

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

**CLASSIFICAÇÃO DE DADOS IMAGEM EM ALTA  
DIMENSIONALIDADE, EMPREGANDO AMOSTRAS SEMI-ROTULADAS  
E ESTIMADORES PARA AS PROBABILIDADES *A PRIORI***

CEL SO ANTONIO LICZBINSKI

Orientador: Prof<sup>o</sup> Dr. Vitor Haertel

Porto Alegre (RS), Agosto de 2007.

Dissertação de Mestrado apresentada  
ao Programa de Pós-Graduação em  
Sensoriamento Remoto, como requisito  
para a obtenção do título de Mestre.

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

**CLASSIFICAÇÃO DE DADOS IMAGEM EM ALTA  
DIMENSIONALIDADE, EMPREGANDO AMOSTRAS SEMI-ROTULADAS  
E ESTIMADORES PARA AS PROBABILIDADES *A PRIORI***

Celso Antonio Liczbinski  
Bacharel em Informática (2001-UNISINOS)

ORIENTADOR: Prof<sup>o</sup> Dr. Vitor Haertel

BANCA EXAMINADORA: Prof<sup>o</sup> Dr. Dante Augusto Couto Barone - UFRGS  
Prof<sup>o</sup> Dr. Jorge Antonio Silva Centeno - UFPR  
Prof<sup>o</sup> Dr. Yosio Edemir Shimabukuro - INPE

Dissertação de Mestrado apresentada ao Programa de Pós-Graduação em Sensoriamento Remoto do Centro Estadual de Pesquisas em Sensoriamento Remoto e Meteorologia da Universidade Federal do Rio Grande do Sul, como requisito parcial para a obtenção do título de Mestre em Sensoriamento, Área de Desenvolvimento de Novas Metodologias.

Porto Alegre  
2007

## **DEDICATÓRIA**

*“Ao mestre da criação pela oportunidade, capacidade e perseverança, e aos meus pais Francisco Liczbinski e Ana Milani, primeiros grandes mestres, que embora distante, neles sempre encontrei inspiração para transformar momentos difíceis em simples realidade.”*

## **AGRADECIMENTOS**

Ao término deste trabalho, quero sinceramente agradecer a todas as pessoas e Instituições que contribuíram para a sua realização, em especial ao Prof<sup>o</sup> Vitor Haertel pela oportunidade e confiança, cuja orientação competente e dedicada ao trabalho e à ciência possibilitou que este estudo se materializasse; aos demais professores e colegas que de alguma forma participaram para a sua realização; aos familiares e amigos, pelo incentivo e, principalmente pela compreensão e paciência em saber esperar, nos diversos momentos em que precisei me ausentar de suas agradáveis companhias; aos funcionários do CEP SRM; à sociedade brasileira, através desta Universidade e do Exército Brasileiro - Diretoria do Serviço Geográfico - 1<sup>a</sup> Divisão de Levantamento; a minha eterna gratidão.

# CLASSIFICAÇÃO DE DADOS IMAGEM EM ALTA DIMENSIONALIDADE, EMPREGANDO AMOSTRAS SEMI-ROTULADAS E ESTIMADORES PARA AS PROBABILIDADES *A PRIORI*\*

Celso Antonio Liczbinski  
Orientador: Prof. Vitor Haertel, PhD.

## RESUMO

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 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 freqüentemente 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 que utilizam amostras não rotuladas e amostras semi-rotuladas para minimizar o problema do tamanho reduzido das amostras de treinamento. Deste modo, técnicas que utilizam amostras semi-rotuladas, 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 foi dado prosseguimento à metodologia investigada por Lemos (2003), o qual implementou a utilização de amostras semi-rotuladas para fins de estimação dos parâmetros do classificador Máxima Verossimilhança Gaussiana (MVG). A contribuição do presente trabalho consistiu na inclusão de uma etapa adicional, introduzindo a estimação das probabilidades *a priori*  $P(\omega_i)$  referentes às classes envolvidas para utilização no classificador MVG. Desta forma, utilizando-se funções de decisão mais ajustadas à realidade da cena analisada, obteve-se resultados mais acurados no processo de classificação. Os resultados atestaram que com um número limitado de amostras de treinamento, técnicas que utilizam algoritmos adaptativos, mostram-se eficientes em reduzir o Efeito de Hughes. Apesar deste Efeito, quanto à acurácia, em todos os casos o modelo quadrático mostrou-se eficiente através do algoritmo adaptativo. A conclusão principal desta dissertação é que o método do algoritmo adaptativo é útil no processo de classificação de imagens com dados em alta dimensionalidade e classes com características espectrais muito próximas.

**Palavras-chave:** Sensoriamento remoto. Reconhecimento de padrões. Probabilidades *a priori*. Amostras semi-rotuladas. Dados imagem em alta dimensionalidade.

---

\*Dissertação de Mestrado em Sensoriamento Remoto, Centro Estadual de Pesquisas em Sensoriamento Remoto e Meteorologia, Curso de Pós-Graduação em Sensoriamento Remoto da Universidade Federal do Rio Grande do Sul. Porto Alegre/RS, (152p.), Agosto, 2007.

# CLASSIFICATION OF HIGH DIMENSIONALITY IMAGE DATA, USING SEMI-LABELED SAMPLES AND ESTIMATION OF THE *A PRIORI* PROBABILITIES\*

Celso Antonio Liczbinski

Advisor: Prof. Vitor Haertel, PhD.

## ABSTRACT

In natural scenes there are some cases in which some of the land-cover classes involved are spectrally very similar, i.e., their first order statistics are nearly identical. In these cases, the more traditional sensor systems such as Landsat-TM and Spot, among others usually result in a thematic image low in accuracy. On the other hand, it is well known that high-dimensional image data allows for the separation of classes that are spectrally very similar, provided that their second-order statistics differ significantly. The classification of high-dimensional image data, however, poses some new problems such as the estimation of the parameters in a parametric classifier. As the data dimensionality increases, so does the number of parameters to be estimated, particularly in the covariance matrix. In real cases, however, the number of training samples available is usually limited preventing therefore a reliable estimation of the parameters required by the classifier. The paucity of training samples results in a low accuracy for the thematic image which becomes more noticeable as the data dimensionality increases. This condition is known as the Hughes Phenomenon. Different approaches to mitigate the Hughes Phenomenon investigated by many authors have been reported in the literature. Among the possible alternatives that have been proposed, the so called semi-labeled samples has shown some promising results in the classification of remote sensing high dimensional image data, such as AVIRIS data. In this dissertation the approach proposed by Lemos (2003) is further investigated to increase the reliability in the estimation of the parameters required by the Gaussian Maximum Likelihood (GML) classifier. In this dissertation, we propose a methodology to estimate the *a priori* probabilities  $P(\omega_i)$  required by the GMV classifier. It is expected that a more realistic estimation of the values for the *a priori* probabilities well help to increase the accuracy of the thematic image produced by the GML classifier. The experiments performed in this study have shown an increase in the accuracy of the thematic image, suggesting the adequacy of the proposed methodology.

**Key-words:** remote sensing. patterns of recognition. *a priori* probability. semi-labeled samples. high dimensionality image data.

---

\*Master of Sciences Dissertation in Remote Sensing, Center for Remote Sensing and Meteorology, Graduate Program in Remote Sensing at Federal University of Rio Grande do Sul. Porto Alegre/RS, (152 p.), Agosto, 2007.

# SUMÁRIO

<b>RESUMO</b> .....	<b>v</b>
<b>ABSTRACT</b> .....	<b>vi</b>
<b>LISTA DE TABELAS</b> .....	<b>ix</b>
<b>LISTA DE FIGURAS</b> .....	<b>xii</b>
<b>LISTA DE GRÁFICOS</b> .....	<b>xv</b>
<b>LISTA DE ABREVIATURAS E SIGLAS</b> .....	<b>xvii</b>
<b>LISTA DE SÍMBOLOS</b> .....	<b>xviii</b>
<b>CAPÍTULO I</b>	
<b>1 INTRODUÇÃO</b> .....	<b>19</b>
1.1 APRESENTAÇÃO DO PROBLEMA .....	19
1.2 SENSORES HIPERESPECTRAIS .....	20
1.2.1 Satélite EO-1.....	21
1.2.1.1 O Sistema HYPERION .....	22
1.2.2 O Sistema Sensor AVIRIS .....	22
1.3 RECONHECIMENTO DE PADRÕES .....	26
1.4 CLASSIFICAÇÃO DE IMAGENS DIGITAIS .....	27
1.4.1 Classificação Multiespectral .....	27
1.5 O PROBLEMA DA DIMENSIONALIDADE DOS DADOS .....	31
1.5.1 Dados de Alta Dimensionalidade .....	31
1.6 PROPOSTA DE UM CLASSIFICADOR ADAPTATIVO.....	34
1.7 OBJETIVOS .....	36
1.8 ESTRUTURA DA DISSERTAÇÃO .....	36
<b>CAPÍTULO II</b>	
<b>2 REVISÃO BIBLIOGRÁFICA</b> .....	<b>38</b>
2.1 MÉTODOS PARA MINIMIZAR OS EFEITOS DO FENÔMENO DE HUGHES..	38
2.1.1 Redução da Dimensionalidade dos Dados .....	38
2.1.1.1 Seleção de Variáveis .....	39
2.1.1.2 Extração de Variáveis .....	40
2.1.1.2.1 <i>Transformada de Componentes Principais</i> .....	41
2.1.1.2.2 <i>Análise canônica</i> .....	42
2.2 MEDIDAS ESTATÍSTICAS DE SEPARAÇÃO ENTRE CLASSES .....	43
2.2.1 Divergência .....	44
2.2.2 Distância de Bhattacharyya .....	46
2.2.2.1 Distância de Bhattacharyya: Forma Gaussiana .....	47
2.3 ANÁLISE DISCRIMINANTE REGULARIZADA (RDA) .....	48
2.4 TÉCNICAS DE INCREMENTO DO NÚMERO DE AMOSTRAS DE TREINAMENTO SEMI-ROTULADAS .....	49
2.5 INFLUÊNCIA DO VOLUME DE AMOSTRAS DE TREINAMENTO NA ESTIMAÇÃO DE PARÂMETROS .....	50

2.6 OS CLASSIFICADORES DE BAYES E DA MÁXIMA VEROSSIMILHANÇA ..	55
2.6.1 O Classificador de BAYES .....	59
2.6.2 Classificador Máxima Verossimilhança Gaussiana .....	60
2.7 LIMIARES .....	62
2.8 O CLASSIFICADOR QUADRÁTICO .....	64
<b>CAPÍTULO III</b>	
<b>3 METODOLOGIA .....</b>	<b>66</b>
3.1 INTRODUÇÃO.....	66
3.2 NOVA METODOLOGIA PROPOSTA .....	68
<b>CAPÍTULO IV</b>	
<b>4 TESTES E EXPERIMENTOS .....</b>	<b>75</b>
4.1 INTRODUÇÃO .....	75
4.2 MATERIAIS .....	75
4.2.1 Equipamentos .....	75
4.2.2 Aplicativos .....	76
4.2.3 Dados .....	76
4.3 CENA DE ESTUDO .....	77
4.3.1 Seleção de Bandas Espectrais .....	79
4.3.2 Seleção de Classes .....	81
4.4 FERRAMENTA CDH - CLASSIFICADOR DE DADOS HIPERESPECTRAIS.	84
4.4.1 Amostras de Treinamento .....	84
4.4.2 Realização de Experimentos .....	86
4.4.2.1 Experimento 1 .....	89
4.4.2.2 Experimento 2 .....	92
4.4.2.3 Experimento 3 .....	94
4.4.2.4 Experimento 4 .....	96
4.4.2.5 Experimento 5 .....	98
4.4.2.6 Experimento 6 .....	100
4.4.2.7 Experimento 7 .....	102
4.5 ANÁLISE DOS EXPERIMENTOS .....	113
<b>CAPÍTULO V</b>	
<b>5 CONCLUSÕES E SUGESTÕES .....</b>	<b>117</b>
5.1 CONCLUSÕES .....	117
5.2 SUGESTÕES .....	119
<b>REFERÊNCIAS BIBLIOGRÁFICAS .....</b>	<b>120</b>
<b>APÊNDICE A - CÓDIGOS FONTE DA FERRAMENTA CDH .....</b>	<b>125</b>



## LISTA DE TABELAS

TABELA 1. Características técnicas do sensor AVIRIS. ....	24
TABELA 2. Comparação entre o sensor hiperespectral AVIRIS e o TM do Landsat. ....	25
TABELA 3. Intervalos das bandas espectrais de alguns sensores orbitais em três regiões do espectro refletivo. ....	26
TABELA 4. Parâmetros necessários para o sensor LandSat-TM e o sensor AVIRIS. ....	32
TABELA 5. Seleção de bandas ruidosas que foram excluídas (34 bandas). ....	80
TABELA 6. Total de <i>pixels</i> extraídos da imagem AVIRIS. ....	82
TABELA 7. Número de amostras das classes selecionadas .....	87
TABELA 8. Experimento com método tradicional, faz uso apenas das amostras de treinamento rotuladas e assume idêntico valor para as probabilidades <i>a priori</i> das classes, com dimensionalidade 40 e tamanho das amostras de treinamento 300 <i>pixels</i> . ....	91
TABELA 9. Experimento com método que incrementa as amostras de treinamento rotuladas com as chamadas amostras semi-rotuladas, e assume idêntico valor para as probabilidades <i>a priori</i> das classes, com dimensionalidade 40 e tamanho das amostras de treinamento 300 <i>pixels</i> . ....	91
TABELA 10. Experimento com método proposto neste estudo, que inclui valores estimados para as probabilidades <i>a priori</i> de cada uma das classes, com dimensionalidade 40 e tamanho das amostras de treinamento 300 <i>pixels</i> . ....	91
TABELA 11. Experimento com método tradicional, faz uso apenas das amostras de treinamento rotuladas e assume idêntico valor para as probabilidades <i>a priori</i> das classes, com dimensionalidade 60 e tamanho das amostras de treinamento 300 <i>pixels</i> . ....	93
TABELA 12. Experimento com método que incrementa as amostras de treinamento rotuladas com as chamadas amostras semi-rotuladas, e assume idêntico valor para as probabilidades <i>a priori</i> das classes, com dimensionalidade 60 e tamanho das amostras de treinamento 300 <i>pixels</i> . ....	93
TABELA 13. Experimento com o método proposto neste estudo, que inclui valores estimados para as probabilidades <i>a priori</i> de cada uma das classes, com dimensionalidade 60 e tamanho das amostras de treinamento 300 <i>pixels</i> . ....	93

TABELA 14. Experimento com método tradicional, faz uso apenas das amostras de treinamento rotuladas e assume idêntico valor para as probabilidades <i>a priori</i> das classes, com dimensionalidade 80 e tamanho das amostras de treinamento 300 <i>pixels</i> .....	95
TABELA 15. Experimento com método que incrementa as amostras de treinamento rotuladas com as chamadas amostras semi-rotuladas, e assume idêntico valor para as probabilidades <i>a priori</i> das classes, com dimensionalidade 80 e tamanho das amostras de treinamento 300 <i>pixels</i> . ....	95
TABELA 16. Experimento com o método proposto neste estudo, que inclui valores estimados para as probabilidades <i>a priori</i> de cada uma das classes, com dimensionalidade 80 e tamanho das amostras de treinamento 300 <i>pixels</i> .....	95
TABELA 17. Experimento com método tradicional, faz uso apenas das amostras de treinamento rotuladas e assume idêntico valor para as probabilidades <i>a priori</i> das classes, com dimensionalidade 40 e tamanho das amostras de treinamento 500 <i>pixels</i> . ....	97
TABELA 18. Experimento com método que incrementa as amostras de treinamento rotuladas com as chamadas amostras semi-rotuladas, e assume idêntico valor para as probabilidades <i>a priori</i> das classes, com dimensionalidade 40 e tamanho das amostras de treinamento 500 <i>pixels</i> . ....	98
TABELA 19. Experimento com o método proposto neste estudo, que inclui valores estimados para as probabilidades <i>a priori</i> de cada uma das classes, com dimensionalidade 40 e tamanho das amostras de treinamento 500 <i>pixels</i> . ....	98
TABELA 20. Experimento com método tradicional, faz uso apenas das amostras de treinamento rotuladas e assume idêntico valor para as probabilidades <i>a priori</i> das classes, com dimensionalidade 60 e tamanho das amostras de treinamento 500 <i>pixels</i> . ....	99
TABELA 21. Experimento com método que incrementa as amostras de treinamento rotuladas com as chamadas amostras semi-rotuladas, e assume idêntico valor para as probabilidades <i>a priori</i> das classes, com dimensionalidade 60 e tamanho das amostras de treinamento 500 <i>pixels</i> . ....	100
TABELA 22. Experimento com o método proposto neste estudo, que inclui valores estimados para as probabilidades <i>a priori</i> de cada uma das classes, com dimensionalidade 60 e tamanho das amostras de treinamento 500 <i>pixels</i> . ....	100
TABELA 23. Experimento com método tradicional, faz uso apenas das amostras de treinamento rotuladas e assume idêntico valor para as probabilidades <i>a priori</i> das classes, com dimensionalidade 80 e tamanho das amostras de treinamento 500 <i>pixels</i> . ....	101

TABELA 24. Experimento com método que incrementa as amostras de treinamento rotuladas com as chamadas amostras semi-rotuladas, e assume idêntico valor para as probabilidades <i>a priori</i> das classes, com dimensionalidade 80 e tamanho das amostras de treinamento 500 <i>pixels</i> . .....	102
TABELA 25. Experimento com o método proposto neste estudo, que inclui valores estimados para as probabilidades <i>a priori</i> de cada uma das classes, com dimensionalidade 80 e tamanho das amostras de treinamento 500 <i>pixels</i> . .....	102

## LISTA DE FIGURAS

FIGURA 1. Comparação entre baixa e alta resolução espectral .....	23
FIGURA 2.a) Espaço imagem. b) Espaço espectral . .... Fonte: Adaptado de Landgrebe (2002).	25
FIGURA 3. Esquema básico para reconhecimento de padrões .....	27
FIGURA 4. Formas de representação de dados multiespectrais, $\lambda$ é o comprimento de onda. .... Fonte: Adaptado de Landgrebe (1997).	28
FIGURA 5. Representação típica do fenômeno de Hughes. .... Fonte: Adaptado de Landgrebe (2002).	33
FIGURA 6. Um procedimento adaptativo de classificação .....	35
FIGURA 7. Distribuição hipotética de duas classes num espaço de duas variáveis. .... Fonte: Adaptado de Richards (1993).	42
FIGURA 8. Funções densidade típicas para o logaritmo da razão de verossimilhança. .... Fonte: Adaptado de Therrien (1989).	45
FIGURA 9. Exemplo típico de separação entre duas funções densidade. .... Fonte: Adaptado de Therrien (1989).	46
FIGURA 10. Efeito do número de amostras na estimação de parâmetros. .... Fonte: Adaptado de Richards (1999).	51
FIGURA 11. Espaço das variáveis para $n = 2$ . ....	53
FIGURA 12. Fronteiras de decisão entre 2 classes. .... Fonte: Adaptado de Richards (1993).	54
FIGURA 13. Fronteiras de decisão para $n = 2$ . ....	54
FIGURA 14. Interpenetração de classes no espaço das variáveis. ....	55
FIGURA 15. A regra de decisão de Bayes. .... Fonte: Adaptado de Fukunaga (1990).	59
FIGURA 16.a) Ilustração de uma classificação pouco precisa para padrões localizados nas caudas das distribuições. b) Uso de limiares para remover a classificação não acurada. ....	62
FIGURA 17. Fluxograma de funcionamento da Ferramenta CDH. ....	71

FIGURA 18. Localização da Universidade Purdue, Estado de <i>Indiana</i> – EUA. ....	77
Fonte: Landgrebe (2002).	
FIGURA 19. AVIRIS: Composição colorida RGB com três bandas espectrais, do segmento utilizado nos experimentos, ( <i>NW Indiana's Indian Pine</i> ) - 220 bandas espectrais.....	78
FIGURA 20. Imagem temática da verdade terrestre ( <i>ground truth</i> ) da FIG. 19, relativa às classes em estudo com as respectivas SIGLAS.....	79
FIGURA 21. Vista de áreas contendo as culturas e milho (a) e soja (b) mostrando a baixa porcentagem que representa frente à área emaciada no momento da coleta dos dados. ....	81
Fonte: Adaptado de Landgrebe (2003).	
FIGURA 22. Imagens temáticas para dados com dimensionalidade igual a 40, tamanho das amostras de treinamento rotuladas igual a 300: (a) amostras de treinamento rotuladas e idêntico valor para as probabilidades <i>a priori</i> das classes; (b) igual a (a), mas com a adição de 50 amostras semi-rotuladas; e (c) igual a (b), porém com a introdução dos valores estimados para as probabilidades <i>a priori</i> de cada classe. ....	90
FIGURA 23. Imagens temáticas para dados com dimensionalidade igual a 60, tamanho das amostras de treinamento rotuladas igual a 300: (a) amostras de treinamento rotuladas e idêntico valor para as probabilidades <i>a priori</i> das classes; (b) igual a (a), mas com a adição de 50 amostras semi-rotuladas; e (c) igual a (b), porém com a introdução dos valores estimados para as probabilidades <i>a priori</i> de cada classe. ....	92
FIGURA 24. Imagens temáticas para dados com dimensionalidade igual a 80, tamanho das amostras de treinamento rotuladas igual a 300: (a) amostras de treinamento rotuladas e idêntico valor para as probabilidades <i>a priori</i> das classes; (b) igual a (a), mas com a adição de 50 amostras semi-rotuladas; e (c) igual a (b), porém com a introdução dos valores estimados para as probabilidades <i>a priori</i> de cada classe. ....	94
FIGURA 25. Imagens temáticas para dados com dimensionalidade igual a 40, tamanho das amostras de treinamento rotuladas igual a 500: (a) amostras de treinamento rotuladas e idêntico valor para as probabilidades <i>a priori</i> das classes; (b) igual a (a), mas com a adição de 50 amostras semi-rotuladas; e (c) igual a (b), porém com a introdução dos valores estimados para as probabilidades <i>a priori</i> de cada classe. ....	96
FIGURA 26. Imagens temáticas para dados com dimensionalidade igual a 60, tamanho das amostras de treinamento rotuladas igual a 500: (a) amostras de treinamento rotuladas e idêntico valor para as probabilidades <i>a priori</i> das classes; (b) igual a (a), mas com a adição de 50 amostras semi-rotuladas; e (c) igual a (b), porém com a introdução dos valores estimados para as probabilidades <i>a priori</i> de cada classe. ....	99

FIGURA 27. Imagens temáticas para dados com dimensionalidade igual a 80, tamanho das amostras de treinamento rotuladas igual a 500: (a) amostras de treinamento rotuladas e idêntico valor para as probabilidades *a priori* das classes; (b) igual a (a), mas com a adição de 50 amostras semi-rotuladas; e (c) igual a (b), porém com a introdução dos valores estimados para as probabilidades *a priori* de cada classe. .... 101

## LISTA DE GRÁFICOS

GRÁFICO 1. Curvas de Resposta Espectral Média para as classes Milho Plantio Direto (preto), Milho Cultivo Mínimo (azul), Soja Plantio Direto (vermelho), Soja Cultivo Mínimo (ciano), Soja Cultivo Convencional (magenta) e Floresta (verde). .....	83
GRÁFICO 2. Acurácia da classificação (usuário), com os valores das probabilidades <i>a priori</i> idênticos para todas as classes, com amostras de treinamento rotuladas de tamanho 200 e incremento de 50 amostras semi-rotuladas, em três iterações. ....	103
GRÁFICO 3. Acurácia da classificação (usuário), com os valores das probabilidades <i>a priori</i> estimados para cada classe, com amostras de treinamento rotuladas de tamanho 200 e incremento de 50 amostras semi-rotuladas, em três iterações. ....	104
GRÁFICO 4. Acurácia da classificação (usuário), com os valores das probabilidades <i>a priori</i> idênticos para todas as classes, com amostras de treinamento rotuladas de tamanho 300 e incremento de 50 amostras semi-rotuladas, em três iterações. ....	105
GRÁFICO 5. Acurácia da classificação (usuário), com os valores das probabilidades <i>a priori</i> estimados para cada classe, com amostras de treinamento rotuladas de tamanho 300 e incremento de 50 amostras semi-rotuladas, em três iterações. ....	106
GRÁFICO 6. Acurácia da classificação final (usuário) com os valores das probabilidades <i>a priori</i> idênticos para todas as classes e, estimados para cada classe, com as amostras de treinamento rotuladas de tamanho 200 e incremento de 50 amostras semi-rotuladas, em três iterações. ....	107
GRÁFICO 7. Acurácia da classificação final (usuário) com os valores das probabilidades <i>a priori</i> idênticos para todas as classes e, estimados para cada classe, com as amostras de treinamento rotuladas de tamanho 300 e incremento de 50 amostras semi-rotuladas, em três iterações. ....	108
GRÁFICO 8. Acurácia da classificação final (usuário) com os valores das probabilidades <i>a priori</i> idênticos para todas as classes e, estimados para cada classe, com as amostras de treinamento rotuladas de tamanho 400 e incremento de 50 amostras semi-rotuladas, em três iterações. ....	109
GRÁFICO 9. Acurácia da classificação final (usuário) com os valores das probabilidades <i>a priori</i> idênticos para todas as classes e, estimados para cada classe, com as amostras de treinamento rotuladas de tamanho 500 e incremento de 50 amostras semi-rotuladas, em três iterações. ....	110

GRÁFICO 10. Acurácia da classificação final (usuário), através dos seguintes métodos: original; e, com os valores das probabilidades <i>a priori</i> : idênticos para todas as classes e, estimados para cada classe, com amostras de treinamento rotuladas de tamanho 300 e incremento de 50 amostras semi-rotuladas, em três iterações. ....	111
GRÁFICO 11. Acurácia da classificação final (usuário), através dos seguintes métodos: original; e, com os valores das probabilidades <i>a priori</i> : idênticos para todas as classes e, estimados para cada classe, com amostras de treinamento rotuladas de tamanho 500 e incremento de 50 amostras semi-rotuladas, em três iterações. ....	112
GRÁFICO 12. Tempo total de processamento para amostras de treinamento de tamanhos 300 e 500 <i>pixels</i> , e incremento de 50 amostras semi-rotuladas. ....	113



## LISTA DE ABREVIATURAS E SIGLAS

ALI	<i>Advanced Land Imager</i>
ASTER	<i>Advanced Spaceborne Thermal Emission and Reflection Radiometer</i>
ARIES	<i>Australian Resource Information and Environment Satellite</i>
AVIRIS	<i>Airbone Visible/Infrared Imaging Spectrometer</i>
$B$	Distância de Bhattacharyya entre as classes $\omega_i$ e $\omega_j$
BLOOC	Estimador de Covariância Bayesiano <i>Leave-one-out</i>
CD's	Números ou contadores digitais. Valor numérico atribuído ao pixel em cada banda da imagem
CDH	Classificador de Dados Hiperespectrais
EM	<i>Expectation Maximization</i>
ETM+	<i>Enhanced Thematic Mapper</i>
EOS	<i>Earth Observing System</i>
FIG	Figura
GRAF	Gráfico
HYPERION	Sensor Hiperespectral
INPE	Instituto Nacional de Pesquisas Espaciais
InSb	<i>Indium-Antimonide</i>
KL	<i>Karhunen-Loève</i>
LANDSAT MSS	<i>Land Remote Sensing Satellite</i>
LOOC	Estimador de Covariância <i>Leave-one-out</i>
MAXVER	Classificador Máxima Verossimilhança
MATLAB	Linguagem Matemática
MODIS	<i>Moderate Resolution Imaging Spectroradiometer</i>
MVG	Máxima Verossimilhança Gaussiana
Multispec	Sistema computacional para processamento de imagens multiespectrais. Foi desenvolvido pela <i>Purdue Research Foundation</i> , Estados Unidos
NASA	<i>National Aeronautic and Space Administration</i>
RDA	Análise Discriminante Regularizada
Si	<i>Silicon</i>
SPOT	<i>Système pour l'Observation de la Terra</i>
SFS	<i>Sequential Forward Selection</i>
SBS	<i>Sequential Backward Selection</i>
SFFS	<i>Sequential Forward Floating Selection</i>
SBFS	<i>Sequential Backward Floating Selection</i>
TAB	Tabela
$T_i$	<i>Threshold</i> – Limiar com objetivo de detectar <i>pixels</i> com baixa probabilidade de pertencer a qualquer uma das classes amostradas
TM	<i>Thematic Mapper</i>

## LISTA DE SÍMBOLOS

$i, J, K$	Índices de referência
$\mu m$	Micrômetro; equivale a $10^{-6}$ m
$nm$	Nanometro
$\chi^2$	Qui-quadrado
$gl$	Grau de liberdade
$\omega_i$	Classe $i$ dentro de uma <i>imagem</i> digital
$P(\omega_i)$	Probabilidade de ocorrência da classe $\omega_i$
<i>Pixel</i>	<i>Picture element</i>
$\mathbf{X}$	<i>Pixel</i> de interesse
$p(\mathbf{X})$	Probabilidade de encontrar um <i>pixel</i> de qualquer classe na posição $\mathbf{X}$
$P(\omega_i \mathbf{X})$	Probabilidade do <i>pixel</i> $\mathbf{X}$ pertencer à classe $\omega_i$
$p(\mathbf{X} \omega_i)$	Probabilidade interna da classe $\omega_i$
$exp$	Função exponencial
$ln$	Logaritmo neperiano (natural)
$\boldsymbol{\mu}_1, \boldsymbol{\mu}_2$	Vetores de médias das classes 1 e 2, respectivamente
$\boldsymbol{\mu}_k$	Vetor de média da classe $\omega_k$
$\boldsymbol{\Sigma}_k$	Matriz de covariância da classe $\omega_k$
$G_i(\mathbf{X})$	Função de discriminação para a classe $i$

# CAPÍTULO I

## INTRODUÇÃO

### 1.1 APRESENTAÇÃO DO PROBLEMA

Dados em alta dimensionalidade oferecem, em princípio, um poder discriminante bem mais elevado do que dados tradicionais em baixa dimensionalidade. Pode-se mostrar que classes espectralmente muito semelhantes, isto é, classes que compartilham de vetores de médias, muito próximos entre si, podem ser separadas com um alto grau de acurácia em espaços de dimensão suficientemente alta (FUKUNAGA, 1990). Neste fato reside uma das motivações para o desenvolvimento de sistemas sensores com um número muito grande de bandas espectrais, conhecidos como sensores hiperespectrais.

Avanços mais recentes na tecnologia de sistemas sensores espectrais permitiram o surgimento de uma nova geração de sistemas sensores, denominados de hiperespectrais, capazes de coletar simultaneamente dados imagem de uma mesma cena em um número muito grande de faixas espectrais. Entre eles podemos citar os sensores AVIRIS (*Airbone Visible Infrared Imager Spectroradiometer*), com 224 bandas espectrais e o HYPERION a bordo do satélite *Hyperspectral Imager (Earth Observing-1)*, com 220 bandas espectrais.

A utilização de dados em alta dimensionalidade deu origem a novos problemas e desafios à comunidade científica. Entre os novos problemas que surgem destaca-se aquele conhecido como o Fenômeno de Hughes. Nos processos envolvendo reconhecimento de padrões em imagens digitais, isto é, no processo de rotulagem dos *pixels* para produção de uma imagem temática, observa-se que a partir

de um número pequeno de variáveis (bandas espectrais, por exemplo), a acurácia da imagem temática produzida por um classificador paramétrico vai aumentando na medida em que novas variáveis são acrescentadas, isto é, na medida em que informação adicional é introduzida no processo. Em um determinado ponto, o valor da acurácia atinge um máximo, passando a declinar na medida em que a dimensionalidade dos dados continua a aumentar. Este comportamento é causado pelo fato de que com o aumento da dimensionalidade dos dados aumenta também o número de parâmetros a serem estimados, especialmente nas matrizes de covariância. Como em situações práticas o número de amostras de treinamento permanece limitado, o valor estimado para os parâmetros do classificador vão tornando-se cada vez menos confiáveis, resultando em uma degradação na performance do classificador.

