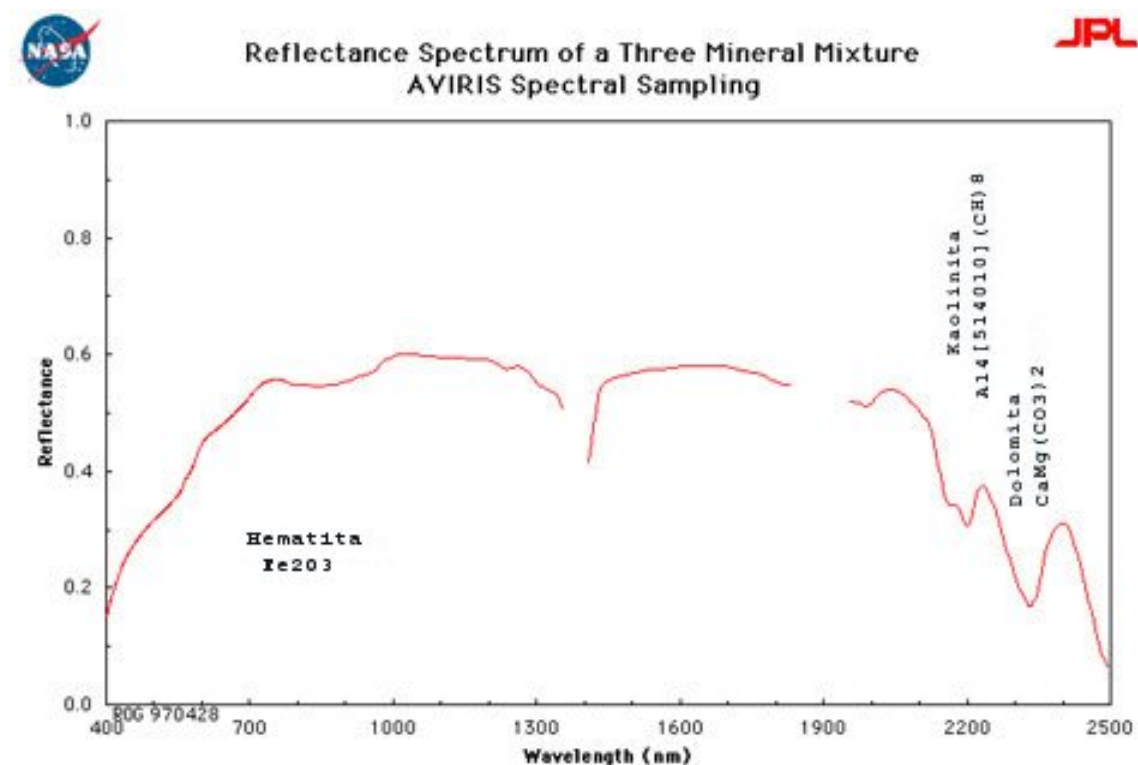


Fig. 12 - Espectro de reflectância de 3 minerais diferentes(Hematita, Kaolinita e Dolomita), segundo sensor AVIRIS

Uma forma interessante de visualizar as informações espectrais das cenas é a chamada *imagem cubo* ou *cubo hiperespectral* gerada pelo sensor AVIRIS, que ilustra a quantidade de informações (volume de dados) que retorna do instrumento. O exemplo da figura 13 é datado em 20 de agosto de 1992, quando foi feito um vôo no avião da NASA ER-2 (figura 7) a uma altitude de 20.000 m (65.000 pés) sobre a região de Moffett Field, Califórnia, a parte final Sul da Baía de São Francisco.

Descrevendo a imagem gerada, na parte superior do cubo foi usada uma imagem “falsa-cor”, feita para acentuar a estrutura da água e dos lagos à direita (água salgada de cultivo de camarão) . Também está visível na imagem o Aeroporto de Moffett Field.

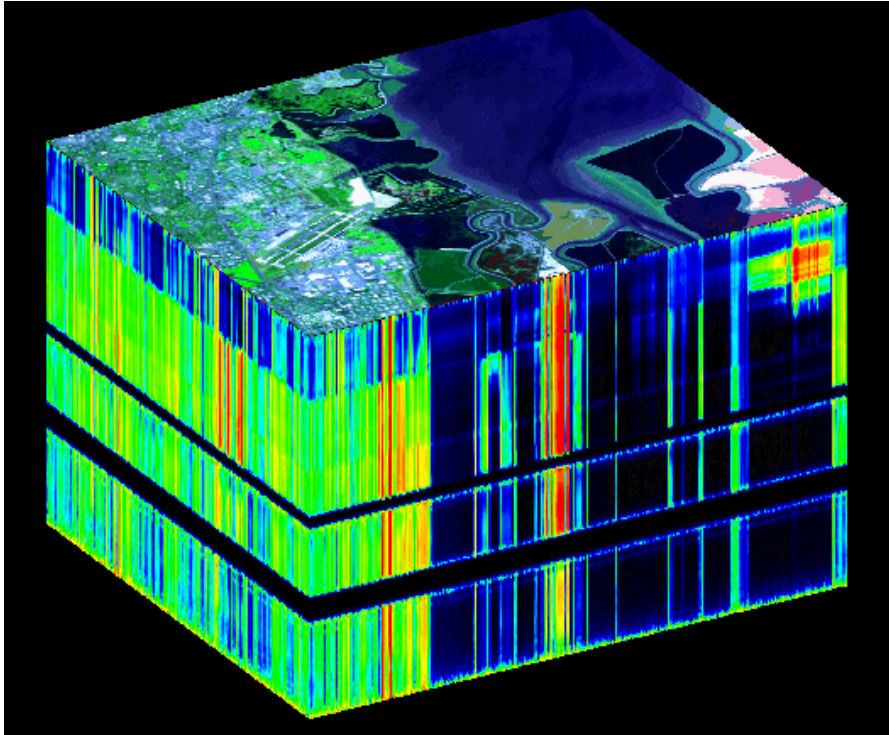


Fig. 13 - Imagem Cubo (Moffett Field)

Os lados do cubo mostram os limites das 224 bandas espectrais do sensor AVIRIS. Nas laterais, na parte superior cubo, é a parte visível do espectro (comprimento de onda de 400nm), e a parte inferior pertence ao Infravermelho médio (comprimento de onda de 2.500nm). São usadas pseudo-cores, entre o preto e o azul (baixa resposta, isto é, muita absorção e pouca reflectância) até o vermelho (alta resposta, isto é, pouca absorção e muita reflectância). As regiões que apresentam duas falhas no cubo são as regiões de absorção da atmosfera; sendo assim, correspondem aos intervalos de comprimento de onda que não chegam ao sensor, consistindo de apenas ruído.

1.3 Reconhecimento de padrões

A utilização dos dados coletados por sistemas sensores normalmente requer o emprego de técnicas de *reconhecimento de padrões*.

Reconhecimento de padrões tem como objetivo classificar objetos de interesse em uma dada classe ou categoria. Os objetos de interesse são genericamente denominados de *padrões* e podem ser caracteres, gráficos, células em biologia, sinais eletrônicos ou qualquer outro objeto que se deseje classificar em uma classe, entre várias classes disponíveis. [THE 89]

Em sensoriamento remoto, normalmente “padrão” vem a ser um pixel individual a ser atribuído a uma das classes definidas pelo analista. Neste contexto, um pixel é representado por um vetor cuja dimensão é igual ao número de bandas espectrais adotadas.

O desenvolvimento de sistemas de reconhecimento de padrões, visualizado na figura 14, essencialmente envolve três aspectos: 1) aquisição de dados através de dispositivos de captação dos dados; 2) representação dos dados na forma de um vetor x ; 3) transformações necessárias e tomada de decisão.

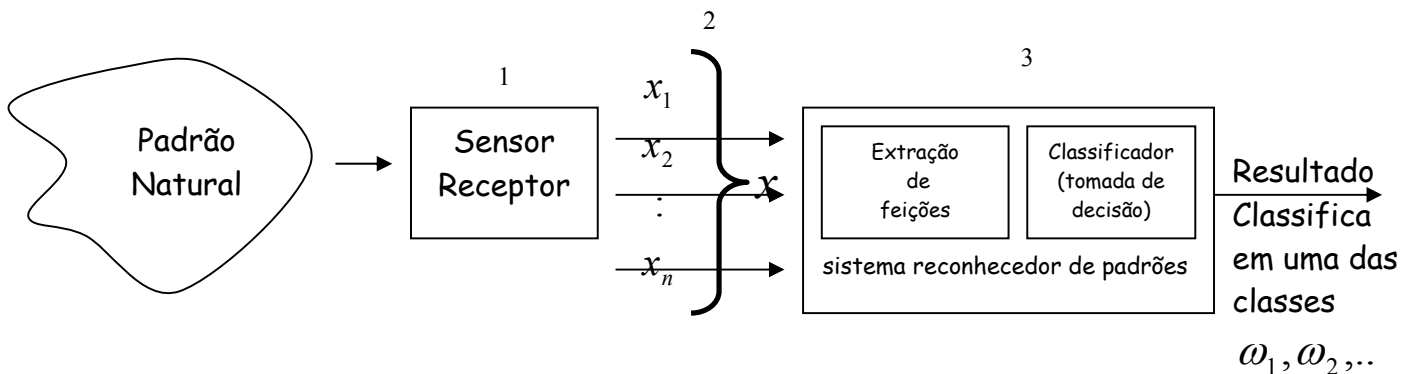


Fig 14 - Esquema básico para reconhecimento de padrões

O domínio do problema dita a escolha do sensor, técnica de pré-processamento, esquema de representação, e o modelo de tomada de decisão [JAI 00].

O aprendizado a partir de um conjunto de amostras (amostras de treinamento) é uma etapa inicial na maioria dos sistemas de reconhecimento de padrões. Os métodos desenvolvidos para fins de reconhecimento de padrões, isto é, classificação de dados em classes previamente estabelecidas, têm sido objeto de estudo por um número grande de pesquisadores. Um levantamento teórico dos vários métodos, hoje em utilização, pode ser encontrado em [JAI 00].

Diferentes técnicas de classificação são usadas dependendo do tipo do modelo escolhido [DUD 01]. Os métodos de classificação podem ser agrupados em duas grandes categorias: métodos supervisionados e métodos não-supervisionados.

Nos métodos da classificação supervisionada, as classes são previamente definidas pelo analista, isto é, definidas ou caracterizadas através das amostras de treinamento. Cada classe pode então ser caracterizada por uma *função decisão* que pode ser: (1) de natureza probabilística ou (2) de natureza determinística.

No primeiro caso, os vetores pixels são considerados variáveis aleatórias e a função decisão fica então associada a uma função densidade de probabilidade que descreve a distribuição dos dados. Os parâmetros são estimados a partir das amostras de treinamento disponíveis. No segundo caso, a função decisão não tem a forma de uma função de probabilidade. Um exemplo determinístico seria uma função polinomial. Neste caso, os coeficientes do polinômio são calculados a partir das amostras de treinamento.

Os passos básicos, para qualquer caso de classificação supervisionada, são:

- 1º. Identificar as classes existentes;
- 2º. Definir a forma geral da função decisão a ser adotada;
- 3º. Selecionar os pixels representativos de cada classe, que serão chamados de *amostras de treinamento*;
- 4º. Estimar os parâmetros ou calcular os coeficientes específicos da função decisão a ser usada;
- 5º. Usando o classificador treinado, classificam-se os pixels presentes na imagem, numa das classes anteriormente identificadas, produzindo-se uma imagem temática que abrange o resultado da classificação, e/ou faz-se uma análise numérica quantitativa dos resultados;

Os métodos não-supervisionados oferecem um outro tipo de abordagem. Em alguns casos, existem problemas na área de reconhecimento de padrões, nas quais a natureza (ou definição) das classes, e mesmo o número de classes presentes são desconhecidos. Neste caso, o problema a ser tratado consiste não somente na classificação propriamente dita, mas também, na própria definição das classes. Ao contrário do método supervisionado, onde tem-se um conhecimento prévio das classes, os métodos não-supervisionados atribuem à técnica ou ao algoritmo escolhido, a tarefa de identificar, por si só, as classes existentes num conjunto de dados.

Uma possível abordagem para resolver um problema desta natureza, freqüentemente utilizada em sensoriamento remoto, consiste em procurar aglomerados (*clusters*) de pontos (*pixels*) no espaço multiespectral. Estes aglomerados são formados (ou definidos) por um

grupo de pontos que podem ser considerados como semelhantes entre si, de acordo com um determinado critério (distância euclidiana no espaço espectral entre pixels componentes, por exemplo). Este processo é denominado de *clustering* ou análise de aglomerados.

Os processos de classificação podem também ser agrupados em duas grandes categorias: métodos paramétricos e métodos não-paramétricos. Nos métodos paramétricos, a forma geral da função densidade de probabilidade que descreve o comportamento dos dados é suposta conhecida. Os parâmetros existentes são estimados a partir das amostras de treinamento disponíveis para cada classe.

Ocorrem casos em que a forma geral da função densidade probabilidade associada aos dados não é conhecida. A alternativa nestes casos consiste na utilização dos chamados métodos não-paramétricos. Como exemplo de métodos não-paramétricos pode-se citar o método ou estimação de Parzen e o método do vizinho mais próximo [THE 89].

O método, dentro da abordagem estatística, mais utilizado em sensoriamento remoto, é o da Máxima Verossimilhança Gaussiana, que é um método supervisionado e paramétrico. É considerado paramétrico, pois o método envolve parâmetros (vetor média e matriz covariância) da distribuição gaussiana multivariada, e é supervisionado, pois estima estes parâmetros através das amostras de treinamento.

Contudo, uma das maiores dificuldades ocorre na estimação dos parâmetros, em especial a matriz covariância. Em dados de alta dimensionalidade, esta estimação torna-se extremamente problemática, devido ao fato de que em situações reais o tamanho das amostras disponíveis nunca é adequado. A alta dimensionalidade dos dados se constitui portanto simultaneamente em um benefício e em um problema. Esta nova situação vem sendo objeto de intenso interesse por parte da comunidade científica internacional.

O que se verifica na prática é que o aumento na dimensionalidade dos dados, isto é, do número de bandas espectrais, resulta inicialmente em um incremento na acurácia da resultante imagem classificada. Este fato seria de se esperar, pois com o aumento do número de bandas, ocorre um incremento na informação disponível. A partir de um certo ponto, entretanto, e utilizando-se as mesmas amostras de treinamento (para fins de estimação dos parâmetros do classificador), a acurácia começa a decrescer com o aumento da dimensionalidade dos dados. Este efeito, conhecido como o *fenômeno de Hughes* ou a *maldição da dimensionalidade (the curse of dimensionality)* deve-se ao aumento do número de parâmetros a serem estimados, especialmente na matriz covariância. Atualmente, pesquisam-se novas formas de minimizar este efeito. Os estudos da utilização de técnicas estatísticas têm envolvido uma gama de estudiosos com o intuito de desenvolver novas ferramentas para manusear estas informações.

Esta dissertação propõe-se a analisar uma das metodologias propostas com a finalidade de diminuir o efeito do fenômeno de Hughes em dados de sensoriamento remoto. As técnicas de regularização da matriz covariância, propostas em diferentes modalidades por autores distintos, oferecem a possibilidade de obter-se uma estimativa mais confiável para a matriz covariância associada a cada classe presente na imagem. Consequentemente, os resultados fornecidos por classificadores que fazem uso deste parâmetro (Máxima Verossimilhança Gaussiana, Distância de Mahalanobis, por exemplo), tornam-se menos vulneráveis aos problemas de estimação envolvendo dados em alta dimensionalidade, permitindo desta forma obter as vantagens potenciais (aumento de informação) proporcionadas pelo maior número de bandas espectrais e minimizando os aspectos negativos associadas ao fenômeno de Hughes.

Nesta dissertação desenvolve-se portanto, uma série de experimentos visando avaliar o efeito das técnicas de regularização da matriz covariância em dados de sensoriamento remoto.

A análise proposta nesta dissertação requer o desenvolvimento de softwares específicos. A elaboração deste software resultou em produto adicional desta dissertação e encontra-se descrito no ANEXO 4.

2 REVISÃO BIBLIOGRÁFICA

2.1 Problema da dimensionalidade

Conforme citado anteriormente, no contexto do sensoriamento remoto, cada pixel em uma cena é representado por um vetor com dimensão igual ao número de bandas espectrais. Deste modo, a dimensionalidade dos dados fica definida pelo número de bandas empregadas. Logo, dados fornecidos por sensores do tipo hiperespectral, caracterizam-se por uma dimensionalidade muito alta.

2.1.1 *Alta dimensionalidade*

Lewotsky relata em seu estudo, que o espectro da reflectância da maioria dos materiais da superfície terrestre apresenta feições ou picos de absorção. O fato destas novas tecnologias de sensores remotos serem capazes de adquirir o espectro detalhado de reflectância de uma determinada área, torna-a uma ferramenta poderosa para o estudo da Terra e do seu meio-ambiente [LEW 94].

Lewotsky apresenta a evolução do imageamento através da espectrometria, sendo este um dos campos do sensoriamento remoto que mais se desenvolveu nas últimas décadas. *Espectrometria* constitui-se em uma ferramenta com amplas possibilidades de aplicação em estudos ambientais. O conceito apresentado por Lewotsky é que a espectrometria pode ser entendida como o imageamento de uma cena com a utilização de um grande número de bandas espectrais discretas e contíguas, onde o espectro bem detalhado da reflectância pode ser obtido. Este tipo de imageamento é também conhecido como *imageamento hiperespectral* [LEW 94].

Xiuping Jia e John A. Richards [JIA 99], na parte introdutória de seu estudo, salientam que a alta resolução espectral que está se tornando disponível com os espectrômetros hiperespectrais, gera uma rica quantidade de informações com respeito à reflectância de alvos na superfície terrestre. Esta quantidade enorme de dados gerados causa, entretanto, alguns novos problemas no processo de análise e visualização da imagem, devido à quantidade de bandas como, por exemplo, o sensor AVIRIS, que gera 224 informações para cada pixel da imagem. O tratamento deste tipo de informação requer técnicas adequadas.

2.1.2 Fenômeno de Hughes, a “maldição da dimensionalidade”

Dados em alta dimensionalidade oferecem novas possibilidades no processo de classificação de imagens. Classes com comportamento espectral muito semelhante, e portanto praticamente indistinguíveis quando se emprega dados tradicionais em baixa dimensionalidade (LandSat, por exemplo), podem passar a ser distinguíveis ou separáveis, com alto grau de acurácia, quando se dispõe de dados em alta dimensionalidade [FUK 90].

A alta dimensionalidade dos dados apresenta, entretanto, novos desafios ao processo de classificação. Possivelmente, o maior desafio neste caso consiste na estimação de parâmetros associados a classificadores estatísticos. O número de parâmetros a serem estimados, na matriz covariância particularmente, cresce rapidamente na medida em que a dimensionalidade dos dados aumenta, como ilustrado na figura 15 e na tabela 2. Se o número das amostras de treinamento não aumenta, a confiabilidade na estimação destes parâmetros decresce. Em outras palavras, quando o tamanho das amostras de treinamento

permanece constante, o acréscimo de bandas espectrais adicionais resulta no aumento da variância, isto é, da incerteza da estimativa dos parâmetros.

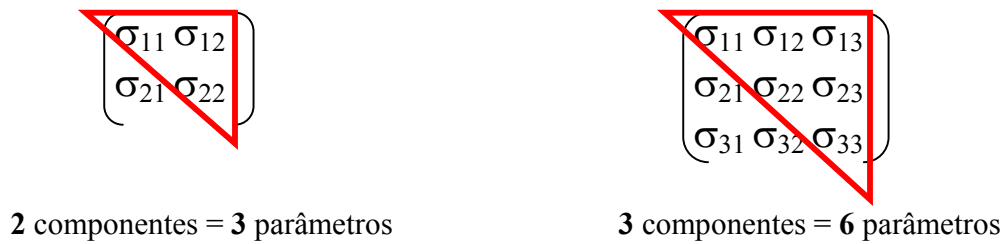


Fig. 15 – Número de parâmetros a serem estimados na matriz covariância

Aplicando a fórmula $np = (n + \frac{n^2 - n}{2})$, onde np é o número total de parâmetros à serem estimados e n a dimensionalidade dos dados, é possível calcular quantos parâmetros são necessários estimar para cada caso específico. Comparando-se um sistema sensor tradicional como o LandSat-TM, usando 6 bandas espectrais, com, por exemplo, 185 bandas do sistema sensor AVIRIS, calcula-se a quantidade de parâmetros necessários para cada caso, conforme mostrado na tabela 2.

Tab. 2 - Parâmetros necessários para o sensor LandSat-TM e o sensor AVIRIS

LANDSAT-TM	AVIRIS
$n = 6$	$n = 185$
$np = (6 + \frac{6^2 - 6}{2})$	$np = (185 + \frac{185^2 - 185}{2})$
$np = 21$	$np = 17205$

Nota-se que no uso do sistema AVIRIS, é necessário estimar 820 vezes mais parâmetros, do que a quantidade de parâmetros necessários no uso de sistemas convencionais como o LandSat-TM.

O fato do aumento do número de parâmetros a serem estimados ocasionar a degradação da acurácia é responsável pelo efeito conhecido como *fenômeno de Hughes* ou a *maldição da dimensionalidade* [TAD 98]. O efeito, demonstrado na figura 16, manifesta-se no comportamento da acurácia resultante de um processo de classificação estatística. Inicialmente, o acréscimo de novas bandas espectrais faz com que a acurácia tenda a aumentar. Em um determinado ponto, um máximo para a acurácia é atingido. A partir daí, um acréscimo adicional do número de bandas resulta em um decréscimo cada vez mais acelerado na acurácia fornecida pelo processo de classificação.

Este problema, bem como possíveis alternativas para mitigá-lo, vem sendo objeto de estudo por parte de vários pesquisadores, como Shahshahani [SHA 94], Hsieh [HSI 98], Tadjudin [TAD 98] e Cortijo F.J. [COR 99].

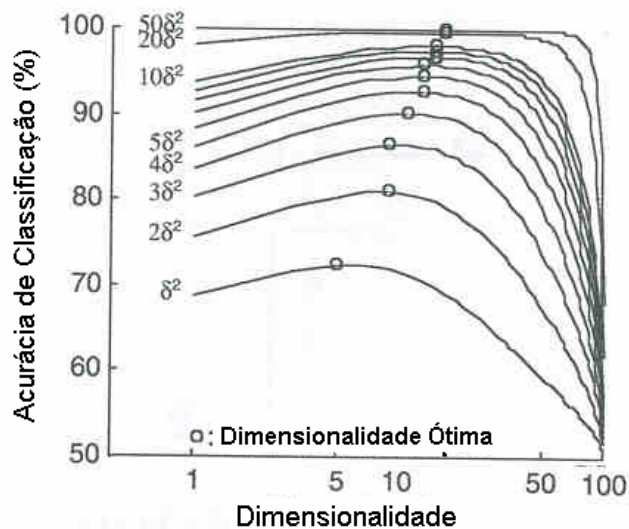


Fig. 16 - Efeito de Hughes. [HSI 98]

Fukunaga [FUK 89] provou que o número de amostras de treinamento necessárias para o classificador é uma função linear da dimensionalidade para um classificador *linear* e uma função quadrática da dimensionalidade para um classificador *quadrático*.

2.2 Métodos para mitigar o fenômeno de Hughes

O método proposto por Shahshahani [SHA 94] testa o uso do método iterativo EM (Expectation Maximization). Este método possibilita incluir no processo de estimação dos parâmetros as chamadas amostras não-rotuladas, isto é, pixels disponíveis na imagem, mas cuja classe não é conhecida. Desta forma aumentaria o número de amostras de treinamento.

Em [HAE 99] identificam-se duas abordagens principais que estão sendo ponto de referência de pesquisa:

A primeira abordagem, denominada *redução da dimensionalidade*, consiste no desenvolvimento de métodos que permitem a redução da dimensionalidade dos dados com a perda mínima de informação (ex.: Projection Pursuit, Transformada de Karhunen-Loeve (Componentes Principais), Transformada Discreta do Coseno, Hashing, e Análise Discriminante “Step Wise” para selecionar as bandas mais úteis para a discriminação) .

Estudos relacionados com técnicas que envolvem redução da dimensionalidade (Jimenez [JIM 99] e Jia [JIA 99], por exemplo) obtiveram resultados notáveis. Neste contexto, o objetivo principal é diminuir o Efeito de Hughes, reduzindo a dimensionalidade dos dados, mas preservando a quantidade máxima de informação possível e, assim, resolvendo em parte o problema do aumento da dimensionalidade e a conseqüente degradação da acurácia. Contudo, o simples fato de existir a perda de informação no

momento da redução da dimensionalidade já resulta num ponto fundamental da análise, visto que, pode ocorrer a perda de informações importantes e que poderiam ser preservadas para auxiliar na discriminação das classes de comportamento espectral muito semelhantes. Outra possível alternativa consiste na parametrização da matriz covariância, por exemplo, a partição da matriz covariância em blocos.