Métodos visando amenizar este problema vêm sendo investigados pela comunidade internacional. As propostas apresentadas inserem-se em três linhas gerais: (1) técnicas de análise discriminante regularizada, (2) métodos para extração/seleção de variáveis e (3) uso de amostras semi-rotuladas.

No presente estudo é investigada uma metodologia que implementa, em um mesmo processo iterativo, amostras semi-rotuladas na estimação dos parâmetros do classificador e estimativas para as probabilidades *a priori* para as classes em consideração. O objetivo consiste em obter imagens temáticas mais acuradas quando utilizando dados em alta dimensionalidade.

## 1.2 SENSORES HIPERESPECTRAIS

Recentes avanços na tecnologia de sensores remotos têm possibilitado a aquisição de dados de melhor resolução espectral do que os coletados por sensores multiespectrais mais tradicionais como, por exemplo, aqueles a bordo dos satélites Landsat 7 e SPOT 4. Em dezembro de 1999, foi lançado pela NASA (*National*

*Aeronautics and Space Administration*), o satélite Terra do programa EOS (*Earth Observing System*). Esse satélite possui cinco instrumentos, dentre os quais o ASTER (*Advanced Spaceborne Thermal Emission and Reflection Radiometer*) com 14 bandas e o MODIS (*Moderate Resolution Imaging Spectroradiometer*) com 36 bandas, posicionadas nas faixas espectrais do visível e infravermelho próximo (400-2.500 nm) e infravermelho termal (8.000 - 15.000 nm). No final do ano 2000, foi lançado o satélite *Hyperspectral Imager (Earth Observing-1)*, com o sensor Hyperion (220 bandas), que constitui parte do programa de tecnologias revolucionárias proposto pela NASA para o terceiro milênio. Já o primeiro satélite comercial de alta resolução espectral, o ARIES (*Australian Resource Information and Environment Satellite*), com 105 bandas posicionadas no intervalo entre 400 e 2.500 nm, foi lançado no ano de 2001.

Essa nova fase de aquisição de dados tem recebido a denominação de Sensoriamento Remoto Hiperespectral ou Espectroscopia de Imageamento.

### 1.2.1 Satélite EO-1

Revolucionários instrumentos de imageamento do solo a bordo do sistema EO-1 coletam cenas multiespectrais e hiperespectrais sobre o curso desta missão (observação da terra) em coordenação com o *Enhanced Thematic Mapper (ETM+)* no *Landsat 7*.

Os sensores nos sistemas Landsat 7 e EO-1 capturam imagens das mesmas áreas (cenas). São três os instrumentos a bordo do sistema EO-1: ALI (*Advanced Land Imager*), *Atmospheric Corretor* e *Hyperion*. Cada um destes instrumentos possui métodos únicos de filtragem para coleta de dados somente em bandas espectrais específicas. As bandas são selecionadas em função das características específicas da superfície ou do solo em aplicações científicas ou comerciais.

### 1.2.1.1 O Sistema HYPERION

O sistema Hyperion dispõe de um imageador de alta resolução capaz de adquirir dados em 220 bandas espectrais (de 0.4 a 2.5  $\mu\text{m}$ ) com uma resolução espacial de 30m. O instrumento pode imagear uma área de 7.5 km por 100 km a cada passagem. Concebido originalmente como parte do instrumento ALI, o Hyperion agora é um instrumento completo e independente a bordo do satélite EO-1.

O sistema Hyperion é o primeiro imageador hiperespectral da NASA operando em nível orbital.

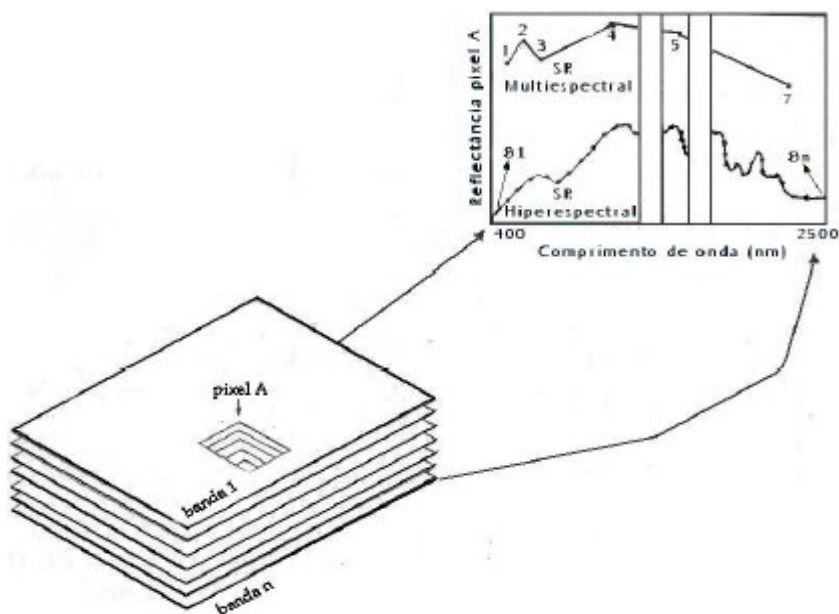
Minerais na superfície da Terra podem ser identificados e novos mapas de minérios podem ser criados visando escolher locais de exploração; inventários florestais podem ser desenvolvidos para regiões remotas como suporte ao gerenciamento e planejamento ecológico.

### 1.2.2 O Sistema Sensor AVIRIS

O AVIRIS foi desenvolvido pela NASA em 1983. Trata-se de um sensor experimental, areotransportado coletando dados em 224 bandas espectrais, no intervalo de 0.4  $\mu\text{m}$  a 2.5  $\mu\text{m}$ . Estes dados são caracterizados, portanto, por uma alta dimensionalidade (alta resolução espectral). A utilização destes dados requer uma abordagem distinta daquelas tradicionalmente empregadas em dados de baixa dimensionalidade (dados LandSat, SPOT, por exemplo), tornando-se necessário desta forma, o desenvolvimento de novas metodologias e técnicas.

Na FIG. 1 observa-se a curva de resposta espectral de um *pixel*, gerada pelos sensores multiespectrais tradicionais e hiperespectrais. A curva de resposta

espectral do *pixel* é muito mais detalhada com as informações obtidas do sensor hiperespectral, do que a curva obtida a partir de dados fornecidos por um sensor multiespectral.



**FIGURA 1 - Comparação entre baixa e alta resolução espectral**

O primeiro vôo do sensor AVIRIS foi realizado em 1987. A plataforma atual do sistema AVIRIS é um avião modelo U-2, tendo sido modificado com a finalidade de adaptação ao sensor. A altitude de vôo é de 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, para fins de pesquisa. No Brasil, já foram feitos vôos em 1995, em Cuiabá, no Pantanal Mato-grossense, em Brasília e na cidade de Porto Nacional. Algumas imagens e composições estão disponíveis no site do INPE. O sistema óptico do sensor AVIRIS consiste de um scanner, 4 espectrômetros e uma fonte de calibração acoplada por fibra óptica. Detectores de linha são usados, necessitando que a varredura seja feita mecanicamente no modo *whisk broom*. Existe um par de detectores 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* óptico para os espectrômetros, onde o sinal é dividido em bandas discretas registradas pelos detectores. Um 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 do infravermelho médio. Os detectores de *Silicon* possuem 32 elementos de linha na porção do visível e as outras três regiões do espectro são registradas por 64 elementos de linha, por detectores de *Indium-Antimonide*.

O sensor AVIRIS usa um espelho *scanner* que varre no sentido longitudinal, produzindo 614 *pixels* para cada varredura dos 224 detectores, cada um cobrindo um intervalo de comprimento de onda (banda espectral), de aproximadamente 10nm, permitindo uma cobertura completa entre 400nm e 2500nm. 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 do solo são registrados a bordo do instrumento juntamente com os dados de engenharia, navegação e 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* que tem aproximadamente o comprimento de 10 Km no terreno. A TAB. 1 mostra as características técnicas do sensor AVIRIS.

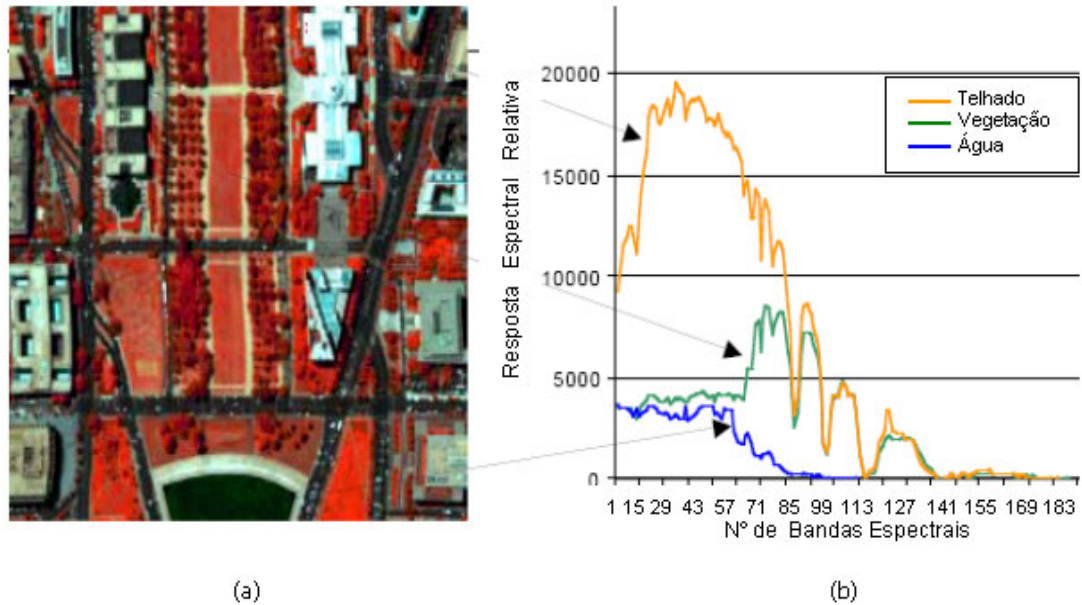
**TABELA 1- Características técnicas do sensor AVIRIS**

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
Detectores para os $\lambda$ 's da faixa do visível da <i>Silicon</i> (Si), e detectores de <i>Indium-Antimonide</i> (InSb) para o infravermelho-próximo
Varredura do tipo <i>Whisk Broom</i>
Taxa de varredura de 12Hz
Detectores frios ( <i>cooled</i> ) de <i>Liquid Nitrogen</i> (LN2)
Canais de bandas nominais de 10 nm, calibradas para 1 nm
Campo de visada total de 30° (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

Fonte: JPL - Jet Propulsion Laboratory. Disponível em: <<http://aviris.jpl.nasa.gov/html>>.



Na FIG. 2 é apresentada uma composição colorida de uma área em Washington, DC, conhecida por “Mall”. Esta composição é feita usando três das 224 bandas coletadas pelo sistema sensor, duas na região do visível (verde e vermelho) e uma terceira na região do infravermelho próximo.



**FIGURA 2.a) Espaço imagem. b) Espaço espectral.**

Fonte: Adaptado de Landgrebe (2002).

Na TAB. 2 são apresentados alguns dados comparativos entre os sensores AVIRIS e Landsat-TM e na TAB. 3 podem ser observados intervalos das bandas espectrais de alguns sensores orbitais em três regiões do espectro refletivo.

**TABELA 2 - Comparação entre o sensor hiperespectral AVIRIS e o TM do Landsat**

SENSOR	FAIXA ESPECTRAL (nm)	BANDA ESPECTRAL	RESOLUÇÃO ESPECTRAL (nm)	LARGURA DE FAIXA IMAGEADA (Km)	RESOLUÇÃO ESPACIAL (m <sup>2</sup> )
AVIRIS	400 a 2500	224	10	11	20
TM	450 a 12500	7	Variável	185	900 e 14400

**TABELA 3 - Intervalos das bandas espectrais de alguns sensores orbitais em três regiões do espectro refletivo.**

SENSOR	VISÍVEL ( $\mu\text{m}$ )			IVP* ( $\mu\text{m}$ )	IVM* ( $\mu\text{m}$ )	
TM/Landsat-4	0.45-0.52	0.52-0.60	0.63-0.69	0.76-0.90	1.55-1.75	2.08-2.35
TM/Landsat-5	0.45-0.52	0.52-0.60	0.63-0.69	0.76-0.90	1.55-1.75	2.08-2.35
CCD/CBERS	0.45-0.52	0.52-0.59	0.63-0.69	0.77-0.89		
ETM+/Landsat-7	0.45-0.52	0.52-0.61	0.63-0.69	0.78-0.90	1.55-1.75	2.09-2.35
ASTER/EOS		0.52-0.60	0.63-0.69	0.76-0.86	1.60-1.70	2.145-2.185 2.185-2.225 2.235-2.285 2.295-2.235 2.360-2.430

\* IVP – Infravermelho Próximo. IVM – Infravermelho Médio.

### 1.3 RECONHECIMENTO DE PADRÕES

O tópico de pesquisa abrangido por este estudo está inserido dentro da área de conhecimento denominada de Reconhecimento de Padrões. Reconhecimento de padrões tem como objetivo classificar objetos de interesse em classes ou categorias.

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 disponíveis (THERRIEN, 1989).

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, ilustrado na FIG. 3, essencialmente envolve três aspectos: 1) aquisição de dados através de dispositivos de captação dos dados; 2) representação dos dados na forma de um vetor  $X$ ; e 3) transformações necessárias à tomada de decisão.

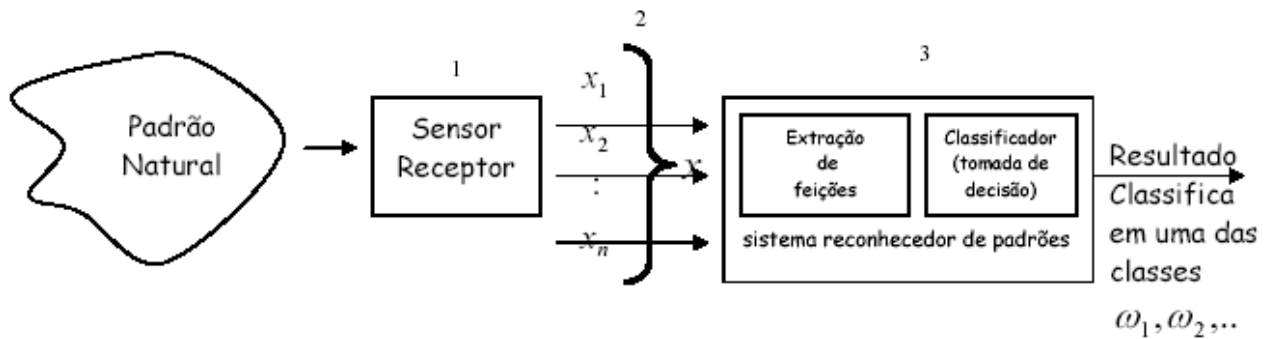


FIGURA 3 – Esquema básico para reconhecimento de padrões

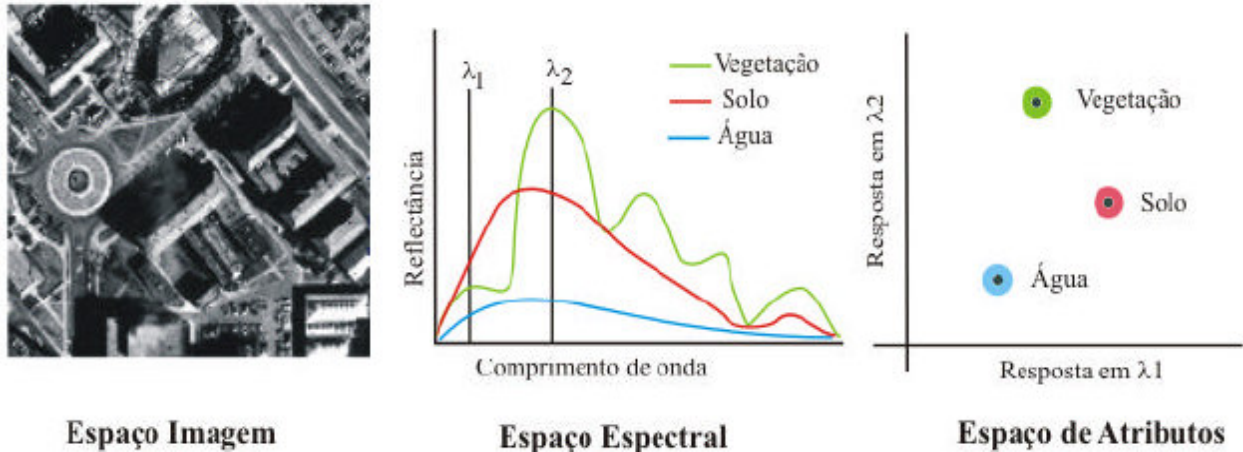
## 1.4 CLASSIFICAÇÃO DE IMAGENS DIGITAIS

A classificação de imagens multiespectrais de Sensoriamento Remoto assistida por computador, trata de atribuir a cada *pixel* da imagem um “rótulo” identificando uma determinada classe de cobertura do solo. Cada *pixel* na imagem é representado por um vetor, com dimensão igual ao número de bandas espectrais utilizado. Cada elemento neste vetor fica definido pelo valor do contador digital na banda correspondente. Neste contexto, a informação utilizada no processo de classificação é de natureza espectral e os rótulos a serem atribuídos a cada *pixel*, referem-se às classes espectrais.

### 1.4.1 Classificação Multiespectral

Segundo Landgrebe (1997) existem três principais maneiras de representar e visualizar quantitativamente os dados de sensores multiespectrais: na

forma de imagem, na forma de curva de respostas espectrais e num espaço de atributos. A FIG. 4 ilustra estas três alternativas.



**FIGURA 4 – Formas de representação de dados multiespectrais,  $\lambda$  é o comprimento de onda.**  
 Fonte: Adaptado de Landgrebe (1997).

A forma de imagem é a maneira mais comum de representação, serve como uma interface homem/dados ajudando o usuário a fazer a conexão entre um *pixel* e a área que ele representa. O espaço espectral descreve a variação da reflectância de uma dada superfície em função do comprimento de onda, resultando na “assinatura” espectral, muito utilizado de maneira didático-ilustrativa. Por último, a forma de representar dados discretizados em determinados comprimentos de onda, no espaço N-dimensional, é a forma mais adequada para análise através de algoritmos (LANDGREBE, 1997).

O processo de classificação de uma imagem digital é normalmente realizado no chamado *espaço das variáveis*, definido por um sistema de eixos ortogonais, cada um associado à resposta espectral em uma banda espectral. O processo de classificação consiste basicamente em particionar o espaço das variáveis, definindo fronteiras entre classes. A classificação é então feita *pixel a pixel*, atribuindo cada *pixel* à classe correspondente à partição do espaço das variáveis na qual o *pixel* está localizado.

Diferentes técnicas de classificação são usadas dependendo do tipo do modelo escolhido. 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 e caracterizadas através de amostras rotuladas denominadas de amostras de treinamento. Cada classe passa, então, a 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 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 presentes na cena;
- 2º) Definir a forma geral da função decisão a ser adotada;
- 3º) Selecionar um conjunto de *pixels* representativos de cada classe que constituirão as *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 por vezes mesmo o número de classes presentes são desconhecidos. Neste último 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 se tem um conhecimento prévio das classes, os métodos não-supervisionados atribuem 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 das variáveis. 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 das variáveis 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 as redes neurais artificiais e o método do vizinho mais próximo.

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, sendo supervisionado, pois estima estes parâmetros através das amostras de treinamento.

## **1.5 O PROBLEMA DA DIMENSIONALIDADE DOS DADOS**

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.

### **1.5.1 Dados de Alta Dimensionalidade**

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 paramétricos. O número de parâmetros a serem estimados, na matriz covariância particularmente, cresce rapidamente na medida em que a dimensionalidade dos dados aumenta, como ilustrado (TAB. 4). Se o número das amostras de treinamento não aumenta, na

mesma proporção, 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 na incerteza da estimativa dos parâmetros e, conseqüentemente, num decréscimo na acurácia da imagem temática produzida pelo classificador.

Por meio da expressão  $np=(n+((n^2-n)/2))$  onde  $np$  é o número total de parâmetros a 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 o sistema sensor AVIRIS com suas 224 bandas, pode-se calcular o número de parâmetros necessários para cada caso, conforme mostrado (TAB. 4).

**TABELA 4 - Parâmetros necessários para o sensor LandSat-TM e o sensor AVIRIS**

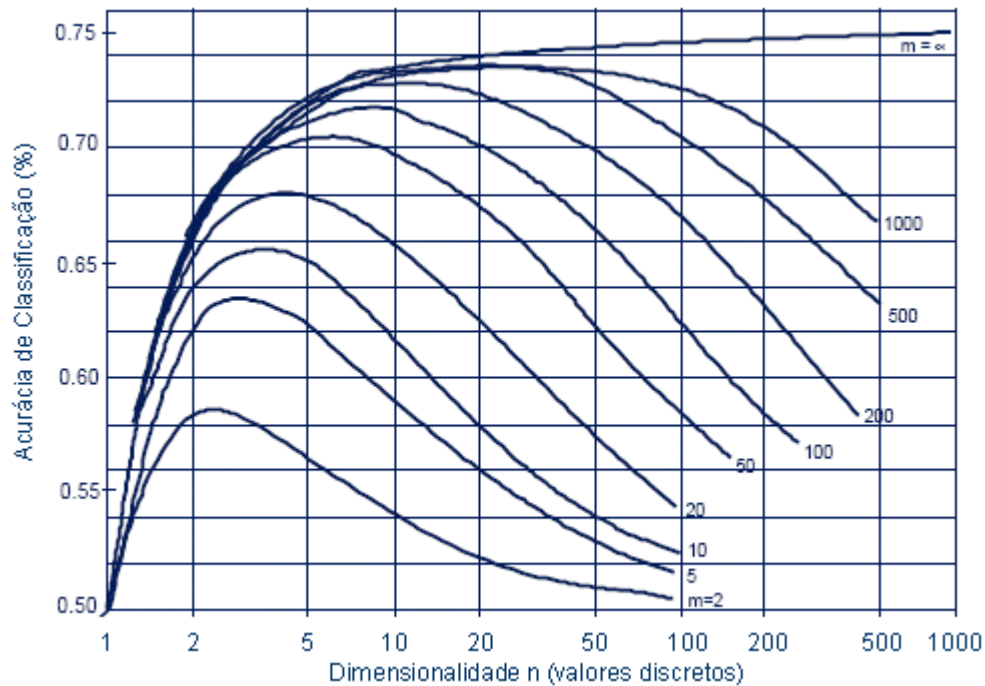
<b>LandSat-TM</b>	<b>AVIRIS</b>
$n = 6$	$n = 224$
$np = (6+((6^2 - 6) / 2))$	$np = ((224+((224^2 - 224) / 2))$
$np = 21$	$np = 25200$

Nota-se que a plena utilização de dados fornecidos pelo sistema AVIRIS, é necessário estimar 1200 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*, também conhecido como a *maldição da dimensionalidade* (TADJUDIN, 1998). O efeito, demonstrado na FIG. 5, 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 fenômeno está ilustrado na FIG. 5 para amostras de vários tamanhos.



**FIGURA 5 - Representação típica do fenômeno de Hughes.**

Fonte: Adaptado de Landgrebe (2002).

A solução para este problema está, atualmente, basicamente centrada em três abordagens: a) pela utilização de técnicas de análise discriminante regularizada (AEBERHARDT, 1990, 1994), (FRIEDMAN, 1989), (ERBERT, 2001); b) redução do número de variáveis fornecidas como entrada no classificador utilizado. Como um super sensor hiperespectral adquire bandas muito próximas e, portanto, correlacionadas, é possível reduzir o número de variáveis sem que haja perda significativa de informação Fukunaga (1990), reduzindo desta forma, os efeitos do fenômeno de Hughes e, conseqüentemente, melhorando os resultados fornecidos pelo classificador em relação àqueles obtidos utilizando diretamente o número total de

bandas originais; e, c) introdução de amostras não rotuladas (isto é, *pixels* cuja classe a que pertencem não é conhecida) no processo de estimação dos parâmetros através de utilização de técnicas como, por exemplo, o algoritmo *Expectation-maximization* (EM), além de outras possíveis.

O problema de redução do número de variáveis (redução na dimensionalidade dos dados) pode ser resolvido por duas abordagens, baseadas na seleção de um subconjunto de variáveis com maior poder discriminante, entre o total de bandas disponíveis, ou na extração de um número limitado de variáveis após um mapeamento num novo espaço de variáveis (SERPICO, 2003).

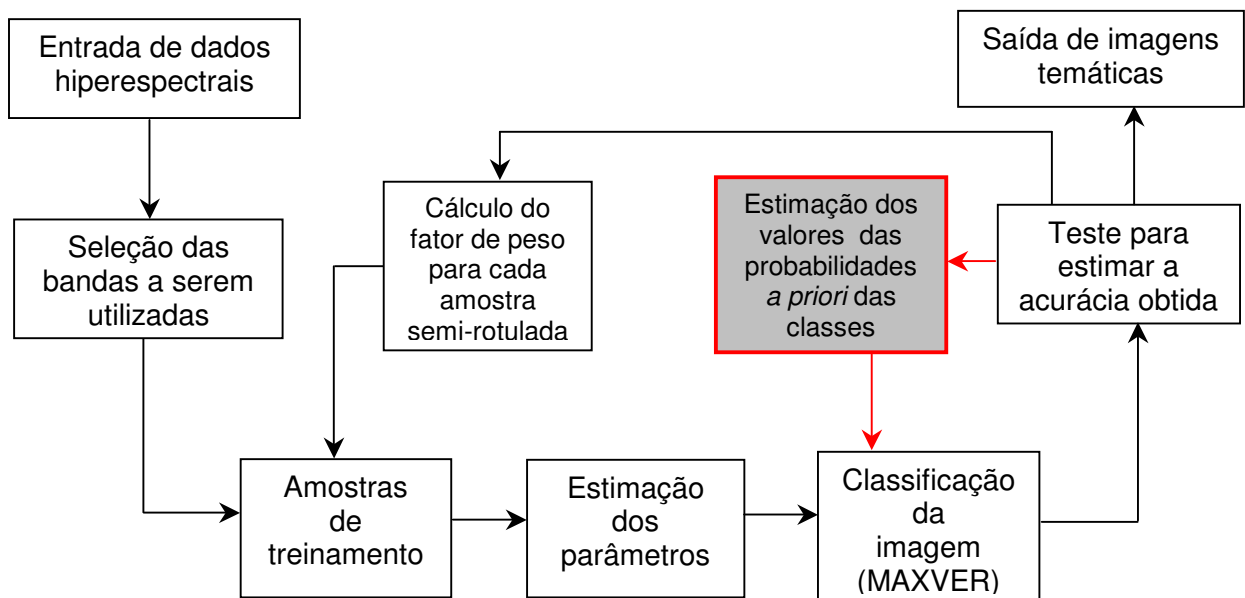
## **1.6 PROPOSTA DE UM CLASSIFICADOR ADAPTATIVO**

A inclusão de amostras semi-rotuladas no processo de estimação dos parâmetros em um classificador paramétrico vem sendo investigada por vários autores.

Lemos (2003), em seu trabalho investigou um procedimento iterativo, aplicado a um classificador quadrático, gerando amostras semi-rotuladas. Diferentemente do trabalho desenvolvido por Jackson (2001), não foi utilizado nenhum método de estimação regularizada dos parâmetros estatísticos. As amostras semi-rotuladas assim geradas são adicionadas, de uma forma ponderada, às amostras rotuladas disponíveis no processo de estimação dos parâmetros do classificador. Tal procedimento permite, portanto, incrementar o número de amostras de treinamento para cada uma das classes envolvidas, e por conseqüência, uma estimativa mais confiável dos parâmetros do classificador pode ser obtida e, finalmente, resultados mais acurados no processo de classificação serão gerados. O início do processo se dá por uma primeira classificação, onde somente as amostras rotuladas são utilizadas na estimação dos parâmetros. Como resultado desta iteração inicial, *pixels* originalmente não rotulados, isto é, não associados a nenhuma das classes existentes, recebem um

rótulo. Desta forma amostras originalmente não rotuladas passam a contar com alguma informação com relação à classe a que pertencem, embora o rótulo assim atribuído no processo de classificação contenha um grau de incerteza. Estas novas amostras recebem a denominação de amostras semi-rotuladas. A estimação dos parâmetros do classificador, a partir da segunda iteração passa a ser feita utilizando o conjunto das amostras rotuladas acrescido de um determinado número de amostras semi-rotuladas. Neste processo, as amostras rotuladas recebem um peso unitário enquanto que as amostras semi-rotuladas recebem um peso variável, no intervalo  $[0,1]$  refletindo o maior ou menor grau de confiança no rótulo que lhes foi atribuído.

O presente estudo consiste em dar continuidade ao trabalho desenvolvido por (LEMOS, 2003). Conforme descrito anteriormente, naquele trabalho foi investigado e testado um método que implementa de uma forma iterativa, a utilização de amostras semi-rotuladas no processo de classificação. A contribuição aqui proposta consiste em adicionar uma etapa complementar que compreende a estimação das probabilidades *a priori* das classes envolvidas (salientado pelas setas vermelhas e sombreamento - FIG. 6), no classificador adaptativo proposto por aquele autor.



**FIGURA 6 - Um procedimento adaptativo de classificação.**

Busca-se, portanto, desenvolver estudos que visam estimar, de uma forma iterativa, valores para as probabilidades *a priori* das classes envolvidas fazendo uso de uma abordagem nas mesmas linhas daquelas propostas por (JACKSON, 2001, 2002), (LEMOS, 2003). Neste trabalho, esta abordagem é implementada ao classificador *Máxima Verossimilhança Gaussiana*, com vistas a aumentar a acurácia da classificação de imagens.

A análise proposta neste estudo requer o desenvolvimento de ferramentas específicas. Para tanto, implementou-se um conjunto de funções no ambiente MATLAB, as quais resultaram em produto adicional desta pesquisa, doravante denominado de CDH - Classificador de Dados Hiperespectrais.

## 1.7 OBJETIVOS

Este estudo tem como principais objetivos:

1. Investigar uma metodologia de fácil uso por parte do usuário, para fins de estimação dos valores das probabilidades *a priori* associadas a cada uma das classes em consideração. Este processo de estimação é implementado conjuntamente com a inclusão de amostras semi-rotuladas em um processo iterativo/adaptativo de classificação.
2. Desenvolver ferramentas computacionais para implementar e testar a Metodologia proposta.

## 1.8 ESTRUTURA DA DISSERTAÇÃO

Esta dissertação está organizada da seguinte maneira:

No primeiro capítulo é feita uma introdução aos objetivos da dissertação.

No segundo capítulo, é apresentada uma revisão bibliográfica sobre problemas associados ao processo de classificação de dados em alta dimensionalidade empregando classificadores paramétricos. São revisadas as abordagens propostas por vários pesquisadores, visando minimizar os efeitos causados pelo fenômeno de Hughes.

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

O quarto capítulo apresenta os testes e experimentos do algoritmo adaptativo através de um estudo de caso em imagens hiperespectrais obtidas pelo sensor AVIRIS.

No quinto capítulo são apresentadas as conclusões e análise final.

O Apêndice A, contém a listagem dos programas computacionais elaborados especificamente para este trabalho.

## **CAPÍTULO II**

### **REVISÃO BIBLIOGRÁFICA**

#### **2.1 MÉTODOS PARA MINIMIZAR OS EFEITOS DO FENÔMENO DE HUGHES**

Conforme mencionado, imagens em alta dimensionalidade, como é o caso de imagens hiperespectrais, permitem uma separação mais acurada de classes espectrais muito semelhantes entre si. A utilização destes dados em classificadores paramétricos apresenta o problema da estimação dos parâmetros com as amostras de treinamento disponíveis. Em situações práticas, o número de amostras de treinamento é usualmente limitado, resultando em uma estimativa pouco confiável dos parâmetros do classificador e conseqüentemente em uma acurácia reduzida na imagem temática produzida.