Uma segunda possível abordagem para este problema consiste no desenvolvimento de métodos para estimação da matriz covariância, mantendo a dimensionalidade dos dados. Este processo é conhecido pela designação genérica de *regularização da matriz covariância* e os métodos de classificação que empregam esta técnica, como *Análise Discriminante Regularizada (RDA)*. Dentre as várias técnicas para regularização da matriz covariância propostas recentemente, destacam-se as seguintes: Hoffbeck e Landgrebe [HOF 96] realizaram estudos de um novo estimador da matriz covariância, para classificadores com tamanho limitado de amostras. Através dos experimentos, os autores mostram que a matriz covariância estimada desta forma pode gerar resultados mais precisos no processo de classificação do que aqueles obtidos utilizando-se processos convencionais em um classificador tanto linear quanto quadrático.

Outras alternativas, como as propostas por Friedman [FRI 89] e Aeberhard et al [AEB 94], são importantes meios de se tentar suavizar o declínio da acurácia gerada pelo Efeito de Hughes, quando comparadas com outras técnicas. Os experimentos realizados com técnicas de regularização da matriz covariância usando diferentes tipos de dados, porém preservando a característica de alta dimensão dos dados, comprovaram que estas técnicas podem ser aplicadas para mitigar a degradação da acurácia com o aumento da dimensionalidade. No estudo de Cortijo [COR 99] foram geradas duas imagens espectrais

sintéticas com seis classes semelhantes, obtendo, com a técnica de regularização, os melhores resultados.

Não foi encontrado trabalho ou estudo algum na literatura onde fosse usada alguma técnica de regularização da matriz covariância em imagens com alvos naturais de culturas agrícolas com comportamento espectral semelhantes, provenientes diretamente de um sensor hiperespectral.

2.3 Métodos de regularização da matriz covariância

O fenômeno de Hughes coloca limites no uso prático das imagens com alta dimensionalidade, pois, como o número de amostras de treinamento disponível é limitado, o problema da estimação da matriz covariância torna-se crítico com o aumento da dimensionalidade.

Neste caso, Friedman[FRI 89] e Hoffbeck[HOF 96], além de outros, sugerem a segunda abordagem, citada no capítulo 2.2, que faz uso de *técnicas de regularização* para estimar a matriz covariância, como um objeto interessante de pesquisa.

Estudos realizados com técnicas de regularização, utilizando dados sintéticos, mostraram resultados muito satisfatórios. Dois destes estudos são utilizados neste trabalho. Inicialmente, a proposta de Friedman [FRI 89] apresenta uma análise dos métodos de Análise Discriminante Linear (LDA), Análise Discriminante Quadrática (QDA) e Análise Discriminante Regularizada (RDA). Os dados sintéticos usados nos experimentos, foram submetidos às três técnicas propostas. Uma vez aplicada as três técnicas no mesmo

conjunto de dados sintéticos, os resultados obtidos por Friedman com o método RDA foram superiores aos métodos QDA e LDA.

O outro estudo que motivou este trabalho, foi o de Aeberhard et al [AEB 93]. Neste trabalho realizaram-se comparações entre 8 métodos estatísticos de classificação (*Análise Discriminante Quadrática(QDA)*, *Análise Discriminante Linear(LDA)*, *Análise Discriminante Regularizada (RDA)*, *K-vizinho mais próximo (KNN)*, *Plano Discriminante de Fisher's(FDP)*, *Fisher-Fukunaga-Koonz (FF)*, *Plano de Fisher-Radius(FR)*, *Plano Variância de Fisher(FVP)*). Foram usados nos experimentos conjuntos de dados sintéticos e reais, e foram comparadas duas abordagens distintas: métodos que classificam usando *dimensionalidade completa* (bandas espectrais disponíveis), e os métodos que inicialmente *reduzem a dimensionalidade*. Os resultados obtidos das simulações sugerem o processo RDA (mantendo a dimensionalidade completa dos dados) como o mais eficiente classificador, e salienta que, na maioria das vezes, a redução da dimensionalidade antes da classificação aumenta o erro na classificação das amostras de teste.

Os estudos de Friedman e de Aeberhard e Coomans, apontam para a aplicabilidade e performance da Análise Discriminante Regularizada (RDA) em ambientes de alta dimensionalidade e a necessidade de desenvolver novas metodologias para o tratamento de dados provenientes de sensores hiperespectrais, como objetos de pesquisa para a comunidade científica. Estas sugestões são os principais motivos deste trabalho, que visa analisar o comportamento do classificador RDA aplicado em dados de natureza real, provenientes do sensor AVIRIS.

3 METODOLOGIA

Os dados em sensoriamento remoto são inicialmente coletados sob forma de um sinal elétrico, isto é, de natureza analógica. Este sinal é posteriormente discretizado (quantizado), gerando, para cada pixel, valores numéricos representativos da radiância refletida (ou emitida) pela região da superfície da Terra compreendida pelo pixel. O processo é visualizado na figura 17.

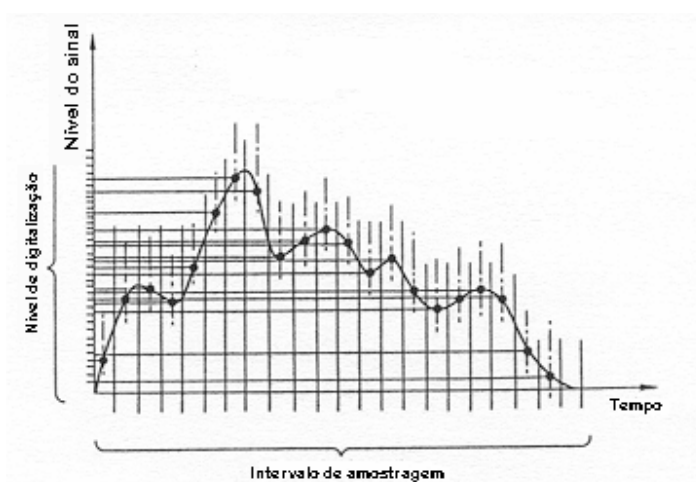


Fig. 17 - Conversão de sinal analógico para contador digital

Esse conjunto de valores numéricos discretos, denominados de contadores digitais (CD), são então organizados na forma de um vetor, onde cada elemento representa uma banda espectral.

No caso do sistema sensor AVIRIS, cada pixel é portanto representado por um vetor com dimensão igual a 224, constituindo-se, portanto, em uma variável no espaço de 224 dimensões (espaço multiespectral). Este vetor padrão X , pode ser representado como:

$$X = \begin{Bmatrix} x_1 \\ x_2 \\ \vdots \\ x_p \end{Bmatrix} \quad (1)$$

onde p é o número total de bandas espectrais. As classes podem ser representadas por

$$\omega_i, i = 1, \dots, k \quad (2)$$

onde k é o número total de classes. Idealmente, cada classe deveria se constituir em um *aglomerado* bem definido de pontos no espaço multiespectral, cada ponto representando um pixel, como mostra a figura 18, para duas dimensões.

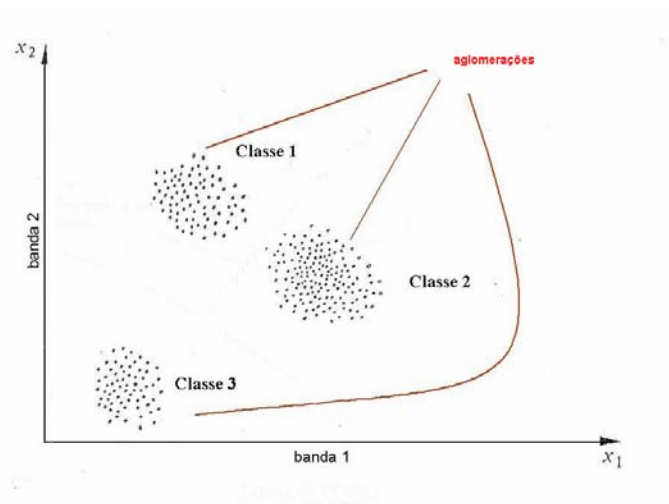


Fig. 18 - Espaço multiespectral para $n=2$

Logo, cada aglomerado está associado a uma classe diferente, e aglomerados distintos representam diferentes classes de padrões. O processo de classificação pode então ser implementado através de uma *função de decisão*, que serve para definir a classe à qual o padrão (pixel) em questão pertence. A cada classe é associada, portanto, uma função

decisão $G_i(X)$. O processo de classificação pode então ser descrito pela seguinte regra: atribui o pixel X à classe ω_i , se

$$G_i(X) > G_j(X), \quad \forall i \neq j \quad (3)$$

Conceitos bem conhecidos da teoria de decisão estatística são usados para definir a forma ou expressão geral para as funções decisão, conforme a natureza dos dados em cada caso. Também, idealmente, seria possível definir fronteiras (hiper-planos ou hiper-superfícies) entre várias classes neste espaço multiespectral, possibilitando desta forma a classificação correta de todos os pixels em uma das classes existentes, como mostra a figura 19.

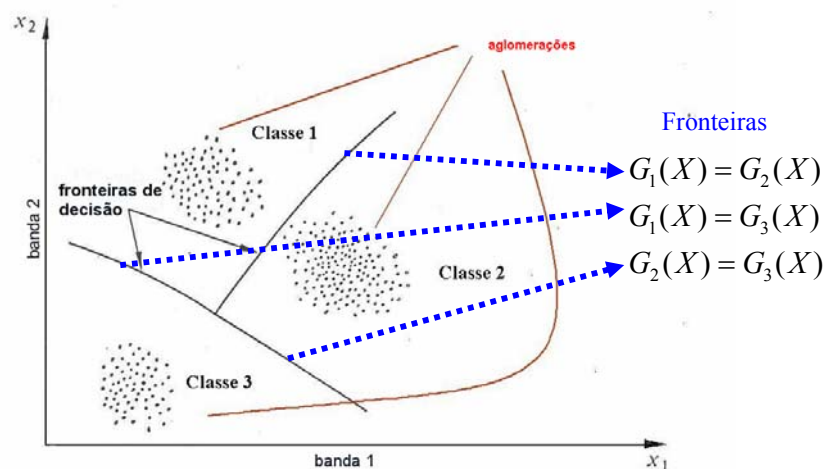


Fig. 19 - Fronteiras de decisão para $n=2$

Em situações reais ocorrem, entretanto, alguns problemas originados principalmente pela dispersão interna em cada aglomerado, conforme definido pela matriz covariância. Esta dispersão causa freqüentemente uma interpenetração entre os diferentes aglomerados (figura 20), tornando inviável o estabelecimento de fronteiras “rígidas”, separando as várias classes no espaço multiespectral. Esta situação requer, portanto, uma abordagem não

determinística ao problema da classificação, Neste caso a abordagem estatística se apresenta como uma solução.

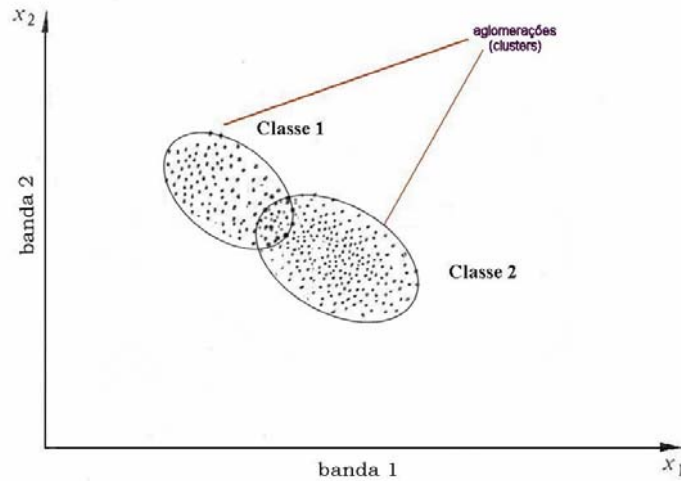


Figura 20 - Interpenetração de classes

A figura 21, mostra duas classes (ω_1, ω_2) com distribuição de probabilidades distintas. As distribuições representam a probabilidade de um pixel pertencer a uma ou outra classe, dependendo da posição do pixel. Existe uma região onde as duas curvas se cruzam, indicando que um determinado pixel tem igual probabilidade de pertencer às duas classes. Nesta situação estabelece-se um critério (fronteira) de decisão que será responsável por determinar à qual classe pertence o pixel.

A fronteira (limite) de decisão neste caso é colocada no ponto em que as duas distribuições se cruzam, de tal forma que um pixel desconhecido nesta posição teria a igual probabilidade de pertencer a ambas as classes.

$$G_i(X) = G_j(X) \quad (4)$$

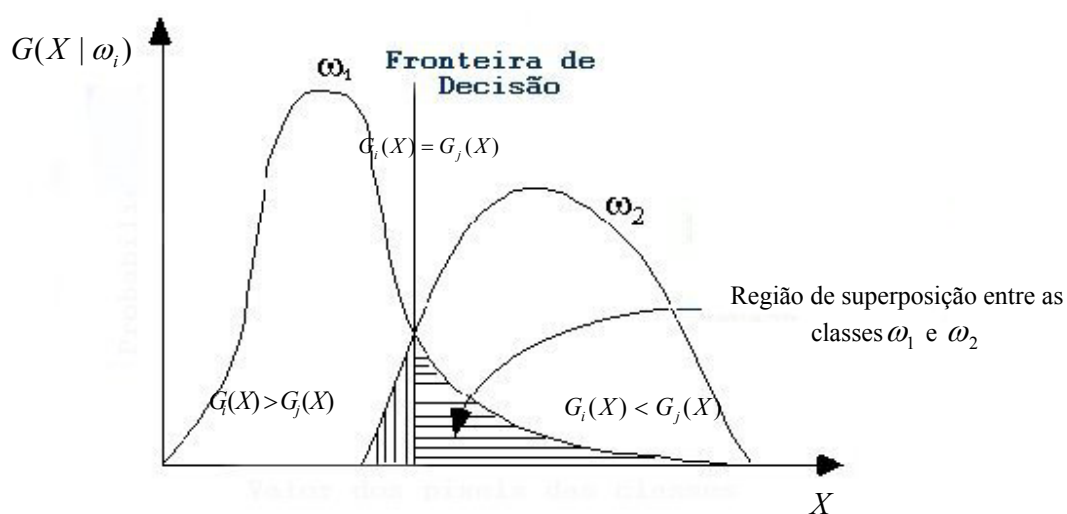


Fig. 21 - Fronteira de decisão entre duas classes

Sing-Tze Bow [BOW 92] considera vantajoso o uso de funções estatísticas discriminantes para classificação em sensoriamento remoto, pois é considerável o conhecimento já existente em áreas como comunicação estatística, teoria da detecção, teoria da decisão, além de outras teorias, e também que este conhecimento é diretamente aplicável em reconhecimento de padrões.

O valor da radiância refletida por pixels pertencentes a uma dada classe apresenta sempre um caráter de variável aleatória. A figura 20 ilustra esta afirmação. Por esta razão, uma abordagem probabilística é mais adequada ao problema da classificação de imagens em sensoriamento remoto.

3.1 Teoria Bayesiana da decisão

Na abordagem estatística, o vetor do pixel X , pertencente a classe ω_i , é visto como uma variável aleatória. Representa-se por $p(X | \omega_i)$ a função densidade probabilidade condicional para cada uma das classes envolvidas. Neste contexto estatístico, estão disponíveis varias funções decisão, sendo as mais utilizadas a função decisão Bayesiana, a função decisão da Máxima Verossimilhança, e a função decisão de Neyman-Pearson, para a estimação das fronteiras de decisão [JAI 00].

Entre os vários classificadores disponíveis, o classificador Bayesiano é o mais utilizado, pelo fato de tratar-se, de um classificador “ótimo”, no sentido de que minimiza a probabilidade de *perda total média condicional*, ou *risco médio*, conforme definido a seguir. O classificador Máxima Verossimilhança, que é possivelmente o mais utilizado em sensoriamento remoto, pode ser visto como um caso particular do classificador Bayesiano.

O desenvolvimento do classificador de Bayes requer as seguintes definições iniciais:

- 1ª) $P(\omega_i)$, é a probabilidade “*a priori*” da classe ω_i . É a probabilidade de ocorrência de um padrão da classe ω_i , independente de qualquer outra informação. Ex: Dados históricos de uma determinada área, mostram que 60% é vegetação, 20% é solo e 20% é água, então:

$$P(\omega_1) = 0.6$$

$$P(\omega_2) = 0.2$$

$$P(\omega_3) = 0.2$$

- 2ª) $p(X | \omega_i)$, é a função densidade probabilidade associada à classe ω_i . É a função de probabilidade que pode ser estimada através das amostras de treinamento.
- 3ª) $p(\omega_i | X)$, é a probabilidade “*a posteriori*” da classe ω_i condicionada ao padrão (pixel) X . Geralmente desconhecida e difícil definição de sua função de probabilidade
- 4ª) $p(X)$, é a função densidade probabilidade para X . É a probabilidade de encontrar um pixel de qualquer classe, na posição X . Pode ser definida por:

$$p(X) = \sum_{i=1}^k p(X | \omega_i) p(\omega_i) \quad (5)$$

O classificador intuitivamente ideal para o papel da função decisão, é definido pela probabilidade condicional *a posteriori*. O teorema de Bayes, que relaciona as quatro funções de probabilidade acima, pode ser expresso por:

$$p(\omega_i | X) = \frac{p(X | \omega_i)p(\omega_i)}{p(X)} \quad , i = 1, \dots, k \quad (6)$$

3.1.1 Classificador de Bayes

A estratégia do classificador Bayesiano “ótimo”, consiste em classificar cada padrão de modo a minimizar a perda média sobre todo o conjunto a ser classificado.

Inicialmente, a função de perda é definida como o custo, perda ou penalidade incorrida quando um algoritmo classifica $X \in \omega_i$ quando, de fato, $X \in \omega_j$. Representa-se esta *função perda* por:

$$C(i, j) \quad i, j = 1, \dots, k \quad (7)$$

Logo, a *perda média condicional ou risco médio*, resultante da classificação de X na classe ω_i , é definida por:

$$L_X(\omega_i) = \sum_{j=1}^k C_{i,j} p(\omega_j | X) \quad (8)$$

O objetivo consiste em classificar X de modo a minimizar esta função perda média condicional. Substituindo (6), na equação (8), tem-se:

$$L_X(\omega_i) = \frac{1}{p(X)} \sum_{j=1}^k C_{i,j} p(X | \omega_j) P(\omega_j) \quad (9)$$

Como o interesse consiste em encontrar a classe que minimiza $L_X(\omega_i)$, e visto que, a probabilidade $p(X)$ é comum para todas as classes, é possível eliminá-la da equação (9), tornando-se:

$$L_X(\omega_i) \propto \sum_{j=1}^k C_{i,j} p(X | \omega_j) P(\omega_j) \quad (10)$$

No processo de classificação, normalmente procura-se maximizar a função decisão $G_i(X)$. No presente contexto, portanto, uma definição conveniente para $G_i(X)$ é:

$$G_i(X) = -L_X(\omega_i) \quad (11)$$

Uma simplificação bastante comum para o valor da função perda $C_{i,j}$, é:

$$\begin{aligned} C_{i,j} &= 0, & \forall i = j \\ C_{i,j} &= 1, & \forall i \neq j \end{aligned} \quad (12)$$

Com esta suposição, a função decisão $G_i(X)$ fica:

$$G_i(X) = -\sum_{\substack{j=1 \\ j \neq i}}^k p(X | \omega_j)P(\omega_j) \quad (13)$$

É possível simplificar a equação (13). Sabendo-se que uma função densidade probabilidade $p(Y)$ qualquer, com condição de $\sum_{j=1}^k P(\omega_j) = 1$, pode ser expressa como

$$p(Y) = \sum_{j=1}^k p(Y | \omega_j)P(\omega_j) \quad (14)$$

ou, isolando-se o termo $j = i$:

$$p(Y) = p(Y | \omega_i)P(\omega_i) + \sum_{\substack{j=1 \\ j \neq i}}^k p(Y | \omega_j)P(\omega_j) \quad (15)$$

ou

$$\sum_{\substack{j=1 \\ j \neq i}}^k p(Y | \omega_j)P(\omega_j) = p(Y) - p(Y | \omega_i)P(\omega_i) \quad (16)$$

então, substituindo (16), a regra na equação (13), onde $G_i(X) = -L_X(\omega_i)$ obtêm-se:

$$G_i(X) = -p(X) + p(X | \omega_i)P(\omega_i) \quad (17)$$

Visto que $p(X)$ é comum para todas as classes, não contribuindo com informação discriminante, pode-se eliminar o termo da expressão, gerando desta forma:

$$G_i(X) = p(X | \omega_i)P(\omega_i) \quad (18)$$

Esta expressão é conhecida como a *função decisão de Máxima Verossimilhança*.

Deste modo, a regra de decisão, também conhecida como *regra de decisão de Máxima Verossimilhança*, fica

$$X \in \omega_i \text{ se } p(X | \omega_i)P(\omega_i) > p(X | \omega_j)P(\omega_j), \text{ para todo } j \neq i \quad (19)$$

$$\text{ou } X \in \omega_i \text{ se } G_i(X) > G_j(X), \text{ para todo } j \neq i \quad (20)$$

3.1.2 Classificador de Máxima Verossimilhança Gaussiana

A efetiva utilização do classificador Máxima Verossimilhança requer que seja conhecida a forma genérica da função densidade de probabilidade $p(X | \omega_i)$. 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 do Limite Central*, pois a radiância refletida por alvos terrestres consistem do somatório de um número grande de outras variáveis aleatórias.

Assumindo N bandas, a função multivariada de Gauss tem a forma

$$p(X | \omega_i) = (2\pi)^{-N/2} |\Sigma_i|^{-1/2} \exp\left\{-\frac{1}{2}(X - \mu_i)^T \Sigma_i^{-1}(X - \mu_i)\right\} \quad (21)$$

onde μ_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.

$$\hat{\mu}_k = \frac{1}{N_k} \sum_{i=1}^{N_k} X_i \quad \hat{\Sigma}_k = \frac{1}{N_k} \sum_{i=1}^{N_k} (X_i - \hat{\mu}_k)(X_i - \hat{\mu}_k)^T \quad (22)$$

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

Como a regra de classificação é a equação (20), qualquer função monotonicamente crescente de (18) mantém inalterado o resultado de (20). Do ponto de vista da eficiência computacional, uma escolha adequada consiste em tomar-se o logaritmo natural. Visto que o termo $-N/2 \ln(2\pi)$ é comum para todo $G_i(X)$, e, não havendo correlação espacial entre pixels vizinhos, a função decisão Máxima Verossimilhança dentro da suposição Gaussiana, pode ser simplificada para:

$$G_i(X) = -\ln|\Sigma_i| - (X - \mu_i)^T \Sigma_i^{-1}(X - \mu_i) + \ln(P(\omega_i)), \quad i=1\dots k \quad (23)$$

Esta função decisão claramente significa um classificador quadrático, isto é, inclui termos X^2 , resultantes do termo $(X - \mu_i)^T \Sigma_i^{-1}(X - \mu_i)$. Neste caso, as fronteiras de decisão assumem a forma de superfícies (ou hiper-superfícies) quadráticas, como elipsóides (ou hiper-elipsóides). As fronteiras de decisão entre as classes ω_i e ω_j são obtidas, simplesmente, igualando-se as funções decisão,

$$\begin{aligned}
 &G_i(X) = G_j(X) \\
 \text{ou} \quad &G_i(X) - G_j(X) = 0
 \end{aligned}
 \tag{24}$$

Este processo pode ser visualizado nas figuras 19 e 21. No caso de uma matriz covariância comum Σ a expressão (23) simplifica-se ainda mais, pois o termo quadrático passa a ser comum a todas as classes, podendo portanto ser cancelado. Neste caso as funções decisão tornam-se funções lineares em X . As superfícies de decisão entre pares de classes tornam-se portanto lineares, descrevendo retas, planos ou hiper-planos.

3.2 Descrição dos classificadores

O processo que implementa a regra de classificação (23), chama-se de *Análise Discriminante Quadrática (QDA)*, visto que estas equações separam as regiões distintas, no espaço multiespectral, correspondente a cada classe, pelas suas fronteiras quadráticas. O processo que implementa a função decisão linear denomina-se de *Análise Discriminante Linear (LDA)*.

Quanto ao classificador adequado, tanto QDA quanto LDA podem fornecer resultados satisfatórios. Dependendo das características das classes, algumas são separáveis linearmente, e outras não.

3.2.1 Classificador quadrático – QDA (Quadratic Discriminant Analysis)

No caso do classificador quadrático, estima-se para cada classe um vetor média e uma matriz covariância próprios, resultando em funções de decisão quadráticas (23).

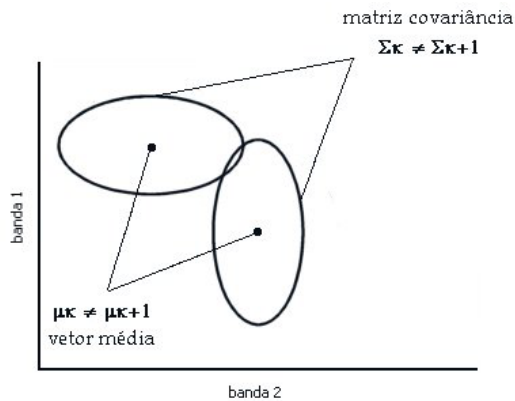


Fig. 22 – Classes com $\mu_k \neq \mu_{k+1}$ e $\Sigma_k \neq \Sigma_{k+1}$

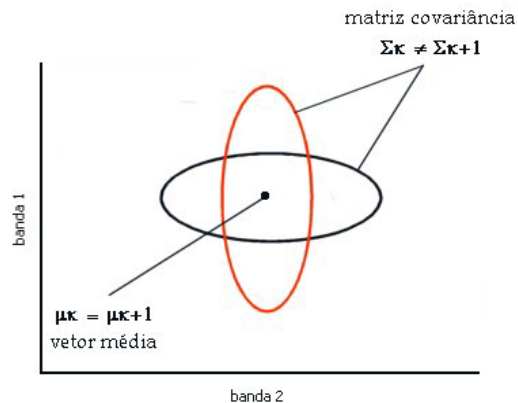


Fig. 23 – Classes com $\mu_k = \mu_{k+1}$ e $\Sigma_k \neq \Sigma_{k+1}$

A forma com que os dados das classes estão dispostos é um fator importante referente às características da classe, e que deve ser usado. Na figura 22 as classes apresentam uma distribuição onde as suas médias são bastante distintas. Esta seria a situação ideal, para fins de classificação, especialmente se a dispersão interna de cada classe (variância) é pequena.

A matriz covariância associada a uma determinada classe contém informações a respeito do formato da distribuição dos pixels da classe, no espaço multiespectral. Desta forma, como apresentado na figura 23, classes que compartilham o mesmo vetor média podem ainda assim serem separadas com razoável grau de precisão, desde que as respectivas matrizes covariância, sejam suficientemente distintas. Pode-se demonstrar [FUK 89] que esta capacidade de separação cresce com a dimensionalidade dos dados.

O classificador QDA aplica a função decisão (23), e classifica o pixel teste X de entrada na classe que maximiza $G_i(X)$.

3.2.2 Classificador linear – LDA (Linear Discriminant Analysis)

O método linear (LDA) pressupõe que a diferença entre as classes está relacionada apenas com o vetor média das classes, assumindo que a matriz covariância é igual para todas as classes.

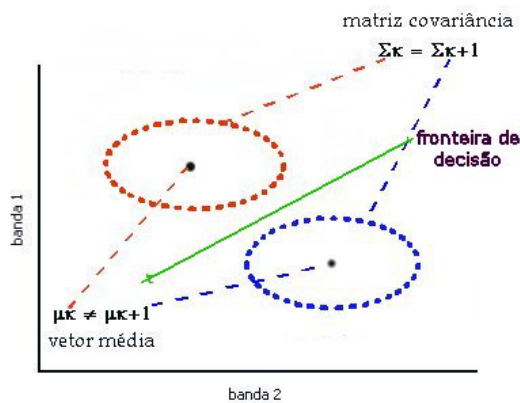


Fig. 24 - Classes com $\mu_k \neq \mu_{k+1}$ e $\Sigma_k = \Sigma_{k+1}$

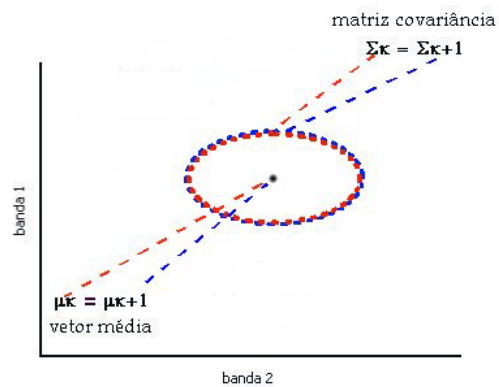


Fig. 25 - Classes com $\mu_k = \mu_{k+1}$ e $\Sigma_k = \Sigma_{k+1}$

As figuras 24 e 25 ilustram este caso. Na primeira situação, supõem-se que as duas classes possuem idêntica matriz covariância, mas distintos vetores médias. A fronteira de decisão fica definida por uma função linear e uma razoável acurácia pode ser obtida, desde que haja uma diferença razoável entre os vetores média de cada classe. Se os vetores médias forem muito próximos ou quase iguais, o classificador linear fornecerá resultados extremamente pobres.

A LDA também assume que as informações das classes seguem uma distribuição normal multivariada (21). O termo quadrático na função decisão (23), torna-se idêntico para todas as funções decisão ($\Sigma_1 = \Sigma_2 = \dots = \Sigma$), podendo portanto ser cancelado. Consequentemente, a fronteira resultante será linear entre as classes; por este motivo usa-se a chamada *Análise Discriminante Linear*. Onde a matriz covariância estimada comum $\hat{\Sigma}$ é usada no lugar da matriz covariância específica de cada classe $\hat{\Sigma}_k$:

$$\hat{\Sigma} = \frac{1}{N} \sum_{k=1}^k N_k \hat{\Sigma}_k \quad N = \sum_{k=1}^k N_k \quad (26)$$

Sendo N o número total de amostras de treinamento. Realizando estas mudanças na equação (21), o classificador linear LDA torna-se

$$G_i(X) = -\frac{1}{2}(X - \hat{\mu}_i)^T \hat{\Sigma}^{-1}(X - \mu_i) + \ln(P(\omega_i)) \quad (27)$$

Assumindo, como freqüentemente é feito em aplicações práticas, que as probabilidades *a priori* $P(\omega_i)$ são idênticas para todas as classes, a função decisão linear (27) simplifica-se para

$$G_i(X) = -(X - \hat{\mu}_i)^T \hat{\Sigma}^{-1}(X - \mu_i) \quad (28)$$

também conhecido como *distância de Mahalanobis*.

Aeberhard em [AEB 94] cita que, pelo motivo de no método LDA ocorrer número muito menor de parâmetros a serem estimados, em alguns casos torna-se preferível usar o método LDA ao método QDA em configurações de alta dimensionalidade. Porém, por possuir fronteiras lineares, o método LDA tem sérias limitações em casos onde a real matriz covariância das classes difere significativamente. Já o processo que utiliza as

funções de decisão quadráticas requer um número maior de amostras do que aquelas baseadas no método LDA[WAL 77], tornando-se o grande problema do classificador de máxima verossimilhança gaussiana, por exemplo.

Uma situação que envolva uma quantidade muito limitada de amostras para estimação de parâmetros pode ser mais eficiente quando se opta por um classificador que não faça uso das informações da matriz covariância, como é o caso do classificador de mínima distância. Isto porque este classificador utiliza as amostras de treinamento apenas para estimar o vetor média das classes, e a classificação é então realizada classificando o pixels na classe com o vetor média mais próximo[RIC 86]. A grande diferença entre o classificador de mínima distância e os classificadores baseados no classificador de máxima verossimilhança gaussiana está no uso das informações contidas na matriz covariância das classes.

Um exemplo de classificador de mínima distância é o classificador euclidiano, o qual considera que as matrizes covariância para todas as classes são diagonais e iguais, e que as variâncias em cada componente são idênticas (29).

$$\hat{\Sigma} = \begin{vmatrix} \sigma^2 & 0 & 0 & 0 \\ 0 & \sigma^2 & 0 & 0 \\ 0 & 0 & \sigma^2 & 0 \\ 0 & 0 & 0 & \sigma^2 \end{vmatrix} = \sigma^2 \begin{vmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{vmatrix} = \hat{\Sigma} = \sigma^2 I \quad \text{para todo } i \quad (29)$$

Nestas circunstâncias, a função discriminante, assumindo a função multivariada gaussiana (21), para o classificador de máxima verossimilhança (23), torna-se:

$$G_i(X) = 1/2 \ln \sigma^{2N} - 1/2 \sigma^2 (X - \mu_i)^T (X - \mu_i) + \ln P(\omega_i) \quad (30)$$

O termo $\ln \sigma^{2N}$ é comum para todas as classes, podendo ser desconsiderado. Deste modo a equação torna-se

$$G_i(X) = -\frac{1}{2\sigma^2}(X - \mu_i)^T(X - \mu_i) + \ln P(\omega_i) \quad (31)$$

Assumindo ainda que as probabilidades *a priori* $P(\omega_i)$ são idênticas para todas as classes, a função decisão simplifica-se, para:

$$G_i(X) = -(X - \mu_i)^T(X - \mu_i) \quad (32)$$

que é a função discriminante que resultará na distância euclidiana entre o vetor X e o vetor média de cada classe. O classificador LDA, expresso na equação (28), supõe que as matrizes covariância das classes são iguais ($\Sigma_1 = \Sigma_2 = \dots = \Sigma$). Com uma segunda suposição adicional, em que as probabilidades *a priori* das classes ($P(\omega_1) = P(\omega_2) = \dots = P(\omega_k) = \frac{1}{k}$) também são idênticas, neste caso o classificador pode ser considerado baseado na distância mínima, normalizada pela matriz covariância.

O raciocínio apresentado até o momento, iniciado pela teoria de decisão de Bayes, onde a estratégia é minimizar a perda média, continuando com o classificador máxima verossimilhança, que é um caso particular do classificador de Bayes, o classificador de máxima verossimilhança gaussiana, que supõe que os dados apresentam a distribuição de probabilidade multivariada de Gauss, o classificador linear, que é uma simplificação do classificador de máxima verossimilhança e por fim o classificador baseado na distância euclidiana mínima, é o conhecimento teórico em que está baseado o presente estudo.

3.3 Método Regularizado – RDA (Regularized Discriminant Analysis)

Técnicas de regularização da matriz covariância surgem como possibilidades para mitigar estes problemas apresentados pelos classificadores que fazem uso da matriz covariância de cada classe, como o classificador máxima verossimilhança Gaussiana.

Conforme mostrado anteriormente na figura 15 e na tabela 2, em métodos paramétricos, o número de parâmetros a serem estimados para o modelo aumenta com a dimensionalidade. A menos que o número de amostras de treinamento seja aumentado coerentemente com o aumento da dimensionalidade, a matriz covariância estimada resultará mal-condicionada, isto é, quando o número de amostras de treinamento em cada classe é pequeno, quando comparado com a dimensionalidade dos dados. Na matriz covariância caracterizada como mal-condicionada, os parâmetros estimados apresentarão uma variância grande [AEB 94]. Uma situação ainda mais desfavorável ocorre quando o número de amostras de treinamento é inferior à dimensionalidade dos dados. Neste caso, a estimativa da matriz covariância resulta singular.

Friedman ressalta que os trabalhos baseados nos estudos de Titterington [TIT 85] e O’Sullivan [OSU 86] mostram que as técnicas de regularização obtiveram alto sucesso na solução de problemas de mal-condicionamento da matriz covariância. Baseando-se ainda em Cornfield [COR 67], ao empregar um método de regularização, procura-se melhorar a estimação dos parâmetros, introduzindo ou forçando uma tendenciosidade adequada em sua estimação, de forma que as estimativas sejam mais “plausíveis fisicamente” [FRI 89].

A *Análise Discriminante Regularizada (RDA)*, foi especialmente desenvolvida para resolver os problemas de mal-condicionamento, e pode ser vista como uma adaptação da QDA.

Assim como o método QDA, o método RDA classifica a amostra teste, na classe cuja função decisão $G_i(X)$ (23) retornar o maior valor numérico. Contudo, enquanto o método QDA usa uma estimação da matriz covariância $\hat{\Sigma}_k$ não-regularizada (não-tendenciosa), o método RDA faz uso da matriz covariância regularizada (tendenciosa) $\hat{\Sigma}_k(\lambda, \gamma)$ no lugar daquela. O processo de regularização é feita em duas etapas distintas, para tentar deixar os valores da matriz covariância mais plausíveis fisicamente. A técnica de regularização da matriz covariância (RDA) apresentada por Friedman faz uso de dois parâmetros (lambda (λ) e gama(γ)) que variam entre 0 e 1. Através da combinação entre λ e γ , obtém-se valores intermediários entre os métodos QDA e LDA. Estas combinações tem por objetivo combinar em diferentes graus tanto a matriz covariância comum das classes, quanto da matriz covariância particular de cada classe, resultando em uma estimativa mais confiável.

Como o número de amostras de treinamento é limitado, as estimativas não-tendenciosas da $\hat{\Sigma}_k$ de cada classe podem não refletir, adequadamente, as similaridades que podem existir entre as estruturas das classes. Nestes casos, a combinação linear da matriz covariância $\hat{\Sigma}_k$ de cada classe e a matriz covariância $\hat{\Sigma}$ comum entre as classes irá produzir uma estimativa melhor da estrutura básica das classes, e, conseqüentemente, um melhor resultado de classificação. Este processo é considerado o 1º passo da regularização, onde $\hat{\Sigma}_k$ é substituída por uma combinação linear $\hat{\Sigma}_k(\lambda)$, entre $\hat{\Sigma}_k$ e a $\hat{\Sigma}$ para cada classe:

$$\hat{\Sigma}_k(\lambda) = \frac{(1-\lambda)\hat{Q}_k + \lambda\hat{Q}}{N_k(\lambda)}, \quad 0 \leq \lambda \leq 1 \quad (29)$$

sendo,

$$N_k(\lambda) = (1 - \lambda)N_k + \lambda N \quad (30)$$

e

$$\hat{Q} = \sum_{k=1}^k \hat{Q}_k \quad \hat{Q}_k = N_k \hat{\Sigma}_k \quad (31)$$

onde $\hat{\Sigma}_k$ representa a matriz covariância própria de cada classe, estimada pela equação (22), $\hat{\Sigma}_k(\lambda)$ representa a matriz covariância regularizada pelo parâmetro λ . N_k é o número de amostras de cada classe, N é o somatório das amostras de todas as classes, Q_k representa a matriz covariância de cada classe $\hat{\Sigma}_k$ ponderada pelo tamanho da amostra correspondente, e Q é o somatório das estimativas Q_k .

A equação (30) tem o objetivo de ajustar o grau de influência do tamanho das amostras em cada classe, levando em consideração o número total de amostras e/ou o tamanho individual de amostras de cada classe.

A equação em (29) implementa um processo semelhante, porém com o termo Q_k , que resulta no 1º passo de regularização da matriz covariância $\hat{\Sigma}_k(\lambda)$. Este processo ajusta o número de amostras (pixels) associados com a matriz covariância regularizada como uma combinação linear do número de amostras de treinamento na classe ω_k e o número total de amostras de treinamento. O parâmetro λ é responsável por controlar o grau de ajuste da estimativa individual da matriz covariância $\hat{\Sigma}_k$ de cada classe, em relação à matriz covariância comum $\hat{\Sigma}$.

Este primeiro procedimento de regularização da matriz covariância normalmente já produz resultados significativos, especialmente quando o tamanho de amostras das classes

individuais é pequeno. Os valores intermediários de λ resultarão em combinações de regularização entre os métodos LDA e QDA.

A regularização realizada pela equação (29) ainda é limitada por duas razões:

- (a) pode não realizar uma regularização suficiente, se o número de amostras é inferior ou comparável à dimensionalidade, e
- (b) ajustar a matriz covariância da classe através da matriz covariância comum pode não ser a forma mais efetiva para corrigir as matrizes $\hat{\Sigma}_k$, por que, em alguns casos [FRI 89], este tipo de variação (*shrinkage*) pode introduzir tendências severas.

Friedman adicionou um segundo passo ao processo de regularização. Este segundo passo tem por objetivo corrigir o fato bem conhecido de que, quando o número de amostras de treinamento é pequeno, os valores estimados para os autovalores tornam-se tendenciosos: os autovalores menores são estimados ainda menores e os maiores ainda maiores. Logo, a equação (32) tenta controlar esta tendência, escalonando a matriz covariância (que já passou pela 1ª etapa de regularização) $\hat{\Sigma}_k(\lambda)$, através da multiplicação da matriz identidade com a média dos autovalores de $\hat{\Sigma}_k(\lambda)$. O efeito deste processo será diminuir os autovalores maiores e aumentar os autovalores menores Friedman. [FRI 89]

$$\hat{\Sigma}_k(\lambda, \gamma) = (1 - \gamma)\hat{\Sigma}_k(\lambda) + \frac{\gamma}{d} \text{tr}[\hat{\Sigma}_k(\lambda)]\mathbf{I}, \quad \begin{array}{l} 0 \leq \lambda \leq 1 \\ 0 \leq \gamma \leq 1 \end{array} \quad (32)$$

onde $\text{tr}[\hat{\Sigma}_k(\lambda)]$ é o *traço* de $\hat{\Sigma}_k(\lambda)$, d é a dimensionalidade dos dados (número de variáveis), e \mathbf{I} é a matriz identidade. O segundo parâmetro de regularização gama (γ) controla portanto esta 2ª etapa do processo de regularização.

Ao aplicar ambos os processos de regularização da matriz covariância, e ao empregar $\hat{\Sigma}_k(\lambda, \gamma)$ no lugar da $\hat{\Sigma}_k$, a função decisão (23) utilizada no classificador quadrático QDA toma a forma:

$$G_i(X) = -(X - \mu_i)^T \hat{\Sigma}_i^{-1}(\lambda, \gamma)(X - \mu_i) - \ln |\hat{\Sigma}_i(\lambda, \gamma)| \quad (33)$$

Aeberhard [AEB 93], salienta que não existe um conhecimento de quais as melhores configurações de parâmetros de λ e γ , para que se tenha uma probabilidade de classificação correta (PCC) maximizada. A avaliação dos resultados do classificador é realizada experimentalmente, com um conjunto de pares de (λ, γ) e então são selecionados os valores onde o resultado tenha a melhor acurácia.

Friedman analisa ainda a série de classificadores obtidos pelo processo de regularização como se estivessem posicionados num cubo, onde o resultado seria conforme a figura 26.

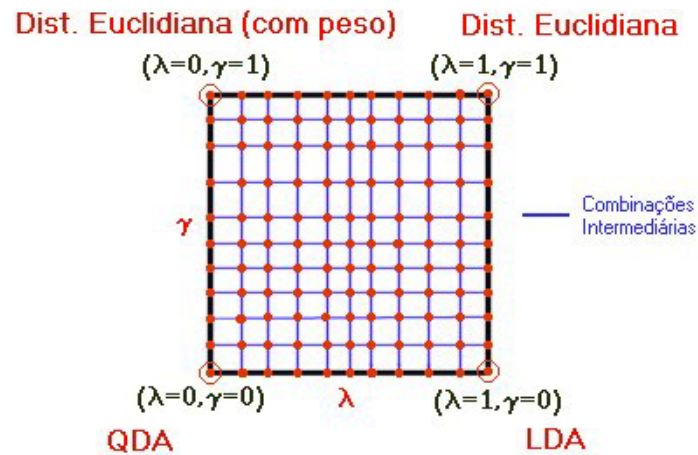


Fig. 26 - Cubo de resultados da combinação de lambda (λ) e gama (γ)

A técnica RDA propicia várias alternativas, onde os 4 cantos definem os extremos do plano (λ, γ) representando procedimentos de classificação, bem conhecidos.

- O canto inferior esquerdo $(\lambda=0, \gamma=0)$ representa a Análise Discriminante Quadrática;
- O canto inferior direito $(\lambda=1, \gamma=0)$ representa a Análise Discriminante Linear;
- O canto superior direito $(\lambda=1, \gamma=1)$ corresponde ao classificador distância euclidiana mínima.
- O canto superior esquerdo $(\lambda=0, \gamma=1)$ que representa o mesmo classificador de média mais próxima, contudo leva em considerações atributos de “pesos”, onde o peso da classe é inversamente proporcional a média da variância dos valores das variáveis das classes;

Mantendo γ fixo em 0 e variando apenas λ , produz-se modelos entre os métodos QDA e LDA. Mantendo λ fixo em 0 e variando apenas γ , visa-se diminuir a tendenciosidade na estimativa dos autovalores a partir das amostras disponíveis.

Baseando-se nas conclusões de Friedman [FRI 89], o método de regularização por ele aplicado tem potencial para (as vezes dramaticamente) aumentar o poder da análise discriminante em configurações em que o tamanho das amostras é pequeno e o número de parâmetros a serem estimados é grande. Os experimentos relatados por Friedman, utilizando dados sintéticos e dados de enologia, mostram casos em que o processo RDA proporcionou ganhos substanciais na exatidão do processo de classificação, e casos em que houve uma pequena perda na acurácia do método RDA em relação aos outros métodos, principalmente quando as situações dos experimentos eram desfavoráveis para o classificador.

4 EXPERIMENTOS

Conforme mencionado anteriormente, este estudo tem por objetivo testar as técnicas de regularização da matriz covariância em dados de sensoriamento remoto. Com esta finalidade, são desenvolvidos experimentos com dados de alta dimensionalidade (dados hiperespectrais). As técnicas de regularização são então implementadas e analisadas. Este capítulo apresenta e analisa os experimentos desenvolvidos.

4.1 Materiais

Para a realização deste projeto, são necessários os seguintes materiais:

1. Equipamentos

- Impressora Jato de Tinta HP DeskJet 720C ;
- PC- AMD k6 II 500 Mhz, RAM 128Mb, HD 500Mb disponível (computador particular – desenvolvimento e implementação)
- PC-Pentium III 500 Mhz, RAM 512Mb, HD 500Mb disponível (CEPSRM/UFRGS - processamento)
- ZipDrive 100Mb;

2. Aplicativos

- Win98, NT 4.0 Workstation, MatLab 5.3, Multispecw32, Office97;

3. Dados

- Imagem AVIRIS cedida para estudo, pelo prof. David Landgrebe – Purdue University, para fins de validação;

4.2 Cena de estudo

Os dados hiperespectrais usados neste experimento são um segmento (435 x 435), num total de 189.225 pixels, de uma cena de dados AVIRIS, tomada em junho de 1992, de uma região agrícola (região experimental de testes) chamada Indian Pine, localizada a noroeste no estado de Indiana – EUA (Fig. 27). A cena observada na figura 28 possui 220 bandas espectrais (das 224 bandas originais do sensor AVIRIS), cobrindo intervalos de comprimentos de onda distanciados uniformemente de modo a cobrir todo o intervalo de comprimento de onda do espectro eletromagnético entre 400 nm e 2.500 nm.

A imagem (Fig. 28) cobre uma área agrícola, com culturas de milho e soja, além de outras. A imagem foi obtida no início da época de crescimento, ocorrendo por isso grande quantidade de solo exposto. Foram ainda utilizadas técnicas de plantio distintas (plantio tradicional e plantio direto). Esta situação resulta em classes de difícil separação.

A verdade terrestre está ilustrada na figura 29, e a curva de resposta espectral média para as cinco classes selecionadas para os experimentos é ilustrada na figura 36.