Métodos que visam reduzir este problema vem sendo objeto de interesse por parte da comunidade internacional. Uma revisão de trabalhos publicados em periódicos internacionais mostra a existência de três vias principais: (1) métodos de redução da dimensionalidade dos dados, (2) métodos de análise discriminante regularizada e (3) utilização de amostras semi-rotuladas.

##### **2.1.1 Redução da Dimensionalidade dos Dados**

Os sistemas sensores atualmente utilizados são projetados para atender as características de um amplo conjunto de alvos de naturezas diversas. Em situações individuais, a utilização no processo de classificação empregando simultaneamente

todas as variáveis originais (dimensionalidade total) ou de um número grande destas, além de geralmente desnecessário pode tornar-se prejudicial à eficiência do processo, devido principalmente às dificuldades que surgem na obtenção de estimativas confiáveis dos parâmetros empregados pelo classificador sempre que dados em alta dimensionalidade são utilizados (LANDGREBE, 2003). É geralmente mais eficiente selecionar um subconjunto composto pelas variáveis mais adequadas a cada caso em particular. Assim uma etapa importante que deve preceder o processo de classificação refere-se à escolha de um subconjunto ótimo ou pelo menos sub-ótimo de variáveis para cada situação em particular. Este problema é conhecido como redução da dimensionalidade dos dados, conforme abordagens feitas nas seções seguintes deste capítulo.

Nas metodologias que visam reduzir a dimensionalidade dos dados, existem duas abordagens distintas: (1) seleção de variáveis e (2) extração de variáveis.

#### 2.1.1.1 Seleção de Variáveis

Possivelmente a maneira mais direta de reduzir o problema da dimensionalidade dos dados, seja simplesmente selecionar um subconjunto de bandas espectrais, contendo aquelas que oferecem melhores condições de separabilidade entre as classes em consideração. Normalmente, uma procura exaustiva torna-se computacionalmente proibitiva, visto que o número de combinações possíveis para a seleção de um subconjunto contendo  $m$  variáveis entre um total de  $n$  bandas pode se tornar muito alto. O número de possíveis combinações ( $N_c$ ) pode ser calculado pelo coeficiente binomial:

$$N_c = \binom{n}{m} = \frac{n!}{(n-m)! m!} \quad (2.1)$$

Desta forma, as técnicas de seleção de variáveis geralmente envolvem um algoritmo de procura associado a uma função critério (SWAIN, 1978), (JAIN, 2000). O algoritmo de procura gera e compara possíveis “soluções” para o problema de seleção de variáveis (isto é, subconjunto de variáveis), aplicando uma função critério como medida da eficácia de cada solução.

Diversas abordagens sub-ótimas para a seleção de variáveis têm sido propostas (BRUZZONE, 2000), (DUDA, 2001). As estratégias mais simples como procura sub-ótimas são os algoritmos “*Sequential Forward Selection*” (SFS) e a “*Sequential Backward Selection*” (SBS) (JAIN, 2000), (PUDIL, 1994). Estes algoritmos identificam iterativamente o melhor subconjunto de variáveis que pode ser obtido pela adição ou remoção respectivamente, de uma feição por vez ao subconjunto selecionado. Infelizmente, ambos os algoritmos têm um sério inconveniente: uma vez que a feição foi selecionada, ela não pode ser descartada (SFS), ou uma vez que a feição foi descartada, ela não pode ser re-introduzida (SBS). Estas peculiaridades caracterizam estes processos como sub-ótimos (SERPICO, 2003).

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

#### 2.1.1.2 Extração de Variáveis

Nesta abordagem, são utilizadas transformações (lineares ou não-lineares) das bandas espectrais originais. Estas transformações são selecionadas de forma que o poder de separação fique concentrado em um número menor de bandas, permitindo assim uma diminuição na dimensionalidade dos dados, minimizando a perda de informações. Na literatura encontram-se citados vários métodos para extração



de variáveis, sendo as mais conhecidas: transformação de componentes principais, também conhecida como transformação de Karhunen-Loève(KL) e análise canônica.

#### 2.1.1.2.1 Transformada de Componentes Principais

A transformada de componentes principais é uma transformação linear que busca mapear os dados do espaço original  $\mathbf{X}$  para um espaço  $\mathbf{Y}$  no qual as variáveis estão decorrelacionadas. Transformações lineares podem ser expressas de uma maneira conveniente em notação matricial:

$$\mathbf{Y} = \mathbf{G} \mathbf{X} \quad (2.2)$$

Sendo  $\mathbf{G}$  a matriz que implementa a transformação. O critério para escolha da matriz  $\mathbf{G}$  consiste na condição de que a matriz de covariância ( $\mathbf{\Sigma}_Y$ ) das variáveis no espaço transformado  $\mathbf{Y}$  deve ser uma matriz diagonal. Pode-se mostrar que a matriz  $\mathbf{G}$  que satisfaz a esta condição é uma matriz cujas colunas são os auto-vetores da matriz de covariância dos dados ( $\mathbf{\Sigma}_X$ ) no espaço original  $\mathbf{X}$ . No espaço  $\mathbf{Y}$  os eixos ficam alinhados segundo a direção dos auto-vetores de  $\mathbf{\Sigma}_X$  e os correspondentes auto-valores exprimindo as variâncias nas direções destes novos eixos. A matriz de covariância  $\mathbf{\Sigma}_Y$  é, portanto uma matriz diagonal com os auto-valores de  $\mathbf{\Sigma}_X$  compondo a diagonal principal.

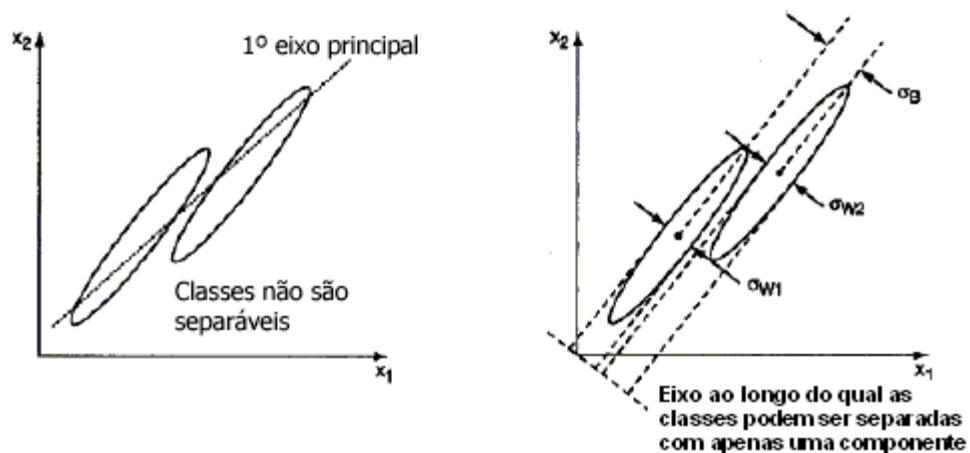
Deve-se aqui ter o cuidado de notar que a transformada de componentes principais é baseada na matriz de covariância global do conjunto dos dados (pixels) envolvendo, portanto todas as classes presentes na imagem. Esta transformação não é, assim, sensível à estrutura de cada classe individualmente e por esta razão não é necessariamente uma transformação que maximiza a separabilidade entre as classes. Em imagens de cenas naturais, esta abordagem geralmente funciona bem como ferramenta de redução da dimensionalidade devido ao fato de que as classes são

geralmente distribuídas na direção do máximo espalhamento dos dados (*maximum data scatter*).

### 2.1.1.2.2 Análise Canônica

A técnica de componentes principais como mecanismo redutor da dimensionalidade dos dados mostra-se, portanto eficiente apenas quando os dados estão distribuídos substancialmente ao longo das primeiras componentes (FIG. 7). Se os dados não apresentarem esta característica, possivelmente não se terá um ganho ao se utilizar esta transformada ao invés dos dados originais.

Em tais condições, é possível que a técnica de análise canônica apresente melhores resultados.



**FIGURA 7 – Distribuição hipotética de duas classes num espaço de duas variáveis.**  
Fonte: Adaptado de Richards (1993).

Neste caso, o critério para seleção da direção dos eixos no espaço transformado consiste na maximização da variância entre as classes (*between-class*

variance) e na minimização da variância média das classes individuais (*average within-class variance*). A variância entre as classes é geralmente estimada pela matriz  $\Sigma_B$ :

$$\Sigma_B = \sum_{i=1}^k P(\omega_i) (\mu_i - \mu_0) (\mu_i - \mu_0)^T \quad (2.3)$$

onde  $\mu_i$  representa o vetor de médias da classe  $\omega_i$ , sendo  $\mu_0$  o vetor média ponderada dos vetores de médias das  $k$  classes.

$$\mu_0 = \sum_{i=1}^k P(\omega_i) \mu_i \quad (2.4)$$

A variância média das classes individuais é estimada pela matriz  $\Sigma_W$ :

$$\Sigma_W = \sum_{i=1}^k P(\omega_i) \Sigma_i \quad (2.5)$$

onde  $\Sigma_i$  representa a matriz de covariância da classe  $\omega_i$ .

Segundo Landgrebe (2003), o critério para definição dos novos eixos consiste na maximização de:

$$\mathbf{J} = tr(\Sigma_W^{-1} \Sigma_B) \quad (2.6)$$

Na FIG. 7 está ilustrada esta abordagem.

## 2.2 MEDIDAS ESTATÍSTICAS DE SEPARAÇÃO ENTRE CLASSES

A metodologia que implementa as técnicas de seleção de variáveis faz uso de medidas estatísticas de distância para selecionar um subconjunto ótimo ou sub-ótimo das bandas espectrais originais para fins de separação entre as classes em consideração.

Nessa seção, será apresentada uma introdução sobre duas medidas de separação entre classes, conhecidas como *Divergência* e *Distância de Bhattacharyya*. Para essa última, é apresentada sua forma geral e a forma que assume quando os dados apresentam uma distribuição normal.

### 2.2.1 Divergência

A função de decisão Máxima Verossimilhança, revista na seção 2.6.2 deste capítulo, para o caso de duas classes pode ser escrita como:

$$X \in \omega_1 \Leftrightarrow l(X) = \frac{p(X / \omega_1)}{p(X / \omega_2)} > \frac{P(\omega_2)}{P(\omega_1)} \quad (2.7)$$

Caso contrário,  $X$  pertence a  $\omega_2$ .

Neste caso, a razão  $P(\omega_2) / P(\omega_1)$  pode ser vista como um limiar. O termo  $l(X)$  é denominado de *razão de verossimilhança* e é a quantidade básica no teste de hipóteses. Para valores de  $l(X)$  maiores do que o limiar,  $X$  é atribuído à classe  $\omega_1$ . Para  $l(X)$  maior do que o limiar,  $X$  é atribuído à  $\omega_2$ .

O mesmo critério é válido para qualquer função monotonicamente crescente de  $l(X)$ :

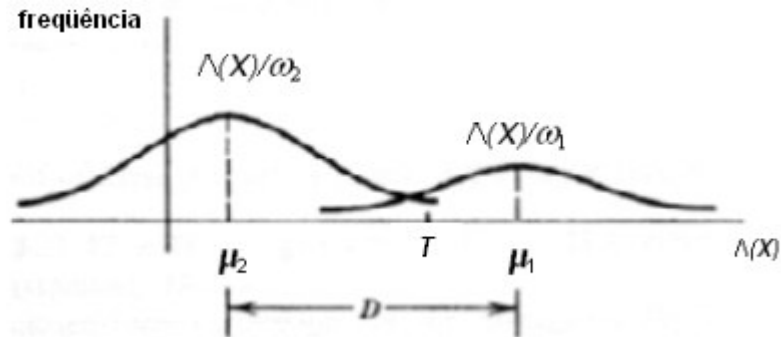
$$\Lambda(X) = \ln \frac{p(X / \omega_1)}{p(X / \omega_2)}, \quad \text{para } \omega_1 > T \text{ e } \omega_2 < T \quad (2.8)$$

Para analisar o comportamento de  $\Lambda(X)$ , que também é uma estatística.

Se:  $X \in \omega_1$ , então  $\Lambda(X) < T$

$X \in \omega_2$ , então  $\Lambda(X) > T$

Essa medida pode ser descrita com o par de funções densidade de probabilidade ilustrados na FIG. 8, conforme exposto por Therrien (1989). Em uma regra de decisão empregando o logaritmo da razão de verossimilhança, comparado a um *limiar*  $T$ , a probabilidade de erro será menor quando as funções densidade forem bem separadas entre si, e maior caso contrário.



**FIGURA 8 - Funções densidade típicas para o logaritmo da razão de verossimilhança.**

Fonte: Adaptado de Therrien (1989).

Uma medida de separação entre classes baseada na diferença entre as médias  $\mu_1$  e  $\mu_2$ , definida por  $D$ , está ilustrada (FIG. 8). Essa quantidade é conhecida como *Divergência*.

A Divergência é formalmente definida pela diferença entre os valores esperados para o logaritmo das razões de verossimilhança entre as duas classes:

$$D = \mu_1 - \mu_2 = E[\Lambda(X)|\omega_1] - E[\Lambda(X)|\omega_2] \quad (2.9)$$

Então, substituindo-se  $\Lambda(\mathbf{X})$  pela sua definição, a Divergência pode ser reescrita como:

$$D = \int_{-\infty}^{\infty} \left( \ln \frac{p(X/\omega_1)}{p(X/\omega_2)} \right) p(X/\omega_1) dX - \int_{-\infty}^{\infty} \left( \ln \frac{p(X/\omega_1)}{p(X/\omega_2)} \right) p(X/\omega_2) dX \quad (2.10)$$

Quando os dados são Gaussianos, a Divergência adquire a forma:

$$D = 1/2 \text{tr}(\Sigma_1^{-1} \Sigma_2 + \Sigma_2^{-1} \Sigma_1 - 2I) + 1/2 (\mu_1 - \mu_2)^T (\Sigma_1^{-1} + \Sigma_2^{-1}) (\mu_1 - \mu_2) \quad (2.11)$$

em que  $\text{tr}$  é o traço de:

$$(\Sigma_1^{-1} \Sigma_2 + \Sigma_2^{-1} \Sigma_1 - 2I); \text{ e } \mathbf{I}, \text{ é a matriz identidade.}$$

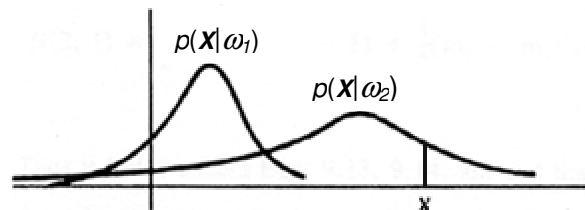
### 2.2.2 Distância de Bhattacharyya

A distância de Bhattacharyya é uma outra distância estatística que pode ser usada na estimação da separabilidade entre um par de classes (LANDGREBE, 2003), (TERRIEN, 1989, 2001), (DUDA, 2001), (FUKUNAGA, 1990).

A forma geral da distância de Bhattacharyya é definida por:

$$B = -\ln \left[ \int_{-\infty}^{\infty} \sqrt{p(\mathbf{X} | \omega_1) p(\mathbf{X} | \omega_2)} d\mathbf{X} \right] \quad (2.12)$$

A FIG. 9 mostra um par típico de funções densidade  $p(\mathbf{X} | \omega_1)$  e  $p(\mathbf{X} | \omega_2)$ . Para interpretar a distância de Bhattacharyya, note que se as funções originais estão bem separadas e a probabilidade de  $\mathbf{X}$  com respeito à classe  $i$  for alta, a probabilidade de  $\mathbf{X}$  com respeito à classe  $j$  será muito próxima de zero.



**FIGURA 9 - Exemplo típico de separação entre duas funções densidade.**

Fonte: Adaptado de Therrien (1989).

A distância de Bhattacharyya é invariante frente a uma transformação linear do vetor  $\mathbf{X}$  e também é aditiva quando os componentes de  $\mathbf{X}$  são independentes, isto é, pode ser expressa como uma soma dos termos similares com cada termo envolvendo somente uma das componentes de  $\mathbf{X}$ . Em adição a isso, se  $J_m(\omega_1, \omega_2)$  representa a distância entre as duas classes baseada em uma variável  $\mathbf{X}$  com  $m$  componentes, então as seguintes propriedades métricas de uma função distância são apropriadas:

$$\begin{aligned} J_m(\omega_1, \omega_2) &> 0; \omega_1 \neq \omega_2 \\ J_m(\omega_1, \omega_1) &= J_m(\omega_2, \omega_2) = 0 \\ J_m(\omega_1, \omega_2) &= J_m(\omega_2, \omega_1) \end{aligned} \quad (2.13)$$

Tais propriedades não satisfazem a desigualdade triangular e, assim não podem ser classificadas como funções verdadeiras de distâncias. Contudo, ambas satisfazem a propriedade adicional:

$$J_m(\omega_1, \omega_2) \leq J_{m+1}(\omega_1, \omega_2) \quad (2.14)$$

### 2.2.2.1 Distância de Bhattacharyya: Forma Gaussiana

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

$$\mathbf{B} = \frac{1}{8}(\mu_1 - \mu_2)^T \left( \frac{\Sigma_1 + \Sigma_2}{2} \right)^{-1} (\mu_1 - \mu_2) + \frac{1}{2} \ln \left( \frac{\left| \frac{1}{2}(\Sigma_1 + \Sigma_2) \right|}{|\Sigma_1|^{1/2} |\Sigma_2|^{1/2}} \right) \quad (2.15)$$

onde  $\mu_1$  e  $\mu_2$  são os vetores média e  $\Sigma_1$  e  $\Sigma_2$  são as matrizes de covariância das classes. Na expressão acima, pode-se observar que a primeira parcela no membro da direita estima a contribuição dos vetores de médias (*Bhatt Mean*) no valor da distância de Bhattacharyya ( $B$ ), enquanto que a segunda parcela (*Bhatt Cov*) estima a contribuição das matrizes de covariância. Note que  $B$  é um número real, variando no intervalo  $[0, \infty]$ .

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

Note que quando as matrizes de covariância para as duas classes são iguais, a distância de Bhattacharyya ( $B$ ), e a distância da Divergência ( $D$ ), são medidas equivalentes, tal que:

$$D = 8B = (\mu_1 - \mu_2)^T \Sigma^{-1} (\mu_1 - \mu_2) \quad (2.16)$$

Essa quantidade é também reconhecida como a distância de Mahalanobis entre duas classes.

### 2.3 ANÁLISE DISCRIMINANTE REGULARIZADA (RDA)

As técnicas de Análise Discriminante Regularizada se constituem em uma segunda possível via para fins de minimizar o problema da incerteza na estimação dos parâmetros do classificador, devido ao número reduzido de amostras de treinamento frente à dimensionalidade dos dados (fenômeno de Hughes).



Nesta abordagem procura-se obter estimativas mais confiáveis para os parâmetros do classificador, introduzindo-se ou forçando-se uma tendenciosidade adequada em sua estimação, de forma que as estimativas sejam mais plausíveis fisicamente (FRIEDMAN, 1989). Os métodos que empregam esta abordagem são genericamente conhecidos como *Análise Discriminante Regularizada* (RDA).

Diferentes abordagens para fins de regularização da matriz covariância estão relatadas na literatura. Os autores Titterington (1985), O'sullivan (1986), Friedman (1989), Hoffbeck, Landgrebe (1996), mostram que as técnicas de regularização obtiveram alto sucesso na solução de problemas de mal-condicionamento da matriz covariância.

## **2.4 TÉCNICAS DE INCREMENTO DO NÚMERO DE AMOSTRAS DE TREINAMENTO SEMI-ROTULADAS**

Uma terceira via para fins de minimização do efeito de Hughes consiste em aumentar o número de amostras de treinamento pela inclusão das chamadas amostras semi-rotuladas. Shahshahani e Landgrebe (1994), propõem o uso do método iterativo EM (*Expectation Maximization*) na geração de amostras de treinamento adicionais, a partir de *pixels* não rotulados, isto é, *pixels* na imagem que não fazem parte da amostra de treinamento.

Jackson (2001, 2002), propõe uma família de estimadores adaptativos de covariância para mitigar o problema causado pelo número limitado e insuficiente de amostras de treinamento na análise de dados hiperespectrais utilizando um classificador quadrático (Máxima Verossimilhança Gaussiana). Nestes estimadores propostos, as amostras semi-rotuladas são incorporadas no processo determinando os melhores parâmetros regularizadores.

Lemos (2003), desenvolveu um estudo sobre um classificador adaptativo baseado na proposta de Jackson (2001), que visa suavizar os efeitos causados por um número insuficiente de amostras de treinamento, fato este que pode degradar severamente a acurácia dos resultados obtidos por um classificador paramétrico utilizando dados com dimensão alta.

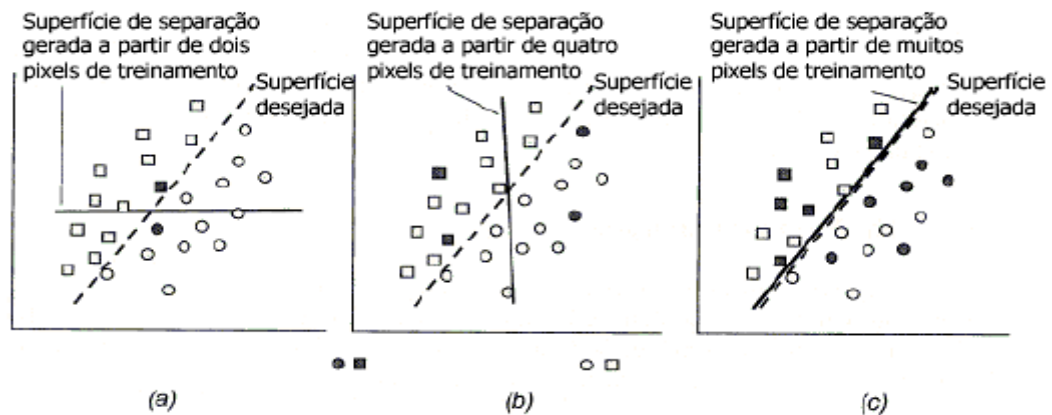
## **2.5 INFLUÊNCIA DO VOLUME DE AMOSTRAS DE TREINAMENTO NA ESTIMAÇÃO DE PARÂMETROS**

O problema que resulta da utilização de um número pequeno de amostras de treinamento na estimação dos parâmetros em um classificador já era bastante conhecido pela comunidade que opera com a classificação de imagens multiespectrais em sensoriamento remoto. Este problema porém, só adquiriu uma notoriedade maior após a disponibilidade de dados em alta dimensionalidade, como os coletados pelos sistemas sensores hiperespectrais.

Resumidamente, uma relação mínima entre o número de amostras de treinamento e o número de bandas espectrais é necessária para assegurar que sejam obtidas estimativas confiáveis para o treinamento do classificador (num processo supervisionado). Na medida em que a dimensionalidade dos dados aumenta, o número de pixels de treinamento necessários por classe também deve ser aumentado para que seja preservado o mesmo grau de confiabilidade nos valores estimados para os parâmetros. Assim, a adição de mais bandas espectrais, como no caso do sensor AVIRIS, não é útil a menos que mais amostras de treinamento (pixels) por classe estejam disponíveis (RICHARDS,1999). Isto demonstra uma das maiores limitações da aplicação de procedimentos usuais de classificação em imagens hiperespectrais.

A FIG.10 ilustra a importância de um número de pixels de treinamento suficiente para assegurar uma estimativa confiável da superfície de separação das classes. Quando um número de pixels muito pequeno é utilizado (a), pode-se obter

uma boa separação dos pixels de treinamento, porém o desempenho do classificador para o conjunto de pixels de teste é pobre. Uma grande quantidade de pixels de treinamento (posicionados aleatoriamente), gera uma superfície que também é satisfatória para as amostras de teste (c).



**FIGURA 10 - Efeito do número de amostras na estimação de parâmetros.**

Fonte: Adaptado de Richards (1999).

Um exemplo simples, baseado apenas na tentativa de terminação da superfície de separação de duas classes, pode ser utilizado para ilustrar o problema. A FIG.10 acima, mostra três diferentes conjuntos de amostras de treinamento para o mesmo conjunto de dados bidimensional. O primeiro tem apenas um *pixel* por classe. Conforme pode ser observado, mesmo que a superfície de separação possa ser encontrada, ela não pode ser acurada. Tendo-se dois pixels por classe FIG.10(b), têm-se melhor estimativa da superfície de separação das classes. Mas é apenas quando o número de pixel por classe é elevado, se comparado ao número de bandas da imagem (neste caso duas) que se obtém uma boa estimativa para o classificador supervisionado, FIG. 10(c).

Em sensoriamento remoto, os dados 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.

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 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  $\mathbf{X}$  pode ser representado como:

$$\mathbf{X} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ \vdots \\ x_p \end{bmatrix} \quad (2.17)$$

Os dados obtidos pelas  $p$  bandas espectrais são comumente representados em um espaço  $p$ -dimensional, sendo cada eixo associado a uma banda espectral. Neste espaço, denominado de *espaço das variáveis* cada padrão (*pixel*) fica representado por um ponto que pode ser convenientemente descrito por um vetor.

As classes podem ser representadas por

$$\omega_i, \quad i = 1, \dots, k \quad (2.18)$$

onde  $k$  é o número total de classes. Idealmente, cada classe deveria se constituir em um *aglomerado* bem definido de pontos no espaço das variáveis, cada ponto representando um *pixel*, como no exemplo para um caso bi-dimensional (FIG. 11).

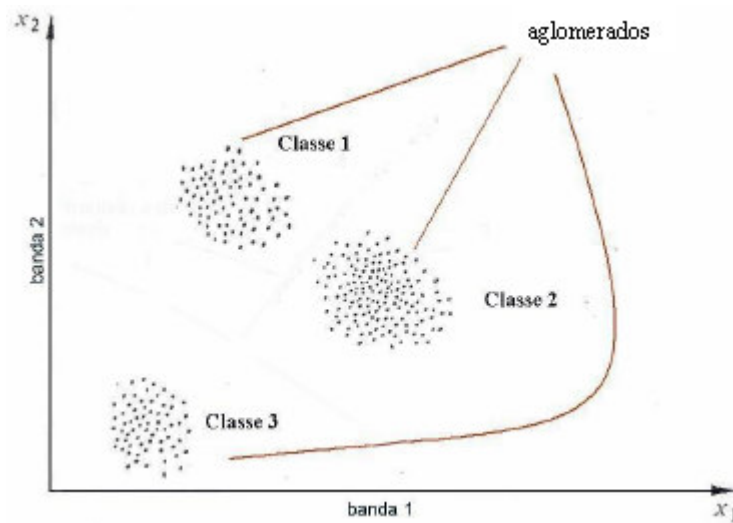
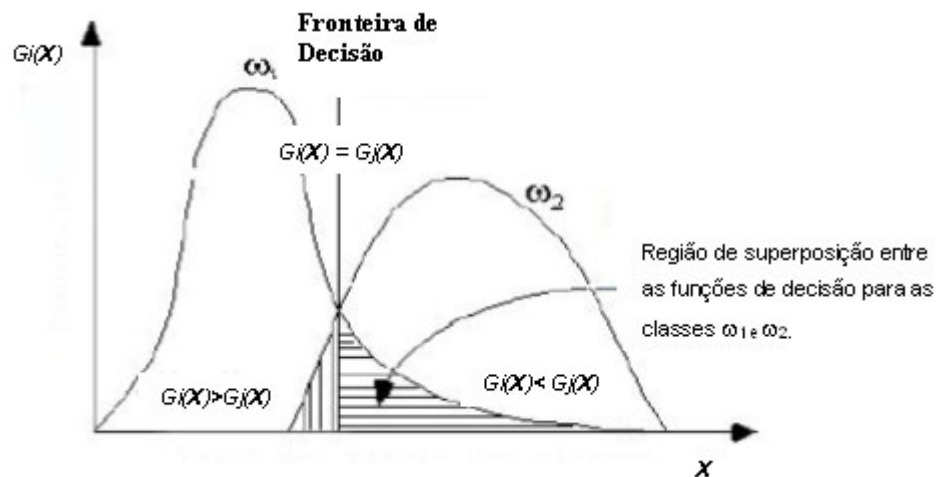


FIGURA 11 - Espaço das variáveis para  $n=2$

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 cada padrão (*pixel*) pertence. A cada classe é associada, portanto, uma função decisão  $G_i(\mathbf{X})$ . O processo de classificação é geralmente descrito pela seguinte regra: atribui o *pixel*  $\mathbf{X}$  à classe  $\omega_i$ , se,

$$G_i(\mathbf{X}) > G_j(\mathbf{X}), \quad \forall i \neq j \quad (2.19)$$

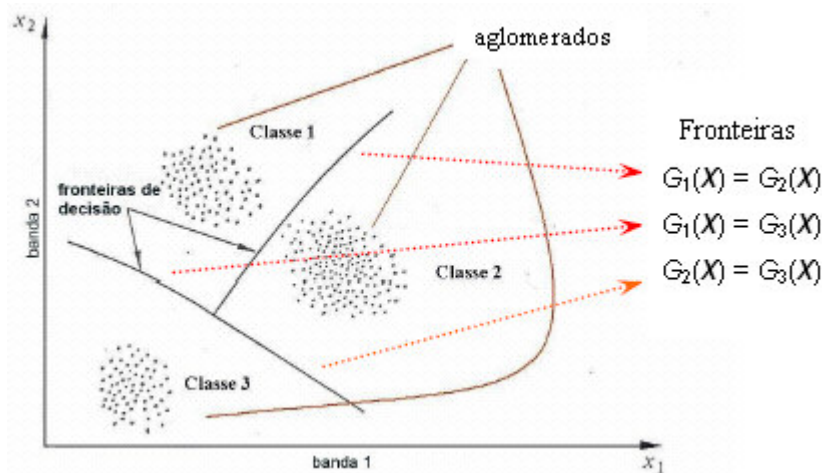
Como a função decisão é uma função probabilística, o *pixel* em teste será classificado na classe para a qual ele apresentar a maior probabilidade de pertencer, como mostra o exemplo (FIG. 12).



**FIGURA 12 - Fronteiras de decisão entre 2 classes**

Fonte: Adaptado de Richards (1993).

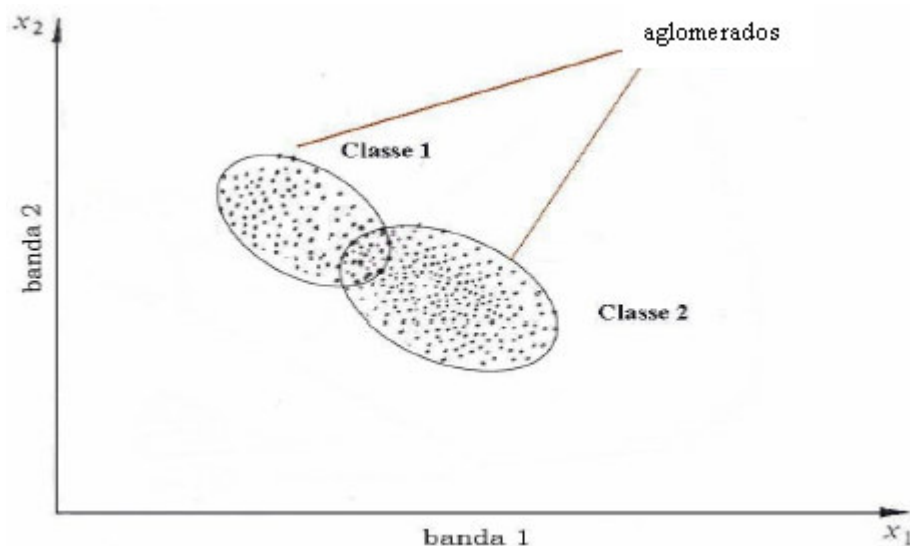
Estas funções de decisão servem também para particionar o *espaço das variáveis*, definindo as sub-regiões associadas a cada uma das classes, conforme os exemplos ilustrados (FIG. 12 e 13).