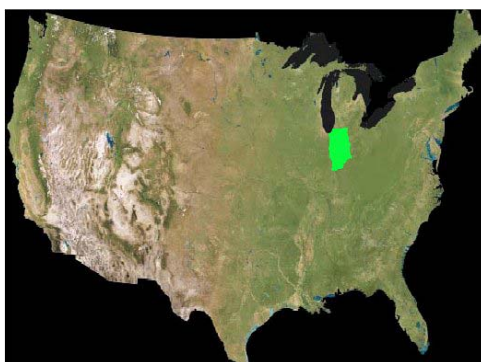


Fig. 27 – Localização do estado de Indiana – EUA (região em verde)



Fig. 28 - AVIRIS: Composição RGB com as bandas 50, 27 e 17, da área experimental de teste, (NW Indiana's Indian Pine) - 220 bandas espectrais

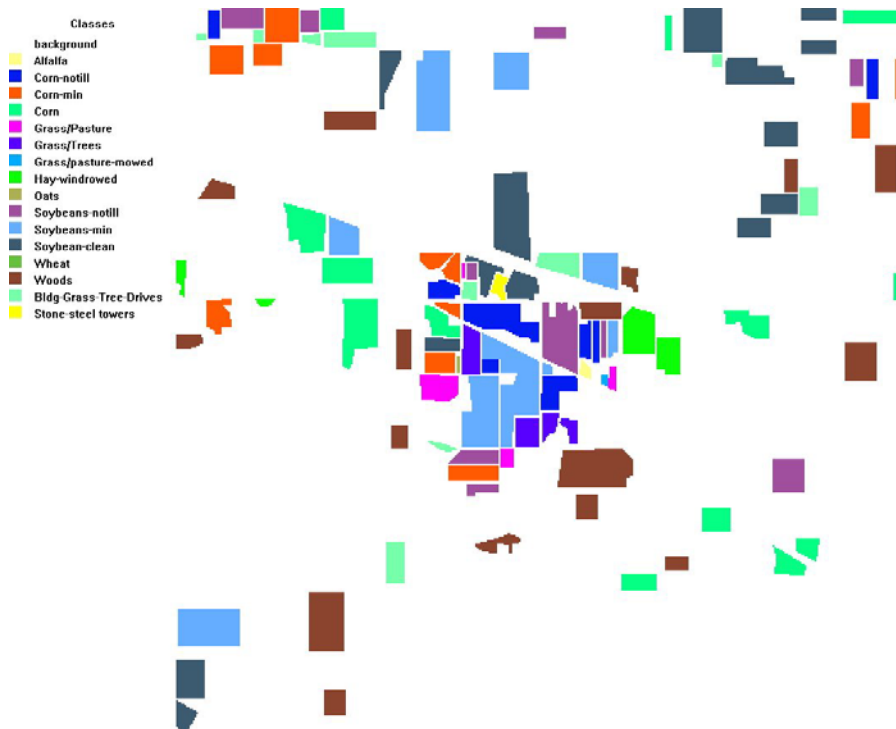


Fig. 29 - Mapa temático da verdade terrestre (ground true) da figura 28

4.2.1 Seleção de Bandas Espectrais

Como etapa inicial para a seleção das amostras das classes, deve-se preparar a imagem, isto é, fazer uma análise preliminar, a fim de que se trabalhe apenas com os dados que realmente tragam alguma informação útil, que caracterize as classes. Desta forma foram eliminadas bandas afetadas pelo efeito da absorção, ocasionado pelos constituintes da atmosfera, tais como vapor d'água, CO₂ e O₃, que apresentam um nível muito alto de ruído. Foram excluídas as seguintes bandas espectrais, conforme estão indicadas nas tabelas 3 e 4.

Tabela 3 – Seleção de bandas ruidosas que foram excluídas (35 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

Tabela 4 – Bandas restantes que serão usadas no estudo (185 bandas)

		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									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												
					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								

A dimensionalidade final a ser usada, é de 185 bandas espectrais, resultando em vetores com 185 valores de contador digital para cada pixel da imagem.

4.2.2 Seleção de Classes

Os testes realizados neste estudo utilizam 5 classes distintas: *Corn* (ou *Corn-Clean* = milho plantio lavrado), *Corn-min*(=milho) , *Soybean-clean* (ou *Soybean* = Soja plantio lavrado), *Soybean-notill* (= Soja plantio direto), *Woods* (= Floresta).



Fig. 30 - Classes selecionadas para o estudo e suas respectivas cores temáticas, segundo figura 29

A escolha destas classes deve-se ao fato de que as quatro primeiras classes apresentam uma resposta espectral muito semelhante, sendo, por esta razão, separáveis com razoável precisão quando são empregados os dados em alta dimensionalidade provenientes dos sensores hiperespectrais.

O motivo da inserção da classe *Woods* nos experimentos deve-se ao fato de que o padrão “floresta” é uma classe bem distinta, se comparada com as outras quatro classes. Este procedimento visa realizar uma análise completa, verificando inclusive, se a acurácia do algoritmo está correta em situações de classificação mais simples, pois a classe *Woods* é espectralmente muito diferente das culturas agrícolas.

As amostras disponíveis para cada uma das classes estão relacionadas na tabela 5, e foram obtidas pela utilização conjunta da imagem AVIRIS (Fig. 28) e da imagem de verdade terrestre (Fig. 29).

Tabela 5 - Total de pixels extraídos da imagem AVIRIS

Nome da Classe	Nº Temático	Total de Pixels por classe
Corn	4	3642
Corn-min	3	2442
Soybean-clean	12	4654
Soybean-notill	10	2264
Woods	14	4664

Definidas as amostras de cada classe, é possível visualizar o comportamento espectral das amostras através das assinaturas espectrais e do gráfico hiperespectral das classes. Para esta finalidade, utilizou-se o software MultispecW32 elaborado pela Universidade Purdue, e algumas rotinas implementadas no software MatLab. Inicialmente a criação de subconjuntos, com a maior quantidade possível de pixels aleatórios, próxima ao número total de amostras disponíveis de cada classe. Este procedimento têm por objetivo formar uma matriz de dimensão $N \times N$ *. O modo como é feito este procedimento, pode ser acompanhado na tabela 6.

Tabela 6 - Subconjuntos $N \times N$

Nome da Classe	Total de Pixels por classe	Maior quantidade de pixels possível, subconjunto $N \times N$	$N \times N$
Corn	3642	3600	60 x 60
Corn-min	2442	2401	49 x 49
Soybean-clean	4654	4624	68 x 68
Soybean-notill	2264	2209	47 x 47
Woods	4664	4624	68 x 68

* Este processo é realizado apenas para ilustrar os dados (pixels) de cada classe na forma de um quadrado, sem maior importância nos experimentos, onde são utilizados todos os dados, sem exceção.

As figuras 31, 32, 33, 34, 35 e 36 ilustram o comportamento espectral das classes. Neste conjunto de figuras, as partes (a) ilustram uma composição RGB (bandas 50-27-17) dos pixels que compõem as amostras de cada uma das classes. As partes (b) apresentam a curva de resposta espectral e as partes (c), o gráfico hiperespectral (Nº de bandas X Quantidade de pixels da amostra) dos pixels em questão.

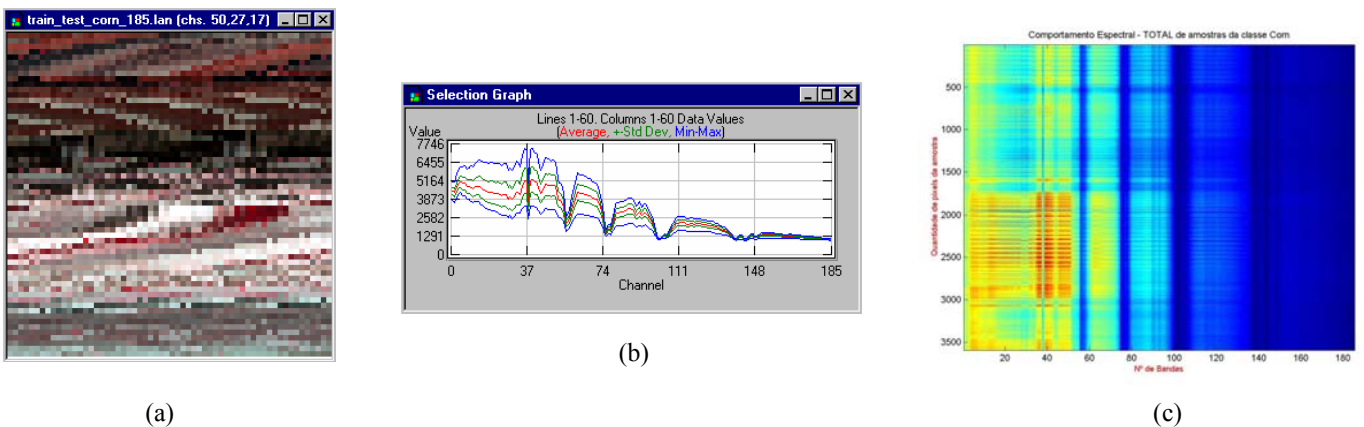


Fig. 31 - As linhas dos gráficos (*Selection Graph*) da assinatura espectral (b), mostram os valores máximo, mínimo, desvio padrão e o valor médio dos pixels pertencentes a classe *Corn*

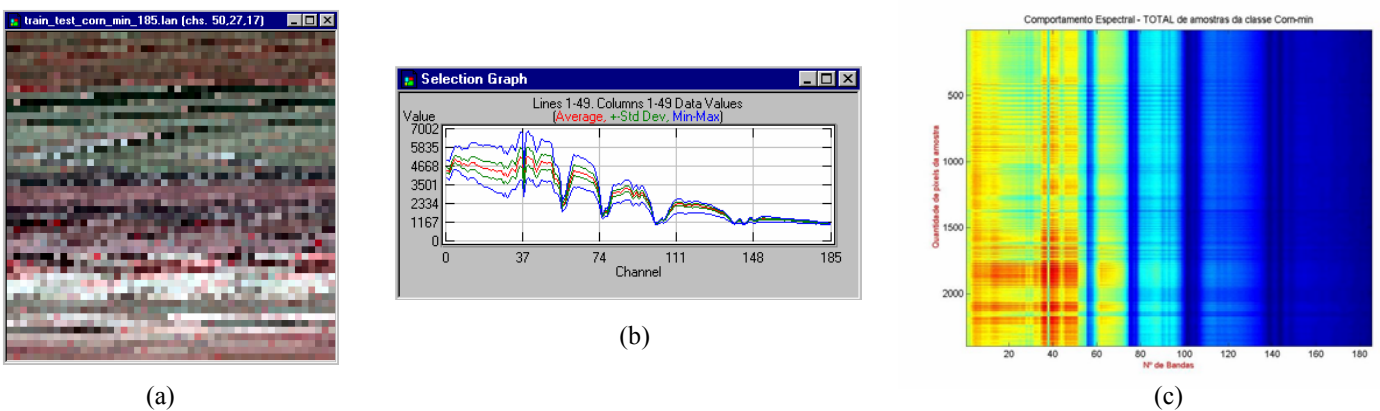
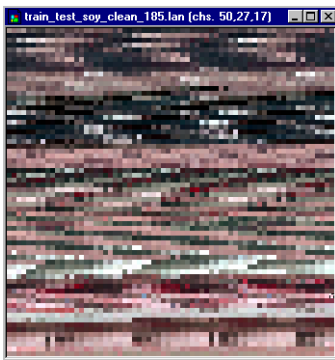
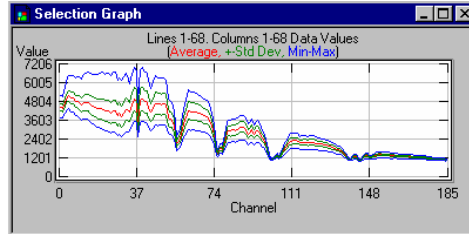


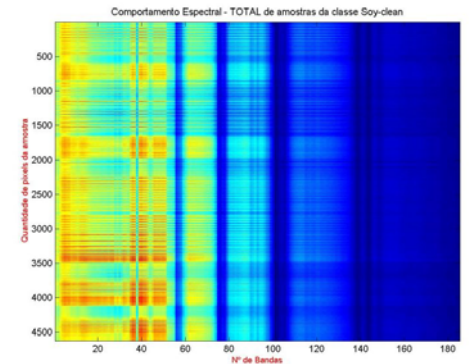
Fig. 32 - As linhas dos gráficos (*Selection Graph*) da assinatura espectral (b), mostram os valores máximo, mínimo, desvio padrão e o valor médio dos pixels pertencentes a classe *Corn-min*



(a)

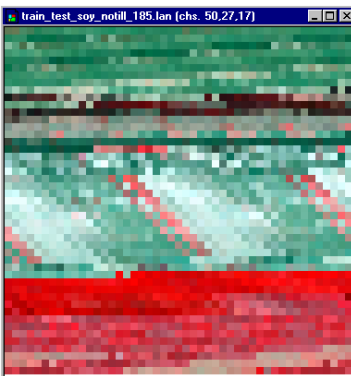


(b)

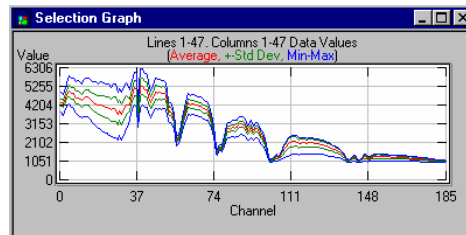


(c)

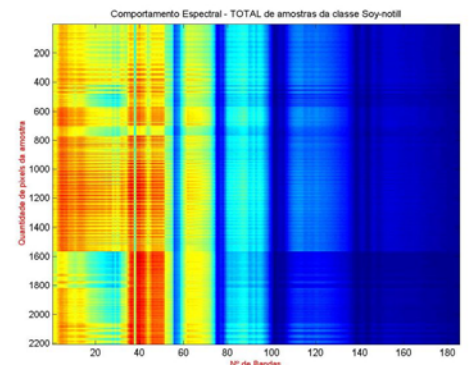
Fig. 33 - As linhas dos gráficos (*Selection Graph*) da assinatura espectral (b), mostram os valores **máximo**, **mínimo**, **desvio padrão** e o valor **médio** dos pixels pertencentes a classe *Soy-clean*



(a)

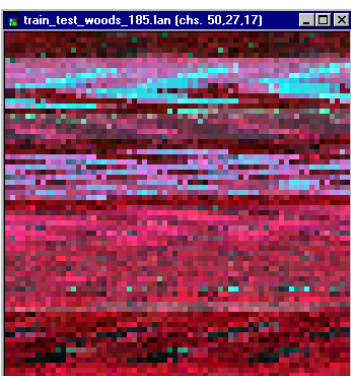


(b)

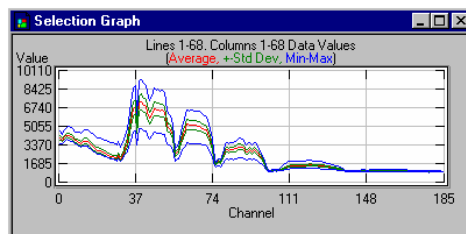


(c)

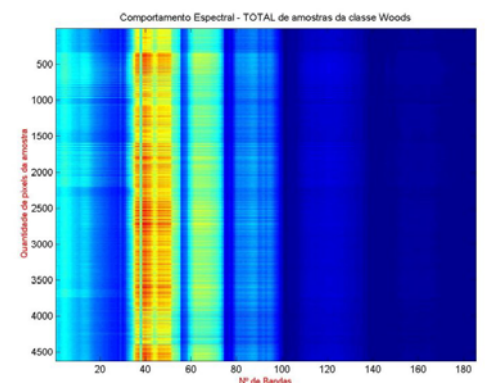
Fig. 34 - As linhas dos gráficos (*Selection Graph*) da assinatura espectral (b), mostram os valores **máximo**, **mínimo**, **desvio padrão** e o valor **médio** dos pixels pertencentes a classe *Soy-notill*



(a)



(b)



(c)

Fig. 35 - As linhas dos gráficos (*Selection Graph*) da assinatura espectral (b), mostram os valores **máximo**, **mínimo**, **desvio padrão** e o valor **médio** dos pixels pertencentes a classe *Woods*

Notam-se dois aspectos principais: a diferença espectral entre as classe Woods em relação as classes Corn, Corn-min, Soy-clean, Soy-notill, e a própria semelhança entre as quatro classes. É possível acompanhar estes dois aspectos, por meio das curvas de resposta espectral na figura 36.

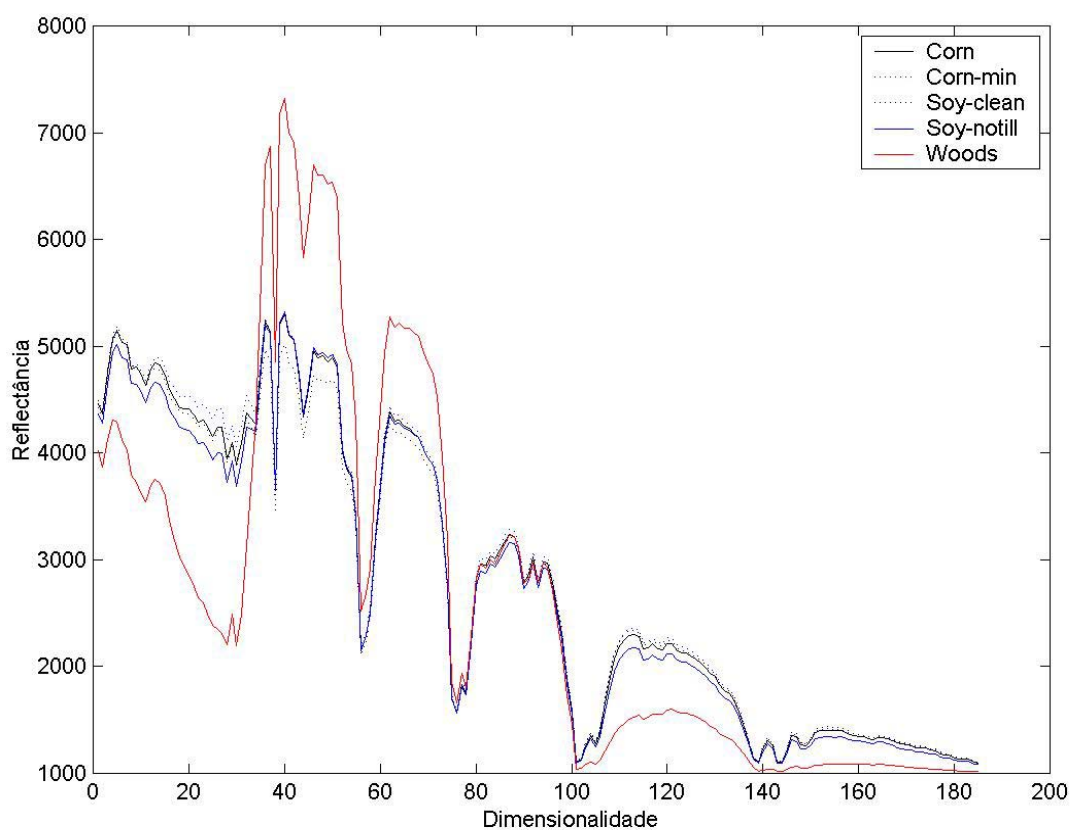


Fig. 36 - Sobreposição das assinaturas espectrais das classes selecionadas

4.2.3 Amostras de treinamento e teste das classes selecionadas

Do conjunto de amostras disponíveis para cada uma das cinco classes, extrai-se dois subconjuntos distintos: um subconjunto para a estimação dos parâmetros (amostras de treinamento) e outro para testar a precisão do processo de classificação (amostras de teste).

A técnica empregada para separar as amostras em dois subconjuntos, cria o contador " i ", que varia de 1 até número total de pixels da classe. Atribui-se os pixels nas posições ímpares (i) do conjunto de amostras da classe ao grupo de treinamento da classe, e os pixels nas posições pares ($i + 1$) do conjunto de amostras da classe para o grupo de teste. Logo, são gerados dois grupos de pixels de tamanho e comportamento idênticos, pois ambos são subconjuntos de um conjunto maior.

A tabela 7 apresenta o tamanho das amostras disponíveis para cada uma das classes.

Tabela 7 - Amostras de treinamento e teste das classes selecionadas

Nome da classe	Total de pixels do conjunto de amostras da classe	Amostras para Treinamento pixels " i " ímpar	Amostras para Teste pixels " i " par
Corn	3642	1821	1821
Corn-min	2442	1221	1221
Soy-clean	4654	2327	2327
Soy-notill	2264	1132	1132
Woods	4664	2332	2332

4.3 Ferramenta ADRIH (*Análise Discriminante Regularizada em Imagem Hiperespectral*)

Esta ferramenta é parte do projeto desenvolvida sobre a plataforma do software MatLab 5.3; logo, não é um arquivo *.EXE (executável). Para poder utilizá-lo é necessário, inicialmente, estar no ambiente do MatLab. O objetivo do ADRIH é analisar passo-a-passo o método e facilitar o manejo dos dados hiperespectrais, bem como na seleção de bandas espectrais, no processamento das informações e na visualização dos resultados em forma de gráficos 2D e 3D, tudo isto tendo-se como pressuposto o método de regularização RDA.

Esta ferramenta foi projetada para a utilização da técnica RDA, porém pode ser facilmente adaptada, ou até incrementada, com novas técnicas, com o objetivo de comparar performances de metodologias diferentes, em ambiente de dados hiperespectrais.

A descrição e o funcionamento do software estão no ANEXO 4.

4.4 Testando o método RDA com dados de sensoriamento remoto

Inicialmente, é necessário que se estipule o número de bandas que devem ser usadas no processamento. Neste trabalho, foram realizados uma grande série de testes com os dados disponíveis, tomando-se a dimensionalidade dos dados como variável independente e a resultante precisão na classificação como variável dependente. O valor da dimensionalidade dos dados variou desde 5 até 185 (bandas espectrais), a intervalos de 5 bandas espectrais. Desta forma é possível identificar o Efeito de Hughes, pois, na medida

em que a dimensionalidade aumenta pela adição de bandas espectrais, no classificador quadrático pode-se visualizar, detalhadamente, o comportamento da acurácia.

Os parâmetros de cada classe, conforme citado anteriormente, são estimados a partir das amostras de treinamento. Os experimentos são realizados para vários pares de valores numéricos dos parâmetros de regularização da matriz covariância, onde λ e γ variam de 0 a 1, em intervalos de 0.1.

Em um primeiro passo, o parâmetro λ regula o grau de influência da matriz covariância de cada classe com o grau de influência da matriz covariância comum. Em um segundo passo, o parâmetro γ tenta contornar a tendenciosidade que o número pequeno de amostras de treinamento introduz nos autovalores da matriz covariância, o que é obtido por meio do escalonamento da matriz covariância $\Sigma_k(\lambda)$ pela matriz identidade ponderada.

4.4.1 *Experimento 01*

Neste experimento, foi utilizada uma quantidade pequena de amostras de treinamento, de tamanho igual a 200 (um pouco maior do que o número de bandas usadas (185)), para todas as classes, conforme apresenta a tabela 8. Assim sendo, o tamanho das amostras de treinamento é muito pequena para estimar μ_k e Σ_k , cerca de, 1.081 vezes o tamanho da dimensionalidade.

Tabela 8 - Quantidade de amostras das classes selecionadas, utilizadas no Experimento 01, para 200 amostras

Nome da classe	Treinameto	Teste
Corn	200	1821
Corn-min	200	1221
Soy-clean	200	2327
Soy-notill	200	1132
Woods	200	2332

Quanto ao número de amostras de teste utilizadas em todos os experimentos, usou-se a quantidade total de pixels de teste de cada classe.

Ao escolhermos $\lambda = 0.0$ e $\gamma = 0.0$, assume-se o comportamento do método QDA para o classificador (Fig. 26). Nesta combinação, conforme as equações (29) e (32), a influência da matriz covariância de cada classe (Σ_k) é total em relação à matriz covariância comum entre as classes (Σ), que é inexistente. O resultado alcançado esta ilustrado por meio de uma tabela de contingência (Tabela 9).

Tabela 9 - Tabela de contingência do método QDA – 200 amostras






	Corn	CornMin	SoyClean	Soynotill	Woods
Corn	921	426	799	303	1
CornMin	337	432	267	172	1
SoyClean	344	154	1055	92	0
Soynotill	218	208	205	534	64
Woods	0	0	0	30	2265
AMOSTRAS:					
treinamento	200	200	200	200	200
teste	1820	1220	2326	1131	2331
% Acertos (Classe)	50.5766%	35.3808%	45.3373%	47.1731%	97.1269%
Bandas →	185				
λ →	0				
γ →	0				

Ao atribuir $\lambda = 1.0$ e $\gamma = 0.0$, assume-se o comportamento do método LDA para o classificador (Fig 26). Nesta combinação, conforme as equações (29) e (32), a influência da matriz covariância de cada classe (Σ_k) é nula em relação à matriz covariância comum (Σ) entre as classes, que neste caso é total. A tabela 10 ilustra os resultados alcançados.

Tabela 10 - Tabela de contingência do método LDA – 200 amostras

	Corn	CornMin	SoyClean	Soynotill	Woods
Corn	946	330	439	79	0
CornMin	387	727	108	121	0
SoyClean	352	56	1744	53	0
Soynotill	135	107	35	878	30
Woods	0	0	0	0	2301
AMOSTRAS:					
treinamento	200	200	200	200	200
teste	1820	1220	2326	1131	2331
% Acertos (Classe)	51.9495%	59.5414%	74.9463%	77.5618%	98.6707%
Bandas → 185					
λ → 1					
γ → 0					

Comparados os resultados das tabelas de contingência 9 e 10, nota-se que a classificação, nesta situação, se comporta de melhor forma com o método LDA, obtendo porcentagens finais de acurácia superiores ao método QDA. Este resultado tende a confirmar o fato de que, em situações nas quais o tamanho das amostras de treinamento é muito pequeno quando comparado com a dimensionalidade dos dados, o classificador linear, geralmente, fornece resultados melhores do que o quadrático. Nota-se que estes resultados foram obtidos para uma dimensionalidade igual a 185 (dimensão máxima). A seguir, são mostrados os principais resultados obtidos nos experimentos, com a seguinte legenda:

-  - Linha correspondente aos resultados obtidos com o método QDA.
-  - Linha correspondente ao resultado obtido com $\lambda = 0.2$ e $\gamma = 0.0$ no método RDA
-  - Linha correspondente à melhor acurácia alcançada. Fornece a combinação de (λ, γ) , com quantas bandas alcançou, e qual foi o valor máximo de acurácia no método RDA.
-  - Linha correspondente à melhor acurácia obtida, com a quantidade total de bandas disponíveis. Fornece a combinação de (λ, γ) , e qual foi o valor máximo de acurácia final no método RDA.
-  - Linha correspondente aos resultados obtidos com o método LDA.

O fato da classe *Woods* possuir um comportamento espectral muito distinto das demais classes causou resultados com acurácia superior a 99% (Fig.37).

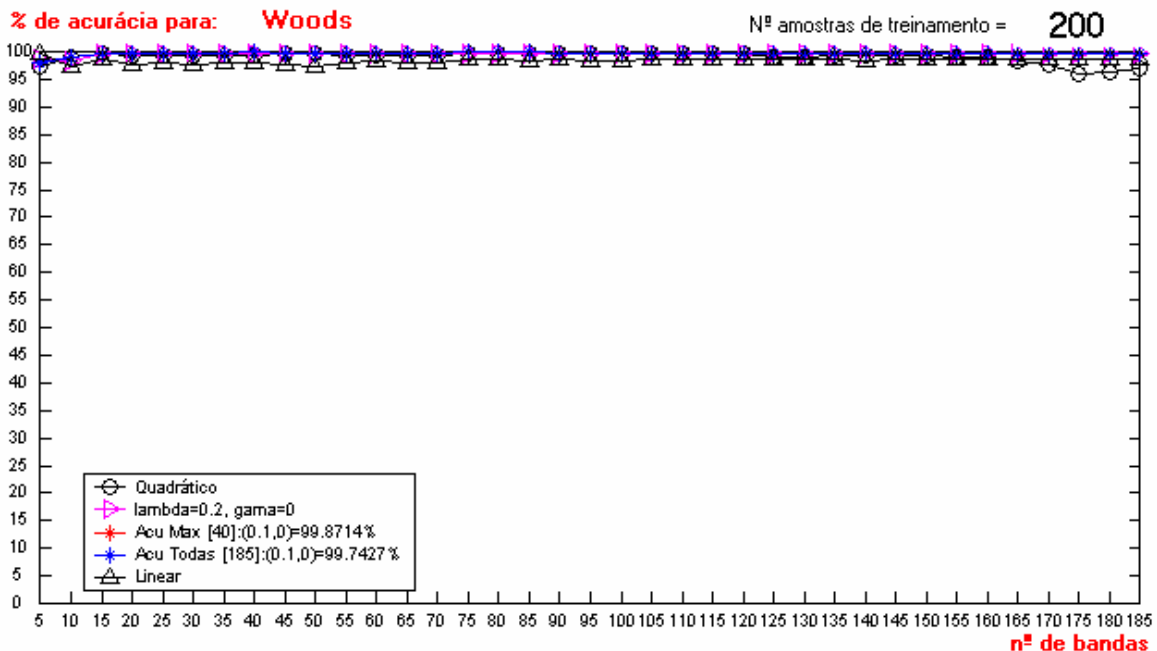


Fig. 37 - Gráfico de acurácia das amostras teste da classe Woods, com estimação de parâmetros usando 200 amostras de treinamento

Como esta classe foi inserida nos testes apenas para avaliar a performance do método em ambiente de fácil classificação, todos os experimentos resultaram em alta acurácia, independente do número de amostras de treinamento e da dimensionalidade dos dados. Sendo assim, não é necessária a inclusão dos demais resultados desta classe dos experimentos, reservando mais atenção e espaço para os resultados mais expressivos.

Resultados obtidos com 200 amostras de treinamento

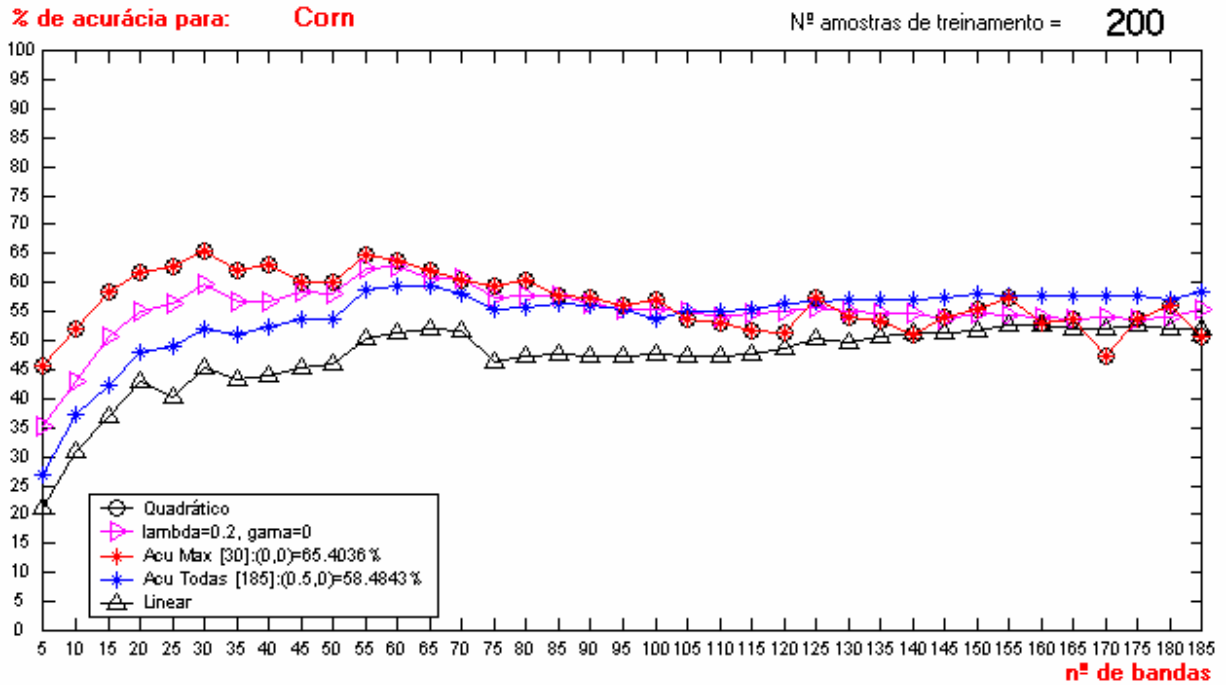


Fig. 38 - Gráfico de acurácia das amostras teste da classe Corn, com estimação de parâmetros usando 200 amostras de treinamento

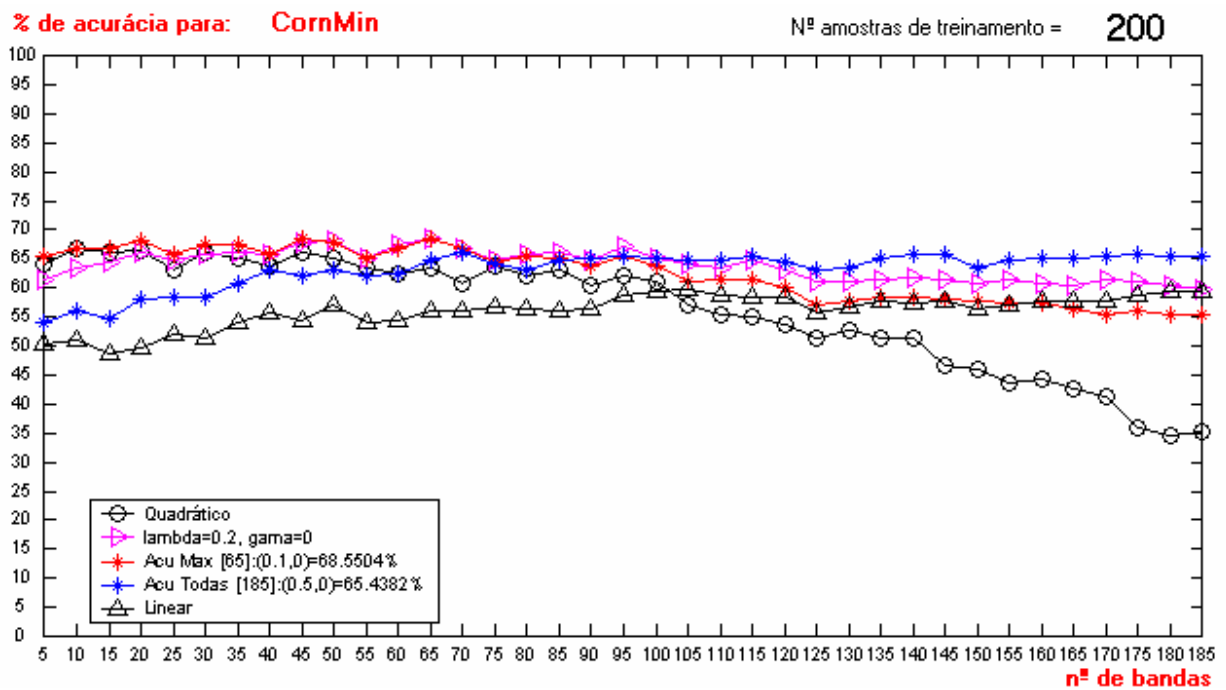


Fig. 39 - Gráfico de acurácia das amostras teste da classe Corn-min, com estimação de parâmetros usando 200 amostras de treinamento

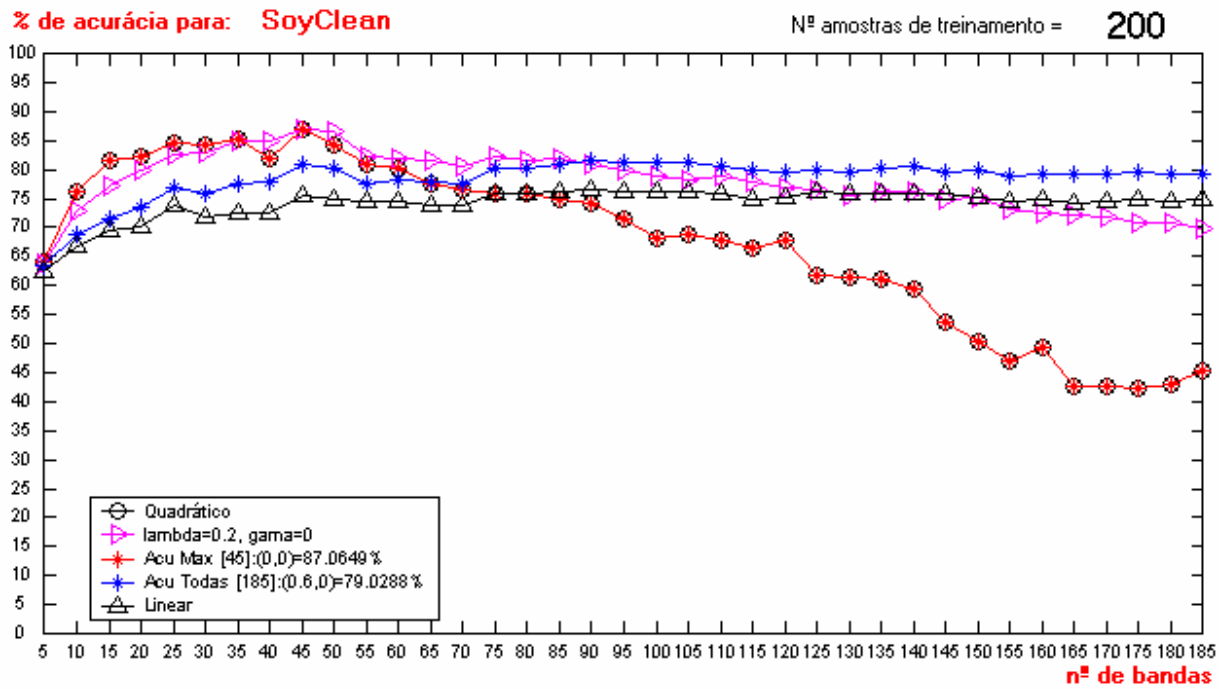


Fig. 40 - Gráfico de acurácia das amostras teste da classe Soy-clean, com estimação de parâmetros usando 200 amostras de treinamento

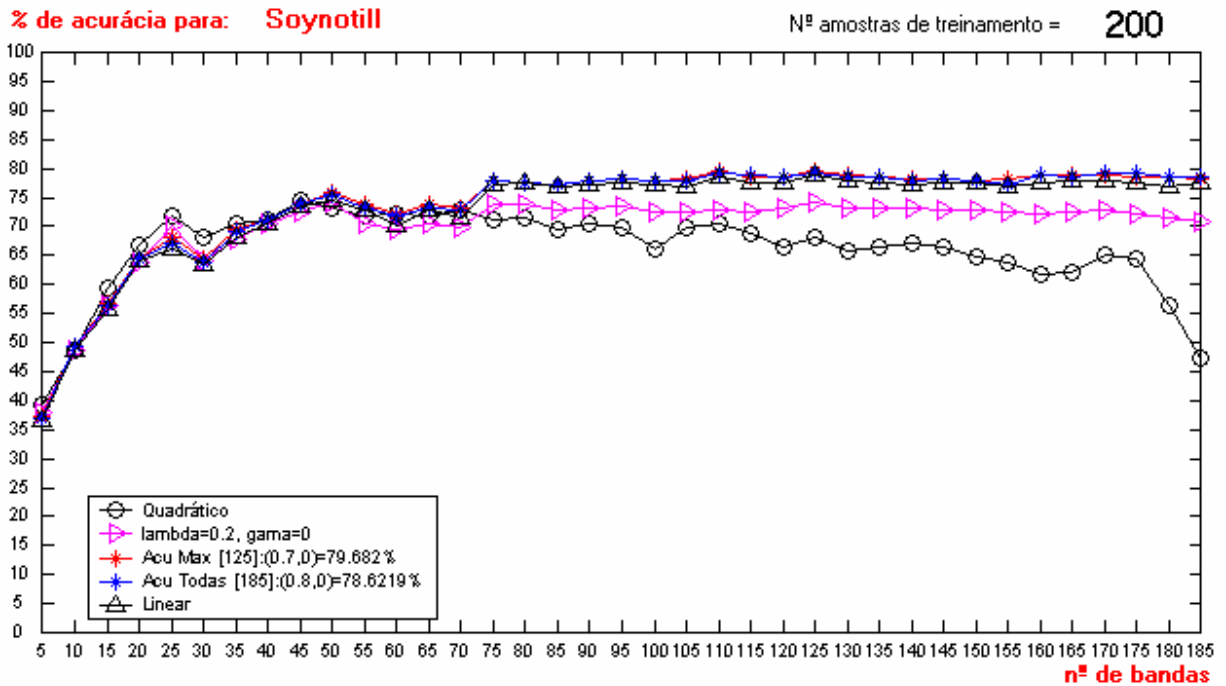


Fig. 41 - Gráfico de acurácia das amostras teste da classe Soy-notill, com estimação de parâmetros usando 200 amostras de treinamento

Tabela 11 - Acurácia máxima dos classificadores com dimensionalidade total (185) para cada classe com 200 amostras de treinamento

Nome da classe	QDA	LDA	RDA	λ	γ	d
Corn	50.57	51.94	58.48	0.5	0.0	185
Corn-min	35.38	59.54	65.43	0.5	0.0	185
Soy-clean	45.33	74.94	79.02	0.6	0.0	185
Soy-notill	47.17	77.56	78.62	0.8	0.0	185

Tabela 12 - Acurácia máxima alcançada, com dimensionalidade “ d ” com 200 amostras

Nome da classe	Maior Acurácia	Modo	λ	γ	d
Corn	65.40	QDA	0.0	0.0	30
Corn-min	68.55	RDA	0.1	0.0	65
Soy-clean	87.06	QDA	0.0	0.0	45
Soy-notill	79.68	RDA	0.7	0.0	125

Os gráficos apresentados, mostram a acurácia para cada classe para diferentes combinações dos parâmetros (λ, γ) no classificador RDA. Para o classificador QDA o Efeito de Hughes ficou bem evidenciado, isto é, a partir de um ponto, à medida que a dimensionalidade dos dados aumenta, a acurácia decresce. Contudo, as melhores taxas de acerto das classes Corn e Soy-clean foram obtidas com o método QDA (sobreposição da linha vermelha sobre a com círculos pretos), atingindo, com 30 bandas, uma acurácia de 65.40%, e com 45 bandas, uma acurácia de 87.06%, respectivamente, porém a acurácia cai drasticamente com o aumento da dimensionalidade dos dados. Já, a classe Corn-min obteve a melhor acurácia com o método RDA com 65 bandas de 68.55% onde $\lambda = 0.1$ e $\gamma = 0.0$, e a classe Soy-notill com 125 bandas obteve acurácia de 79.68% com $\lambda = 0.7$ e $\gamma = 0.0$.

Outra característica muito importante é que o método LDA, apesar de ser inferior ao método QDA neste tipo de ambiente, aumenta a sua acurácia à medida em que aumenta o número de bandas espectrais utilizadas, fazendo com que a sua acurácia final seja muito superior ao método QDA, apesar de o método QDA atingir uma taxa de acerto maior, e com um número menor de bandas.

A inclusão da combinação de $\lambda = 0.2$ e $\gamma = 0.0$ nos gráficos tem por objetivo ilustrar que com valores pequenos de regularização da matriz covariância das classes, é possível obter uma melhora na acurácia de classificação. Em todas as classes, nota-se a diferença dos gráficos de acurácia do método QDA em relação a combinação de λ e γ . São duas linhas completamente diferentes, onde, com apenas uma pequena influência do parâmetro λ , é possível mitigar o Efeito de Hughes que ocorre no método QDA.

4.4.2 Experimento 02

Neste experimento, foram acrescentados mais 300 amostras de treinamento aleatórias para cada classe, totalizando agora, 500 pixels de treinamento para a estimação dos parâmetros (μ_k e Σ_k). Conforme ilustrado na tabela 13

Tabela 13 - Quantidade de amostras das classes selecionadas, utilizadas no Experimento 02, para 500 amostras

Nome da classe	Qtde pixel treinameto	Qtde pixel teste
Corn	500	1821
Corn-min	500	1221
Soy-clean	500	2327
Soy-notill	500	1132
Woods	500	2332

Assim como no Experimento 01, os resultados obtidos pelos classificadores QDA e LDA estão dispostos nas tabelas de contingência 14 e 15, geradas pela ferramenta ADRIH.

Tabela 14 - Tabela de contingência do método QDA – 500 amostras

	Corn	CornMin	SoyClean	Soynotill	Woods
Corn	1248	166	333	114	0
CornMin	268	879	105	89	0
SoyClean	179	48	1813	39	0
Soynotill	125	127	75	888	4
Woods	0	0	0	1	2327
AMOSTRAS:					
treinamento	500	500	500	500	500
teste	1820	1220	2326	1131	2331
% Acertos (Classe)	68.5338%	71.9902%	77.9115%	78.4452%	99.7856%

Bandas → 185
 λ → 0
 γ → 0

Tabela 15 - Tabela de contingência do método LDA – 500 amostras

	Corn	CornMin	SoyClean	Soynotill	Woods
Corn	1103	297	383	73	0
CornMin	301	800	73	119	0
SoyClean	322	37	1828	44	0
Soynotill	94	86	42	895	14
Woods	0	0	0	0	2317
AMOSTRAS:					
treinamento	500	500	500	500	500
teste	1820	1220	2326	1131	2331
% Acertos (Classe)	60.5711%	65.5201%	78.5561%	79.0636%	99.3568%

Bandas → 185
 λ → 1
 γ → 0

Resultados obtidos com 500 amostras de treinamento

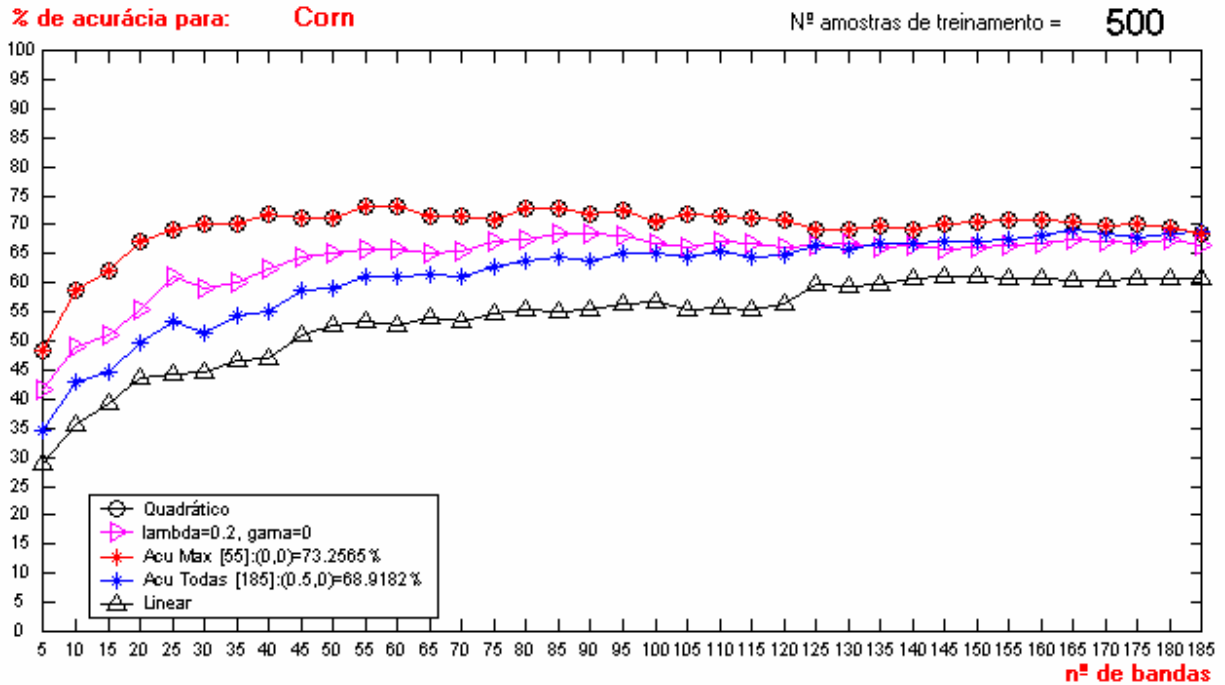


Fig. 42 - Gráfico de acurácia das amostras teste da classe Corn, com estimação de parâmetros usando 500 amostras de treinamento

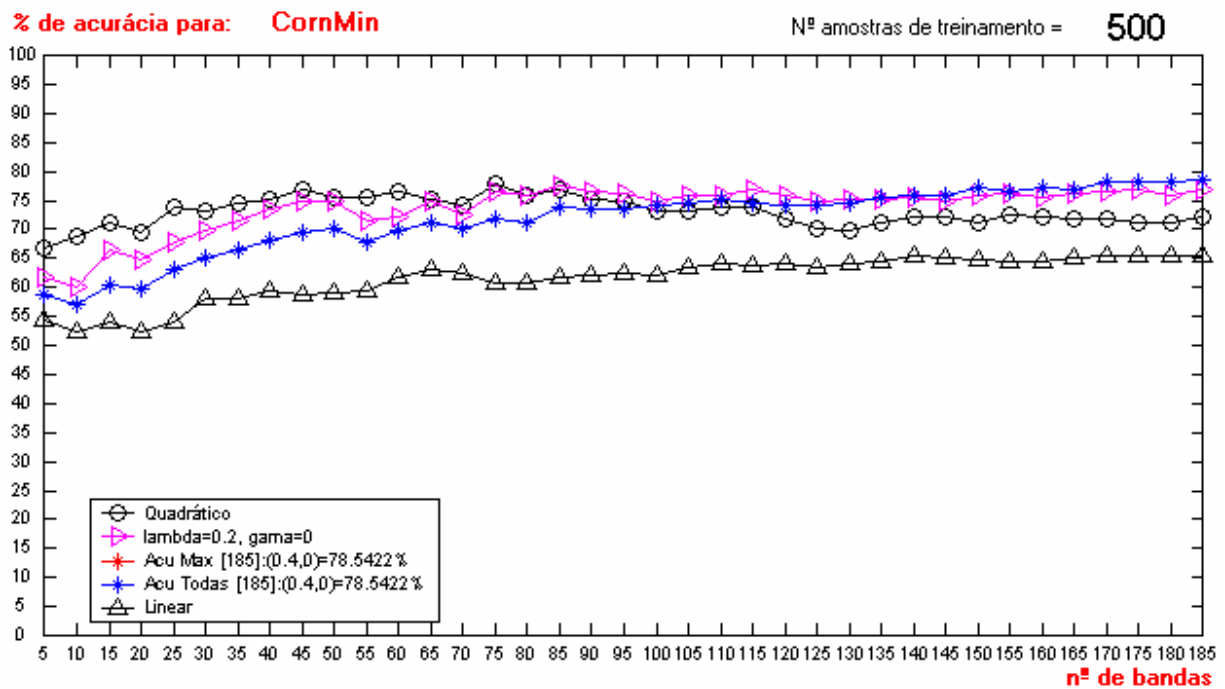


Fig. 43 - Gráfico de acurácia das amostras teste da classe Corn-min, com estimação de parâmetros usando 500 amostras de treinamento