**FIGURA 13 - 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 superposição ou interpenetração entre os diferentes aglomerados, FIG.14, tornando inviável o estabelecimento de fronteiras rígidas, separando as várias classes no espaço das

variáveis. Esta situação requer, portanto, uma abordagem não determinística ao problema da classificação. Em outras palavras, as funções decisão devem ser de natureza probabilística, de modo a permitir uma modelagem mais adequada do comportamento espectral das classes em consideração, isto é, da distribuição destas classes no *espaço das variáveis*. Nesta abordagem, o vetor que representa cada *pixel* no espaço das variáveis é entendido como uma variável aleatória  $p$ -dimensional, cujo comportamento pode ser descrito por uma função densidade de probabilidade. Neste particular, entre as várias abordagens possíveis, merece destaque àquela adotada pelo classificador *Bayesiano*.



**FIGURA 14 - Interpenetração de classes no espaço das variáveis.**

Neste capítulo, é inicialmente revisto o problema de decisão envolvendo duas classes, o qual pode ser posteriormente estendido para o problema multi-classe.

## **2.6 OS CLASSIFICADORES DE BAYES E DA MÁXIMA VEROSSIMILHANÇA**

Na abordagem estatística, o vetor do pixel  $\mathbf{X}$ , pertencente à classe  $\omega$ , é visto como uma variável aleatória, representando-se por  $p(\mathbf{X}/\omega)$  a função densidade de

probabilidade condicional para cada uma das classes envolvidas. Neste contexto estatístico, várias funções de decisão foram propostas por diferentes investigadores. Entre eles, o classificador de Bayes apresenta características que o tornam um classificador “ótimo”, no sentido de que minimiza a perda média que se incorre ao rotular um pixel como pertencente a uma determinada classe.

Os seguintes conceitos são necessários para o entendimento do classificador de Bayes:

1ª)  $P(\omega_i)$ , é a probabilidade *a priori* da classe  $\omega_i$ . É a probabilidade de ocorrência de um padrão da classe  $\omega_i$ , independente de qualquer outra informação. Em se tratando de cenas cobrindo, por exemplo, áreas agrícolas, se os dados históricos indicam ser de 30% da área total a fração ocupada por um determinado tipo de cultura agrícola (ex.: milho), então o valor da probabilidade *a priori* para esta classe pode ser estimado pelo valor desta fração, isto é,  $P(\omega_i) = 0.3$ , e assim, respectivamente para todas as demais classes envolvidas no processo de classificação.

2ª)  $p(\mathbf{X}/\omega_i)$  é a função densidade de probabilidade associada à classe  $\omega_i$ .

3ª)  $P(\omega_i/\mathbf{X})$  é a probabilidade *a posteriori* da classe  $\omega_i$  condicionada ao padrão (pixel)  $\mathbf{X}$ .

4ª)  $p(\mathbf{X})$ , é a função densidade de probabilidade para  $\mathbf{X}$ . É a probabilidade de encontrar um pixel de qualquer classe, na posição  $\mathbf{X}$ .

A estratégia do classificador *Bayesiano* consiste em classificar cada padrão de modo a minimizar a perda média condicional que se incorre ao rotular um pixel como pertencente a uma dada classe (RICHARDS, 1993).



Inicialmente, é definida a função de perda como sendo o custo, perda ou penalidade incorrida quando o classificador decide  $\mathbf{X} \in \omega_i$ , quando a decisão correta seria  $\mathbf{X} \in \omega_j$ . Representa-se esta *função perda* por:

$$C(i, j) \quad i, j = 1, \dots, k \quad (2.20)$$

Logo, a perda média condicional ou risco médio que se incorre ao decidir rotular  $\mathbf{X}$  na classe  $\omega_i$ , é definida por:

$$L_X(\omega_i) = \sum_{j=1}^k C_{i,j} p(\omega_j | X) \quad (2.21)$$

A probabilidade *a posteriori*  $P(\omega_j | \mathbf{X})$  pode ser expressa de uma forma mais conveniente empregando-se o teorema de Bayes:

$$P(\omega_j | \mathbf{X}) = p(\mathbf{X} | \omega_j) P(\omega_j) / p(\mathbf{X}) \quad (2.22)$$

O objetivo consiste em classificar  $\mathbf{X}$  de modo a minimizar esta função perda média condicional (2.21). É nesse sentido que o classificador *Bayesiano* é denominado de “ótimo”. Substituindo (2.22), na equação (2.21), têm-se:

$$L_X(\omega_i) = \frac{1}{p(\mathbf{X})} \sum_{j=1}^k C_{i,j} p(\mathbf{X} | \omega_j) P(\omega_j) \quad (2.23)$$

Como o interesse consiste em atribuir o *pixel*  $\mathbf{X}$  à classe para a qual  $L_X(\omega_i)$  é mínimo, e visto que, a probabilidade  $p(\mathbf{X})$  é comum para todas as funções decisão, pode-se eliminá-la da equação (2.23), tornando-se:

$$L_X(\omega_i) \propto \sum_{j=1}^k C_{i,j} p(\mathbf{X} | \omega_j) P(\omega_j) \quad (2.24)$$

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

$$G_i(X) = -L_X(\omega_i) \quad (2.25)$$

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

$$\begin{aligned} C_{i,j} &= 0, & \forall i = j \\ C_{i,j} &= 1, & \forall i \neq j \end{aligned} \quad (2.26)$$

Com esta suposição, a função decisão  $G_i(\mathbf{X})$  fica:

$$G_i(X) = -\sum_{\substack{j=1 \\ j \neq i}}^k p(X | \omega_j)P(\omega_j) \quad (2.27)$$

É possível simplificar a equação (2.27). Sabendo-se que uma função densidade de probabilidade  $p(\mathbf{Y})$  qualquer, com condição de  $\sum_{j=1}^k P(\omega_j) = 1$  pode ser expressa como:

$$p(Y) = \sum_{j=1}^k p(Y | \omega_j)P(\omega_j) \quad (2.28)$$

ou:

$$p(Y) = p(Y | \omega_i)P(\omega_i) + \sum_{\substack{j=1 \\ j \neq i}}^k p(Y | \omega_j)P(\omega_j) \quad (2.29)$$

ou

$$\sum_{\substack{j=1 \\ j \neq i}}^k p(Y | \omega_j)P(\omega_j) = p(Y) - p(Y | \omega_i)P(\omega_i) \quad (2.30)$$

então, substituindo (2.30), a regra na equação (2.27), onde  $G_i(\mathbf{X}) = -L_X(\omega_i)$  obtém-se:

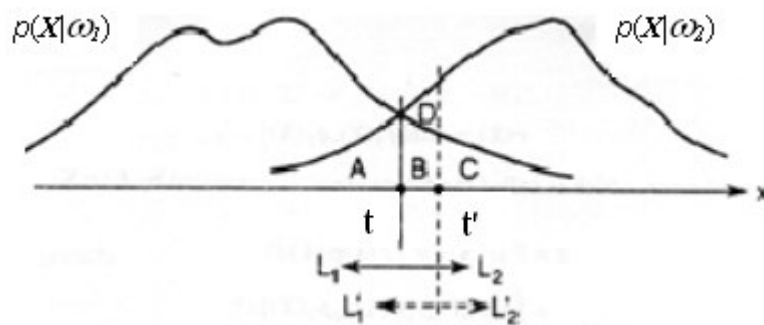
$$G_i(X) = -p(X) + p(X|\omega_i) P(\omega_i) \quad (2.31)$$

Finalmente, 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:

$$G_i'(X) = p(X|\omega_i) P(\omega_i) \quad (2.32)$$

### 2.6.1 O Classificador de Bayes

O classificador Bayesiano é um classificador *ótimo* no sentido de que minimiza a probabilidade de erro. Esta probabilidade mínima de erro é conhecida como *erro de Bayes*. A FIG. 15 ilustra o caso unidimensional dessa regra, onde o erro bayesiano, dado a fronteira de decisão  $t$  é o menor possível, conforme a ilustração,  $\varepsilon = A+B+C$ . Caso seja tomado qualquer outro valor para fronteira de decisão, digamos  $t'$ , a probabilidade de erro é aumentada pela quantidade  $D$ , sendo assim maior que o erro de Bayes (FUKUNAGA, 1990).



**FIGURA 15 - A regra de decisão de Bayes**

Fonte: Adaptado de Fukunaga (1990).

O classificador implementa uma abordagem probabilística ao processo de classificação. A probabilidade de erro neste caso pode ser estimada por:

$$\varepsilon = P(\omega_1) \int_{L_2} p(X / \omega_1) dX + P(\omega_2) \int_{L_1} p(X / \omega_2) dX = P(\omega_1) \varepsilon_1 + P(\omega_2) \varepsilon_2 \quad (2.33)$$

onde  $L_1$  e  $L_2$  são as regiões de domínio das classes  $\omega_1$  e  $\omega_2$ , respectivamente.

### 2.6.2 Classificador Máxima Verossimilhança Gaussiana

A função de decisão (2.32) é referida também por vários autores por “Máxima Verossimilhança”. Este classificador pode ser visto como um caso particular do classificador *Bayesiano*, no qual se atribuí para as funções de perda (2.20) os valores definidos em (2.26). Também, é freqüente supor que as probabilidades *a priori* para todas as classes em consideração têm idêntico valor, em virtude das dificuldades naturais em estimá-los de forma confiável, o que normalmente se faz através de dados históricos, senão diante aos altos custos operacionais quando é necessária a coleta de informações diretamente em campo. A regra de classificação fica, portanto:

$$X \in \omega_i \text{ se } p(X | \omega_i)P(\omega_i) > p(X | \omega_j)P(\omega_j), \text{ para todo } j \neq i \quad (2.34)$$

A efetiva implementação da função decisão de máxima verossimilhança, conforme definida na equação (2.34), requer que seja adotada uma forma específica para a função densidade de probabilidade  $p(\mathbf{X}|\omega)$ . Entre as inúmeras possibilidades, a experiência com dados de sensoriamento remoto tem mostrado que a função de Gauss é adequada na modelagem destes dados. Este fato, ratificado experimentalmente, pode ser mais bem explicado pelo *Teorema Central do Limite*. A radiância registrada pelo sistema sensor é na verdade o somatório de um grande número de outras variáveis aleatórias. Embora estas variáveis possivelmente não sigam a distribuição normal, o seu somatório tenderá a se aproximar daquela distribuição.

Assumindo-se agora, que serão utilizadas  $N$  bandas espectrais, a função densidade de probabilidade em (2.34) fica:

$$p(X | \omega_i) = (2\pi)^{\frac{-N}{2}} |\Sigma_i|^{-\frac{1}{2}} \exp\left\{-\frac{1}{2} (X - \mu_i)^T \Sigma_i^{-1} (X - \mu_i)\right\} \quad (2.35)$$

onde  $\mu_i$  representa o vetor de média e  $\Sigma_i$  representa a matriz covariância da classe  $\omega_i$ . Em situações práticas estes parâmetros são estimados a partir do conjunto das amostras de treinamento disponíveis para cada classe, como segue:

$$\mathbf{M}_k = \frac{1}{N_k} \sum_{i=1}^{N_k} \mathbf{X}_i \quad \mathbf{S}_k = \frac{1}{N_k} \sum_{i=1}^{N_k} (\mathbf{X}_i - \mathbf{M}_k)(\mathbf{X}_i - \mathbf{M}_k)^T \quad (2.36)$$

onde  $N_k$  é o número de amostras de treinamento da classe  $\omega_k$ . Como a regra de classificação é a equação (2.19), qualquer função monotonicamente crescente de (2.34) mantém inalterado o resultado de (2.19). 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  $G_i(\mathbf{X})$ , a função decisão Máxima Verossimilhança dentro da suposição Gaussiana, pode ser simplificada para:

$$G_i(X) = -\ln|\Sigma_i| - (X - \mu_i)^T \Sigma_i^{-1} (X - \mu_i) + \ln(P(\omega_i)), \quad i=1 \dots k \quad (2.37)$$

Esta função decisão claramente representa um classificador quadrático, isto é, inclui termos  $\mathbf{X}^2$ , resultantes do  $(X - \mu_i)^T \Sigma_i^{-1} (X - \mu_i)$ .

Neste caso, as fronteiras de decisão assumem a forma de superfícies (ou hipersuperfícies) quadráticas, como elipsóides ou (hiper-elipsóides). As fronteiras de decisão entre as classes  $\omega_i$  e  $\omega_j$  são obtidas, simplesmente, igualando-se as funções decisão,

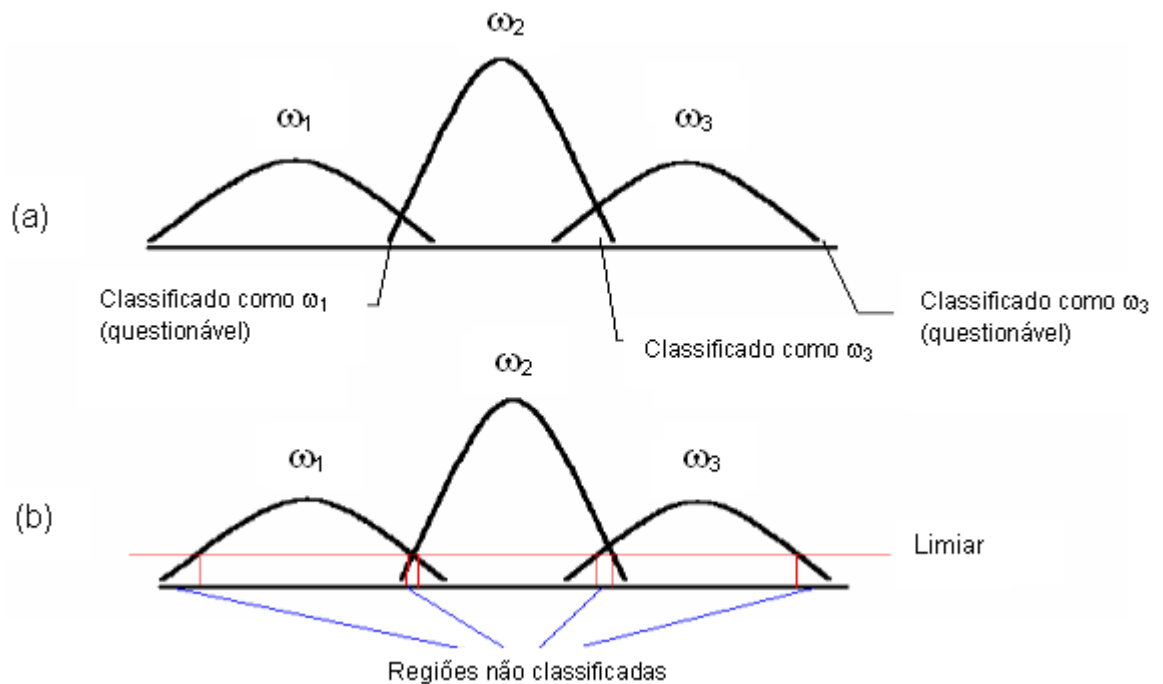
$$\begin{aligned} G_i(X) &= G_j(X) \\ \text{ou} \quad G_i(X) - G_j(X) &= 0 \end{aligned} \quad (2.38)$$

Em algumas situações é aceitável assumir que as matrizes covariância das classes em consideração são suficientemente semelhantes, permitindo a adoção

de uma matriz covariância comum  $\Sigma$ . Neste caso, a expressão (2.37) 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.

## 2.7 LIMIARES

Está implícito no desenvolvimento anterior, que os *pixels* em cada ponto do espaço das variáveis, serão classificados como pertencentes a uma das classes disponíveis  $\omega_i$ , desconsiderando o fato de que alguns *pixels* têm probabilidade bastante reduzida de pertencerem a qualquer uma das classes, conforme ilustrado na FIG. 16 para o caso de dados unidimensionais.



**FIGURA 16.a) Ilustração de uma classificação pouco precisa para padrões localizados nas caudas das distribuições.**  
**b) Uso de limiares para remover a classificação não acurada.**

Na FIG.16(a), está ilustrado o resultado de uma classificação não acurada.

A regra de classificação, conforme (2.19) requer aqui, entretanto, uma consideração adicional. Em alguns casos, a probabilidade de um *pixel* pertencer à classe para a qual o valor da correspondente função decisão é máximo, pode ser pequena (FIG.16(a)). Em situações como a ilustrada na FIG.16(a), a situação mais provável é a de que o *pixel* na verdade pertença a uma outra classe que por não ter sido amostrada, não foi gerada a correspondente função decisão. Nesta condição, a decisão mais adequada consiste em manter o *pixel* como não classificado. A detecção de *pixels* nesta situação pode ser feita pela introdução de um limiar  $T_i$ . Este limiar tem por objetivo detectar pixels com baixa probabilidade de pertencer a qualquer uma das classes amostradas e deixá-lo, portanto, como não classificado:

$$X \in \omega_i \text{ se } G_i(X) > G_j(X) \text{ para } j \neq i \text{ e } G_i > T_i \quad (2.39)$$

A partir das funções definidas em (2.35) e (2.37), uma classificação é, portanto adequada, se:

$$\ln p(\omega_i) - \frac{1}{2} \ln |\Sigma_i| - \frac{1}{2} (X - \mu_i)^t \Sigma_i^{-1} (X - \mu_i) > T_i \quad (2.40)$$

isto é:

$$G_i(X) = - (X - \mu_i)^T \Sigma_i^{-1} (X - \mu_i) - \ln |\Sigma_i| + 2 \ln P(\omega_i) \quad (2.41)$$

Sabe-se que o termo da esquerda em (2.41) tem uma distribuição Qui-quadrado ( $\chi^2$ ) com N graus de liberdade, desde que  $X$  tenha uma distribuição normal. O índice N refere-se à dimensionalidade do espaço das variáveis (número de bandas). Conhecendo-se a função densidade de probabilidade deste termo, torna-se possível estabelecer um teste de hipóteses e, portanto um limiar, de forma a excluir valores com baixa probabilidade de ocorrência. Dessa maneira, *pixels* que certamente não

pertencem a nenhuma das classes selecionadas, não serão erroneamente classificados e assim não influenciarão negativamente no resultado final.

Como exemplo Richards (1986), considera-se aqui o limiar para o sensor Landsat MSS (com quatro bandas espectrais somente), sendo, portanto, equivalentes aos graus de liberdade ( $gl=4$ ), caso em que num processo de classificação no qual 5% dos *pixels* com menor probabilidade de pertencer à classe selecionada serão deixados como não classificados. O valor da distribuição Qui-quadrado ( $\chi^2$ ) para um nível de significância de 95% é igual a 9.488. O valor do limiar fica então:

$$T_i = -4,744 - \frac{1}{2} \ln |\Sigma_i| + \ln P(\omega_i) \quad (2.42)$$

## 2.8 O CLASSIFICADOR QUADRÁTICO

O processo que implementa a regra de classificação (2.37), constitui-se em um *Classificador Quadrático*, visto que as funções discriminantes utilizadas são quadráticas em  $\mathbf{X}$ . O processo que implementa a função decisão linear constitui-se em um *Classificador Linear*.

No caso do classificador quadrático *Máxima Verossimilhança Gaussiana*, estima-se para cada classe um vetor média e uma matriz covariância próprios, resultando em funções de decisão quadráticas (2.37).

Aeberhard (1994) cita que, pelo motivo de no método linear ocorrer número muito menor de parâmetros a serem estimados, em alguns casos torna-se preferível usar o classificador linear ao classificador quadrático em espaços de alta dimensionalidade (o que implica em um número elevado de parâmetros a serem estimados), especialmente quando o número de amostras de treinamento disponíveis é



pequeno. Porém, por possuir fronteiras lineares, o classificador linear tem sérias limitações em casos onde as matrizes covariâncias das classes diferem significativamente.

No caso extremo de uma situação que envolva uma quantidade muito limitada de amostras de treinamento, pode resultar mais eficiente empregar um classificador que não faça uso das informações da matriz covariância, como é o caso do classificador 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 *pixel* na classe com o vetor média mais próximo (RICHARDS, 1986).

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, e por fim o classificador de máxima verossimilhança gaussiana, que supõe que os dados apresentam a distribuição de probabilidade multivariada de Gauss, é o conhecimento teórico em que está baseado o presente estudo.

## CAPÍTULO III

### METODOLOGIA

#### 3.1 INTRODUÇÃO

No processo de classificação de dados em imagens de cenas naturais, vem se utilizando basicamente de dados fornecidos por sistemas sensores que operam com um número relativamente reduzido de bandas espectrais. Sistemas sensores como Landsat-TM e SPOT têm sido amplamente utilizados na obtenção de informações em áreas diversas tais como florestas, agricultura, geologia, além de várias outras. Na maioria destes casos, as classes de interesse apresentam uma distribuição normal multivariada e o processo de classificação pode ser feito utilizando o bem conhecido classificador quadrático Máxima Verossimilhança Gaussiana (MVG):

$$G_i(\mathbf{X}) = -(\mathbf{X} - \boldsymbol{\mu}_i)^T \boldsymbol{\Sigma}_i^{-1} (\mathbf{X} - \boldsymbol{\mu}_i) - \ln |\boldsymbol{\Sigma}_i| + 2 \ln P(\omega_i) \quad (3.1)$$

onde  $\omega_i$  representa uma classe em particular,  $\mathbf{X}$  um pixel genérico a ser classificado,  $\boldsymbol{\mu}_i$  o vetor de médias desta classe,  $\boldsymbol{\Sigma}_i$  a matriz de covariância da classe,  $P(\omega_i)$  a correspondente probabilidade *a priori* desta classe, e  $G_i(\mathbf{X})$  a função discriminante MVG associada a classe  $\omega_i$ .

Na prática, esta abordagem tem se mostrado satisfatória, produzindo imagens temáticas com um aceitável grau de acurácia. Ocorrem, entretanto, algumas situações em que algumas ou todas as classes envolvidas, são espectralmente muito semelhantes entre si, isto é, os vetores de médias são quase idênticos. Nestes casos, dados em baixa dimensionalidade como os fornecidos por sensores tradicionais,

produzem uma imagem temática com baixa acurácia. Uma possível solução para problemas desta natureza é aquela fornecida por uma geração mais recente de sistemas sensores que coletam dados em alta dimensionalidade (dados hiperespectrais). O sistema AVIRIS, por exemplo, coleta dados de cenas naturais em 224 bandas espectrais, oferecendo novas possibilidades para a separação de classes espectralmente muito semelhantes entre si. Pode-se mostrar que classes que compartilham do mesmo vetor de médias podem, entretanto, serem separadas com alto grau de acurácia em um espaço em alta dimensionalidade, desde que suas matrizes de covariância difiram entre si, substancialmente (FUKUNAGA, 1990).

Conforme mencionado nos capítulos anteriores, a limitação prática que dificulta a implementação de dados em alta dimensionalidade, consiste na estimação do grande número de parâmetros que resultam de dados em alta dimensionalidade. O número de amostras de treinamento, normalmente disponível em situações práticas é insuficiente para uma estimativa confiável deste número crescente de parâmetros, resultando em estimativas pouco confiáveis para os parâmetros do classificador, o que por sua vez, compromete o desempenho do classificador e, conseqüentemente, a acurácia da imagem temática produzida.

Entre as abordagens visando a minimização deste problema, e propostas pelos vários pesquisadores que atuam na área de Reconhecimento de Padrões, uma refere-se à técnica de incremento no número de amostras de treinamento disponíveis (amostras rotuladas) por meio das chamadas amostras semi-rotuladas. Amostra semi-rotulada é a designação genérica dada às amostras cuja classe a que pertencem (rótulo), foi obtida por métodos de inferência e não por observação direta como é o caso das amostras rotuladas. Ao contrário das amostras rotuladas, cujo rótulo é conhecido, as amostras semi-rotuladas são rotuladas com um certo grau de incerteza, o que se traduz pelo peso reduzido a elas atribuído no processo de estimação dos parâmetros do classificador.

Um outro problema que surge ao empregar-se um classificador estatístico como a MVG, refere-se à estimação dos valores para as probabilidades *a priori* associadas à cada classe. Em situações práticas, a estimação dos valores para estas probabilidades é de difícil implementação prática. Por esta razão, o procedimento mais comumente adotado, consiste em atribuir um único valor para as probabilidades *a priori* de todas as classes envolvidas. Este não é obviamente, um procedimento adequado, pois introduz no classificador valores que não correspondem à realidade, provocando um decréscimo na sua performance, e uma correspondente degradação na acurácia da imagem temática produzida.

### 3.2 NOVA METODOLOGIA PROPOSTA

Neste estudo é investigada e testada uma proposta visando uma solução prática envolvendo ambos os problemas acima mencionados. Na metodologia apresentada neste trabalho, propõe-se um algoritmo iterativo, através do qual é extraído um determinado número de amostras semi-rotuladas, com o objetivo de incrementar o reduzido número de amostras rotuladas disponíveis. Neste mesmo processo iterativo, são também estimados valores para as probabilidades *a priori* das classes em consideração. Espera-se desta forma, incrementar a acurácia nas imagens temáticas produzidas pelo classificador (3.1).

A metodologia proposta está sucintamente descrita no fluxograma ilustrado (FIG. 17). O procedimento é essencialmente formado pela adição da etapa em destaque (sombreamento), no classificador *Máxima Verossimilhança Gaussiana*.

O processo inicia com uma primeira classificação, onde as probabilidades *a priori* serão estimadas e atualizadas a cada iteração da seguinte forma:

1- Na primeira iteração, são utilizados valores iguais para as probabilidades *a priori* de todas as classes. Esta aproximação quase sempre é feita, devido à dificuldade prática na obtenção de valores mais realistas para estas probabilidades. Conforme Lemos (2003), nesta primeira iteração os parâmetros são estimados somente a partir das amostras rotuladas disponíveis;

2- Como resultado desta classificação inicial (de baixa confiabilidade), resulta uma imagem temática (baixa acurácia). Desta imagem é extraído um certo número de amostras semi-rotuladas e estimados os respectivos pesos, que estimam o grau de confiança no acerto da classificação para cada pixel individual, para utilização junto com as amostras rotuladas iniciais em uma nova estimação dos parâmetros do classificador. Com base nesta imagem temática, são também atualizadas as probabilidades *a priori*. Na metodologia proposta, o valor das probabilidades *a priori* associadas a cada uma das classes individuais, é estimado pela fração dos pixels rotulados na respectiva classe. Estes valores estimados são atualizados a cada iteração.

3- Uma nova classificação é feita, utilizando estas novas estimativas para os parâmetros do classificador e os novos valores para as probabilidades *a priori*. A nova imagem classificada, assim obtida, deve ser mais confiável que a anterior;

4- Este processo repete-se de uma forma iterativa, atualizando-se a cada etapa tanto os novos valores estimados para os parâmetros do classificador, como também para as probabilidades *a priori*, gerando-se desta forma imagens classificadas cada vez mais acuradas;

5- O critério de parada para este processo iterativo, ocorre quando o número de pixels que trocaram de classe na iteração anterior em relação à seguinte for menor que 5%.

As amostras rotuladas recebem peso igual à unidade, indicando uma total confiabilidade na classe atribuída a cada amostra individual. As amostras semi-rotuladas recebem uma ponderação proporcional ao grau de pertinência à classe vencedora no processo de classificação e contido no intervalo  $[0,1]$ . O cálculo dos pesos para as amostras semi-rotuladas é realizado por meio da expressão (3.2) e a estimação ponderada dos parâmetros (vetor média e matriz covariância) pelas expressões (3.3) e (3.4).

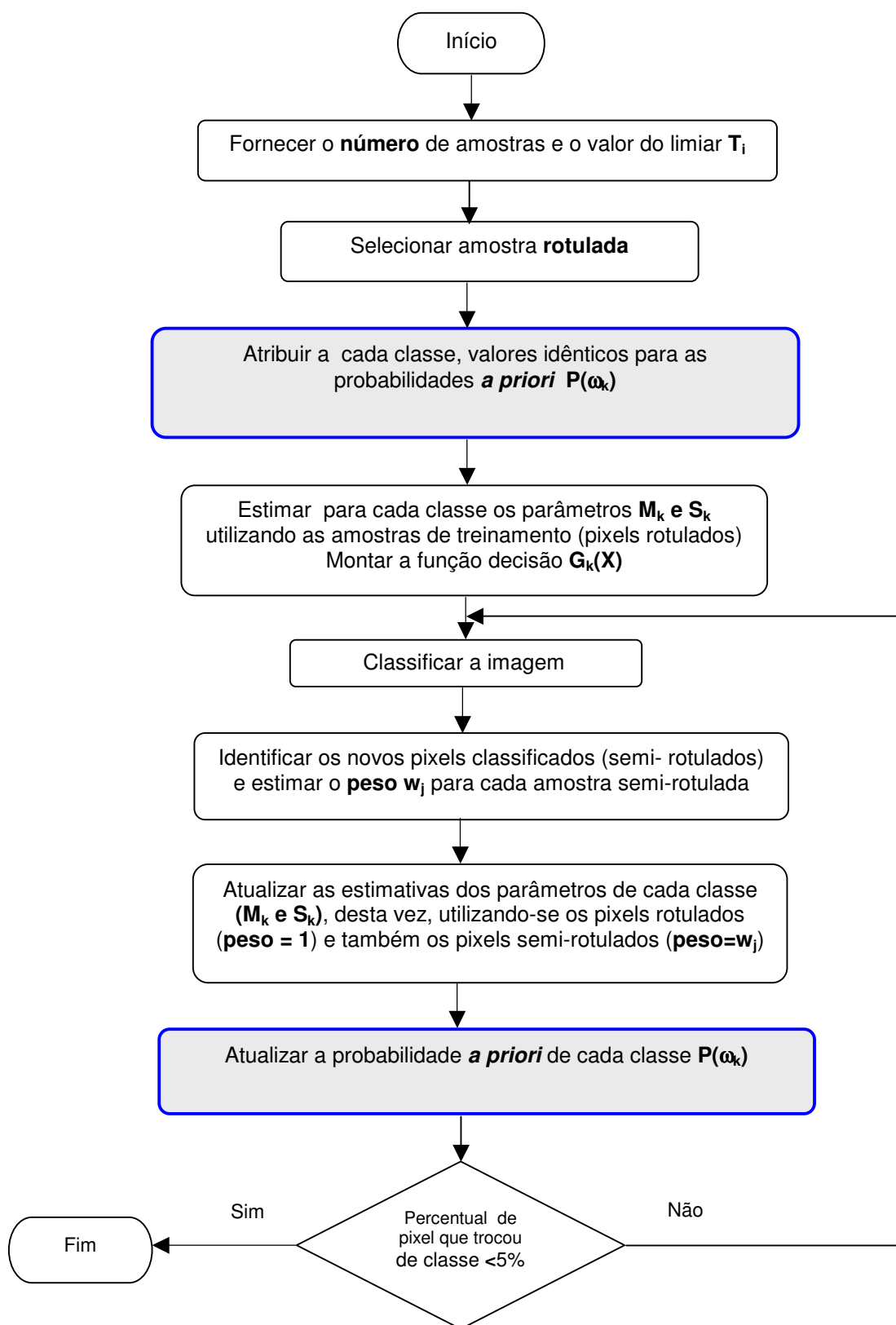


FIGURA 17 - Fluxograma de funcionamento da Ferramenta CDH.

Neste contexto, implementa-se um classificador adaptativo, cujas características devem produzir resultados mais acurados, devido ao aumento no número de amostras de treinamento, resultante da adição de amostras semi-rotuladas ao conjunto inicial de amostras rotuladas e da introdução de valores mais realistas na estimação das probabilidades *a priori*  $P(\omega)$ . Este é um processo iterativo, o qual inicia-se com uma primeira estimativa baseada apenas nas amostras de treinamento e valores idênticos para as probabilidades *a priori* associadas a cada uma das classes, sendo os passos seguintes repetidos a cada iteração, utilizando-se também as amostras semi-rotuladas e valores estimados para as probabilidades *a priori*.

Neste estudo propõe-se a seguinte expressão para estimar os pesos ( $W$ ) para as amostras semi-rotuladas:

$$\begin{array}{ccccccc} \omega_1 & \omega_2 & \omega_3 & \dots & \omega_i & \dots & \omega_k \\ | & | & | & & | & & | \\ G_1(X) & G_2(X) & G_3(X) & & \mathbf{G}_i(\mathbf{X}) & & G_k(X) \end{array}$$

$$W_i = \frac{1/\mathbf{G}_i(\mathbf{X})}{1/G_1(X) + 1/G_2(X) + 1/G_3(X) + \dots + 1/\mathbf{G}_i(\mathbf{X}) + \dots + 1/G_k(X)} \quad (3.2)$$

Sendo  $G_i(\mathbf{X})$  o valor que a função de decisão Máxima Verossimilhança Gaussiana (3.1), associada à classe  $\omega_k$  assume para o pixel  $\mathbf{X}$  a ser classificado, e  $k$  o número total de classes em consideração. Note que de acordo com (3.1), as funções de decisão  $G_i(\mathbf{X})$  são todas negativas, sendo que a função vencedora em valor absoluto é a menor de todas.