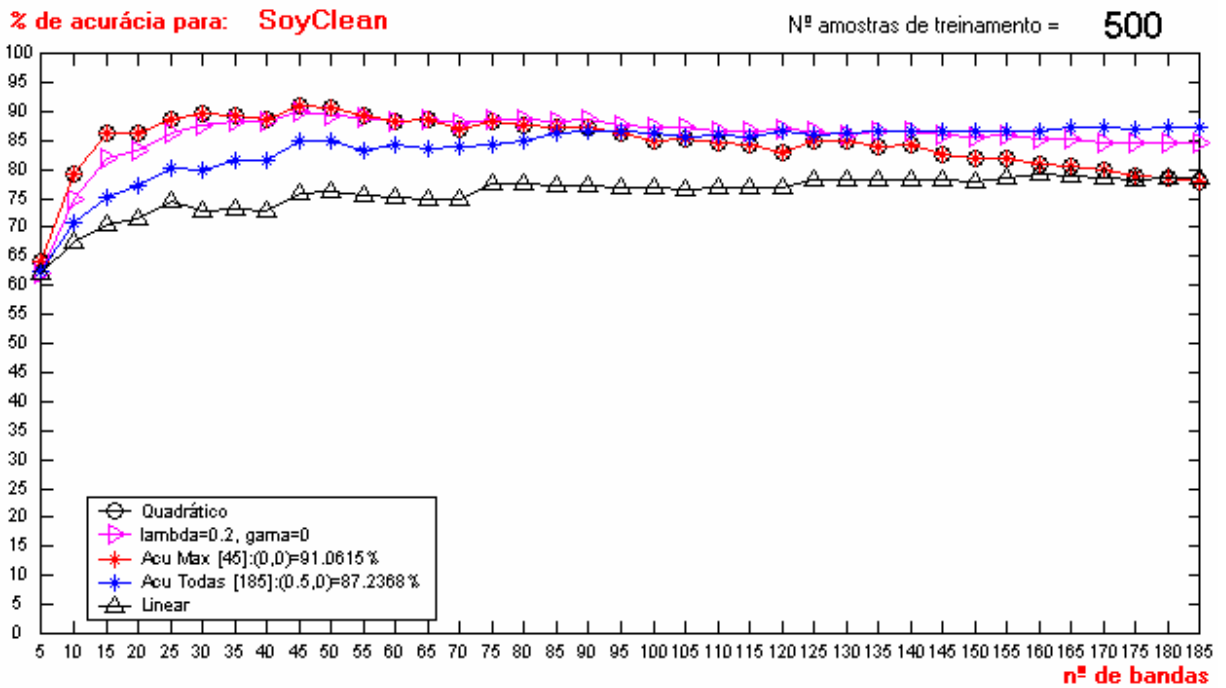


Fig. 44 - Gráfico de acurácia das amostras teste da classe Soy-clean, com estimação de parâmetros usando 500 amostras de treinamento

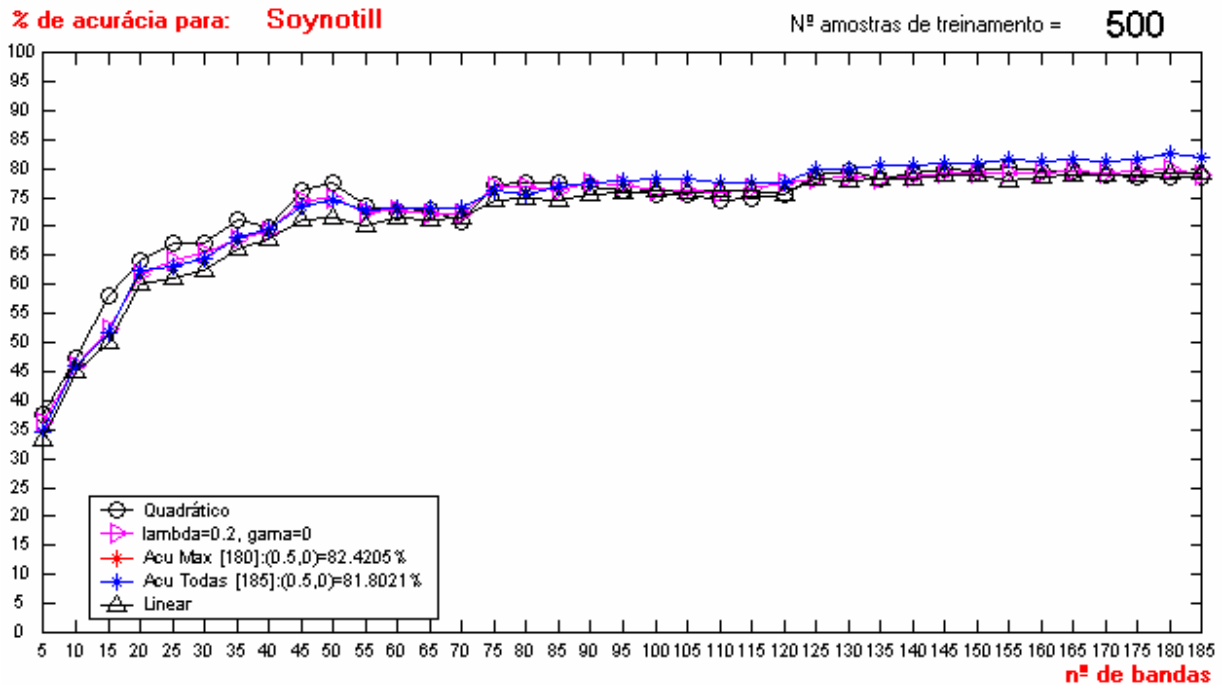


Fig. 45 - Gráfico de acurácia das amostras teste da classe Soy-notill com estimação de parâmetros usando 500 amostras de treinamento

Tabela 16 - Acurácia máxima dos classificadores com dimensionalidade total (185) para cada classe com 500 amostras

Nome da classe	QDA	LDA	RDA	λ	γ	d
Corn	68.53	60.57	68.91	0.5	0.0	185
Corn-min	71.99	65.52	78.54	0.4	0.0	185
Soy-clean	77.91	78.55	87.23	0.5	0.0	185
Soy-notill	78.44	79.06	81.80	0.5	0.0	185

Tabela 17 - Acurácia máxima alcançada, com dimensionalidade “d”, com 500 amostras

Nome da classe	Maior Acurácia	Modo	λ	γ	d
Corn	73.25	QDA	0.0	0.0	55
Corn-min	78.54	RDA	0.4	0.0	185
Soy-clean	91.06	QDA	0.0	0.0	45
Soy-notill	82.42	RDA	0.5	0.0	180

Como seria de esperar, este experimento 02 ilustra o aumento na acurácia dos resultados da classificação, devido ao aumento no tamanho das amostras de treinamento. As características apresentadas no Experimento 01, contudo, ainda são bem visíveis, onde as mesmas classes obtiveram uma melhor acurácia com o método QDA. Corn, com 55 bandas espectrais, obteve 73.25% de precisão e Soy-clean, com 45 bandas espectrais, obteve 91.06% de precisão. As classes restantes obtiveram com o método RDA resultados superiores ao método QDA, Corn-min com a combinação de $\lambda = 0.4$ e $\gamma = 0.0$ atingiu com 185 bandas espectrais um total de 78.54% na precisão de acerto e a classe Soy-notill com $\lambda = 0.5$ e $\gamma = 0.0$ atingiu com 180 bandas espectrais, 82.42%.

Com o aumento do número de amostras de treinamento, a precisão fornecida pelo classificador linear (LDA) decresceu, em comparação àquela fornecida pelo classificador quadrático (QDA). Este comportamento era esperado, pois, à medida em que o número de amostras de treinamento aumenta, melhoram as estimativas dos parâmetros utilizados no classificador quadrático. O fenômeno de Hughes, entretanto, é ainda plenamente visível, o que se explica pelo fato de o número de amostras de treinamento ser pequeno quando comparado com a dimensionalidade dos dados. Nota-se que ainda neste caso, o processo de regularização da matriz covariância demonstrou sua utilidade.

Neste experimento, embora tenha sido utilizado um número maior de amostras de treinamento do que no Experimento 01, ainda assim a relação entre o número de amostras de treinamento e a dimensionalidade dos dados é pequena (Cap. 2.1). Este fato não deve ser esquecido ao analisar-se os resultados obtidos.

4.4.3 Experimento 03

Neste experimento, usou-se o número total de amostras de treinamento disponíveis para cada classe. Portanto, aumentou-se substancialmente o número das amostras de treinamento. Nota-se ainda que, ao contrário dos experimentos anteriores, o tamanho das amostras de treinamento varia de uma classe para outra (Tabela 18).

Tabela 18 - Quantidade de amostras das classes selecionadas, utilizadas no Experimento 03, com tamanhos variados

Nome da classe	Qtde pixel treinameto	Qtde pixel teste
Corn	1821	1821
Corn-min	1221	1221
Soy-clean	2327	2327
Soy-notill	1132	1132
Woods	2332	2332

Neste experimento está portanto presente uma característica que não ocorreu nos experimentos anteriores, que é o tamanho das amostras de treinamento diferentes de cada classe. Este fator irá influenciar nas equações (29), (30), (31) e (32).

A relação, no Experimento 03, entre o número de amostra e a dimensionalidade para cada classe individual, está relacionada na tabela 19.

Tabela 19 - Relação entre o número de amostras de treinamento com a dimensionalidade

Nome da classe	Qtde pixels treinameto	Relação do nº de amostras com a dimensionalidade
Corn	1821	9.84 x
Corn-min	1221	6.6 x
Soy-clean	2327	12.57 x
Soy-notill	1132	6.11 x
Woods	2332	12.60 x

Tabela 20 - Tabela de contingência do método QDA – Amostras de tamanho diferentes

	Corn	CornMin	SoyClean	Soynotill	Woods
Corn	1579	142	65	112	0
CornMin	80	981	5	67	0
SoyClean	143	60	2247	53	0
Soynotill	18	37	9	898	1
Woods	0	0	0	1	2330
AMOSTRAS:					
treinamento	1821	1221	2327	1132	2332
teste	1820	1220	2326	1131	2331
% Acertos (Classe)	86.7106%	80.344%	96.5621%	79.3286%	99.9142%
Bandas → 185					
λ → 0					
γ → 0					

Ou seja, o número de amostras aumenta bastante em relação aos outros testes, permitindo uma melhor estimação dos parâmetros do classificador, e, como consequência, uma melhor precisão nos resultados da classificação. Os resultados estão apresentados nas tabelas de contingência 20 e 21.

Tabela 21 - Tabela de contingência do método LDA – Amostras de tamanho diferentes

	Corn	CornMin	SoyClean	Soynotill	Woods
Corn	1193	325	365	68	0
CornMin	236	802	53	107	0
SoyClean	317	36	1880	46	0
Soynotill	74	57	28	910	9
Woods	0	0	0	0	2322
AMOSTRAS:					
treinamento	1821	1221	2327	1132	2332
teste	1820	1220	2326	1131	2331
% Acertos (Classe)	65.5135%	65.6839%	80.7907%	80.3887%	99.5712%
Bandas → 185					
λ → 1					
γ → 0					

Os gráficos resultantes da acurácia do métodos QDA, LDA e RDA no experimento 03, são ilustrados na figuras 46, 47, 48 e 49.

Resultados obtidos com diferentes tamanhos de amostras de treinamento

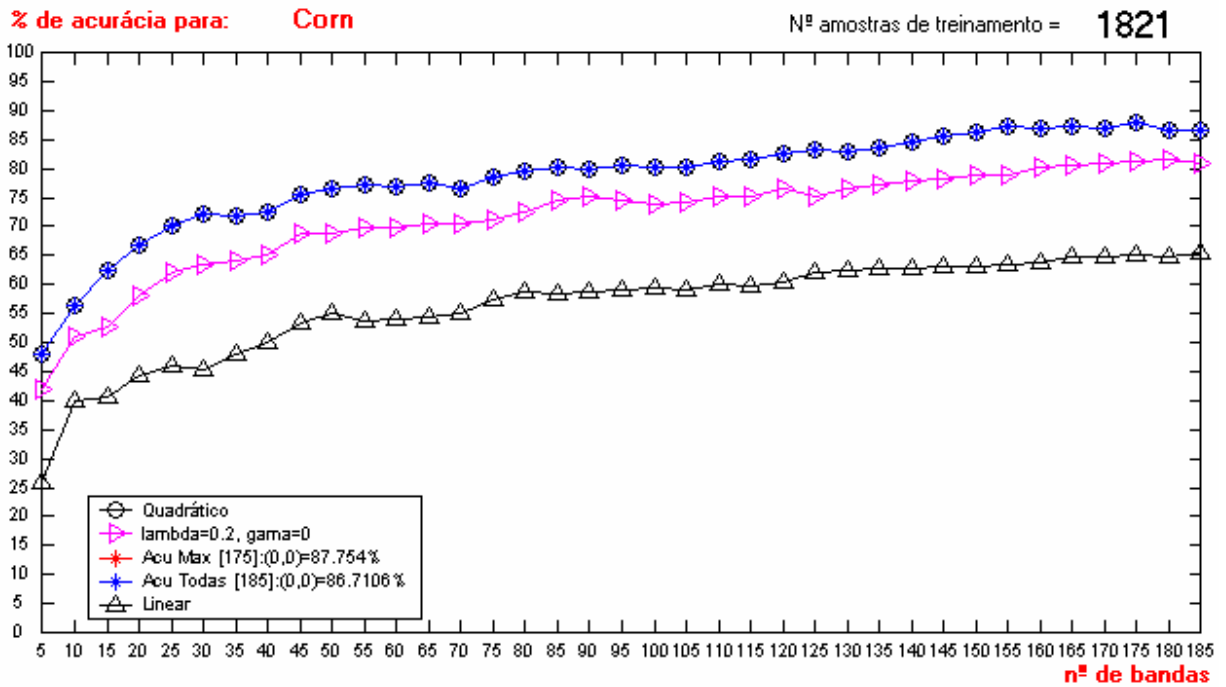


Fig. 46 - Gráfico de acurácia das amostras teste da classe Corn, com estimação de parâmetros usando 1821 amostras de treinamento

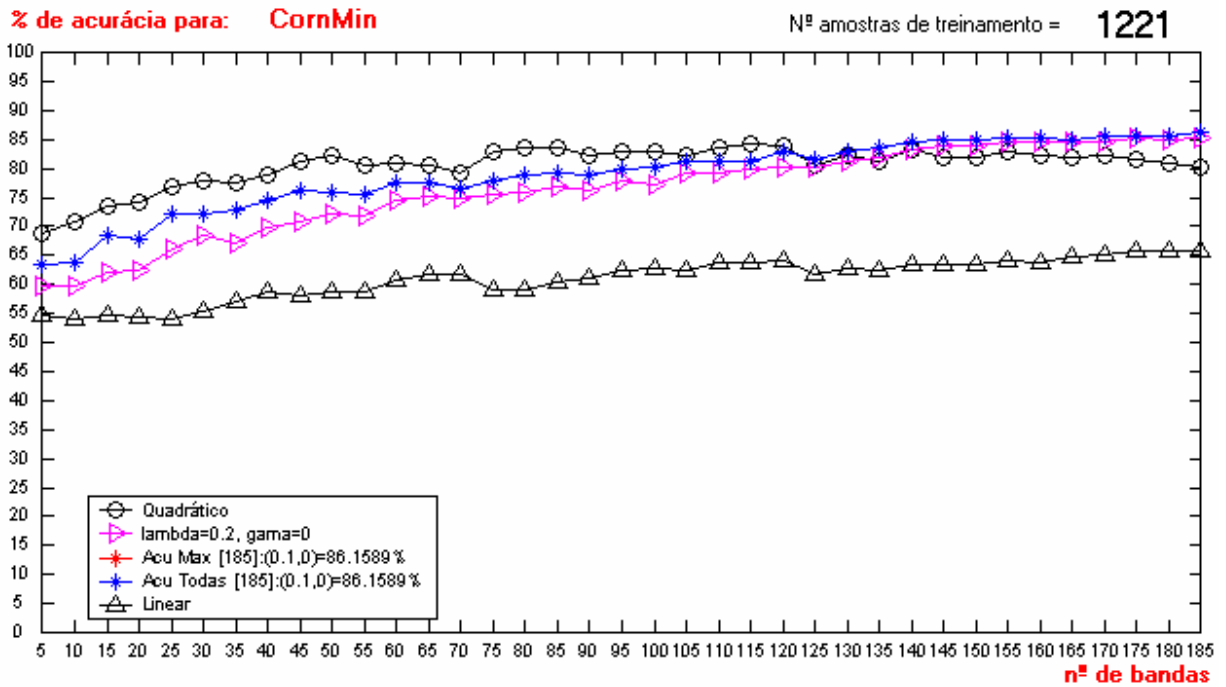


Fig. 47 - Gráfico de acurácia das amostras teste da classe Corn-min, com estimação de parâmetros usando 1221 amostras de treinamento

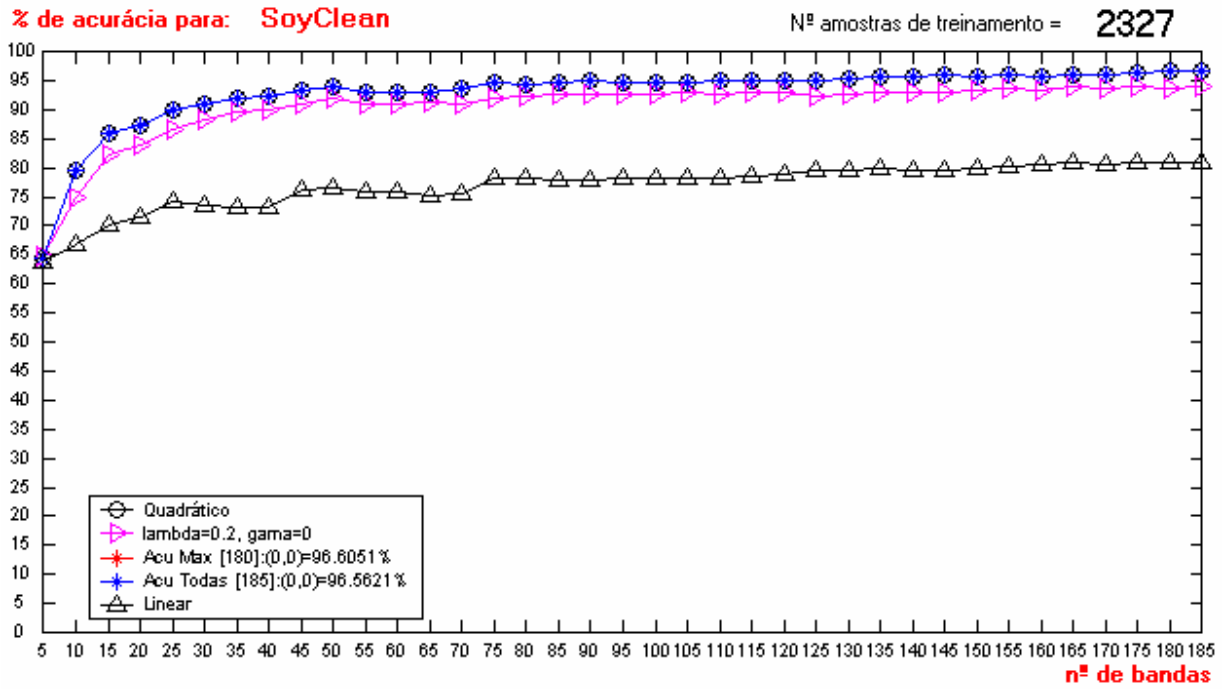


Fig. 48 - Gráfico de acurácia das amostras teste da classe Soy-clean, com estimação de parâmetros usando 2327 amostras de treinamento

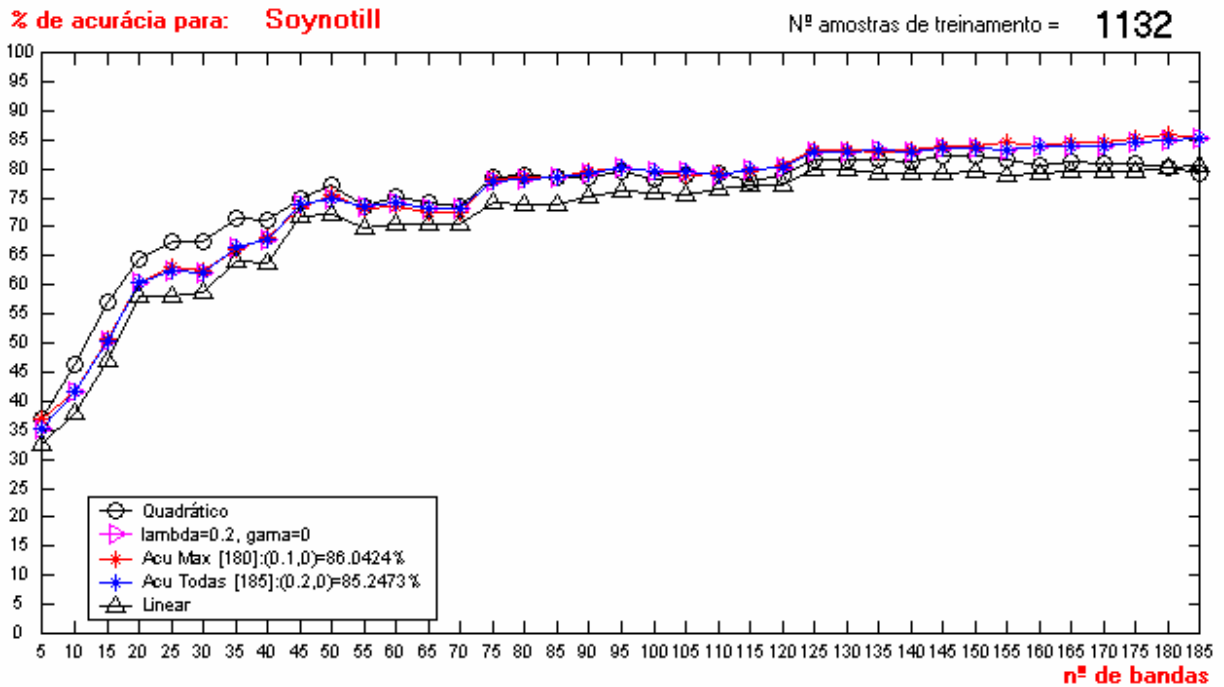


Fig. 49 - Gráfico de acurácia das amostras teste da classe Soy-notill, com estimação de parâmetros usando 1132 amostras de treinamento

Tabela 22 - Acurácia máxima dos classificadores com dimensionalidade total (185) para cada classe

Nome da classe	QDA	LDA	RDA	λ	γ	d
Corn	86.71	65.51	86.71	0.0	0.0	185
Corn-min	80.34	65.68	86.15	0.1	0.0	185
Soy-clean	96.56	80.79	96.56	0.0	0.0	185
Soy-notill	79.32	80.38	85.24	0.2	0.0	185

Tabela 23 - Acurácia máxima alcançada, com “d” dimensionalidade

Nome da classe	Maior Acurácia	Modo	λ	γ	d
Corn	87.75	QDA	0.0	0.0	175
Corn-min	86.15	RDA	0.1	0.0	185
Soy-clean	96.60	QDA	0.0	0.0	180
Soy-notill	86.04	RDA	0.1	0.0	180

Nota-se que no modelo QDA, o Efeito de Hughes ainda está presente, como seria de esperar nas classes para as quais o tamanho de amostras de treinamento é menor (Corn-min e Soy-notill). Neste contexto, verifica-se que o objetivo proposto pelo método RDA foi alcançado: a acurácia máxima nestas classes foi de 86,15% e 86,04%, respectivamente. Nota-se que estes resultados foram obtidos com uma pequena contribuição fornecida pelo parâmetro λ , isto é, o valor de regularização λ foi de 0.1.

Já nas outras duas classes Corn e Soy-clean, cujas amostras de treinamento são maiores que as anteriores, o comportamento da precisão do processo de classificação é distinto. A precisão aumenta com a dimensionalidade dos dados, atingindo um patamar para um número de bandas espectrais entre 115 e 150. Neste contexto, os valores de $\lambda = 0$ e $\gamma = 0$

para o método RDA, que neste caso é o classificador quadrático, obtiveram os melhores resultados: para Corn, com 175 bandas, uma acurácia de 87.75% e para Soy-Clean, com 180 bandas, 96,60% de acurácia.

O aumento de amostra de treinamento colabora para uma melhor estimação de parâmetros, e conseqüentemente uma melhor acurácia nos experimentos. Deste modo, a porcentagem de acertos aumenta muito em relação aos demais experimentos. Isto comprova que, para configurações que possuam um número pequeno de amostras para estimação de parâmetros, o classificador RDA é uma alternativa plausível para o problema da redução da acurácia, provocada pelo Efeito de Hughes.

5 CONCLUSÕES/ SUGESTÕES

Atualmente, um grande desafio em classificação de imagens é separar classes que apresentam vetores média iguais ou muito próximos (soja e milho, por exemplo). Uma solução para este problema consiste na utilização de dados em alta dimensionalidade. Pode-se provar [FUK 90] que é possível separar classes espectralmente muito semelhantes entre si, desde que se utilize dados em alta dimensionalidade e, ainda, que as matrizes covariância entre as classes envolvidas sejam razoavelmente distintas. Esta possível solução para melhoria no processo de classificação de imagens digitais apresenta entretanto uma dificuldade. Na medida em que a dimensionalidade dos dados aumenta, aumentam também o número de parâmetros a serem estimados para utilização do classificador. Em situações reais, entretanto, o número de amostras de treinamento disponíveis para estimação de parâmetros é muito pequena, insuficiente para uma estimação acurada dos parâmetros. Este fato resulta numa performance insatisfatória do classificador. Sabe-se, através de estudos [AEB 94], que métodos que reduzem a dimensionalidade, como métodos de extração de feições, tendem a obter resultados inferiores, se comparados aos resultados obtidos em experimentos usando técnicas de regularização, que fazem uso de toda a informação disponível. Esta dissertação testa um processo de se mitigar este problema de classificação associada à alta dimensionalidade através da regularização da matriz covariância das classes pelo método RDA e aplicada a dados de sensoriamento remoto.

Com base nos resultados expostos e análises realizadas nos experimentos apresentados no capítulo 4, conclui-se que:

- O método de regularização (RDA) mostrou ser eficiente no processo de classificação de dados de sensoriamento remoto em alta dimensão (imagens hiperespectrais).
- Mesmo com tamanho pequeno de amostras de treinamento no experimento 01, nota-se a vantagem do uso do método RDA. Em todos os demais casos, mesmo naqueles em que não foi o processo que forneceu a melhor acurácia na classificação, o método RDA pelo menos suavizou, consideravelmente, o Efeito de Hughes.
- Como é de conhecimento que o aumento de amostras de treinamento no experimento 02 melhora a acurácia de classificação, nota-se que o processo RDA melhora a acurácia na medida em que a dimensionalidade aumenta. Isto não ocorre com o método QDA, o qual se deteriora significativamente com o incremento de bandas espectrais ocasionado pelo Efeito de Hughes.
- O classificador RDA mostrou-se adequado para fins de classificação em sensoriamento remoto, principalmente em classes de comportamento espectral muito semelhantes, e que são de difícil discriminação quando no uso de técnicas convencionais.
- Os experimentos colocaram em evidência um fato com respeito ao comportamento do parâmetro γ : os melhores resultados em termos de acurácia no processo de classificação foram aqueles para os quais $\gamma = 0$. Em outras palavras, ao contrário dos resultados apresentados por Aeberhard [AEB 94], os experimentos realizados neste estudo, e que envolvem classes espectralmente muito semelhantes, sugerem que o parâmetro γ tem pouca relevância no processo de regularização. Uma possível

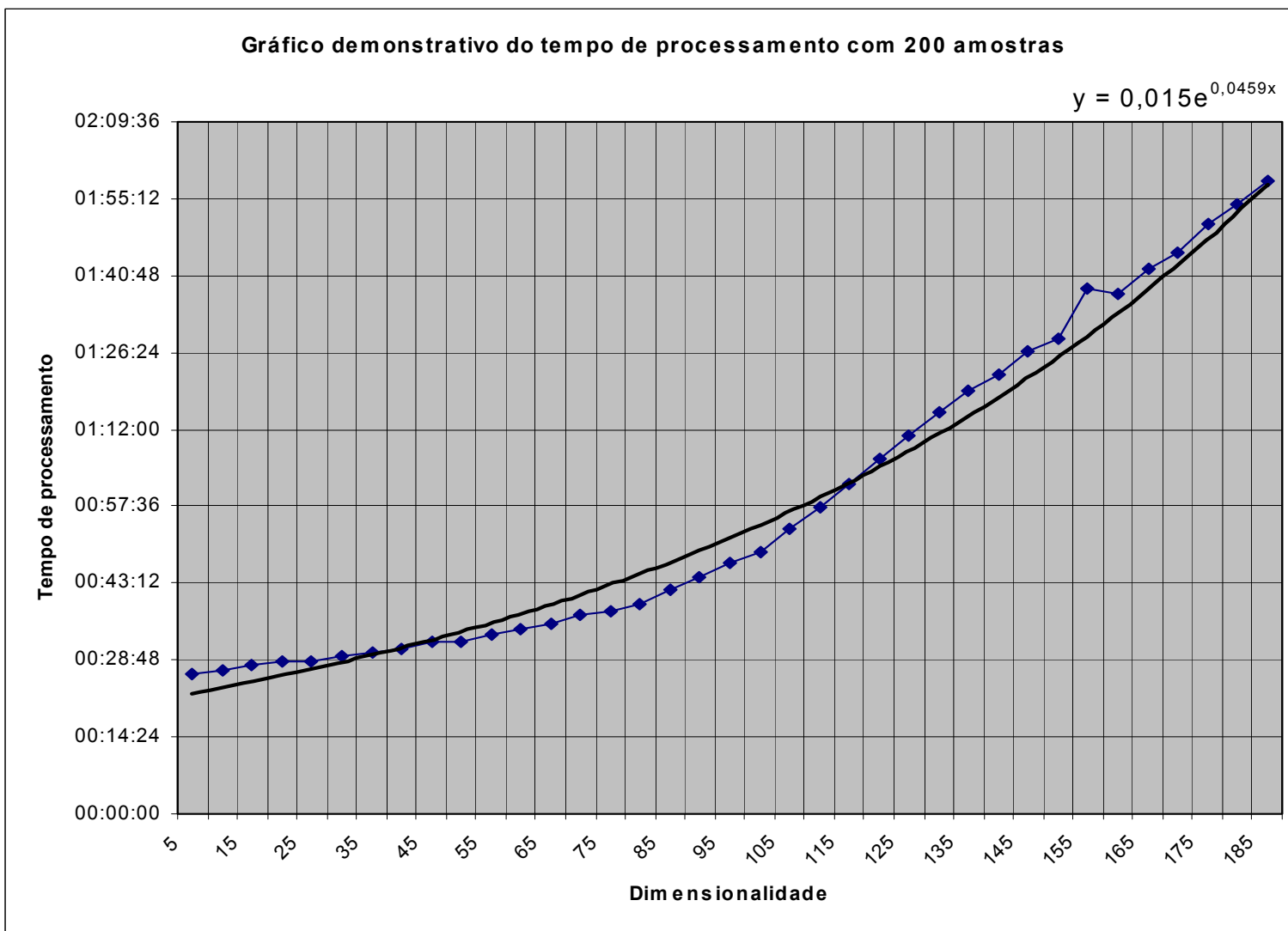
explicação para este inesperado resultado consiste no seguinte: o parâmetro γ tende a contrabalançar a tendenciosidade na estimação dos autovalores da matriz covariância. Sempre que o número de amostras é pequeno, quando comparado com a dimensionalidade dos dados, os menores autovalores são estimados ainda menores, o oposto ocorrendo com os autovalores maiores. Em outras palavras, a aplicação de γ tende a tornar o elipsóide Gaussiano menos elíptico, isto é, menos alongado e com uma dimensão um pouco maior na direção delgada. É um fato bem conhecido que em sensoriamento remoto, as classes estão freqüentemente distribuídas ao longo da direção de máxima dispersão dos dados. Incidentalmente, esta é a razão pela qual a Transformada de Karhunen-Loeve (Componentes Principais), que é baseada na matriz covariância comum, funciona geralmente bem como um redutor de dimensionalidade. Em outras palavras, os autovetores associados aos maiores autovalores em cada classe são aproximadamente paralelos entre si. Nestas condições, e sendo os pontos centrais dos hiperelipsóides (vetor média) muito próximos entre si, o alargamento do hiperelipsóides individuais tende a aumentar a região de superposição e portanto o erro de Bayes. Este fato, peculiar as classes de sensoriamento remoto envolvidas neste estudo pode ser uma explicação do comportamento do parâmetro γ no processo de regularização da matriz covariância neste particular estudo, diferentemente do comportamento deste parâmetro reportado por outros autores.

- O custo computacional para encontrar as melhores configurações de λ e γ no processamento das amostras nos experimentos, vai crescendo à medida que aumenta a dimensionalidade, como pode ser observado nos anexos 1,2,3.

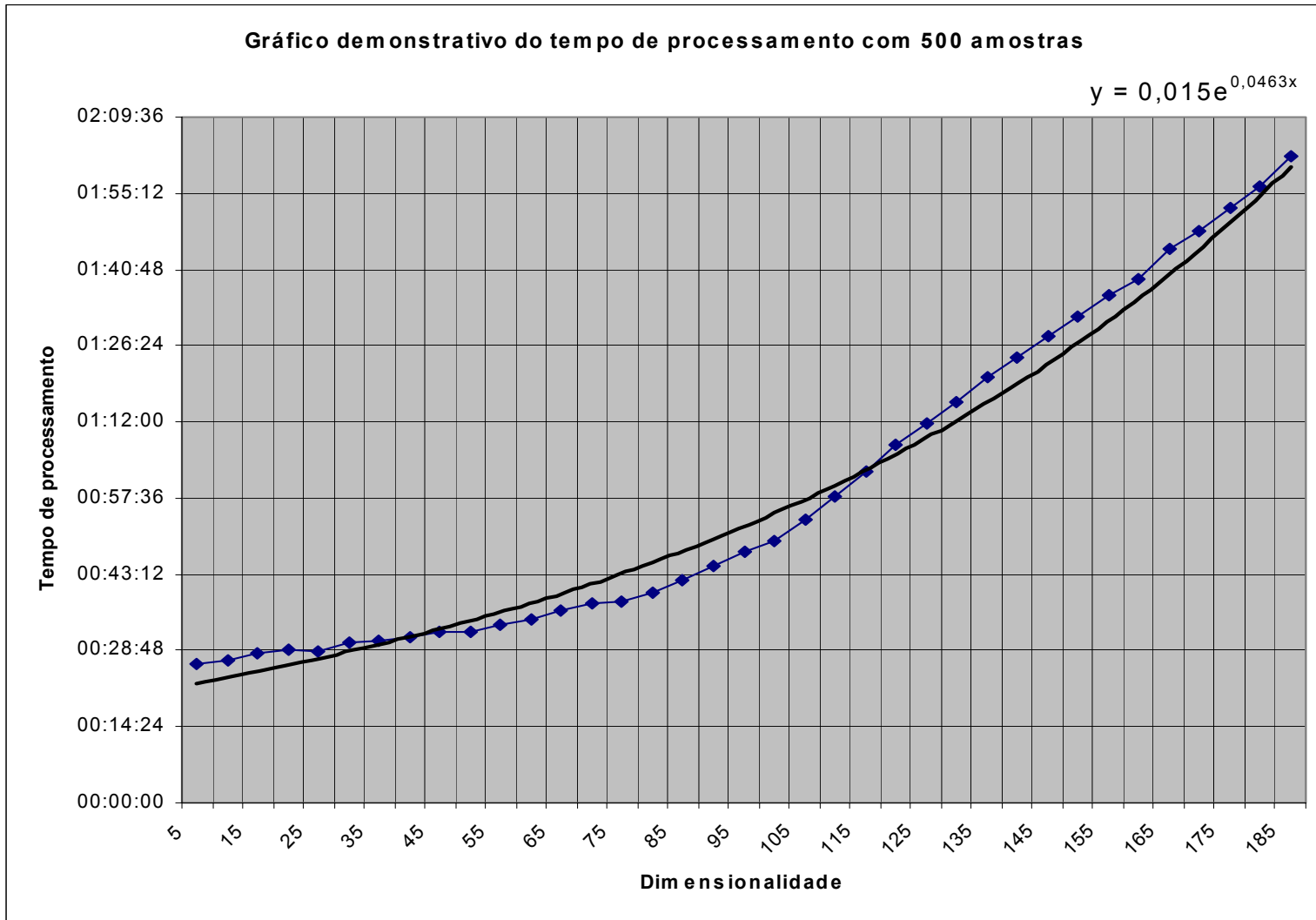
Como sugestões para trabalhos futuros, pode-se citar:

- Buscar um melhor entendimento com relação à contribuição do parâmetro γ para a separação de classes com comportamento espectral muito semelhantes usando dados reais em alta dimensionalidade de sensoriamento remoto (imagens hiperespectrais).
- Estudar um método mais rápido para obtenção das melhores configurações de λ e γ , a fim de se aplicar o método RDA, com mais rapidez.
- Otimização do software ADRIH, pela adição de novas técnicas para fins de comparação de performance dos resultados obtidos, como por exemplo: EM, discriminação logística, método do semi-variograma e outros.

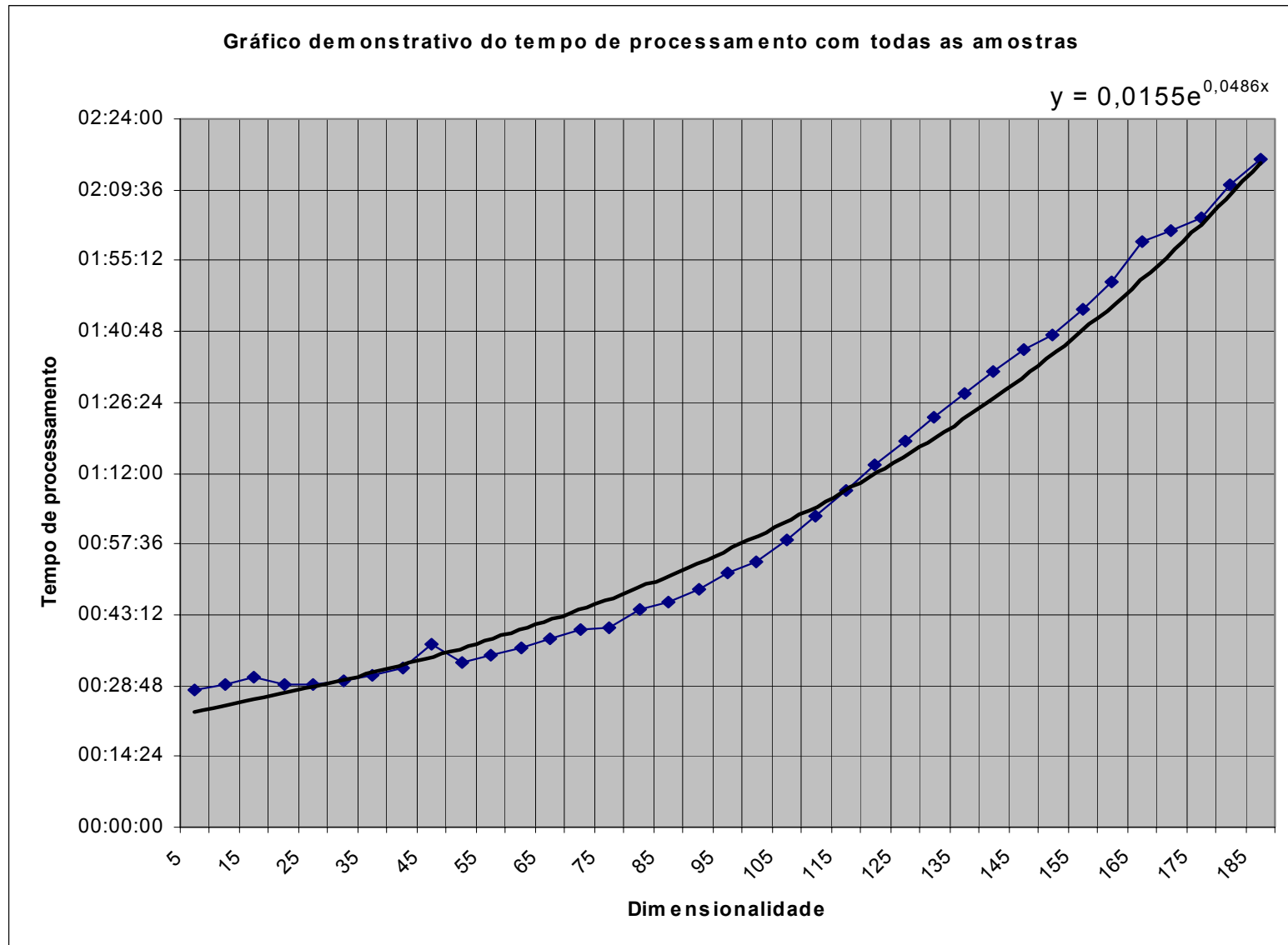
ANEXO 1



ANEXO 2



ANEXO 3



ANEXO 4

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

MESTRADO

MANUAL ADRIH

(Análise Discriminante Regularizada em Imagens Hiperespectrais)

POR
MAURO ERBERT

PROF. PhD VITOR HAERTEL
ORIENTADOR

PORTO ALEGRE,
2001

SUMÁRIO

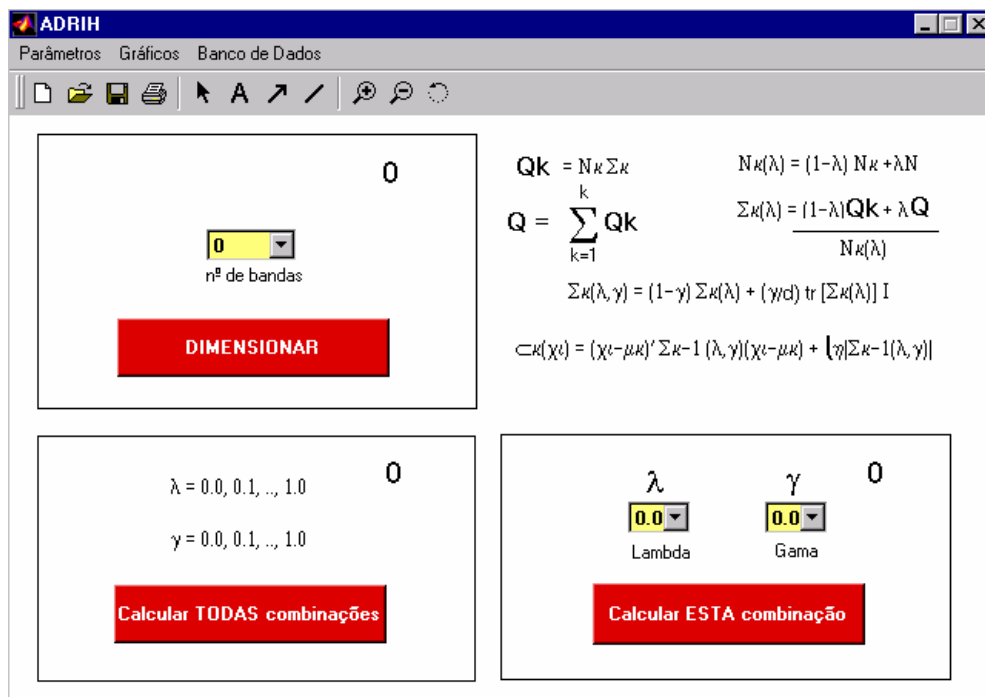
APRESENTAÇÃO DO SOFTWARE ADRIH.....	108
JANELA PRINCIPAL	109
MENUS.....	110
FUNCIONAMENTO	113
EXEMPLO DE FUNCIONAMENTO	113
OBSERVAÇÕES	114

APRESENTAÇÃO DO SOFTWARE ADRIH

Este manual tem por finalidade apresentar ao usuário o funcionamento do software ADRIH desenvolvido em MATLAB 5.3, no laboratório de Sensoriamento Remoto do Centro Estadual de Pesquisas em Sensoriamento Remoto e Meteorologia – CEP SRM/UFRGS.

Inicialmente o projeto visava implementar um algoritmo capaz de processar os dados e analisar a acurácia de resultados obtidos, usando a técnica de regularização da matriz covariância em dados de natureza real, provenientes do sensor hiperespectral AVIRIS. Contudo, a medida que a necessidade exigia, foram necessários incrementos, o que acarretou num conjunto maior de técnicas e algoritmos. Para melhor gerenciar estes algoritmos, e conseqüentemente otimizar o processo de obtenção de resultados, partiu-se para a utilização do ambiente gráfico disponibilizado pelo MATLAB, chamado *guide*. O ambiente gráfico do *guide*, é um ambiente orientado a eventos, com a possibilidade de criação de botões, menus, barras, inclusão de gráficos, e etc.

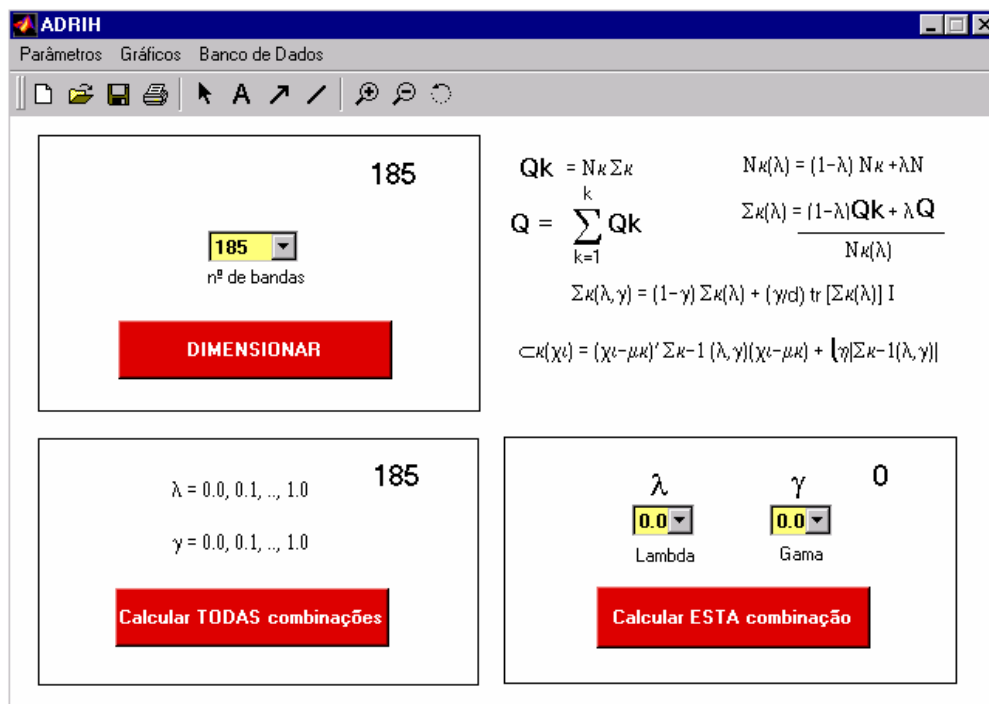
A tela inicial do programa ADRIH é a seguinte:



JANELA PRINCIPAL

Setores da janela principal do ADRIH :

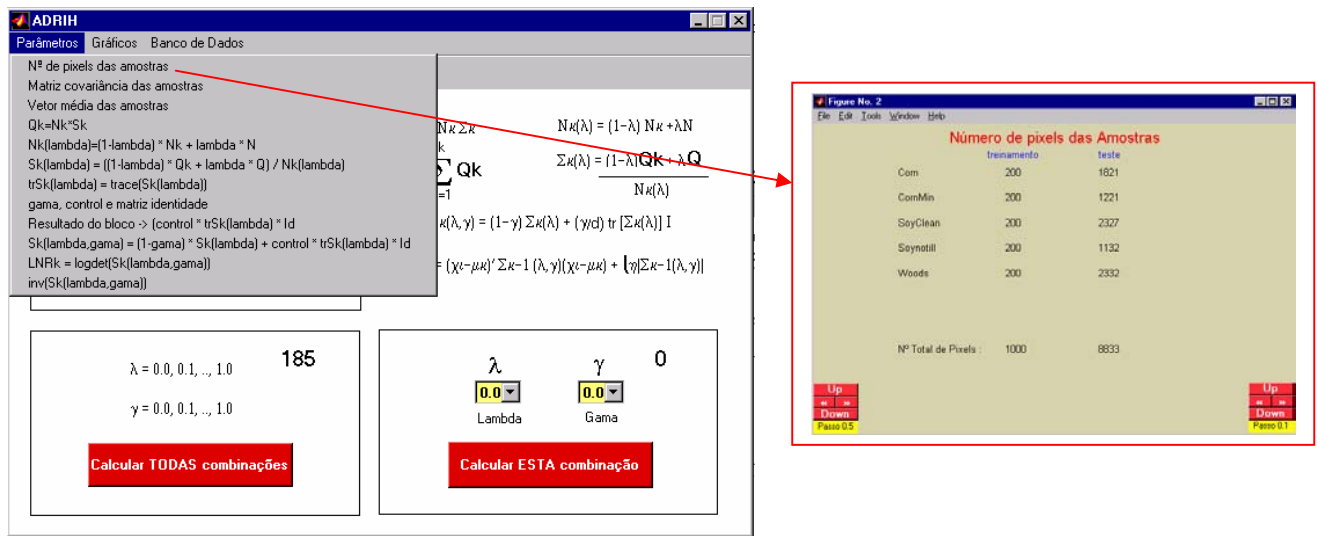
- 1°. Superior esquerdo: serve para selecionar o número de bandas que serão utilizadas no processamento. Quando o programa é executado no MatLab, ele automaticamente verifica qual a dimensionalidade atual selecionada nos arquivos das amostras de treinamento. Para mudar, basta selecionar na caixa a quantidade de bandas desejadas e clicar <<Dimensionar>>;
- 2°. Superior direito: serve apenas como informativo para o usuário a respeito da evolução do método aplicado;
- 3°. Inferior esquerdo: automaticamente, o programa mostra qual foi o último tamanho de dimensionalidade processado com esta função. Ao clicar o botão, o programa calcula todas as interações entre λ e γ na dimensão escolhida;
- 4°. Inferior direito: Também, automaticamente, o programa mostra qual foi o último tamanho de dimensionalidade processado (com esta função). Escolhe-se a combinação de λ e γ desejada, a dimensionalidade atual e calcula esta combinação. Ao final será mostrada a Tabela de Contingência da performance na classificação das classes. Este processo é bom para verificação caso-a-caso.