Os vetores de médias e as matrizes de covariância para cada classe são estimados por:



$$M_k = \frac{1}{N_k} \sum_{i=1}^{N_k} W_i X_i \quad (3.3)$$

$$S_k = \frac{1}{\left(\sum_{i=1}^{N_k} W_i\right)^{-1}} \sum_{i=1}^{N_k} W_i (X_i - M_k) (X_i - M_k)^T \quad (3.4)$$

para  $i = 1, 2, \dots, k$ , sendo  $K$  o número total de classes.

onde  $M_k$  e  $S_k$  são os vetores média e a matriz covariância na classe  $\omega_k$ ,  $W_i$  o valor ponderado para a amostra  $X_i$ , associado à classe  $\omega_k$  e  $N_k$  é o número de amostras de treinamento da classe

Estimando e atualizando o valor para as probabilidades *a priori*:  $P(\omega_1), \dots, P(\omega_2), \dots, P(\omega_k)$  – fração da área classificada para cada uma das classes, ponderada pelo grau de pertinência de cada pixel individual à classe para a qual ele foi atribuído na presente iteração:

$$\begin{array}{ccccccccc} \omega_1 & & \omega_2 & & \omega_3 & \dots & \omega & \dots & \omega_k \\ | & & | & & | & & | & & | \\ P(\omega_1) & & P(\omega_2) & & P(\omega_3) & & P(\omega) & & P(\omega_k) \end{array}$$

$$P(\omega_1) = \frac{\text{Nº total de pixels classificados em } \omega_1}{\text{Nº total de pixels na imagem temática}}$$

⋮

$$P(\omega_k) = \frac{\text{N}^\circ \text{ total de pixels classificados em } \omega_k}{\text{N}^\circ \text{ total de pixels na imagem temática}}$$

Resumindo-se, o valor estimado da probabilidade *a priori* ( $P$ ), associado à classe  $\omega_k$ , pode ser expresso da seguinte forma:

$$P(\omega_j) = \frac{N_j}{N_1 + N_2 + N_3 + \dots + N_j + \dots + N_k} \quad (3.5)$$

onde  $N_j$  é o valor correspondente ao número de *pixels* ( $\mathbf{X}$ ), associados à classe  $\omega_j$  e  $k$  o número total de classes em consideração.

A classificação da imagem será executada da por meio da função de decisão do classificador *Máxima Verossimilhança Gaussiana* (3.1). Utilizar-se-á o vetor média e a matriz covariância ponderados (3.3) e (3.4) Lemos(2003), incrementando-se a probabilidade *a priori* (3.5), fruto do presente estudo.

## CAPÍTULO IV

### TESTES E EXPERIMENTOS

#### 4.1 INTRODUÇÃO

Neste capítulo são apresentados e discutidos os experimentos realizados com a finalidade de ilustrar e validar a metodologia proposta no Capítulo 3. Nesta metodologia é estudado um processo iterativo de classificação no qual a cada iteração são introduzidas amostras semi-rotuladas para fins de atualização dos valores estimados para os parâmetros do classificador e são também atualizadas as estimativas para as probabilidades *a priori* das classes presentes na cena.

#### 4.2 MATERIAIS

Para a realização deste projeto, foram utilizados os seguintes materiais:

##### 4.2.1 Equipamentos

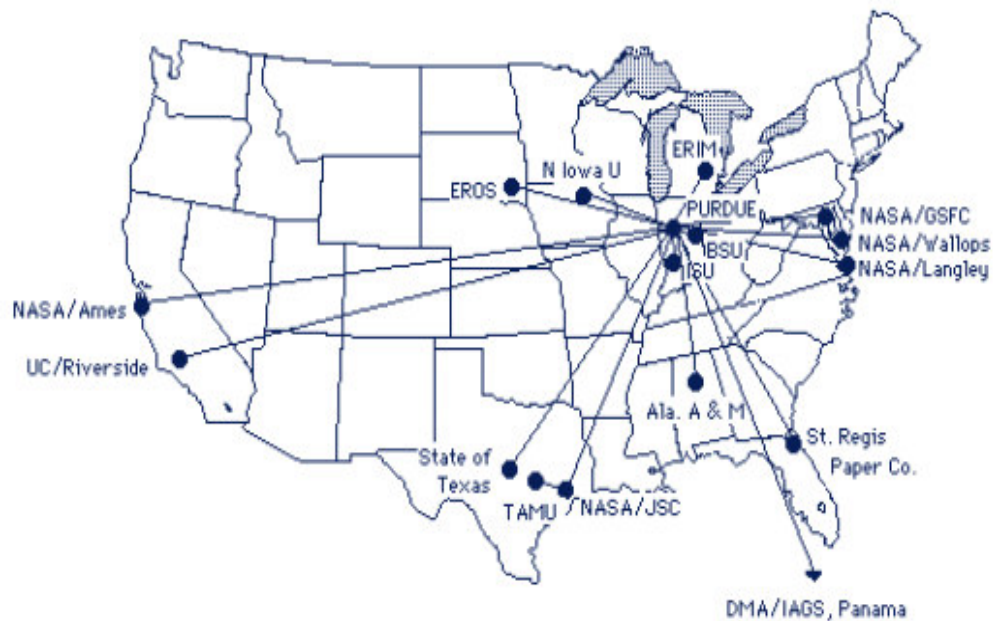
- Microcomputador Pentium III 2 Ghz, RAM 512Mb, HD 80Gb, gravador de DVD e CD;
- Impressora Jato de Tinta HPdeskjet 3820.

#### 4.2.2 Aplicativos

- Microsoft Windows XP;
- Microsoft Office 2000;
- Pacote Estatístico SPSS;
- MATLAB 6.5;
- MultiSpecW32.

#### 4.2.3 Dados

Para fins de lustração e validação da metodologia proposta neste estudo, são apresentados e discutidos neste capítulo uma série de experimentos empregando dados de imagens em alta dimensionalidade. As imagens utilizadas aqui, foram coletadas pelo sensor AVIRIS, compreendendo 220 bandas espectrais abrangendo uma região do espectro eletromagnético entre 0.4  $\mu\text{m}$  e 2.5  $\mu\text{m}$  de comprimento de onda, cobrindo uma área de testes localizada no Estado de Indiana e com dados de verdade terrestre, obtidos por técnicos da Universidade Purdue (FIG. 18). Esta área teste inclui campos agrícolas com características espectrais muito semelhantes entre si, e que por esta razão são difíceis de serem separados com dados tradicionais em média ou baixa dimensionalidade, como por exemplo, dados Landsat-TM ou SPOT, o que a torna especialmente adequada para estudos que empregam dados em alta dimensionalidade.



**FIGURA 18 - Localização da Universidade Purdue, Estado de *Indiana* – EUA.**

Fonte: Landgrebe (2002).

O software necessário para fins de implementação específica da metodologia proposta, foi desenvolvido no ambiente MATLAB. Para fins de implementação de processos mais correntes de tratamento e visualização dos dados, foi empregado o software MultiSpec, desenvolvido pela Universidade Purdue.

### 4.3 CENA DE ESTUDO

Os dados hiperespectrais usados neste experimento são um segmento (154x120), contendo, portanto 18.480 *pixels*, FIG. 18, obtido a partir de outro 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. 18). Os dados imagem cobrindo a referida cena possuem 220 bandas espectrais (das 224 bandas originais do sensor AVIRIS), cobrindo intervalos de comprimentos de onda distanciados

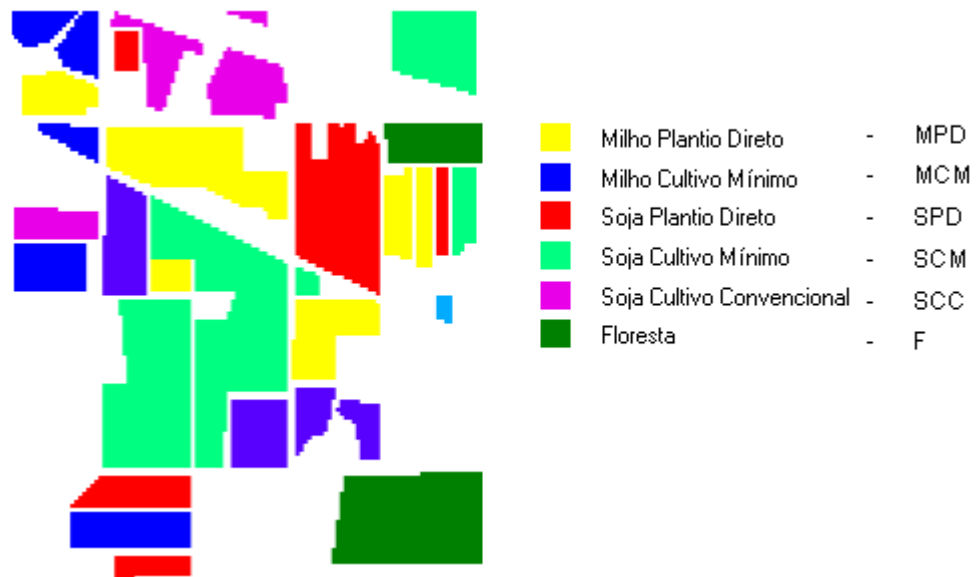
uniformemente de modo a cobrir todo o intervalo de comprimento de onda do espectro eletromagnético entre 0.4  $\mu\text{m}$  e 2.5  $\mu\text{m}$ .

A imagem FIG. 19, cobre uma área agrícola com culturas de milho e soja, além de outras. Foram ainda utilizadas técnicas de plantio distintas (plantio convencional e plantio direto). A imagem foi obtida no início da época de crescimento, ocorrendo por isso grande quantidade de solo exposto. Esta situação resulta em classes espectralmente muito semelhantes, constituindo-se num grande desafio para o classificador.

As FIG. 19 e 20 ilustram, respectivamente, uma composição colorida (RGB) produzida com três bandas espectrais da imagem analisada e o mapa temático com a verdade terrestre de ambos os segmentos.



**FIGURA 19 - AVIRIS: Composição colorida RGB com três bandas espectrais, do segmento utilizado nos experimentos, (NW Indiana's Indian Pine) - 220 bandas espectrais.**



**FIGURA 20 - Imagem temática da verdade terrestre (*ground truth*) da FIG. 19, relativa às classes em estudo com as respectivas SIGLAS.**

#### 4.3.1 Seleção de Bandas Espectrais

Para a seleção das amostras das classes, como sendo a etapa inicial, deve-se preparar a imagem, isto é, fazer uma análise preliminar, a fim de que se trabalhe apenas com os dados que realmente possam oferecer alguma informação útil que caracterize as classes. Desta forma foram eliminadas, através de uma análise visual, bandas afetadas pelo efeito da absorção ocasionada por alguns constituintes da atmosfera, tais como vapor d'água, CO<sub>2</sub> e O<sub>3</sub>, que resultam em imagens com um nível muito alto de ruído. Foram excluídas as seguintes bandas espectrais, conforme está indicado (TAB. 5).

**TABELA 5 - Seleção de bandas ruidosas que foram excluídas (34 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	136	138	139	140
141	142	143	144	145	146	147	148	149	150	150	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	221	222	223	224

A dimensionalidade final a ser usada é, portanto, igual a 190, resultando em vetores com 190 valores de contador digital para cada *pixel* da imagem.

A partir de então, o critério para a seleção das bandas a serem utilizadas em cada um dos experimentos desenvolvidos neste estudo foi determinado pela razão entre as 190 bandas remanescentes (*nr\_total\_bandas*) e o número de bandas desejado para o experimento em questão (*bandas\_experimento*), dado pela seguinte expressão:

$$\text{intervalo\_bandas} = \text{nr\_total\_bandas} / \text{bandas\_experimento}$$

onde, o valor inteiro “*intervalo\_bandas*”, corresponde ao fator determinante do intervalo a ser adotado para selecionar de maneira uniforme o conjunto de bandas desejado em cada experimento, cobrindo assim todo o espectro eletromagnético, conforme descrito na seção 4.3.

Por exemplo, aplicando-se esta expressão em MATLAB para um experimento com a dimensionalidade 60, teremos:

$$\text{intervalo\_bandas} = \text{div}(\text{nr\_total\_bandas}, \text{bandas\_experimento});$$



Substituindo,  $\text{intervalo\_bandas} = \text{div}(190,60)$ ; tomando-se a parte inteira resultante, neste caso 3, obtemos então o total de bandas para o experimento ( $\text{nr\_bandas\_experimento}$ ), da seguinte forma:  $\text{nr\_bandas\_experimento} = \text{div}(190,3)$ ; resultando 63.33, onde serão descartadas as bandas excedentes, neste caso a partir de 60.

#### 4.3.2 Seleção de Classes

No recorte da imagem original, convenientemente selecionado para a realização dos experimentos desenvolvidos neste estudo, visando à validação da metodologia aqui proposta, encontram-se seis classes temáticas distintas: milho plantio direto (MPD), milho em cultivo mínimo (MCM), soja plantio direto (SPD), soja em cultivo mínimo (SCM), soja em cultivo convencional (SCC) e floresta(F).



**FIGURA 21** - Vista de áreas contendo as culturas de milho (a) e soja (b), mostrando a baixa porcentagem que representa frente à área emaciada no momento da coleta dos dados.

Fonte: Adaptado de Landgrebe (2003).

Com exceção da classe floresta, as demais cinco classes são espectralmente muito semelhantes, constituindo-se em um desafio ao classificador. A utilização no processo de classificação, de dados tradicionais em média ou baixa dimensionalidade (dados Landsat-TM ou SPOT, por exemplo) resulta em imagens temáticas com baixa acurácia (LANDGREBE, 2003).

A imagem foi tomada no início da época de crescimento das culturas de soja e milho, onde apenas aproximadamente 5% da área está efetivamente coberta pela vegetação, sendo os restantes 95% composto por solo exposto e resíduo de colheitas anteriores (FIG. 21). Além disto, variadas práticas de cultivo foram aplicadas na região, onde a superfície apresenta diferentes quantidades de resíduo do ano precedente.

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

Utilizando os dados disponíveis de verdade terrestre, foram construídos para cada uma das seis classes de informação, dois conjuntos de *pixels*: o primeiro para fins de treinamento e o segundo para fins de teste. Como as áreas de plantio ocupadas por estas classes de informação não são iguais, o tamanho das amostras é variável de uma classe para outra. A TAB. 6 ilustra o número total de *pixels* disponíveis.

**TABELA 6 - Total de *pixels* extraídos da imagem AVIRIS**

<b>Nome da classe</b>	<b>Identificador numérico</b>	<b>Total de <i>pixels</i> da amostra</b>
Milho Plantio Direto (MPD)	2	1434
Milho Cultivo Mínimo (MCM)	3	834
Soja Plantio Direto (SPD)	10	1180
Soja Cultivo Mínimo (SCM)	11	2468
Soja Cultivo Convencional (SCC)	12	662
Floresta (F)	14	1160

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

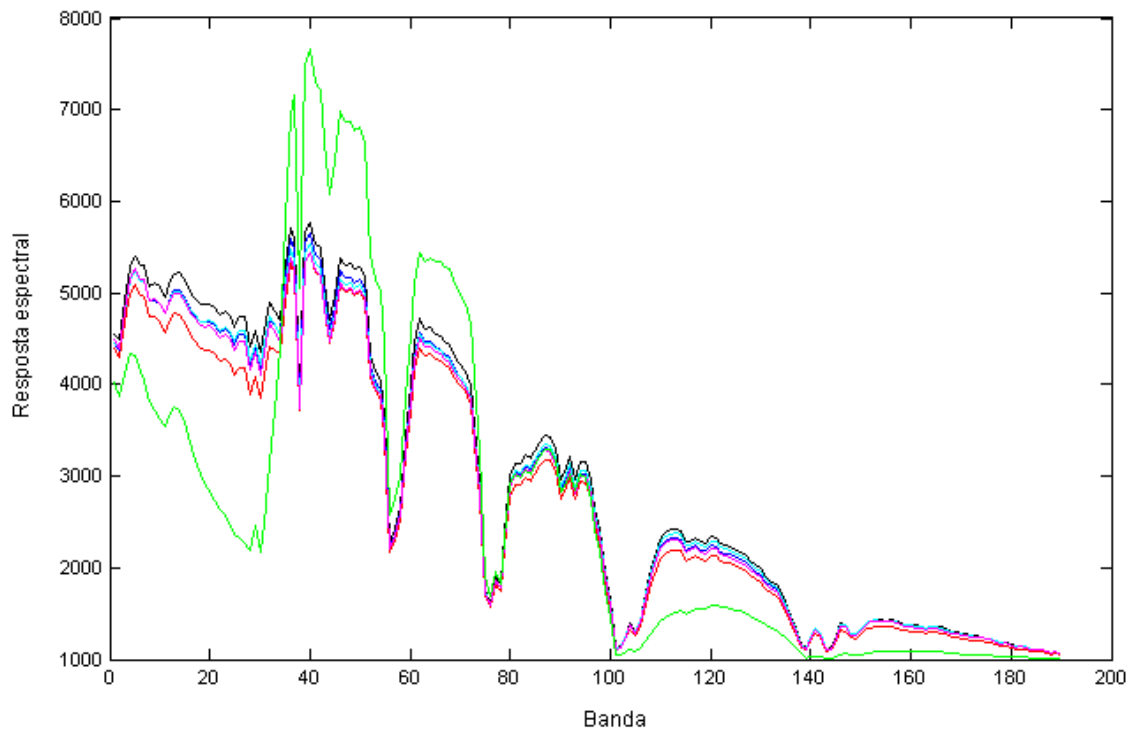
O motivo da inserção da classe floresta nos experimentos, deve-se ao fato de que o padrão floresta é uma classe espectralmente bem distinta, se comparada com as outras cinco classes representando as culturas agrícolas. Este procedimento visa realizar uma análise completa, verificando inclusive, o comportamento da metodologia proposta em situações de classificação mais simples.

Uma vez definidas as amostras de cada classe, é possível visualizar o comportamento espectral das mesmas, através das curvas de resposta espectral das classes. Para esta finalidade, utilizou-se o software MultispecW32 elaborado pela Universidade Purdue, e algumas funções desenvolvidas no ambiente MATLAB.

O GRAF.1 ilustra o comportamento espectral das classes, onde se verificam dois aspectos principais: a diferença espectral da classe *floresta* com relação às demais classes, e a alta semelhança entre as outras cinco classes (variações das culturas de *milho* e *soja*).

**GRÁFICO 1**

**Curvas de Resposta Espectral Média para as classes Milho Plantio Direto (preto), Milho Cultivo Mínimo (azul), Soja Plantio Direto (vermelho), Soja Cultivo Mínimo (ciano), Soja Cultivo Convencional (magenta) e Floresta (verde).**



#### 4.4 FERRAMENTA CDH – CLASSIFICADOR DE DADOS HIPERESPECTRAIS

Para fins de implementação da metodologia proposta neste estudo, foi desenvolvida uma ferramenta denominada de Classificador de Dados Hiperespectrais (CDH), cujo código fonte encontra-se no Apêndice A. O CDH implementa, em um processo iterativo, a inclusão de amostras semi-rotuladas para fins de estimação dos parâmetros do classificador e também o valor estimado para as probabilidades *a priori* das classes envolvidas na função de decisão MVG. Desenvolvido em ambiente MATLAB 6.5, o CDH analisa todos os passos executados pelo método e facilita o manuseio dos dados hiperespectrais, bem como a seleção de bandas espectrais, monitora o processamento das informações e a visualização dos resultados em forma de gráfico, matriz de erro e de imagens temáticas de verdade terrestre.

##### 4.4.1 Amostras de Treinamento

Do conjunto das amostras disponíveis para cada classe individual foram extraídos dois subconjuntos: um subconjunto com amostras de treinamento e um segundo com amostras de teste. Com a finalidade de evitar tendenciosidades, devido às variações que ocorrem em uma mesma classe ao longo da área de estudo, na formação dos dois subconjuntos as amostras individuais (*pixels*) foram extraídas alternadamente do conjunto com a totalidade das amostras.

No contexto desta dissertação, as amostras de treinamento assim selecionadas são denominadas de “amostras rotuladas”, isto é, são amostras obtidas a partir dos dados disponíveis de verdade terrestre. Não existe, portanto, dúvidas quanto à pertinência da amostra (*pixel*) à classe na qual está rotulada. No processo iterativo, conforme a metodologia proposta, a cada etapa o conjunto original com as amostras de treinamento rotuladas, é incrementado por um número adicional de amostras extraídas pelo próprio processo e denominadas de amostras semi-rotuladas. O número de

amostras semi-rotuladas agregadas ao conjunto das amostras de treinamento a cada etapa do processo, é determinado através da utilização de um valor predefinido.

Com a finalidade de testar a eficácia do algoritmo proposto neste estudo, uma série de testes com os dados disponíveis foi realizada, tomando-se a dimensionalidade dos dados como variável independente e a resultante acurácia na classificação como variável dependente. O valor da dimensionalidade dos dados variou desde 10 até 190 bandas, selecionadas ao longo das 190 disponíveis, cobrindo todo o intervalo de comprimento de onda do espectro eletromagnético. Portanto, ao utilizar-se esse método para a seleção de bandas espectrais, a cada experimento têm-se um conjunto próprio destas, que não necessariamente contém aquelas utilizadas nos outros experimentos. Desta maneira é possível analisar, para diferentes tamanhos de amostras de treinamento, o comportamento da acurácia na imagem classificada em função da dimensionalidade dos dados.

Com vistas à apresentação dos resultados, foi feita uma seleção de experimentos considerados mais elucidativos, dentre os vários executados, durante a fase de testes do classificador adaptativo, presentes na seção 4.4.2 deste trabalho.

Em relação aos experimentos, o resultado da classificação inicial (obtido utilizando-se somente as amostras rotuladas) e da classificação final (obtido com a utilização das amostras rotuladas conjuntamente com as amostras semi-rotuladas, e com a inclusão dos valores da estimativa das probabilidades *a priori*, ao final do processo iterativo), é ilustrado através das imagens temáticas produzidas, dos gráficos da acurácia obtida, e das matrizes de erro onde pode ser observado o número total de *pixels* (rotulados e semi-rotulados) que foram classificados em cada uma das classes selecionadas para o experimento, na classificação final (última iteração), bem como a acurácia<sup>1</sup> que se obteve para cada uma das classes após a utilização do classificador adaptativo.

---

<sup>1</sup>A razão das amostras (rotuladas e semi-rotuladas) classificadas corretamente em cada classe, pelo total de amostras classificadas na respectiva classe.

No classificador utilizado nestes experimentos, foi incluído um limiar, baseado na distribuição Qui-quadrado ( $\chi^2$ ) a um nível de significância de 95%. A introdução de um limiar tem por objetivo identificar *pixels* com baixa probabilidade de pertencer a qualquer uma das classes em consideração e que, portanto, devem permanecer como não rotulados.

#### 4.4.2 Realização de Experimentos

Estes experimentos visam quantificar numericamente os resultados de desempenho das metodologias propostas, especialmente no que tange o comportamento da acurácia no processo iterativo de classificação de imagens digitais com a introdução de amostras semi-rotuladas na estimação dos parâmetros do classificador, e também dos valores estimados para as probabilidades *a priori* nas funções de decisão utilizadas no classificador. A eficiência computacional do método proposto neste estudo também é avaliada.

Inicialmente, procedeu-se uma classificação do tipo supervisionada, utilizando-se o classificador Máxima Verossimilhança Gaussiana (na suposição de que os dados apresentam uma distribuição gaussiana multivariada) com o número de bandas espectrais variando de 10 até 190, a intervalos de 10 bandas. Nestes experimentos foram utilizados conjuntos de amostras de treinamento de diferentes tamanhos.

Com a finalidade de tornar os resultados obtidos para as várias classes comparáveis entre si, foram utilizados subconjuntos de treinamento e de teste de mesmo tamanho para todas as classes em estudo. Estes subconjuntos foram extraídos do conjunto de amostras disponíveis (TAB. 7).

**TABELA 7 - Número de amostras das classes selecionadas**

<b>Nome da classe</b>	<b>Amostras de treinamento</b>	<b>Amostras de teste</b>
Milho Plantio Direto (MPD)	717	717
Milho Cultivo Mínimo (MCM)	417	417
Soja Plantio Direto (SPD)	590	590
Soja Cultivo Mínimo (SCM)	1234	1234
Soja Cultivo Convencional (SCC)	331	331
Floresta (F)	580	580

Na fase de experimentos, utilizaram-se quantidades de amostras de treinamento de mesmo tamanho para todas as classes obtidas conforme descrito na seção 4.3.2, e com os seguintes tamanhos: 200, 300, 400 e 500 *pixels*. Note-se que o valor mínimo admissível para as amostras de treinamento é igual à dimensionalidade dos dados mais um. Um valor inferior resultará que a matriz de covariância será singular e, portanto não utilizável no classificador MAXVER, (LANDGREBE, 2003). O tamanho das amostras de treinamento, foi escolhido deliberadamente pequeno com relação à dimensionalidade dos dados para desta forma, melhor evidenciar os problemas que ocorrem em situações reais. A quantidade de amostras disponíveis para as classes “MCM” e “SCC”, está aquém daquela das demais classes, diante dos experimentos em que o tamanho das amostras de treinamento empregado foi de 400 e 500. Contudo, ainda que esse procedimento não seja estatisticamente recomendado, essa prática não configurou tendências que revertessem em prejuízos para as referidas classes, durante o processo de classificação, como se vê nos resultados apresentados neste capítulo.

A cada iteração o conjunto das amostras de treinamento disponíveis (amostras rotuladas) é incrementado com um certo número de amostras semi-rotuladas. Nestes experimentos, utilizou-se como incremento, uma quantidade equivalente a 50 amostras semi-rotuladas, aquelas que apresentavam o maior peso.

O processo iterativo é encerrado quando o número de *pixels* que trocam de classe entre duas iterações sucessivas é inferior a 5% do total de *pixels* na imagem temática.

Vários experimentos foram feitos implementando a metodologia proposta, e empregando amostras de treinamento com diferentes tamanhos, tanto com valores idênticos, como com a inclusão dos valores das estimativas das probabilidades *a priori* das classes.

Destacam-se os valores estimados para a acurácia na classificação inicial e na final, com os valores estimados para as probabilidades *a priori* incluídos na função decisão. Isto se torna bastante evidente, através do exposto nos GRAF. 2 a 11, expostos a seguir.

As imagens temáticas FIG. 22 a 27, ilustram os resultados finais obtidos em cada uma das seguintes opções de classificação:

- (a) metodologia tradicional, que para fins de estimação dos parâmetros do classificador faz uso apenas das amostras de treinamento rotuladas e assume idêntico valor para as probabilidades *a priori* das classes;
- (b) metodologia que incrementa as amostras de treinamento rotuladas com as chamadas amostras semi-rotuladas;
- (c) metodologia proposta neste estudo, que inclui valores estimados para as probabilidades *a priori* de cada uma das classes.

Para uma discussão mais detalhada dos resultados obtidos com a aplicação da metodologia proposta, foram selecionados subconjuntos dos experimentos realizados, empregando dados com dimensionalidade igual a 40, 60 e 80.



Nos experimentos ilustrados nas FIG. 22 a 27, foram utilizadas amostras de treinamento com 300 e 500 *pixels* rotulados. Nas TAB. 8 a 25, são apresentadas as correspondentes matrizes de erro.

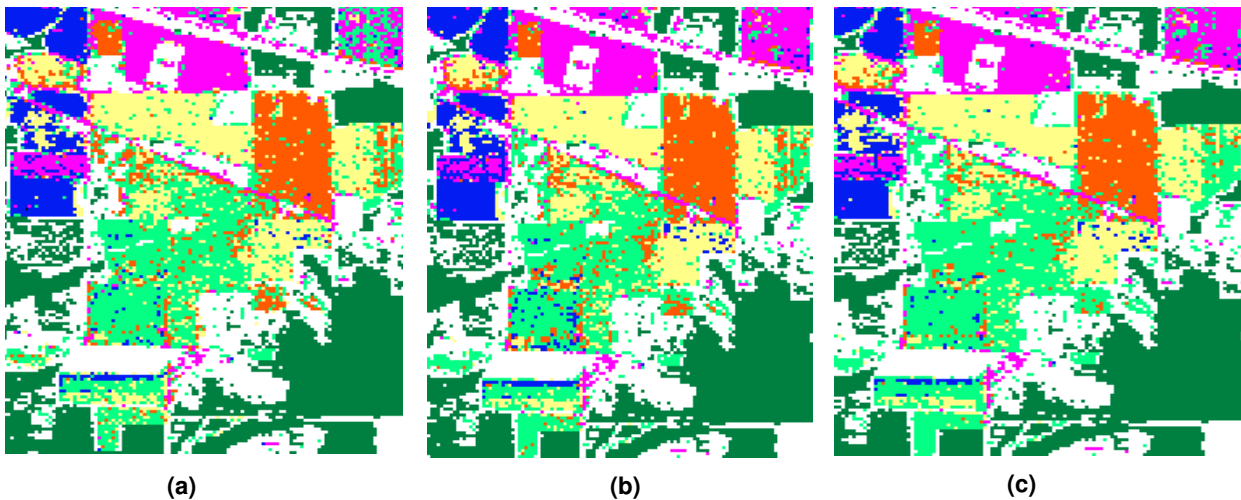
Para a correta interpretação do conteúdo das Matrizes de Erro, contidas nestas tabelas, conforme em Richards (1993), a terminologia empregada é definida a seguir:

- a) Acurácia do Produtor: – é a razão das amostras classificadas corretamente em cada classe, pelo total de amostras de teste da respectiva classe;
- b) Acurácia do Usuário: – é a razão das amostras classificadas corretamente em cada classe, pelo total de amostras classificadas na respectiva classe;
- c) Acurácia Geral: – traduz a razão do total de amostras classificadas corretamente em cada classe, ou seja, a soma dos valores da diagonal principal, pelo total das amostras de teste.

As imagens temáticas que serão apresentadas a seguir têm como legenda a FIG. 20, apresentada na seção 4.3.

#### 4.4.2.1 Experimento 1

Para a execução deste experimento foram utilizados dados com dimensionalidade igual a 40, tamanho das amostras de treinamento rotuladas 300 com incremento de 50 amostras semi-rotuladas. As imagens temáticas mostradas na FIG. 22 ilustram os resultados finais obtidos em cada uma das opções de classificação. As TAB. 8 a 10, apresentam os resultados relativos a cada uma destas classificações.



**FIGURA 22-** Imagens temáticas para dados com dimensionalidade igual a 40, tamanho das amostras de treinamento rotuladas igual a 300: (a) amostras de treinamento rotuladas e idêntico valor para as probabilidades *a priori* das classes; (b) igual a (a), mas com a adição de 50 amostras semi-rotuladas; e (c) igual a (b), porém com a introdução dos valores estimados para as probabilidades *a priori* de cada classe.

Na FIG.22, podem ser visualizados os erros de classificação, principalmente em relação às classes SCM, SCC e MPD, e as regiões da imagem atinentes às classes SPD e MCM, em que o classificador não logrou o êxito esperado.

Nos experimentos 2 e 3, FIG. 23 e 24, nos quais foi aumentada a dimensionalidade dos dados para 60 e 80 e mantido o tamanho das amostras de treinamento em 300, apesar de ter havido evolução na acurácia do usuário em praticamente todas as classes, persistiram os erros e omissões de classificação, produzidos pelos três métodos no experimento 1.

As TAB. 8 a 10, apresentam os valores da distribuição de classes para a área selecionada nas classificações finais pelos três métodos. Nota-se que em apenas algumas regiões correspondentes às classes SCM e SCC, como se vê na imagem temática FIG. 22, não houve a evolução esperada dos valores de acurácia do usuário, quando os métodos são comparados.

TABELA 8

Experimento com método tradicional, faz uso apenas das amostras de treinamento rotuladas e assume idêntico valor para as probabilidades *a priori* das classes, com dimensionalidade 40 e tamanho das amostras de treinamento 300 pixels.

Classes da Imagem de Referência	Acurácia do Produtor	Amostras de Teste	Limiar (Ti)	Classes da Imagem Classificada					
				MPD	MCM	SPD	SCM	SCC	F
Milho Plantio Direto-MPD	80.4	715	2	575	18	33	77	12	0
Milho Cultivo Mínimo-MCM	67.5	416	1	39	281	4	77	15	0
Soja Plantio Direto-SPD	80.1	482	108	41	0	386	53	2	0
Soja Cultivo Mínimo-SCM	61.6	1226	8	159	43	137	755	132	0
Soja Cultivo Convencional-SCC	86.4	331	0	4	25	3	13	286	0
Floresta-F	100.0	580	0	0	0	0	0	0	580
Totais de Amostras		3750	119	818	367	563	975	447	580
Acurácia do Usuário(%)				70.3	76.6	68.6	77.4	64.0	100
Acurácia Geral(%)	76.3								

TABELA 9

Experimento com método que incrementa as amostras de treinamento rotuladas com as chamadas amostras semi-rotuladas, e assume idêntico valor para as probabilidades *a priori* das classes, com dimensionalidade 40 e tamanho das amostras de treinamento 300 pixels.

Classes da Imagem de Referência	Acurácia do Produtor	Amostras de Teste	Limiar (Ti)	Classes da Imagem Classificada					
				MPD	MCM	SPD	SCM	SCC	F
Milho Plantio Direto-MPD	81.1	715	2	580	12	35	74	14	0
Milho Cultivo Mínimo-MCM	68.3	416	1	38	284	3	80	11	0
Soja Plantio Direto-SPD	80.7	482	108	36	0	389	55	2	0
Soja Cultivo Mínimo-SCM	62.8	1226	8	154	31	126	770	145	0
Soja Cultivo Convencional-SCC	87.0	331	0	4	24	3	12	288	0
Floresta-F	100.0	580	0	0	0	0	0	0	580
Totais de Amostras		3750	119	812	351	556	991	460	580
Acurácia do Usuário(%)				71.4	80.9	70.0	77.7	62.6	100
Acurácia Geral(%)	77.1								

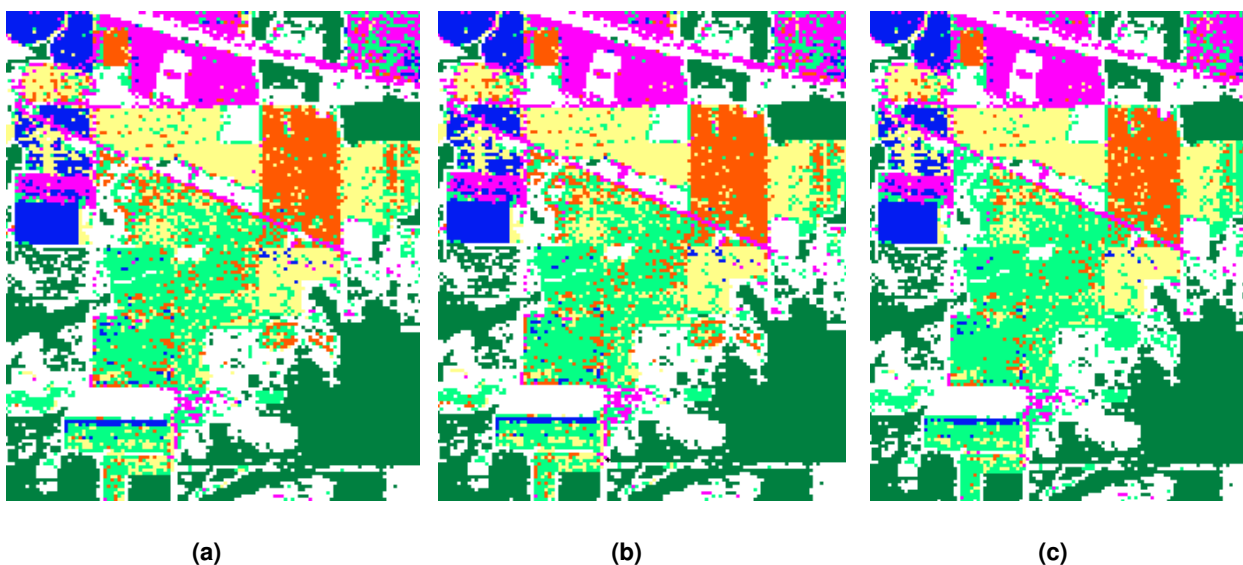
TABELA 10

Experimento com o método proposto neste estudo, que inclui valores estimados para as probabilidades *a priori* de cada uma das classes, com dimensionalidade 40 e tamanho das amostras de treinamento 300 pixels.

Classes da Imagem de Referência	Acurácia do Produtor	Amostras de Teste	Limiar (Ti)	Classes da Imagem Classificada					
				MPD	MCM	SPD	SCM	SCC	F
Milho Plantio Direto-MPD	81.8	715	2	585	10	29	79	12	0
Milho Cultivo Mínimo-MCM	66.1	416	1	38	275	4	87	12	0
Soja Plantio Direto-SPD	78.4	482	108	38	0	378	64	2	0
Soja Cultivo Mínimo-SCM	68.1	1226	8	150	20	91	835	130	0
Soja Cultivo Convencional-SCC	86.4	331	0	4	21	3	17	286	0
Floresta-F	100.0	580	0	0	0	0	0	0	580
Totais de Amostras		3750	119	815	326	505	1082	442	580
Acurácia do Usuário(%)				71.8	84.4	74.9	77.2	64.7	100
Acurácia Geral(%)	78.4								

#### 4.4.2.2 Experimento 2

Para a execução deste experimento foram utilizados dados com dimensionalidade igual a 60, tamanho das amostras de treinamento rotuladas 300, com incremento de 50 amostras semi-rotuladas. As imagens temáticas mostradas na FIG. 23 ilustram os resultados finais obtidos em cada uma das opções de classificação. As TAB. 11 a 13 apresentam os resultados relativos a cada uma destas classificações.

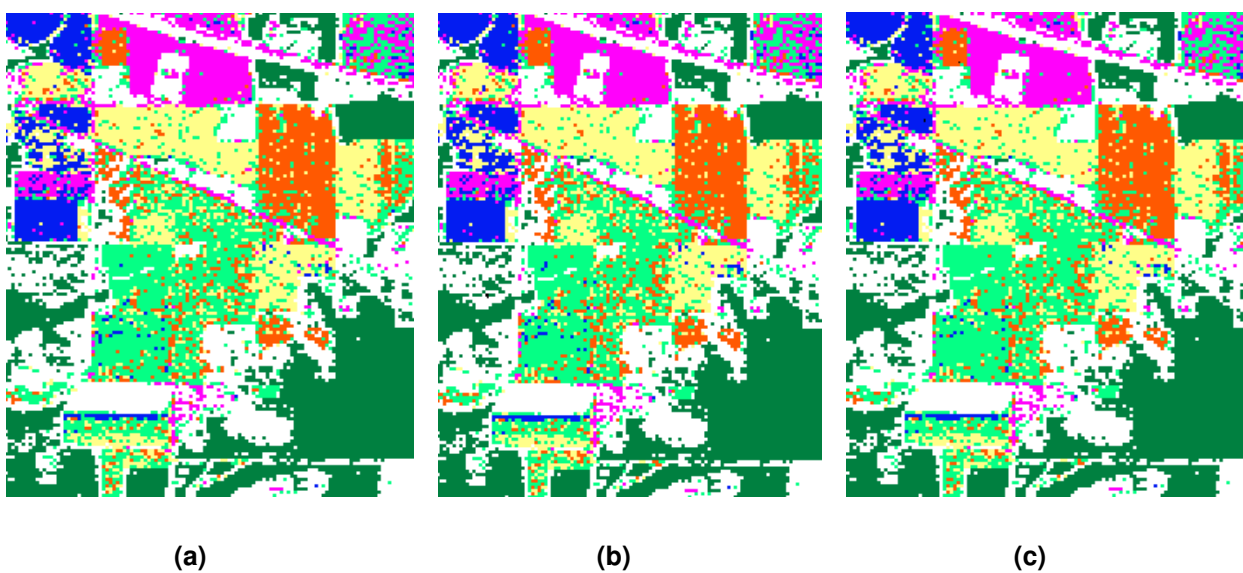


**FIGURA 23-** Imagens temáticas para dados com dimensionalidade igual a 60, tamanho das amostras de treinamento rotuladas igual a 300: (a) amostras de treinamento rotuladas e idêntico valor para as probabilidades *a priori* das classes; (b) igual a (a), mas com a adição de 50 amostras semi-rotuladas; e (c) igual a (b), porém com a introdução dos valores estimados para as probabilidades *a priori* cada classe.



#### 4.4.2.3 Experimento 3

Para a execução deste experimento foram utilizados dados com dimensionalidade igual a 80, tamanho das amostras de treinamento rotuladas 300, com incremento de 50 amostras semi-rotuladas. As imagens temáticas mostradas na FIG. 24 ilustram os resultados finais obtidos em cada uma das opções de classificação. As TAB. 14 a 16 apresentam os resultados relativos a cada uma destas classificações.

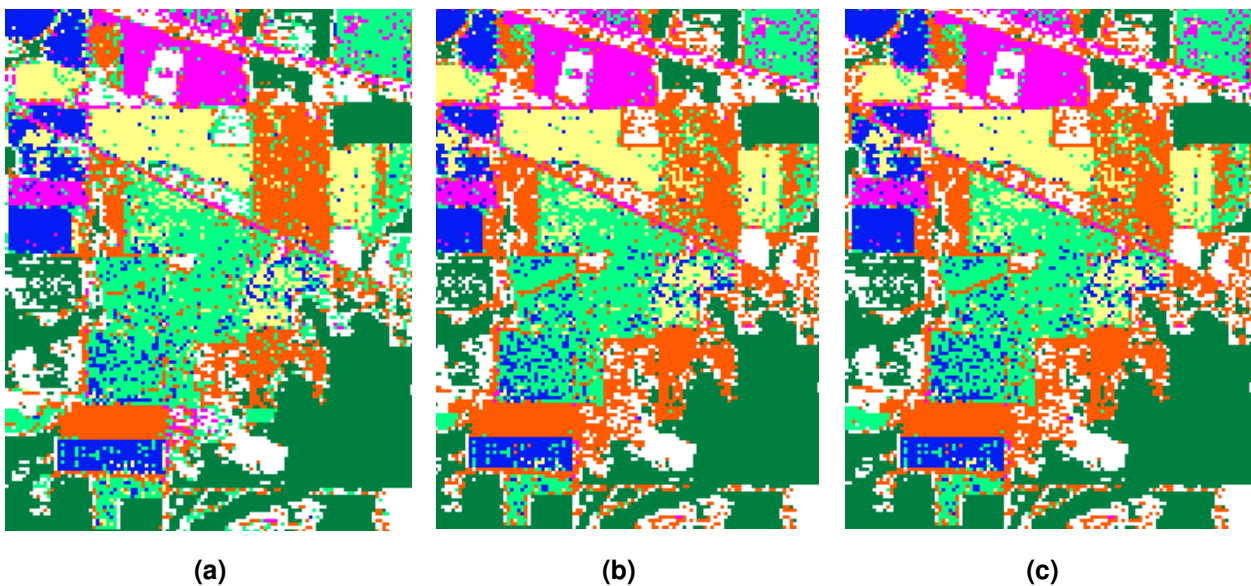


**FIGURA 24-** Imagens temáticas para dados com dimensionalidade igual a 80, tamanho das amostras de treinamento rotuladas igual a 300: (a) amostras de treinamento rotuladas e idêntico valor para as probabilidades *a priori* das classes; (b) igual a (a), mas com a adição de 50 amostras semi-rotuladas; e (c) igual a (b), porém com a introdução dos valores estimados para as probabilidades *a priori* de cada classe.



#### 4.4.2.4 Experimento 4

Para a execução deste experimento foram utilizados dados com dimensionalidade igual a 40, tamanho das amostras de treinamento rotuladas 500, com incremento de 50 amostras semi-rotuladas. As imagens temáticas mostradas na FIG. 25 ilustram os resultados finais obtidos em cada uma das opções de classificação. As TAB. 17 a 19 apresentam os resultados relativos a cada uma destas classificações.



**FIGURA 25-**Imagens temáticas para dados com dimensionalidade igual a 40, tamanho das amostras de treinamento rotuladas igual a 500: (a) amostras de treinamento rotuladas e idêntico valor para as probabilidades *a priori* das classes; (b) igual a (a), mas com a adição de 50 amostras semi-rotuladas; e (c) igual a (b), porém com a introdução dos valores estimados para as probabilidades *a priori* de cada classe.

As imagens da FIG. 25 destacam onde houve melhoras significativas, através dos três métodos, demonstrando também, os locais onde foram introduzidos erros de classificação.

É bastante visível a grande melhora alcançada em todo o processo de classificação nestes experimentos (4, 5 e 6), comparando-se as respectivas imagens temáticas com aquelas dos experimentos anteriores (FIG. 22, 23 e 24).





TABELA 18

Experimento com método que incrementa as amostras de treinamento rotuladas com as chamadas amostras semi-rotuladas, e assume idêntico valor para as probabilidades *a priori* das classes, com dimensionalidade 40 e tamanho das amostras de treinamento 500 pixels.

Classes da imagem de Referência	Acurácia do Produtor	Amostras de Teste	Limiar (Ti)	Classes da Imagem Classificada					
				MPD	MCM	SPD	SCM	SCC	F
Milho Plantio Direto-MPD	78.5	716	1	<b>562</b>	63	12	71	8	0
Milho Cultivo Mínimo-MCM	82.5	417	0	14	<b>344</b>	1	31	27	0
Soja Plantio Direto-SPD	74.9	589	1	38	19	<b>441</b>	88	2	1
Soja Cultivo Mínimo-SCM	61.6	1234	0	101	198	62	<b>760</b>	113	0
Soja Cultivo Convencional-SCC	91.5	331	0	3	8	3	14	<b>303</b>	0
Floresta-F	100.0	580	0	0	0	0	0	0	<b>580</b>
Totais de Amostras		3867	2	718	632	519	964	453	581
Acurácia do Usuário(%)				<b>78.3</b>	<b>54.4</b>	<b>85.0</b>	<b>78.8</b>	<b>66.9</b>	<b>99.83</b>
Acurácia Geral(%):	<b>77.3</b>								

TABELA 19

Experimento com o método proposto neste estudo, que inclui valores estimados para as probabilidades *a priori* de cada uma das classes, com dimensionalidade 40 e tamanho das amostras de treinamento 500 pixels.

Classes da Imagem de Referência	Acurácia do Produtor	Amostras de Teste	Limiar (Ti)	Classes da Imagem Classificada					
				MPD	MCM	SPD	SCM	SCC	F
Milho Plantio Direto-MPD	77.4	716	1	<b>554</b>	52	11	92	7	0
Milho Cultivo Mínimo-MCM	81.8	417	0	13	<b>341</b>	1	44	18	0
Soja Plantio Direto-SPD	71.3	589	1	44	16	<b>420</b>	106	2	1
Soja Cultivo Mínimo-SCM	67.9	1234	0	97	153	49	<b>838</b>	97	0
Soja Cultivo Convencional-SCC	90.3	331	0	4	6	3	19	<b>299</b>	0
Floresta-F	100.0	580	0	0	0	0	0	0	<b>580</b>
Totais de Amostras		3867	2	712	568	484	1099	423	581
Acurácia do Usuário(%)				<b>77.8</b>	<b>60.0</b>	<b>86.8</b>	<b>76.3</b>	<b>70.7</b>	<b>99.83</b>
Acurácia Geral(%):	<b>78.4</b>								

#### 4.4.2.5 Experimento 5

Para a execução deste experimento foram utilizados dados com dimensionalidade igual a 60, tamanho das amostras de treinamento rotuladas 500, com



TABELA 21

Experimento com método que incrementa as amostras de treinamento rotuladas com as chamadas amostras semi-rotuladas, e assume idêntico valor para as probabilidades *a priori* das classes, com dimensionalidade 60 e tamanho das amostras de treinamento 500 *pixels*.

Classes da Imagem de Referência	Acurácia do Produtor	Amostras de Teste	Limiar (Ti)	Classes da Imagem Classificada					
				MPD	MCM	SPD	SCM	SCC	F
				Milho Plantio Direto-MPD	76.7	716	1	<b>549</b>	65
Milho Cultivo Mínimo-MCM	86.1	417	0	10	<b>359</b>	3	26	19	0
Soja Plantio Direto-SPD	74.1	590	0	54	13	<b>437</b>	84	2	0
Soja Cultivo Mínimo-SCM	65.1	1234	0	112	174	69	<b>803</b>	76	0
Soja Cultivo Convencional-SCC	88.5	330	1	4	11	6	17	<b>292</b>	0
Floresta-F	100.0	580	0	0	0	0	0	0	<b>580</b>
Totais de Amostras		3867	2	729	622	523	1016	397	580
Acurácia do Usuário(%)				<b>75.3</b>	<b>57.7</b>	<b>83.6</b>	<b>79.0</b>	<b>73.6</b>	<b>100</b>
Acurácia Geral(%):	<b>78.1</b>								

TABELA 22

Experimento com o método proposto neste estudo, que inclui valores estimados para as probabilidades *a priori* de cada uma das classes, com dimensionalidade 60 e tamanho das amostras de treinamento 500 *pixels*.

Classes da Imagem de Referência	Acurácia do Produtor	Amostras de Teste	Limiar (Ti)	Classes da Imagem Classificada					
				MPD	MCM	SPD	SCM	SCC	F
				Milho Plantio Direto-MPD	76.0	716	1	<b>544</b>	60
Milho Cultivo Mínimo-MCM	86.8	417	0	10	<b>362</b>	2	31	12	0
Soja Plantio Direto-SPD	72.9	590	0	54	11	<b>430</b>	93	2	0
Soja Cultivo Mínimo-SCM	69.9	1234	0	104	154	55	<b>862</b>	59	0
Soja Cultivo Convencional-SCC	87.0	330	1	4	12	6	21	<b>287</b>	0
Floresta-F	100.0	580	0	0	0	0	0	0	<b>580</b>
Totais de Amostras		3867	2	716	599	501	1103	368	580
Acurácia do Usuário(%)				<b>76.0</b>	<b>60.4</b>	<b>85.8</b>	<b>78.2</b>	<b>78.0</b>	<b>100</b>
Acurácia Geral(%):	<b>79.3</b>								

#### 4.4.2.6 Experimento 6

Para executar este experimento foram utilizados dados com dimensionalidade igual a 80, tamanho das amostras de treinamento rotuladas 500, com



**TABELA 24**

Experimento com método que incrementa as amostras de treinamento rotuladas com as chamadas amostras semi-rotuladas, e assume idêntico valor para as probabilidades *a priori* das classes, com dimensionalidade 80 e tamanho das amostras de treinamento 500 *pixels*.

Classes da Imagem de Referência	Acurácia do Produtor	Amostras de Teste	Limiar (Ti)	Classes da Imagem					
				Classificada					
				MPD	MCM	SPD	SCM	SCC	F
Milho Plantio Direto-MPD	79.6	716	1	<b>570</b>	47	8	89	2	0
Milho Cultivo Mínimo-MCM	84.2	417	0	16	<b>351</b>	0	36	14	0
Soja Plantio Direto-SPD	75.4	590	0	42	8	<b>445</b>	94	0	1
Soja Cultivo Mínimo-SCM	70.9	1234	0	104	139	70	<b>875</b>	46	0
Soja Cultivo Convencional-SCC	87.6	330	1	3	8	6	24	<b>289</b>	0
Floresta-F	100.0	580	0	0	0	0	0	0	<b>580</b>
Totais de Amostras		3867	2	735	553	529	1118	351	581
Acurácia do Usuário(%)				<b>77.6</b>	<b>63.5</b>	<b>84.1</b>	<b>78.3</b>	<b>82.3</b>	<b>99.83</b>
Acurácia Geral(%):	<b>80.4</b>								

**TABELA 25**

Experimento com o método proposto neste estudo, que inclui valores estimados para as probabilidades *a priori* de cada uma das classes, com dimensionalidade 80 e tamanho das amostras de treinamento 500 *pixels*.

Classes da Imagem de Referência	Acurácia do Produtor	Amostras de Teste	Limiar (Ti)	Classes da Imagem					
				Classificada					
				MPD	MCM	SPD	SCM	SCC	F
Milho Plantio Direto-MPD	79.3	716	1	<b>568</b>	47	7	93	1	0
Milho Cultivo Mínimo-MCM	83.7	417	0	16	<b>349</b>	0	38	14	0
Soja Plantio Direto-SPD	74.2	590	0	41	8	<b>438</b>	102	0	1
Soja Cultivo Mínimo-SCM	73.4	1234	0	102	123	62	<b>906</b>	41	0
Soja Cultivo Convencional-SCC	86.4	330	1	5	9	6	25	<b>285</b>	0
Floresta-F	100.0	580	0	0	0	0	0	0	<b>580</b>
Totais de Amostras		3867	2	732	536	513	1164	341	581
Acurácia do Usuário(%)				<b>77.6</b>	<b>65.1</b>	<b>85.4</b>	<b>77.8</b>	<b>83.6</b>	<b>99.83</b>
Acurácia Geral(%):	<b>80.8</b>								

#### 4.4.2.7 Experimento 7

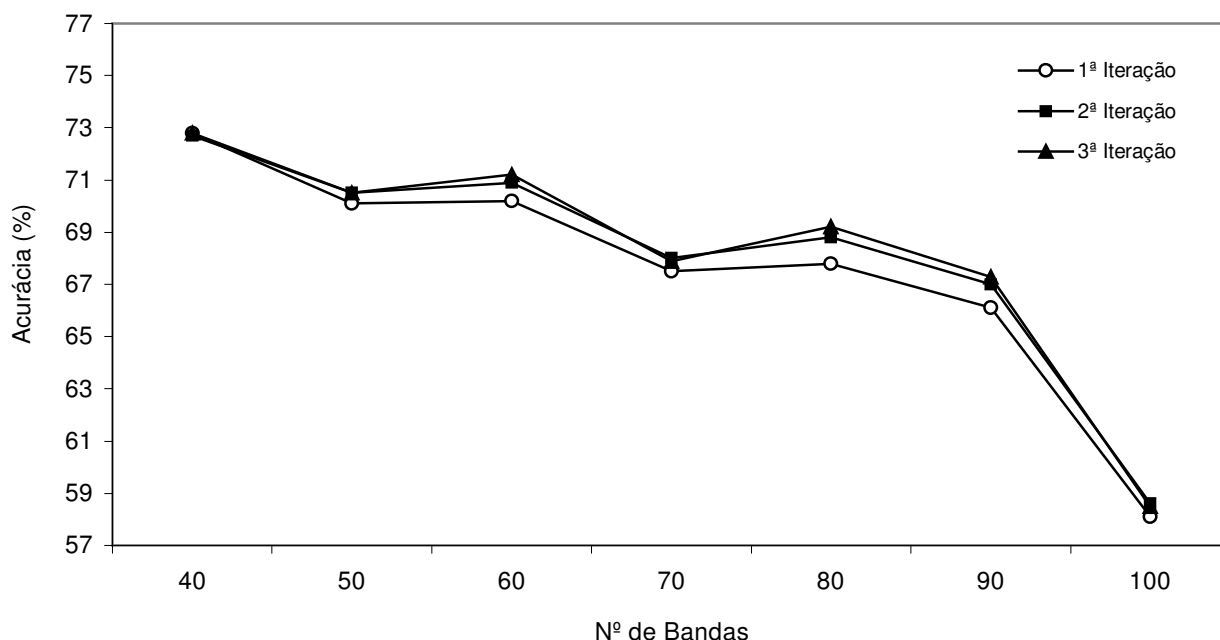
Os efeitos do parâmetro (tamanho das amostras de treinamento) com relação à acurácia do processo de classificação, também foram testados (GRAF. 2 a 11).

Excetuando-se os GRAF. 2 a 5 em que se deu o início a partir de 40 bandas, considerando-se que nos extremos não têm sentido fazer novas iterações, nos GRAF. 6 a 11 são apresentados os resultados relativos aos experimentos, cuja dimensionalidade dos dados está compreendida entre 10 e 100. Para valores ainda mais altos da dimensionalidade dos dados, não foi mais possível minimizar os efeitos do fenômeno de Hughes, usufruindo-se apenas do conjunto de amostras de treinamento disponível (amostras rotuladas).

Nos GRAF. 2 a 5, o classificador executou três iterações em cada experimento, sendo o processo interrompido quando o número de *pixels* que trocou de classe na iteração anterior e a seguinte foi inferior a 5%. Convém lembrar que em cada iteração, as amostras de treinamento são incrementadas com 50 amostras semi-rotuladas disponíveis para cada classe, sempre aquelas cujo peso é maior, procedimento este adotado para todos os experimentos realizados neste estudo.

**GRÁFICO 2**

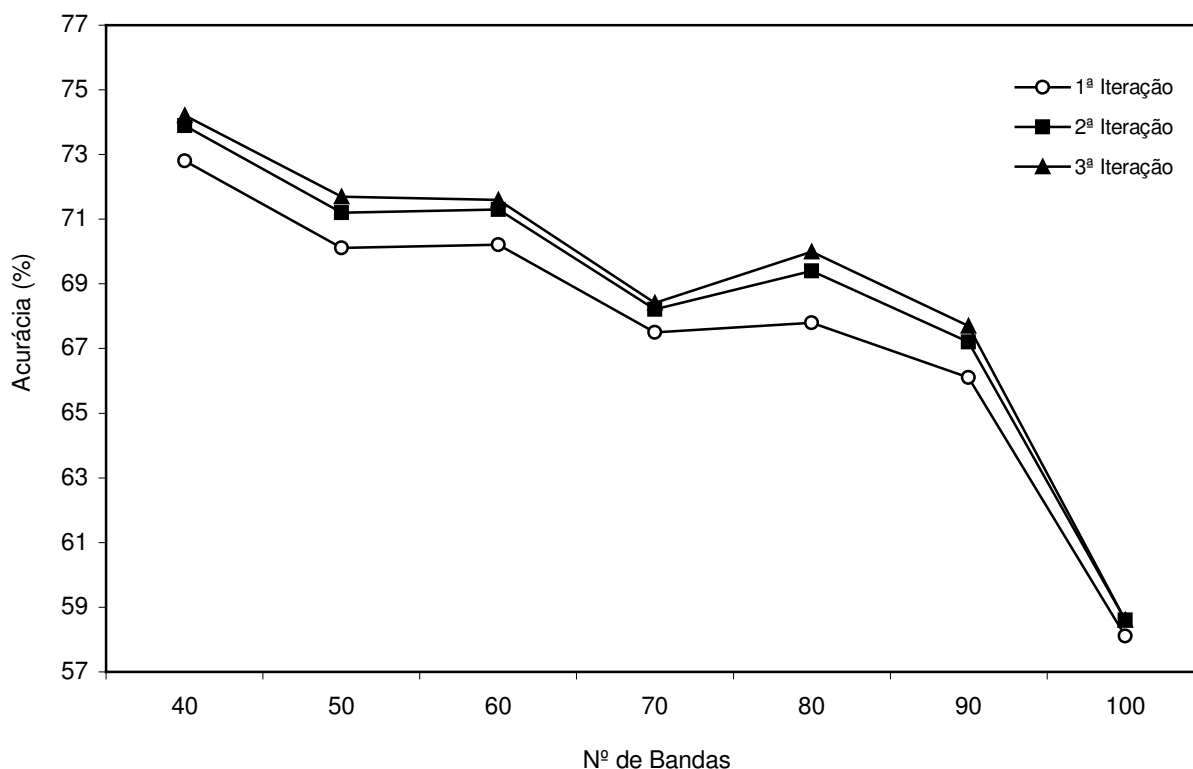
**Acurácia da classificação (usuário), com os valores das probabilidades *a priori* idênticos para todas as classes, com amostras de treinamento rotuladas de tamanho 200 e incremento de 50 amostras semi-rotuladas, em três iterações.**



Como se vê no GRAF. 2, neste experimento em que foram utilizados valores das probabilidades *a priori* idênticos para todas as classes e as amostras de treinamento rotuladas de tamanho 200, verifica-se que apesar das acurácias das iterações apresentadas serem consideradas boas, especialmente para a dimensionalidade 40 (~73%), havendo razoável melhora entre a primeira e a última iteração em todas as dimensionalidades utilizadas, o método não logrou o êxito esperado e, por vezes não convergiu, atingindo a dimensionalidade 100 com valor de aproximadamente 58%, em função do tamanho das amostras de treinamento utilizado nas classes em estudo e a dimensionalidade dos dados, evidenciando-se o efeito do fenômeno de Hughes.

**GRÁFICO 3**

**Acurácia da classificação (usuário), com os valores das probabilidades *a priori* estimados para cada classe, com amostras de treinamento rotuladas de tamanho 200 e incremento de 50 amostras semi-rotuladas, em três iterações.**

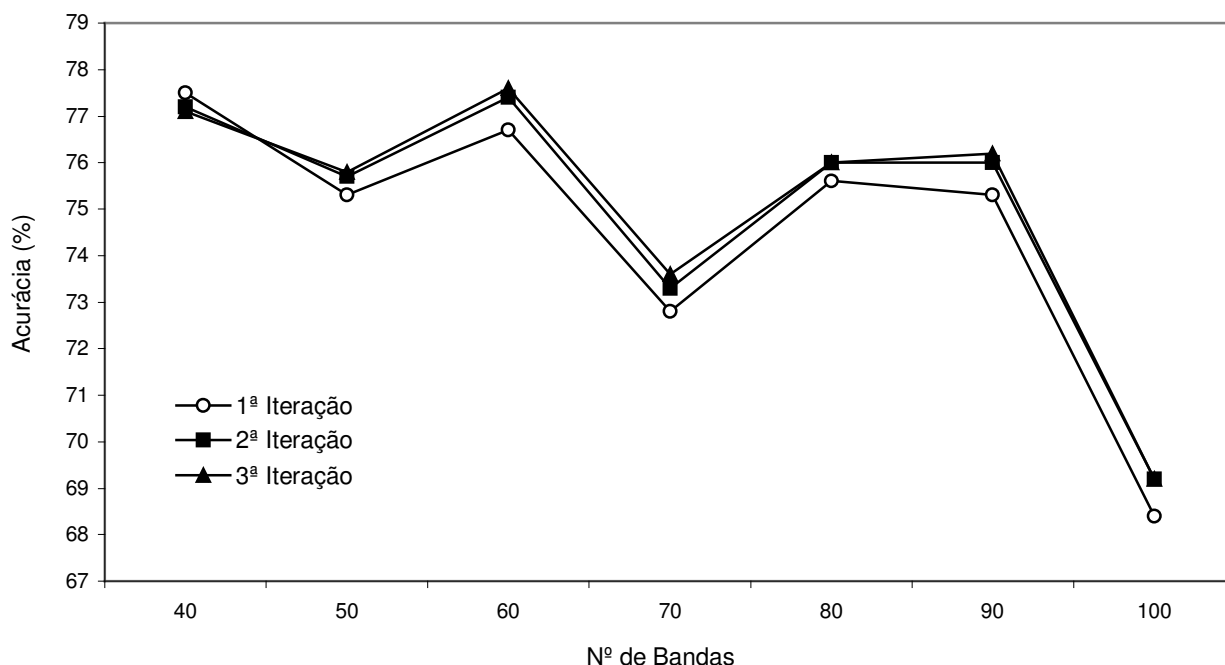




Neste experimento, onde foram estimados os valores das probabilidades *a priori* para cada classe e usadas amostras de treinamento rotuladas de tamanho 200, percebe-se uma sensível elevação da acurácia (~74%) para a dimensionalidade 40 e boa melhora da primeira para a última iteração, comparando-se com o GRAF. 2, contudo, ainda não foi possível suavizar o efeito do fenômeno de Hughes, onde a acurácia manteve-se descendente até o valor aproximado de 58% para a dimensionalidade 100.