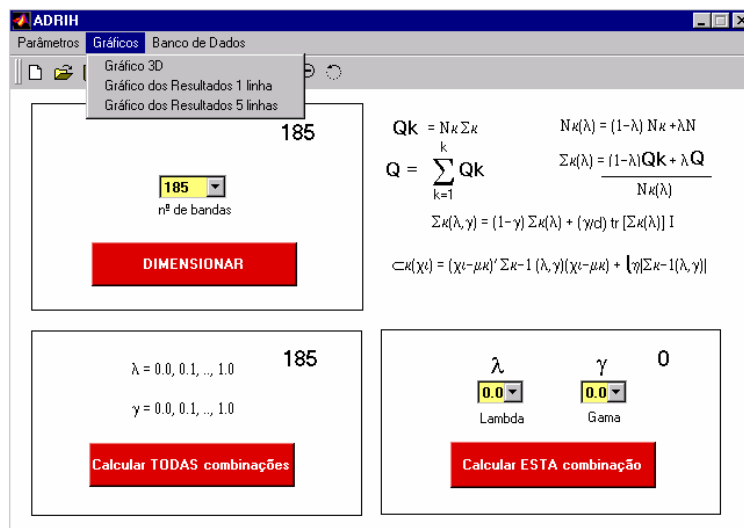
MENUS

Os menus do ADRIH compreendem 3 opções diferentes:

- O 1º menu (*Parâmetros*) disponibiliza funções que habilitam o usuário visualizar as etapas da evolução, do processo, no algoritmo e na geração dos parâmetros necessários para o método (nº pixels, μ , Σ , traço, matriz inversa), como mostrado na janela superior à direita. Esta janela serve para o usuário acompanhar o que está sendo feito passo-a-passo e identificar o que está acontecendo no processo, com a aplicação de λ e γ na equação;

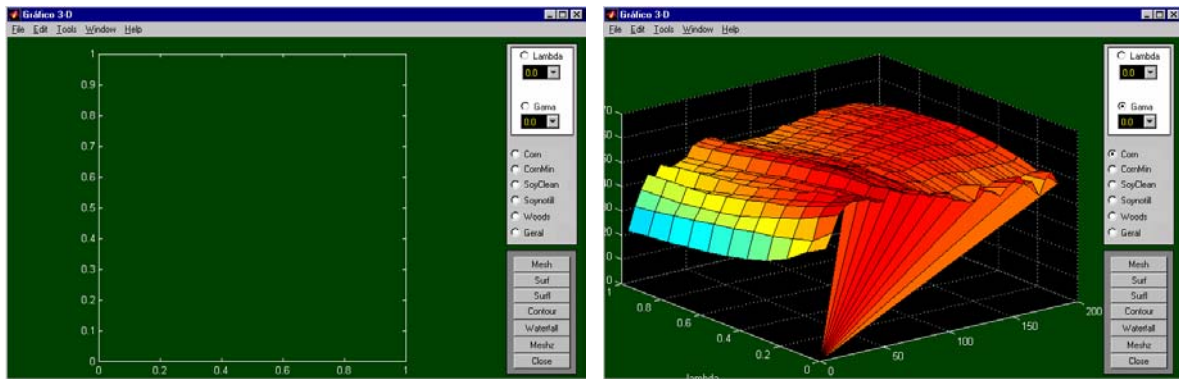


- O 2º menu (*Gráficos*) é usado para a visualização dos resultados obtidos, na forma de gráficos 2D e 3D. As opções “b” e “c” estão disponíveis ((a) *Gráfico 3D* (em construção), (b) *Gráfico dos resultados 1 linha*, (c) *Gráfico dos resultados 5 linhas*). Na opção (a) tem-se uma noção tri-dimensional dos resultados, em todas as combinações de lambda e gama. Em (b) é possível verificar o comportamento da curva de acurácia dos resultados no decorrer do aumento da dimensionalidade, e na opção (c), obter uma melhor noção de como se comportam os resultados da acurácia dos métodos RDA.



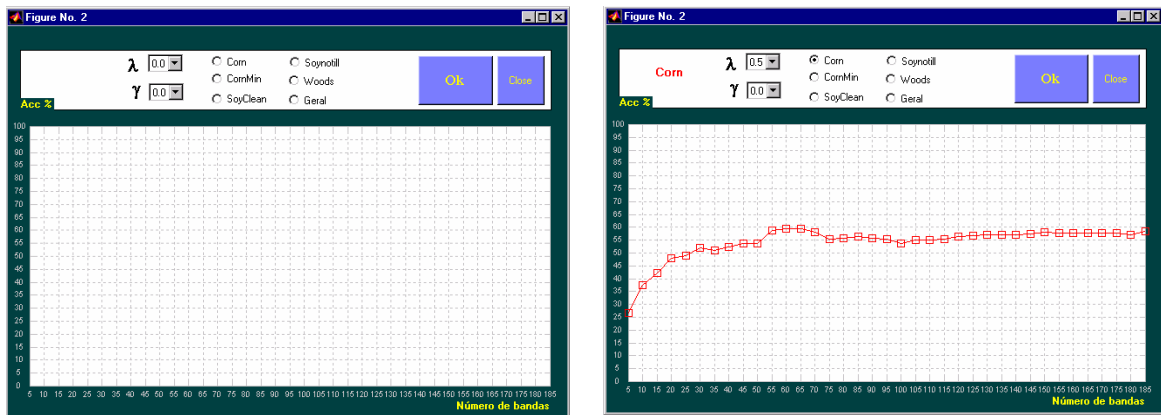
No modo 3D, o usuário deve escolher um ponto fixo de lambda ou gama, uma classe e clicar “Ok”. Será gerado um gráfico com todos os resultados entre o ponto-fixe de lambda ou gama, com todas as outras combinações do parâmetros não-selecionado, em todas as dimensionalidades. Identifica-se o Efeito de Hughes em 3D.

(a) visão inicial da janela, e visão com o gráfico 3D



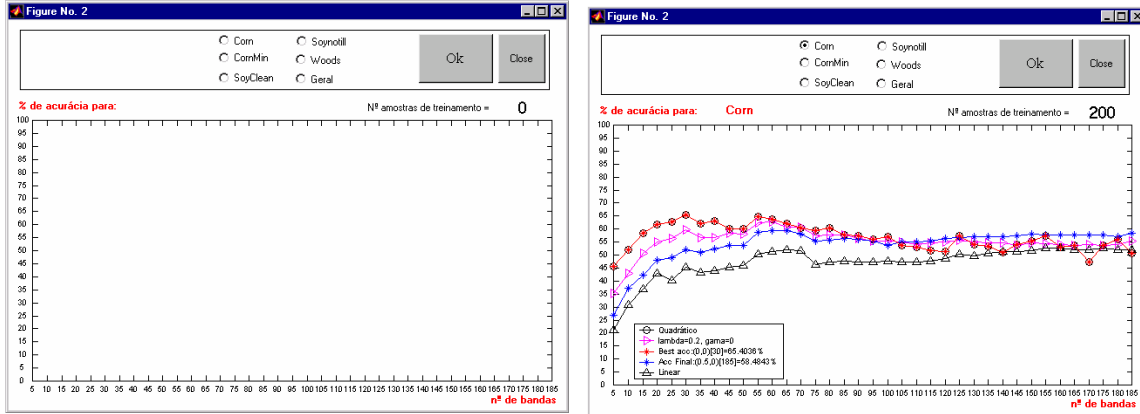
No modo 2D da opção (b), o usuário deve escolher uma combinação específica de λ e γ , selecionar uma classe e clicar “Ok”. Será gerado um gráfico com a combinação desejada em toda a dimensionalidade. Serve para a identificação do Efeito de Hughes.

(b) visão inicial da janela e visão com o gráfico 2D de uma linha



No modo 2D da opção (c), o usuário deve escolher uma classe e clicar “Ok”. Será gerado um gráfico com 5 curvas de acurácia: QDA, Combinação baixa de λ e γ , Melhor acurácia em toda a extensão da dimensionalidade, Melhor acurácia ao final com o número total de bandas, e a LDA. Serve para a identificação da mitigação do Efeito de Hughes.

(c) visão inicial da janela, e visão com o gráfico 2D de 5 linhas



O 3º menu (*Banco de Dados*), disponibiliza 2 opções. A 1ª opção inicializa (zera) a base de dados dos resultados das interações. Deste modo, toda vez que se trocar as classes ou se usar a 2ª opção do menu *Banco de Dados*, para mudar o número de amostras de treinamento para estimação de parâmetros, será necessário inicializar a base de dados.

ADRIH

Parâmetros Gráficos Banco de Dados

Inicializa Banco de Dados
Troca o nº de amostras de treinamento

185
nº de bandas
DIMENSIONAR

$Q_k = N_k \Sigma_k$ $N_k(\lambda) = (1-\lambda) N_k + \lambda N$
 $Q = \sum_{k=1}^k Q_k$ $\Sigma_k(\lambda) = \frac{(1-\lambda) Q_k + \lambda Q}{N_k(\lambda)}$
 $\Sigma_k(\lambda, \gamma) = (1-\gamma) \Sigma_k(\lambda) + (\gamma d) \text{tr} [\Sigma_k(\lambda)] I$
 $C_k(\chi_i) = (\chi_i - \mu_k)' \Sigma_k^{-1}(\lambda, \gamma) (\chi_i - \mu_k) + \frac{1}{2} \ln |\Sigma_k^{-1}(\lambda, \gamma)|$

$\lambda = 0.0, 0.1, \dots, 1.0$ 185
 $\gamma = 0.0, 0.1, \dots, 1.0$
Calcular TODAS combinações

λ γ 0
 0.0 0.0
 Lambda Gama
Calcular ESTA combinação

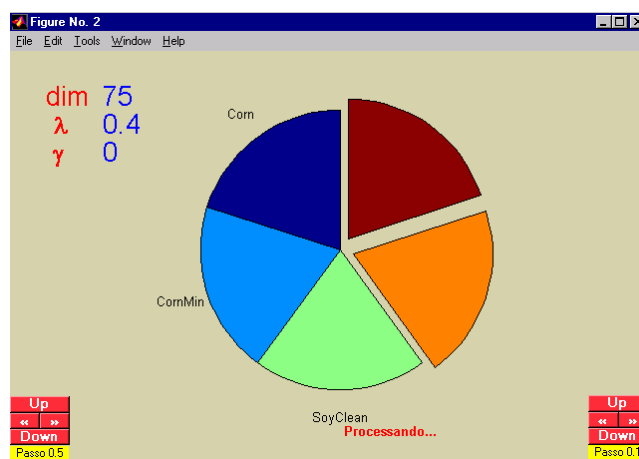
FUNCIONAMENTO

O funcionamento do ADRIH, é realizado em 3 etapas:

1. Preparo das informações, condicionando-as para o uso do software, ou seja, num formato de matriz [pixel, bandas];
2. Limpeza do sistema, através da inicialização da base de dados (menu “Banco de Dados”), preparando, deste modo, o armazenamento dos resultados a serem obtidos;
3. Manuseio das informações, segundo os processos do método RDA;
4. Seleção do número de bandas desejadas;
5. Escolha do modo a ser aplicado, isto é, se o modo deve calcular todas as combinações possíveis de lambda e gama, ou selecionar a combinação desejada (sempre será válido o último processamento realizado, independente de qual modo tenha sido escolhido);
6. À medida que são realizados os processamento dos dados, é possível acompanhar os resultados através das funções do gráficos disponíveis pelo sistema;

EXEMPLO DE FUNCIONAMENTO

O exemplo a seguir demonstra o processamento de uma combinação, onde $\lambda = 0.4$ e $\gamma = 0$, com dimensionalidade = 75. Para acompanhar a execução do algoritmo, é visualizado um círculo, como demonstrado abaixo, onde cada pedaço do círculo representa uma classe. A medida que há o término de processamento do conjunto de amostras de teste da classe, este pedaço junta-se com os outros; logo, os pedaços que estão disjuntos representam as classes que ainda não terminaram de processar.



Ao final do processamento, o programa gera uma tabela de contingência, que serve para visualizar os resultados da acurácia do classificador, com a configuração escolhida.

	Corn	CornMin	SoyClean	Soynotill	Woods
Corn	1224	265	188	81	0
CornMin	194	838	53	125	0
SoyClean	347	47	2062	46	0
Soynotill	55	70	23	879	7
Woods	0	0	0	0	2324
AMOSTRAS:					
treinamento	1821	1221	2327	1132	2332
teste	1820	1220	2326	1131	2331
% Acertos (Classe)	67.2158%	68.6323%	88.6119%	77.6502%	99.6569%
Bandas →	75				
λ →	0.4				
γ →	0				

OBSERVAÇÕES

Sendo o software ADRIH um sistema experimental, existem determinadas limitações, bem como:

- À medida que a dimensionalidade dos dados aumenta, a visualização dos resultados intermediários, como vetor média e matriz covariância, por exemplo, ficam prejudicados;
- Necessidade de uma melhor estruturação das opções do sistema de geração de gráficos 3D;

BIBLIOGRAFIA

- [AEB 94] AEGERHARD, S. and COOMANS, D. e De Vel, O. Comparative Analysis of Statistical Pattern Recognition Methods in High Dimensional Settings. **Pattern Recognition**, Vol. 27, nº 8, pp. 1065-1077, 1994.
- [ANG 00] ANGELO, N. P. **Aplicação de filtros de Gabor no processo de classificação de imagens digitais com base em atributos de textura**, Centro Estadual de Pesquisa em Sensoriamento Remoto e Meteorologia – CEPSRM, UFRGS, Porto Alegre, 2000
- [BIT 01] BITTENCOURT H. R. **Comparação da Discriminação Logística com o Método da Máxima Verossimilhança Gaussiana na Classificação de Imagens Digitais**, Centro Estadual de Pesquisa em Sensoriamento Remoto e Meteorologia – CEPSRM, UFRGS, Porto Alegre, 2001
- [BOW 92] BOW, S.T., **Pattern Recognition and Image Preprocessing**, Department of Electrical Engineering, Northern Illinois University, DeKalb, Illinois, Marcel Dekker Inc, New York – Basel – Hong Kong, 1992.
- [CAR 99] CARR, J. R. and MATANAWI, K., Correspondence Analysis for Principal Components Transformation of Multispectral and Hyperspectral Digital Images, **Photogrammetric Engineering & Remote Sensing**, August 1999.
- [COR 67] CORNFIELD, J., Discriminant Functions, **Review of the International Statistical Institute**, 35, 142-153, 1967.
- [COR 99] CORTIJO F.J. and DE LA BLANCA N. P., The Performance of Regularized Discriminant Analysis Versus Non-parametric Classifiers Applied to High-Dimensional Image Classification, **International Journal of Remote Sensing**, vol.20, nº 17, 3345-3365, 1999.
- [CRO 93] CROSTA, A. P., **Processamento Digital de Imagens de Sensoriamento Remoto**, Instituto de Geociências, Dep. de Metalogêneses e geoquímica, Unicamp, Campinas, SP, 1993.
- [DUD 01] DUDA, R. O., HART P.E., STORK D.G., **Pattern Classification**, 2ª ed., John Wiley & Sons Inc. EUA, 2001.
- [FRE 62] FREUND J.E., **Mathematical Statistics**, New Jersey, Prentice Hall, 1962
- [FRI 89] FRIEDMAN, J. H., Regularized Discrimination Analysis, **Journal of the American Statistical Association**, Theory and Methods, Vol. 84, Nº 405, 1989.

- [FUK 89] FUKUNAGA, K. and HAYES R.R., Effects of Sample Size in Classifier Design, **IEEE Transaction Pattern Analysis and Machine Intelligence**, vol. 11, n°8, pp. 873-885, Aug. 1989.
- [FUK 90] FUKUNAGA, K., **Introduction to Statistical Pattern Recognition**, 2^a ed., School of Electrical Engineering, Purdue University, West Lafayette, Academic Press Inc. 1990.
- [GOE 85] GOETZ, A.F.H., VANE, G., SOLOMON, J.E., ROCK, B.N., Imaging Spectrometry for Earth Remote Sensing, **Science**, 228, no 4704, pp. 1147-1153, 1985.
- [HAE 99] HAERTEL V. and LANDGREBE, D.A., On the Classification of Classes with Nearly equal Spectral Response in Remote Sensing Hyperspectral Image Data, **IEEE Transactions on Geoscience and Remote Sensing**, vol. 37, n°5, pg 2374-2386, september 1999.
- [HAE 00] HAERTEL V., Considerations about Classification of Remote Sensing Digital Image Data in High Dimensional Settings, **IX Simposio Latinoamericano de Perception Remota**, Puerto Iguazú, Misiones, Argentina, Noviembre de 2000.
- [HOF 96] HOFFBECK J.P. and LANDGREBE D. A., Covariance Matrix Estimation and Classification With Limited Training Data, **IEEE Transactions on Pattern Analysis and Machine Intelligence**, vol. 18, n° 7, july 1996.
- [HSI 98] HSIEH, P. and LANDGREBE, D.A., Classification of High Dimensional Data, **TR-ECE 98-4**, School of Elec. And Comp. Eng., Purdue University , May 1998.
- [HUG 68] HUGHES, G. F. On the mean accuracy of statistical pattern recognizers, **IEEE – Transactions an Information Theory**, vol. II – 14, N° 1, pp. 55-65, 1968.
- [JAI 00] JAIN, A., DUIN, R.P.W. and MAO, J., Statistical Pattern Recognition: A Review, **IEEE Transactions on Pattern Analysis and Machine Intelligence**, Vol.22, n° 1, pg 4-37, 2000.
- [JIA 94] JIA, X. and RICHARDS, J., Efficient Maximum Likelihood Classification for Imaging Spectrometer Data Sets, **IEEE Transactions on Geoscience and Remote Sensing**, vol. 32, n°2, pg. 274-280, march 1994.
- [JIA 99] JIA, X. and RICHARDS, J., Segmented Principal Components Transformation for Efficient Hyperspectral Remote-Sensing Image Display and Classification, **IEEE Transactions on Geoscience and Remote Sensing**, vol. 37, n°1, pg. 538-542, January 1999.
- [JIM 96] JIMENEZ, L. O. and LANDGREBE, D.A., High Dimensional Feature Reduction Via Projection Pursuit, **TR-ECE 96-5**, School of Elec. And Comp. Eng., Purdue University , April 1996.

- [JIM 99] JIMENEZ, L. O. and LANDGREBE, D.A., Hyperspectral data Analysis and Supervised Feature Reduction Via Projection Pursuit, **IEEE Transactions on Geoscience and Remote Sensing**, vol. 37, n°6, pg. 2653-2667, november 1999.
- [LAN 97] LANDGREBE, D.A., **Multispectral Data Analysis: A signal Theory Perspective**, School of Elec. And Comp. Eng., Purdue University, July 1997.
- [LANa 97] LANDGREBE, D. A. , On Hyperspectral Data and its Analysis, **EE 577 Notes**, March, 31 1997.
- [OSU 86] O’SULLIVAN, F. , A Statistical Perspective on Ill-Posed Inverse Problems, **Statistical Science**, 1, 502-527, 1986.
- [RIC 86] RICHARDS, J.A., **Remote Sensing Digital Image Analysis – An Introduction**, Centre for Remote Sensing, and School of Electrical Engineering and Computer Science, University of New South Wales, Australia, Springer-Verlag Berlin Heidelberg, 1986
- [SHA 94] SHAHSHAHANI, B. M. and LANDGREBE D. A., 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, n° 5, 1087-1095, sept. 1994
- [SIG 90] SIGILLITO, V. & HUTTON, L. **Case Study II: Radar Signal Processing**. In: Neural Networks PC Tools - A Practical Guide. Eberhart, R. & Dobbins, R. Academic Press, San Diego, 1990.
- [STE 93] STEFFEN, C. A. et al, Tutorial 2 - Radiometria Espectral, **VII Simpósio Brasileiro Sensoriamento Remoto**, Curitiba – Brasil, 1993.
- [TAD 98] TADJUDIN, S. and LANDGREBE, D.A., Classification of High Dimensional Data with Limited Training Samples, **TR-ECE 98-8**, School of Elec. And Comp. Eng., Purdue University, May 1998.
- [THE 89] THERRIEN, C. W., **Decision Estimation and Classification – An Introduction to Pattern Recognition and Related Topics**, Naval Postgraduate School, John Wiley & Sons Inc. EUA, 1989.
- [TIT 85] TITTERINGTON, D. M., Common Structure of Smoothing Techniques in Statistics, **International Statistical Review**, 53, 141-170, 1985.
- [TOU 74] TOU J. and GONZALES R. C., **Pattern Recognition Principles**, 3^a ed. 1979, Advanced Book Program, Massachusetts, Reading: Addison-Wesley Publishing Company Inc. 1974.

Referências eletrônicas (Internet):

- [CAR 97] CARDON A., MÜLLER D. N., **Introdução às Redes Neurais Artificiais**, disponível em: <http://www.ulbra.tche.br/~danielnm/ia/rna/rna.html>
- [INP 00] INPE - Instituto de Pesquisas Espaciais – **Simulações com o sensor AVIRIS no Brasil**, disponível em : <http://www.inpe.br/programas/cbers/english/imagens.html>
- [INP 01] INPE - Instituto de Pesquisas Espaciais - **Introdução ao Sensoriamento Remoto**, disponível em <http://www.dpi.inpe.br/spring/usuario/sensorr.htm#espectro>
- [JPL 00] JPL – Jet Propulsion Laboratory – **NASA – What’s AVIRIS?** disponível em <http://makalu.jpl.nasa.gov/>
- [LEW 94] LEWOTSKY, K, Hyperspectral imaging: evolution of imaging spectrometry, **OE Reports**, November 1994 issue, disponível em http://www.spie.org/web/oer/november/image_spectro.html
- [ORB 01] ORBIMAGE – Orbital Imaging Corporation, disponível em: <http://www.orbimage.com/index.html>
- [PRÄ 01] PRÄSS. A. R. PRASStek -**TEACHING BY INTERNET** Porto Alegre, RS – Brasil, disponível em : http://www.terra.com.br/fisicanet/cursos/ondas_eleto_magneticas/espectro_eleto_magnetico.html