**GRÁFICO 4**

**Acurácia da classificação (usuário), com os valores das probabilidades *a priori* idênticos para todas as classes, com amostras de treinamento rotuladas de tamanho 300 e incremento de 50 amostras semi-rotuladas, em três iterações.**

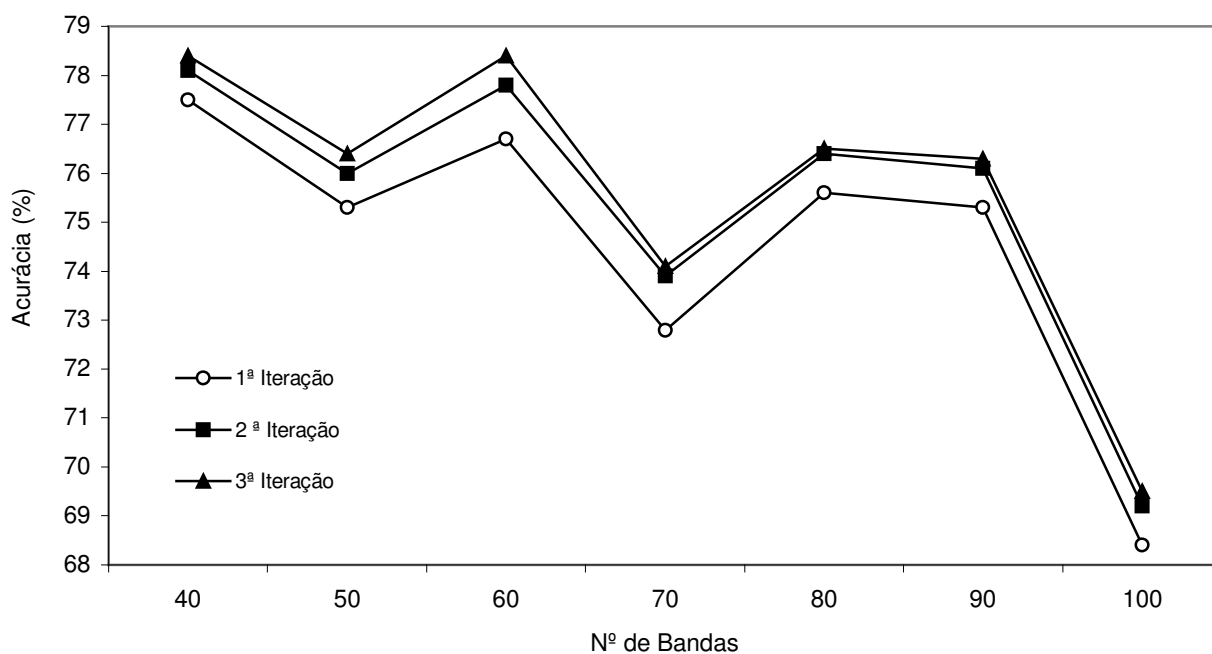


No GRAF. 4, onde foram utilizados valores das probabilidades *a priori* idênticos para todas as classes e as amostras de treinamento rotuladas de tamanho 300, comparando-se com o GRAF. 2, nota-se o grande aumento dos picos de acurácia em relação a todos os conjuntos de bandas utilizadas, alcançando-se valores acima de 77% e a melhora da primeira para a última iteração, com isto foi possível suavizar o

efeito do fenômeno de Hughes, até a dimensionalidade 90, a partir da qual a acurácia decresceu sensivelmente, tendo ainda assim, atingindo o valor aproximado de 69%, relativo à dimensionalidade 100.

**GRÁFICO 5**

**Acurácia da classificação (usuário), com os valores das probabilidades a priori estimados para cada classe, com amostras de treinamento rotuladas de tamanho 300 e incremento de 50 amostras semi-rotuladas, em três iterações.**

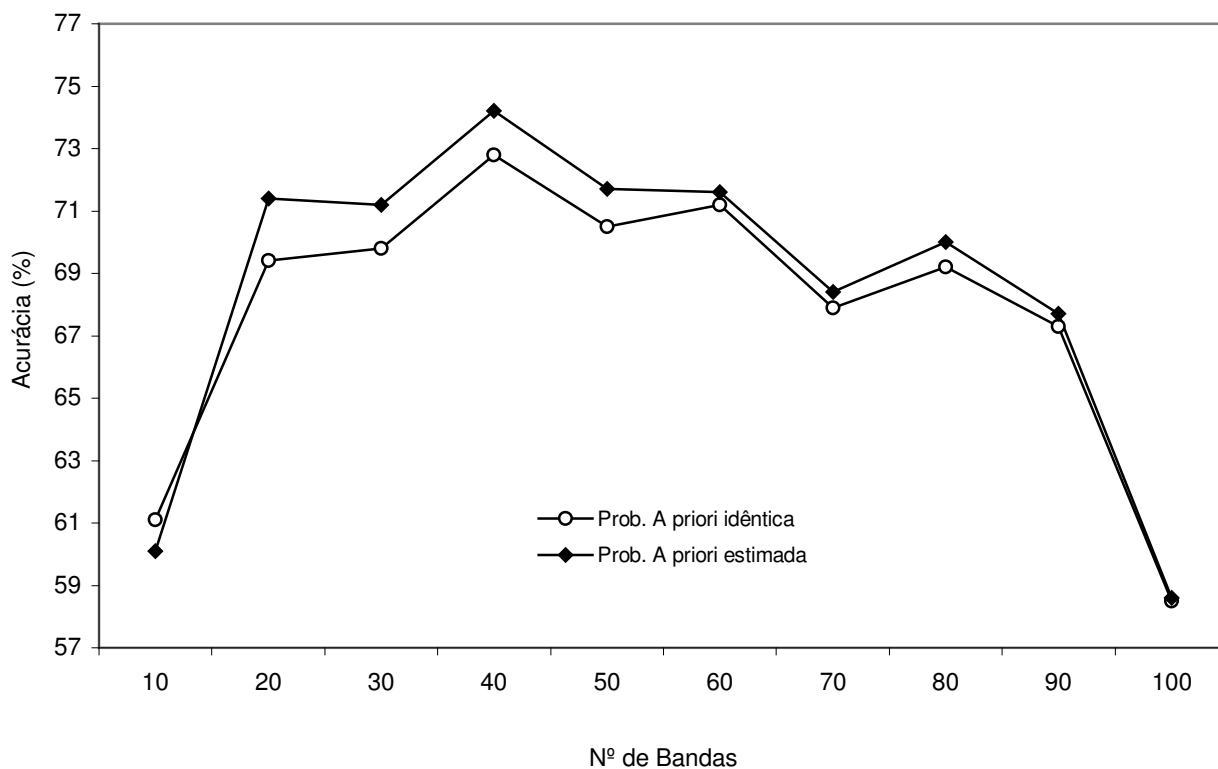


Neste experimento, em que foram estimados os valores das probabilidades *a priori* para cada classe e as amostras de treinamento rotuladas de tamanho 300, comparando-se com o GRAF. 3, percebe-se o aumento dos picos de acurácia em relação a todos os conjuntos de bandas utilizadas e a melhora da primeira para a última iteração, com isto foi possível suavizar os efeitos do fenômeno de Hughes, até a dimensionalidade 90, a partir da qual a acurácia apresentou grande decréscimo, atingindo ainda assim o valor aproximado de 69% relativo à dimensionalidade 100.

Nota-se que ao serem introduzidos os valores estimados para as probabilidades *a priori* das classes, a diferença entre as iterações em todas as dimensionalidades é aumentada, o que melhora consideravelmente a estabilidade dos resultados, ou seja, prevalecem estas diferenças mais acentuadas, vistas especialmente nos GRAF. 3 e 5, através da manutenção do afastamento e a uniformidade no desenvolvimento das linhas correspondentes a cada uma das três iterações.

**GRÁFICO 6**

**Acurácia da classificação final (usuário) com os valores das probabilidades *a priori* idênticos para todas as classes e, estimados para cada classe, com as amostras de treinamento rotuladas de tamanho 200 e incremento de 50 amostras semi-rotuladas, em três iterações.**

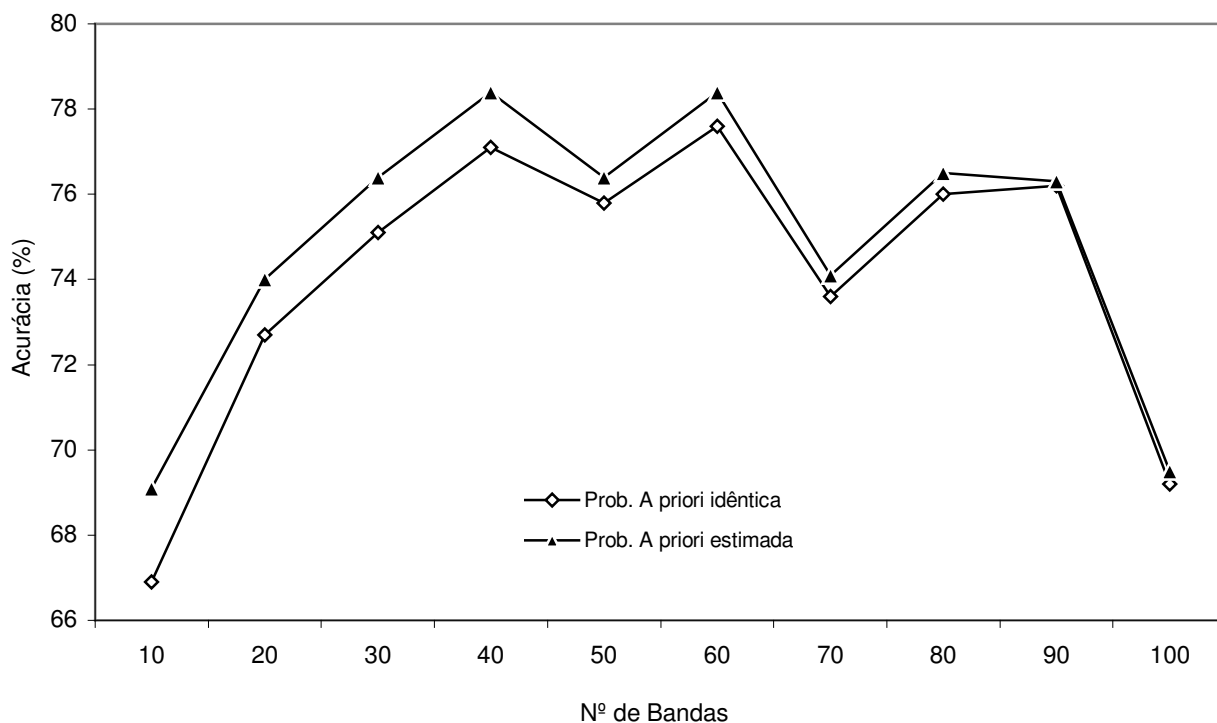


Como se vê no GRAF. 6, usando-se amostras de treinamento de tamanho 200 para todas as classes em estudo, destacam-se os picos de acurácia para

a classificação final da imagem ao longo de todo o experimento, quando os valores estimados para as probabilidades *a priori* foram incluídos na função decisão, ainda que, a partir da dimensionalidade 40, torna-se evidente os efeitos do fenômeno de Hughes.

**GRÁFICO 7**

**Acurácia da classificação final (usuário) com os valores das probabilidades *a priori* idênticos para todas as classes e, estimados para cada classe, com as amostras de treinamento rotuladas de tamanho 300 e incremento de 50 amostras semi-rotuladas, em três iterações.**

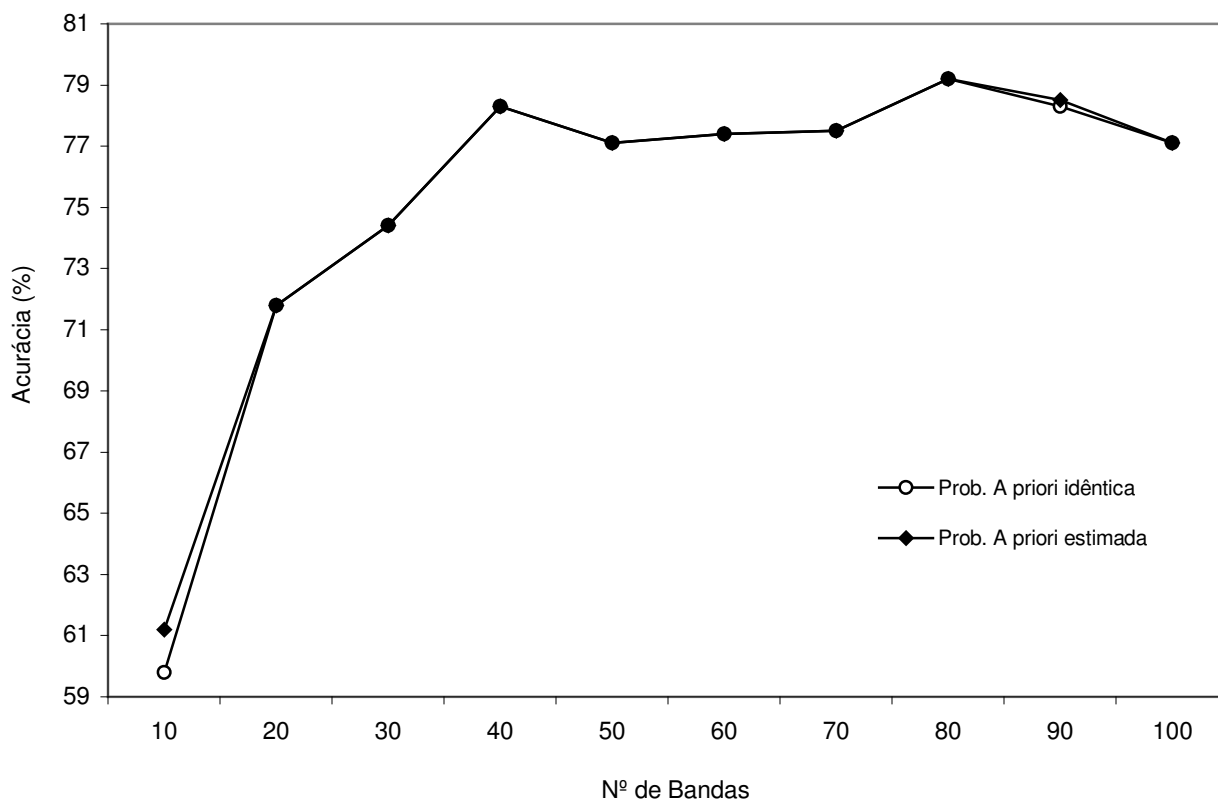


Observando-se o GRAF. 7, onde foram usadas amostras de treinamento de tamanho 300 para todas as classes em estudo, destaca-se o aumento dos picos de acurácia para a classificação final da imagem, praticamente ao longo de todo o experimento, quando os valores estimados para as probabilidades *a priori* foram incluídos na função decisão, onde apesar de apresentar pequenas oscilações ocasionando a redução da acurácia, especialmente em relação às dimensionalidades

70 e 100, até a dimensionalidade 90, também houve considerável suavização dos efeitos do fenômeno de Hughes.

**GRÁFICO 8**

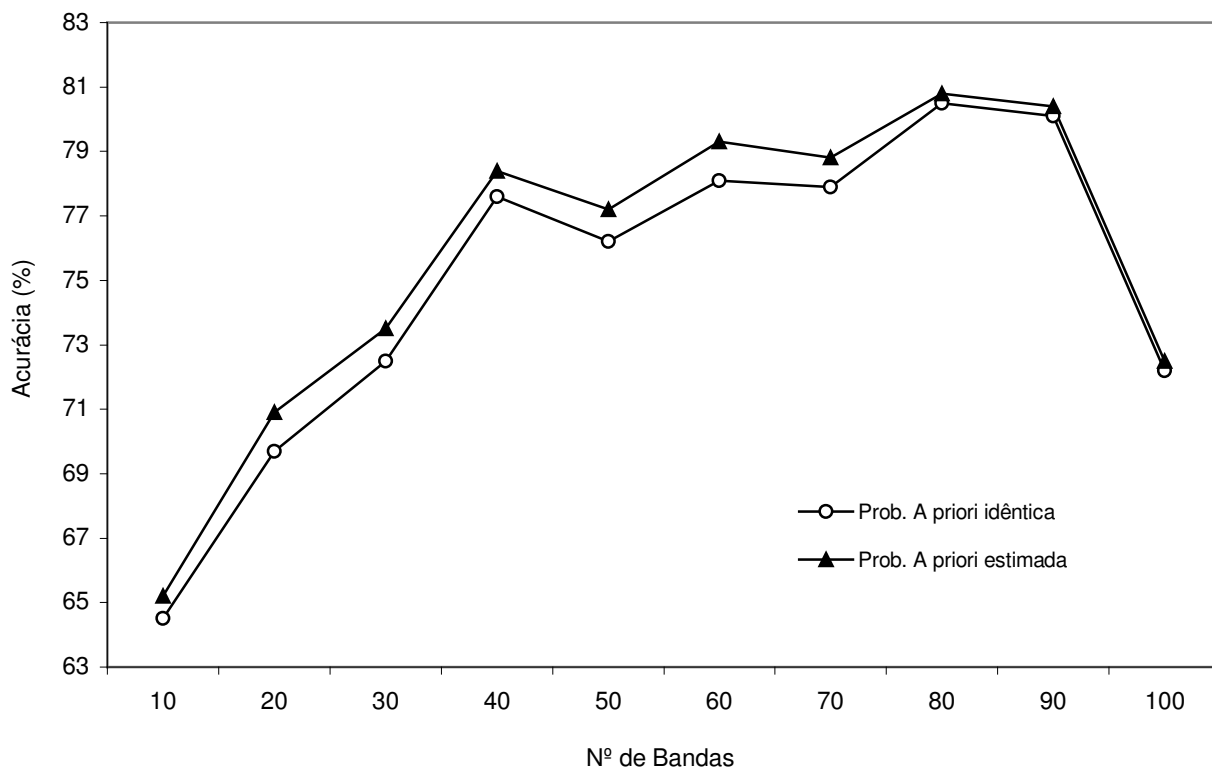
**Acurácia da classificação final (usuário) com os valores das probabilidades *a priori* idênticos para todas as classes e, estimados para cada classe, com as amostras de treinamento rotuladas de tamanho 400 e incremento de 50 amostras semi-rotuladas, em três iterações.**



No GRAF. 8, em que empregou-se amostras de treinamento de tamanho 400 para todas as classes em estudo, ainda que praticamente não haja diferenças, destaca-se a manutenção do aumento dos picos de acurácia para a classificação final da imagem ao longo de todo o experimento, quando os valores estimados para as probabilidades *a priori* foram incluídos na função decisão. Também, percebe-se nitidamente, que os efeitos do fenômeno de Hughes foram mitigados.

GRÁFICO 9

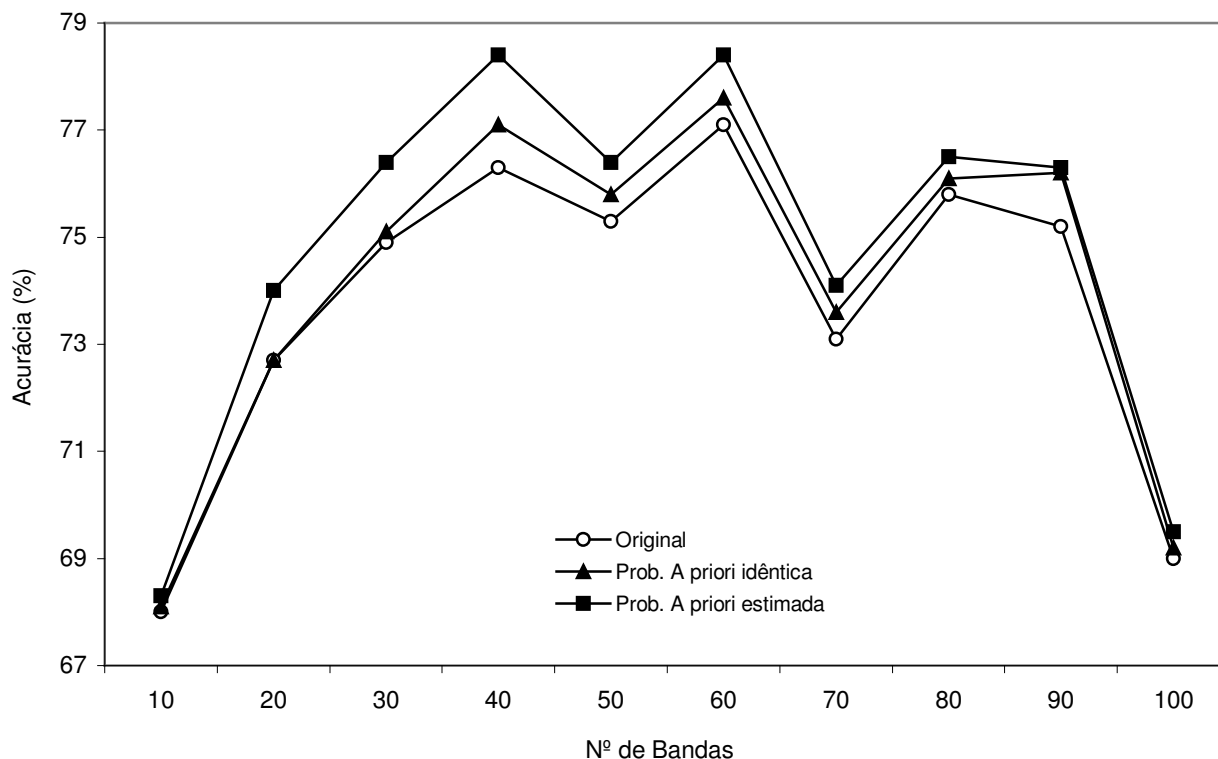
Acurácia da classificação final (usuário) com os valores das probabilidades *a priori* idênticos para todas as classes e, estimados para cada classe, com as amostras de treinamento rotuladas de tamanho 500 e incremento de 50 amostras semi-rotuladas, em três iterações.



Neste experimento, onde o tamanho das amostras de treinamento foi 500 para todas as classes em estudo, destaca-se a manutenção do aumento dos picos de acurácia para a classificação final da imagem ao longo de todo o experimento, quando os valores estimados para as probabilidades *a priori* foram incluídos na função decisão. Também se verifica no GRAF. 9 que até a dimensionalidade 90, os efeitos do fenômeno de Hughes foram perfeitamente suavizados.

GRÁFICO 10

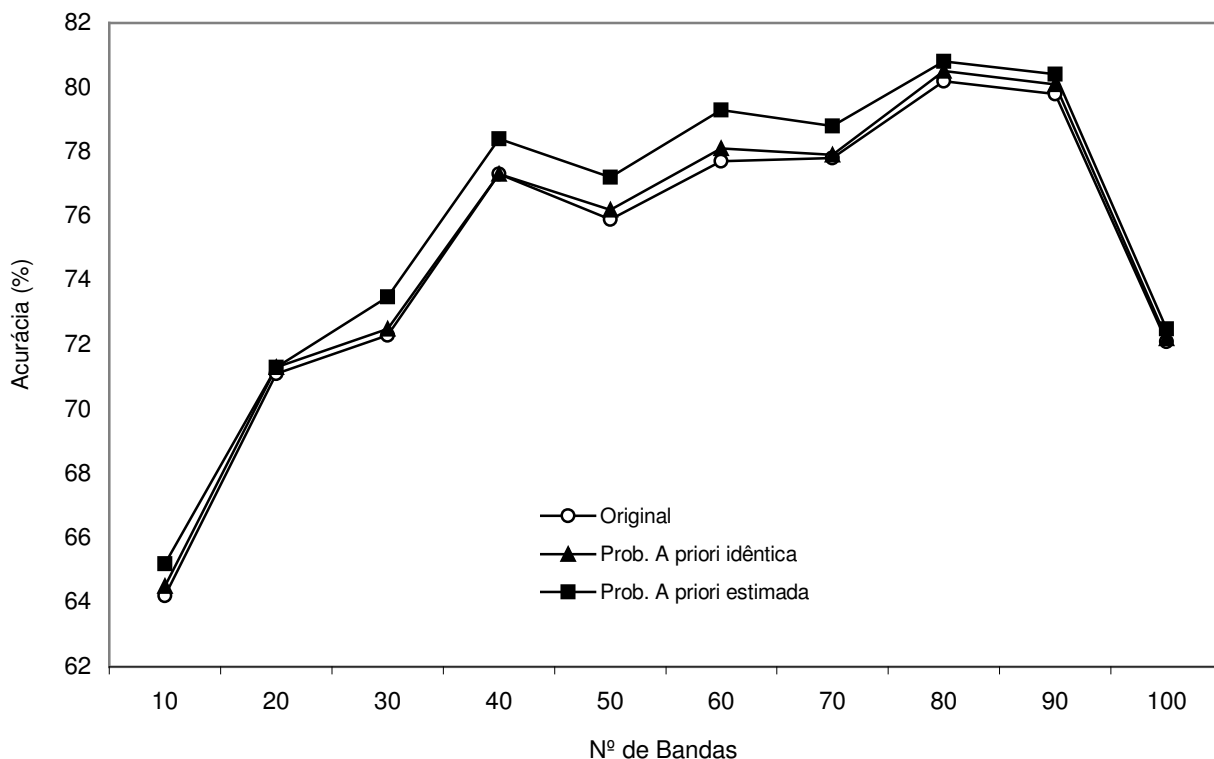
Acurácia da classificação final (usuário), através dos seguintes métodos: original; e, com os valores das probabilidades *a priori*: idênticos para todas as classes e, estimados para cada classe, com amostras de treinamento rotuladas de tamanho 300 e incremento de 50 amostras semi-rotuladas, em três iterações.



Analisando-se o GRAF. 10, em cujo experimento utilizou-se amostras de treinamento de tamanho 300 para todas as classes envolvidas, percebe-se o aumento dos picos de acurácia para a classificação final da imagem, praticamente ao longo de todo o experimento, principalmente daqueles gerados pelo método proposto, ou seja, quando os valores estimados para as probabilidades *a priori* foram incluídos na função decisão, onde apesar de ocorrer em todos os métodos pequenas oscilações ocasionando a redução da acurácia, especialmente em relação às dimensionalidades 70 e 100, excetuando-se o método convencional, até a dimensionalidade 90 os efeitos do fenômeno de Hughes foram razoavelmente suavizados.

GRÁFICO 11

Acurácia da classificação final (usuário), através dos seguintes métodos: original; e, com os valores das probabilidades *a priori*: idênticos para todas as classes e, estimados para cada classe, com amostras de treinamento rotuladas de tamanho 500 e incremento de 50 amostras semi-rotuladas, em três iterações.



Neste último experimento GRAF. 11, onde foram utilizadas amostras de treinamento de tamanho 500 para todas as classes em questão, destaca-se a manutenção do aumento dos picos de acurácia para a classificação final da imagem ao longo de todo o experimento, principalmente quando foi utilizado o método proposto, onde se vê claramente, que até a dimensionalidade 90 os efeitos do fenômeno de Hughes foram completamente minimizados.

Nota-se também, que da dimensionalidade 10 até a 80 GRAF. 10 e da dimensionalidade 20 até a 90 GRAF. 11, mesmo que as diferenças dos valores dos picos de acurácia gerados pelo método proposto não sejam tão grandes, ainda assim,

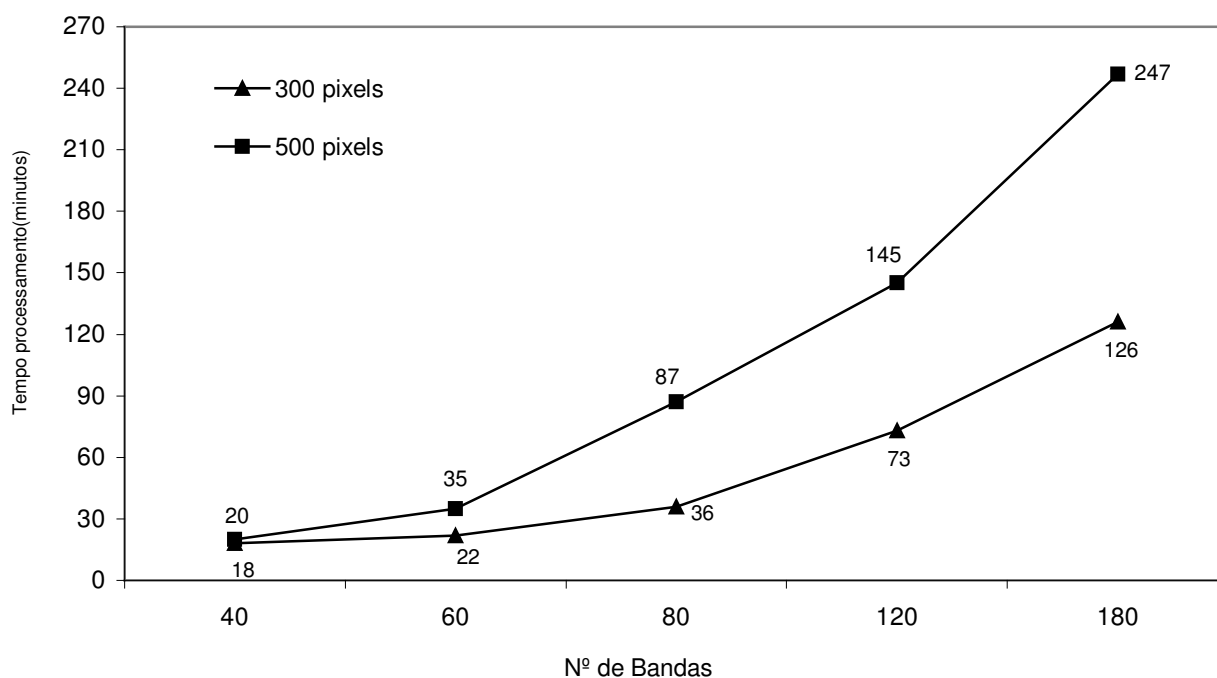


estas são bastante expressivas e constantes em relação aos outros dois métodos, sendo que nestes últimos elas às vezes se confundem.

No GRAF.12 são demonstrados os tempos de processamento, empregando-se o equipamento e os aplicativos descritos na seção 4.2, por ocasião das classificações feitas nos experimentos considerados mais importantes, ilustrados nas FIG. 22 a 27, onde foram utilizadas amostras de treinamento com 300 e 500 *pixels* rotulados.

**GRÁFICO 12**

**Tempo total de processamento para amostras de treinamento de tamanhos 300 e 500 *pixels*, e incremento de 50 amostras semi-rotuladas.**



#### 4.5 ANÁLISE DOS EXPERIMENTOS

As imagens constantes das FIG. 22 a 27, geradas através da aplicação dos três métodos (seção 4.4.2), não causam grande impacto quanto à diferença visual,

entretanto, o objetivo é demonstrar individualmente a qualidade da definição geométrica referente às áreas representadas pelas classes em estudo e, permitir analisar onde foram introduzidos possíveis erros e onde houve melhoras significativas ao longo do processo de classificação.

Considerando-se que foram utilizadas as mesmas classes e igual tamanho das amostras rotuladas e semi-rotuladas em cada série de experimentos, é possível então comparar entre si os resultados da “Acurácia Geral”, TAB. 8 a 25, onde se verifica além da evolução dos métodos, o alto grau de eficácia do classificador adaptativo num processo iterativo de classificação de dados hiperespectrais, sendo conceituados como “Muito Bom” e “Excelente”, quando comparados com valores do índice  $kappa^2$ .

Os experimentos desenvolvidos mostram que a acurácia do usuário obtida no processo de classificação apresenta um comportamento esperado. Com exceção da classe Floresta, TAB. 8 a 25, a qual se manteve sempre com valores de acurácia superiores a 99%, as demais apresentaram valores crescentes com o aumento da dimensionalidade, atingindo um máximo para em seguida decrescer, na medida em que a dimensionalidade dos dados continua a aumentar, numa clara ilustração do fenômeno de Hughes.

Os GRAF. 2 a 11, demonstram o comportamento da acurácia a cada iteração em função da dimensionalidade dos dados. Para uma melhor avaliação da contribuição resultante da metodologia proposta neste estudo, nestas ilustrações são apresentados os resultados dos três métodos para classificação, descritos na seção 4.4.2.

Em relação aos GRAF. 2 a 5, os quais mostram a variação da acurácia com o número de iterações, observa-se que o acréscimo na acurácia resultante da

---

<sup>2</sup> A análise sobre o coeficiente *Kappa* ( $k^{\wedge}$ ) determina estatisticamente o quanto uma matriz de erro difere significativamente de outra (BISHOP et al., 1975 apud CONGALTON & GREEN, 1999).

inclusão nas funções de decisão dos valores estimados para as probabilidades *a priori* é mais notável quando o número de amostras rotuladas é maior.

O valor estimado para a “Acurácia Geral” mostrou uma correlação positiva com o tamanho das amostras de treinamento. Também, conforme o esperado, o pico na curva de acurácia causado pelo fenômeno de Hughes, tende a ocorrer para valores mais altos da dimensionalidade dos dados, na medida em que o tamanho das amostras de treinamento aumenta. Os efeitos do tamanho das amostras na acurácia do processo de classificação é bastante conhecido. Uma amostra maior permite uma estimativa mais confiável dos parâmetros do classificador, em especial da matriz covariância, resultando em valores mais altos para a acurácia dos resultados.

Observando-se a acurácia alcançada nos experimentos concentrados nos GRAF. 2 a 11, nota-se um aumento no valor estimado para a acurácia, especialmente quando os valores estimados para as probabilidades *a priori* são incluídos na função decisão, o resultado da classificação melhora consideravelmente a cada nova iteração do classificador, num comparativo direto com aqueles resultados provenientes quando adotado um valor único, uniforme para as probabilidades *a priori* de todas as classes presentes.

Ainda em referência às possíveis análises, sobre os gráficos nas figuras citadas no parágrafo anterior, podem ser feitas as seguintes constatações quanto aos resultados:

- 1) Nos experimentos com dimensionalidade de 40, 60 e 80, a acurácia da classificação inicial (utilizando-se somente as amostras de treinamento rotuladas), decresce com o aumento da dimensionalidade, indicando a ocorrência do Efeito de Hughes;
- 2) Inserindo-se as amostras semi-rotuladas e a inclusão dos valores da estimativa das probabilidades *a priori*, a partir da segunda iteração a

acurácia da classificação começa a aumentar, indicando que a influência do Efeito de Hughes começa a ser aliviada consideravelmente;

- 3) A acurácia da classificação final (utilizando-se amostras rotuladas e semi-rotuladas de mesmo tamanho para todas as classes envolvidas e a inclusão dos valores da estimativa das probabilidades *a priori*) em todos os experimentos é similar, indicando que sua performance é comparável.

Como se vê para configurações que possuam um número pequeno de amostras para estimação de parâmetros, o classificador com algoritmo adaptativo constitui-se em uma alternativa viável para o problema da redução da acurácia, provocada pelo Efeito de Hughes.

## CAPÍTULO V

### CONCLUSÕES E SUGESTÕES

#### 5.1 CONCLUSÕES

Nesta dissertação foi apresentada e testada uma metodologia, na qual é proposto um método iterativo para estimação de valores para as probabilidades *a priori* das classes, em um processo de classificação. Neste mesmo processo iterativo, são ainda introduzidas amostras semi-rotuladas para fins de estimação dos parâmetros em um classificador quadrático (vetor de médias e matrizes de covariância das classes em consideração). A metodologia proposta é dirigida ao processo de classificação de dados em alta dimensionalidade (imagens hiperespectrais). Como é bem conhecido, o atrativo da utilização destes dados reside na capacidade de separar classes com características espectrais muito semelhantes (FUKUNAGA, 1990). Esta vantagem fica, entretanto, prejudicada sempre que o número de amostras de treinamento disponíveis não é suficiente, fato este que ocorre com freqüência em situações reais. Lemos (2003), Jackson (2001, 2002), Shashahani (1994), mostraram que este problema pode ser contornado com sucesso, quando se incrementa o número de amostras de treinamento (amostras rotuladas) com amostras adicionais, não-rotuladas. A adição de amostras não-rotuladas, mostrou que a estimativa dos parâmetros requeridos pelo classificador quadrático tornam-se mais confiáveis, resultando em um incremento na acurácia da imagem temática produzida pelo classificador. A contribuição principal deste estudo consiste da introdução no processo iterativo acima, de uma etapa adicional referente à estimação das probabilidades *a priori* para as classes envolvidas no processo de classificação. Vários testes foram desenvolvidos para testar a metodologia proposta, utilizando dados em alta dimensionalidade fornecidos por um sistema sensor experimental do tipo hiperespectral (AVIRIS),

coabrindo um campo de teste com verdade terrestre confiável. Nestes testes foram empregados dados espectrais relativos a seis diferentes classes, sendo cinco destas com características espectrais muito semelhantes e, portanto, de difícil separação quando se utiliza dados convencionais em baixa dimensionalidade, como por exemplo dados LANDSAT ou SPOT. Com a finalidade de simular as condições encontradas em situações reais, foram utilizadas amostras de treinamento em número reduzido quando comparado com a dimensionalidade dos dados. Situações como estas tornam difícil uma estimativa confiável dos parâmetros do classificador (vetor de médias e matriz de covariâncias) pelo processo tradicional, empregando apenas as amostras rotuladas disponíveis. Nestes experimentos, especial atenção foi dada à investigação com relação ao incremento na acurácia da imagem temática, resultante da introdução de valores estimados para as probabilidades *a priori*.

Os experimentos desenvolvidos tendem a confirmar a eficácia da metodologia proposta. Os resultados dos experimentos mostram que a metodologia proposta neste trabalho resulta a cada iteração, em um incremento na acurácia da imagem temática produzida, convergindo para um patamar mais elevado do que o obtido pelo processo tradicionalmente utilizado, quando somente as amostras rotuladas são empregadas para fins de estimação dos parâmetros do classificador e a informação referente às probabilidades *a priori* é descartada do processo. Também, os experimentos mostram que com esta metodologia é possível mitigar os efeitos do fenômeno de Hughes, deslocando o pico na curva de acurácia para valores mais altos da dimensionalidade dos dados. Isto significa que a informação adicional resultante da inclusão de um número maior de bandas espectrais no processo de classificação, pode ser utilizada para obter-se uma imagem temática mais acurada.

Os experimentos mostram ainda, que os valores estimados para os parâmetros e para as probabilidades *a priori* tendem a convergir.

Apesar dos resultados promissores, o classificador adaptativo proposto tem limitações. Em particular, em condições muito desfavoráveis em que a

dimensionalidade dos dados é muito maior que o número de amostras de treinamento, a classificação inicial pode ser muito imprecisa. Como resultado o classificador adaptativo proposto pode não convergir.

## 5.2 SUGESTÕES

Como sugestão para futuros desenvolvimentos neste tópico, sugere-se que sejam investigadas outras possíveis abordagens para fins de introdução de amostras semi-rotuladas no processo de classificação, e para estimação das probabilidades *a priori* das classes presentes na cena. São aqui também sugeridos estudos adicionais para melhor determinar a possível introdução de tendenciosidade no processo proposto para estimação das probabilidades *a priori*.

Estudar um método que combine um classificador adaptativo com vários métodos de estimação regularizada de covariância como, por exemplo, Análise Discriminante Regularizada (RDA), Estimador de Covariância *Leave-one-out* (LOOC), Estimador de Covariância Bayesiano *Leave-one-out* (BLOOC).

Otimização da ferramenta CDH, pela adição de novas técnicas para fins de comparação de performance dos resultados obtidos, como por exemplo: EM, RDA, LOOC, BLOOC e outros.

Desenvolver ferramentas computacionais mais eficientes com vistas agilizar o processo de classificação, em face do alto grau de complexidade que somado ao grande volume de informações, característica intrínseca deste tipo de dados imagem e em face dos recursos computacionais, normalmente disponíveis ao usuário em geral, contribuem para a degradação da performance do classificador, evitando-se assim que haja solução de continuidade, em que possa desestimular o seu uso, ou comprometer a viabilidade da sua aplicação em larga escala.

## REFERÊNCIAS BIBLIOGRÁFICAS

AEBERHARD, S.; COOMANS, D.; De Vel, O. **Comparative Analysis of Statistical Pattern Recognition Methods in High Dimensional Settings**. Pattern Recognition, v. 27, n. 8, p. 1065-1077, 1994.

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.

CONGALTON, R. G.; GREEN, K. **Assessing the Accuracy of Remotely Sensed Data: Principles and Practices**. Lewis Publishers. Denver. EUA. 1999. 180 p. ISBN 0873719867.

CORNFIELD, J. **Discriminant Functions**. Review of the International Statistical Institute, 35, 142-153, 1967.

CORTIJO F.J.; DE LA BLANCA N. P. **The Performance of Regularized Discriminant Analysis Versus Non-parametric Classifiers Applied to High-Dimensional Image Classification**. International Journal of Remote Sensing, v.20, n. 17, 3345-3365, 1999.

ERBERT, M. **Uso da Análise discriminante regularizada no reconhecimento de padrões em imagens digitais hiperespectral de sensoriamento remoto**. 2001. 118 f. Dissertação (Mestrado em Sensoriamento Remoto) – Programa de Pós-Graduação em Sensoriamento Remoto e Meteorologia, Centro Estadual de Pesquisas em Sensoriamento Remoto e Meteorologia, Universidade Federal do Rio Grande do Sul, Porto Alegre, RS, 2001.

FRIEDMAN, J. H. **Regularized Discrimination Analysis**. Journal of the American Statistical Association, Theory and Methods, v. 84, n. 405, 1989.

FUKUNAGA, K.; HAYES R. R. **Effects of Sample Size in Classifier Design**. IEEE Transaction Pattern Analysis and Machine Intelligence, v.11, n. 8, p. 873-885, August, 1989.

FUKUNAGA, K. **Introduction to Statistical Pattern Recognition**. 2. ed., School of Electrical Engineering, Purdue University, West Lafayette, Academic Press Inc. 1990.

GREEN, R. O. et al. **Imaging Spectroscopy and the Airborne Visible/ Infrared Imaging SPectrometer (AVIRIS)**. *Remote Sensing of Environment* 64:227-248, 1998.

HAERTEL V.; 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, v. 37, n. 5, p. 2374-2386, September, 1999.



HAERTEL V. **Considerations about Classification of Remote Sensing Digital Image Data in High Dimensional Settings**. IX Simpósio Latinoamericano de Perception Remota, Puerto Iguazú, Misiones, Argentina, Noviembre, 2000.

HOFFBECK J.P.; LANDGREBE D. A. **Covariance Matrix Estimation and Classification With Limited Training Data**. IEEE Transactions on Pattern Analysis and Machine Intelligence, v. 18, n. 7, July, 1996.

HSIEH, P.; LANDGREBE, D.A. **Classification of High Dimensional Data**. TR-ECE 98-4, School of Elec. And Comp. Eng., Purdue University , May, 1998.

HUGHES, G. F. **On the mean accuracy of statistical pattern recognizers**. IEEE Transactions an Information Theory, v. 11 - 14, n. 1, p. 55- 65, 1968.

HYPERION – **Sensor Hyperion/Satélite EO-1**. Disponível em: <<http://www.envi.com.br/sensor/satelites/eo1/Hyperion/hyperion.html>>. Acesso em: 5 jul. 2007.

INPE - Instituto Nacional de Pesquisas Espaciais. Divisão de Sensoriamento Remoto. **Discriminação de Variedades de Cana-de-açúcar com Dados Hiperespectrais do Sensor HYPERION/EO-1**. Disponível em:<[www.rbc.ufrj.br/\\_pdf\\_57\\_2005/57\\_1\\_02](http://www.rbc.ufrj.br/_pdf_57_2005/57_1_02)>. Acesso em: 8 jul. 2007.

INPE - Instituto Nacional de Pesquisas Espaciais. **Simulações com o sensor AVIRIS no Brasil**. Disponível em: <<http://www.inpe.br/programas/cbers/english/imagens.html>>. Acesso em: 14 ago. 2006.

JAIN, A.; DUIN, R.P.W.; MAO, J. **Statistical Pattern Recognition: A Review**. IEEE Transactions on Pattern Analysis and Machine Intelligence, v. 22, n. 1, p. 4-37, 2000.

JACKSON, Q.; LANDGREBE, D.A. **An Adaptive Classifier Design for High-Dimensional Data Analysis with a Limited Training Data Set**. School of Electrical &Computer Engineering Purdue University, West Lafayette, 2001.

JACKSON, Q.; LANDGREBE, D.A. **An Adaptive Method for Combined Covariance Estimation and Classification**. School of Electrical &Computer Engineering Purdue University, West Lafayette, 2002.

JIA, X.; RICHARDS, J. **Efficient Maximum Likelihood Classification for Imaging Spectrometer Data Sets**. IEEE Transactions on Geoscience and Remote Sensing, v. 32, n. 2, p. 274-280, March, 1994.

JIA, X.; RICHARDS, J. **Segmented Principal Components Transformation for Efficient Hyperspectral Remote-Sensing Image Display and Classification**. IEEE Transactions on Geoscience and Remote Sensing, v. 37, n. 1, p. 538-542, January, 1999.

JIMENEZ, L. O.; LANDGREBE, D.A. **High Dimensional Feature Reduction Via Projection Pursuit**. TR-ECE 96-5, School of Elec. And Comp. Eng., Purdue University, April, 1996.

JIMENEZ, L. O.; LANDGREBE, D.A. **Hyperspectral Data Analysis and Supervised Feature Reduction Via Projection Pursuit**. IEEE Transactions on Geoscience and Remote Sensing, v. 37, n. 6, p. 2653-2667, November, 1999.

JPL - Jet Propulsion Laboratory – **NASA - What's AVIRIS?** Disponível em: <<http://aviris.jpl.nasa.gov/html>>. Acesso em: 20 mar. 2007.

KITTLER, J. **“Feature Set Search Algorithm”**. Pattern Recognition and Signal Processing, pp. 41-60, C.H.Chen, Sijthoff and Noordhoff, Alpena and den Rijn, Netherlands, 1978.

LANDGREBE, D. A. **Signal Theory Methods in Multispectral Remote Sensing**. Wiley Interscience, 2003.

LANDGREBE, D. A. **Hyperspectral Image Data Analysis as a High Dimensional Signal Processing Problem**. Reprinted from Special Issue of the IEEE Signal Processing Magazine, v. 19, n. 1 p. 17-28, January, 2002.

LANDGREBE, D. A. **Multispectral Data Analysis: A signal Theory Perspective**. School of Elec. And Comp. Eng., Purdue University, July, 1997.

LANDGREBE, D. A. **Evolution of Landsat data analysis**. Photogrammetric Engineering and Remote Sensing, v. 63, n. 7, p. 859-867, July, 1997.

LANDGREBE, D. A. **On Hyperspectral Data and its Analysis**. EE 577 Notes, March, 31, 1997.

LEE, C.; LANDGREBE, D. A. **“Analyzing high dimensional multispectral data”**. IEEE Transactions on Geoscience and Remote Sensing, v. 31, n. 4, p. 792-800, July, 1993.

LEE, C.; LANDGREBE, D. A. **“Bayes Error Evaluation of the Gaussian ML Classifier”**. IEEE Transactions on Geoscience and Remote Sensing, v. 38, n. 3, p. 147-1475, 2000.

LEMOS, D. M. **Estudo sobre um Classificador Adaptativo para Análise de Dados de Alta Dimensionalidade com um Conjunto Limitado de Dados de Treinamento**. 2003. 89 f. Dissertação (Mestrado em Sensoriamento Remoto) – Programa de Pós-Graduação em Sensoriamento Remoto e Meteorologia, Centro Estadual de Pesquisas em Sensoriamento Remoto e Meteorologia, Universidade Federal do Rio Grande do Sul, Porto Alegre, RS, 2003.

LEWOTSKY, K. **Hyperspectral imaging: evolution of imaging spectrometry**. Disponível em: <[http://www.spie.org/web/oer/november/image\\_spectro.html](http://www.spie.org/web/oer/november/image_spectro.html)>. Acesso em: 25 April, 2007.

MATLAB. **Guia do usuário**. São Paulo: Makron Books do Brasil Editora Ltda., 1999.

MULTISPEC. **Manual**. School of Electrical and Computer Engineering of Purdue University.

PUDIL, P.; NOVOVICOVA J.; KITTLER J. “**Floating Search Methods in Feature Selection**”. Pattern Recognition Letters, v. 15, p. 1119-1125, 1994.

RICHARDS, J.A. **Remote Sensing Digital Image Analysis. An Introduction, Second, revised and Enlarged Edition, and School of Electrical Engineering and Computer Science**. University of New South Wales, Austrália, Springer-Verlag Berlin Heidelberg, 1986 e 1993.

SERPICO, S. B. et al. “**A comparison of feature reduction techniques for classification of hyperspectral remote-sensing data**”. Image and Signal Processing of Remote Sensing VIII, Sebastiano B. Serpico, Editor, Proceedings of SPIE, v. 4885, 2003.

SERPICO, S. B.; L. Bruzzone, “**A New Search Algorithm for Feature Selection in Hyperspectral Remote Sensing Images**”. IEEE Transaction on Geoscience and Remote Sensing. Especial Issue on Analysis of Hiperpectral Image Data, v. 39, n. 7, p. 1360-1367, July, 2001.

SERPICO, S. B.; L. Bruzzone. “**Extension of the Jeffreys-Matusita distance to multiclass cases for feature selection**”. IEEE Transactions on Geoscience and Remote Sensing, v. 33, n. 6, p. 1318-1321, November, 1995.

SHAHSHAHANI, B. M.; 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, v. 32, n. 5, 1087-1095, September, 1994.

STEARNS, S. T. “**On Selection Features For pattern Classifiers**”. Third Int. Conf. On Pattern Recognition, Colorado(CA). 1976. p. 71-75

SABINS, F.F.Jr. **Remote Sensing: principles and interpretation**. California, USA. W.H. Freeman and Company., 1978. 426 p.

SCHOWENGERDT, R.A. **Remote Sensing Models and Methods for Image Processing**. London, Academic Press, 1997. 521 p.

SLATER, P.N. **Remote Sensing, optics and optical systems**. New York, Addison-Wesley, 1980.

SWAIN, P.H.; DAVIS, S.M. *Remote sensing: a quantitative approach*. Lafayette, McGraw-Hill. 1978.

TADJUIIN, S.; 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.

THERRIEN, C. W. **Decision Estimation and Classification, An Introduction to Pattern Recognition and Related Topics**. John Wiley & Sons, 1989.

TOU J.; GONZALES R. C. **Pattern Recognition Principles**. 3. ed. 1979, Advanced Book Program, Massachusetts, Reading: Addison-Wesley Publishing Company Inc. 1974.

## **APÊNDICE A - CÓDIGOS FONTE DA FERRAMENTA CDH**

Neste Apêndice é apresentada de maneira simplificada, a listagem das rotinas do programa computacional em linguagem MATLAB, desenvolvidas especificamente para a realização deste trabalho. Uma versão digital do presente trabalho, encontra-se disponível no disco compacto (CD-ROM) que acompanha esta dissertação.

```

% =====
%                               C D H - CLASSIFICADOR DE DADOS HIPERESPECTRAIS
% =====
% AUTOR – MESTRANDO   : Celso Antonio Liczbinski
% ORIENTADOR          : Profº Dr. Vitor Haertel
% INSTITUIÇÃO         : Universidade Federal do Rio Grande do Sul - Centro Estadual de
%                               Pesquisas em Sensoriamento Remoto e Meteorologia
% =====
% LOCAL : Porto Alegre – RS                                DATA : 20 de Agosto de 2007
% =====
% OBJETIVO:  VALIDAÇÃO DA NOVA METODOLOGIA DE CLASSIFICAÇÃO DE DADOS IMAGEM
%                               EM ALTA DIMENSIONALIDADE, EMPREGANDO AMOSTRAS SEMI-ROTULADAS E
%                               ESTIMADORES PARA AS PROBABILIDADES A PRIORI
% =====

% =====
%                               >>>>>>>>> PROGRAMA PRINCIPAL <<<<<<<<<<<<<<<<<<<<<<
% =====
% Definição de variaveis, constantes, estruturas de dados e funcoes
% -----
clc;
w          = {};
acuINT     = {};
auxACUint  = [];
matCONFUSAO = {};
IMG        = [];
Ti         = [];
porcTROCA  = [];
teste      = [];
i          = [];
j          = [];
NInt       = [];
QuiQuadrado = [];
GL         = [];
intTROCA   = [];

%Controle Temporal
%-----
iniCLASSIF = [];
fimCLASSIF = [];
timeCLASSIF = [];
iniINCLUI  = [];
fimINCLUI  = [];
timeINCLUI = [];
iniCALCULO = [];
fimCALCULO = [];
timeCALCULO = [];

% =====
% Inicio do programa principal
% =====
% 1a. Parte - Fornecer o numero de bandas
% - Fornecer o valor do limiar T(%)
% [Ti, IMG] = extraiLimiar(IMG, w, QuiQuadrado, opPPriori);

```

```

% =====
% 2a. Parte - Selecionar Amostras Rotuladas
% - Carrega as amostras do arquivo
% - Amostra de Teste
% =====
[teste, opPPriori, porcTROCA] = telaConfigPADRAO;
[GL, QuiQuadrado] = telaBANDAS;1
opINCLUSAO = telaPIXELS;
Inicio = clock;
clc;
disp(sprintf('Numero de bandas: %d', GL));
disp(sprintf('Limiar(qui-quadrado = 0.05 ): %f', QuiQuadrado));
IMG = CarregaImagem(GL, teste);
w = CarregaClasses(GL, teste);
disp('2a. Parte');
disp('Efetuada a carga da Imagem e das Classes');
disp('=====');

% =====
% 3a. Parte - Estimar os parametros estatisticos Mk e Sk.
% - Montar a funcao decisao G(X).
% - Calculo do Limiar (T).
% =====
disp('3a. Parte');
disp('Calculando a Mk e Sk padrao');
for i=1 : prod(size(w))
    w{i}.Mk = mean(w{i}.Rotulados);
    w{i}.Sk = cov(w{i}.Rotulados);
    w{i}.PPriori = ProbPriori(IMG.Imagem, w{i}.Rotulados);
end
disp('Efetuado o calculo da Prob. Priori, Mk e Sk padrao');

% Calculo do Limiar(T)
%*****
NInt = 0;
disp('Efetuando a Pre-Classificação...');
[NumP, IMG, w] = Classificador(IMG, w, NInt);
disp('Pre-Classificação efetuada!');
disp('Extraindo o Limiar...');
iniLIMIAR = clock;
[Ti, IMG] = extraiLimiar(IMG, w, QuiQuadrado, opPPriori);
fimLIMIAR = clock;
timeLIMIAR = etime(fimLIMIAR,iniLIMIAR)
disp('=====');
disp('Efetuada a extração do Limiar!');

%*****
% Inicio Iteração %
%*****
disp('Entrando na Iteração');
disp('=====');
NInt = 1;
TROCA = 1;
intTROCA = [intTROCA TROCA]; % Guarda o valor da TROCA em cada iteração
while( TROCA > porcTROCA )
    disp('=====');
    disp('Iteração: ');

```

```

disp(NInt);
disp('=====');

% =====
% 4a. Parte - Passagem do Classificador
% =====
disp('4a. Parte');
disp('Classificando a Imagem');
iniCLASSIF(NInt,:) = clock;
[NumP, IMG, w] = Classificador(IMG, w, NInt, opPPriori);
fimCLASSIF(NInt,:) = clock;
timeCLASSIF(NInt,:) = etime(fimCLASSIF(NInt,:),iniCLASSIF(NInt,:));
disp('Imagem Classificada');
disp('=====');

% =====
% 5a. Parte - Identificar novos pixels classificados: SR
% - Estimar o Peso de cada Pixel SR.
% =====
disp('5a. Parte');
disp('Estimando o Peso');
w = EstimaPeso(w); % Estima o peso para cada pixels SR
disp('Incluindo Semi-Rotulados');
iniINCLUI(NInt,:) = clock;
w = IncluiSemiRot(w,opINCLUSAO,NInt); %Acrescenta pixels SR a classe w(i)
fimINCLUI(NInt,:) = clock;
timeINCLUI(NInt,:)= etime(fimINCLUI(NInt,:),iniINCLUI(NInt,:));
disp('=====');

% =====
% 6a. Parte - Atualizar as estimativas dos valores dos parametros
% Mk e Sk de cada classe - w(i)
% - Calculo da Probabilidade a Priori para cada classe
% OBS: Desta vez utilizando-se os pixels rotulados e tambem os SR
% =====
disp('6a. Parte');
disp('Calculo da Prob. Priori, Mk e Sk estimados');
iniCALCULO(NInt,:) = clock;
w = calculoEst(IMG.Imagem, w);
fimCALCULO(NInt,:) = clock;
timeCALCULO(NInt,:)= etime(fimCALCULO(NInt,:),iniCALCULO(NInt,:));
disp('Efetuado o calculo da Prob. Priori, Mk e Sk');
disp('=====');

% =====
% 7a. Parte - Fim da iteraçao
% Total de pixels que trocaram de classe deve ser menor que 5%
% =====
disp('7a. Parte');
disp('Fim da Iteraçao');
GravaImagem(IMG,w, NInt);
z{NInt} = w;
TROCA = TrocaClasse(size(IMG.Imagem,1),w,NInt);
intTROCA = [intTROCA TROCA];
NInt=NInt+1;
end % Fim do WHILE
% Fim da Iteraçao

```



```

matCONFUSAO = geraMatERRO(IMG.Classe, prod(size(w)));
Fim = clock;
TempoDeProcessamento = etime(Fim,Inicio)
save Variaveis;

% =====
% FUNÇÃO INTERFACE DE CONFIGURAÇÃO DO CLASSIFICADOR
% =====
function [retTESTE,renClasses,retOPPPriori,retPORCTroca]=telaConfigPADRAO;
OPCAO      =[];
numCLASSES =[];
clc;
disp(' ');
disp('=====');
disp(' CONFIGURAÇÕES DO CLASSIFICADOR ');
disp('=====');
disp(' ');
disp('Deseja alterar as configurações');
disp('pre-definidas < S/N > ?');
disp(' ');
OPCAO = input('>> ', 's');
if (OPCAO == 'S') | (OPCAO == 's')
    clc;
    disp('=====');
    disp(' Digite a nova configuração ');
    disp('=====');
    disp('Rodar em modo teste (S/N)? ');
    OPCAO = input('>> ', 's');
    if (OPCAO == 'S') | (OPCAO == 's')
        teste = true;
        disp(' Classes disponiveis: ');
        disp(' ');
        disp('-> Milho cultivado minimo');
        disp('-> Milho cultivado convencional');
        disp('-> Soja plantio direto);
        disp('-> Soja cultivado minimo');
        disp('-> Soja cultivado convencional');
        disp('-> Floresta');
        teste = true;
        disp('Confirme as amostras a serem carregadas:');
        OPCAO = input('-Milho cultivado minimo(Corn Min) (S/N): ', 's');
        if (OPCAO == 'S') | (OPCAO == 's')
            numCLASSES = [numCLASSES 2];
        end
        OPCAO = input('-Milho cultivado convencional(Corn) (S/N): ', 's');
        if (OPCAO == 'S') | (OPCAO == 's')
            numCLASSES = [numCLASSES 3];
        end
        OPCAO = input('-Soja plantio direto(Soy Notill) (S/N): ', 's');

        if (OPCAO == 'S') | (OPCAO == 's')
            numCLASSES = [numCLASSES 4];
        end
        OPCAO = input('-Soja cultivado minimo(Soy Min) (S/N): ', 's');
        if (OPCAO == 'S') | (OPCAO == 's')
            numCLASSES = [numCLASSES 5];
        end
    end
end

```

```

end
OPCAO = input('-Soja cultivo convencional(Soy Clean) (S/N): ','s');
if (OPCAO == 'S') | (OPCAO == 's')
    numCLASSES = [numCLASSES 6];
end
OPCAO = input('-Floresta(Woods) (S/N): ','s');
if (OPCAO == 'S') | (OPCAO == 's')
    numCLASSES = [numCLASSES 7];
end
else
    teste = false;
end
clc;
disp('=====');
disp(' OPÇÃO DE CLASSIFICAÇÃO ');
disp('=====');
OPCAO = input('(C)om ou (S)em Estimativa Prob. Priori ? ','s');
if (OPCAO == 'C') | (OPCAO == 'c')
    opPPriori = true;
else
    opPPriori = false;
end
clc;
disp('=====');
disp('          CONDIÇÃO DE CLASSIFICAÇÃO          ');
disp('=====');
disp('Qtde(%)de pixels que trocaram de classe');
disp('Padrao = 0.05');
disp(' ');
disp('Nova porcentagem (S/N)?');
disp(' ');
OPCAO = input('>> ','s');
if (OPCAO == 'S') | (OPCAO == 's')
    porcTROCA = input('Digite o novo valor: ');
while (porcTROCA < 0) | (porcTROCA > 1)
    disp('Valor invalido!');
    disp('-----');
    disp('O valor deve estar de acordo com o intervalo abaixo:');
    disp('0 < porcentagem <= 1 ');
    porcTROCA = input('Digite novamente: ');
end
else
    porcTROCA = 0.05;
end
else
    teste = false;
    numCLASSES = 0;
    opPPriori = true;
    porcTROCA = 0.05;
end
retTESTE = teste;
reNClasses = numCLASSES;
retOPPPriori = opPPriori;
retPORCTroca = porcTROCA;

```

```

% =====
% FUNÇÃO INTERFACE DE DEFINIÇÃO DO N° BANDAS (GL)
% E DO QUI-QUADRADO DE ALGUMAS DIMENSIONALIDADES
% =====
% Atribuição do valor do Qui-quadrado(95%) para as
% dimensionalidades específicas. Para fins genéricos
% usar a própria função Qui-quadrado.
%=====
function [retGL, retQuiQuadrado] = telaBANDAS;
OPCAO      = [];
GL         = [];
QuiQuadrado = [];
clc;
disp(' ');
disp('=====');
disp('Deseja digitar o Nro. de Bandas ?');
disp('=====');
disp(' ');
disp('< S > Atribui o valor do Limiar em 5%(qui-quadrado)');
disp(' ');
disp('< N > Numero de bandas da imagem padrao');
disp('      e limiar correspondente. ');
disp(' ');
OPCAO = input('(S/N): ', 's');
if (OPCAO == 's') | (OPCAO == 'S')
    GL = 0;
    QuiQuadrado = 0;
    while GL > 40
        disp('Definindo o Qui-quadrado');
        GL = round(input('Digite o numero de bandas: '));
        if (GL >= 40) & (GL < 50)
            QuiQuadrado = 55.7585;
        elseif (GL >= 50) & (GL < 60)
            QuiQuadrado = 67.5048;
        elseif (GL >= 60) & (GL < 70)
            QuiQuadrado = 79.0819;
        elseif (GL >= 70) & (GL < 80)
            QuiQuadrado = 90.5312;
        elseif (GL >= 80) & (GL < 90)
            QuiQuadrado = 101.879;
        elseif (GL >= 90) & (GL < 100)
            QuiQuadrado = 113.145;
        elseif GL >= 100
            QuiQuadrado = 124.342;
        else
            disp('Valor Invalido!');
            disp('-----');
        end
    end
end
% Definindo o N° de Bandas espectrais
while (GL ~= 40) & (GL ~= 190)
    disp('N° de Bandas entre 10 e 190');
    GL = input('Bandas: ');
    if (GL <= 40) & (GL >= 190)
        disp('Valor Invalido!');
        disp('-----');
    end
end

```

```

        else
            GL = 190;
            QuiQuadrado = 124.342;
        end
retGL          = GL;
retQuiQuadrado = QuiQuadrado;

%=====
% FUNÇÃO PARA O PREPARO DA ÁREA EXPERIMENTAL DA IMAGEM
%=====
% Recorta uma area experimental da imagem original
% Compoee amostras com bandas distribuidas pela imagem
%=====
% PARAMETROS
% -----
% LINHA = [linINI : linFINAL]
% Intervalo de linha o que sera retirado da imagem original
% - linINI   = linha inicial a ser retirada da imagem total
% - linFINAL = linha final a ser retirada da imagem total
% COLUNA = [colINI : colFINAL]
% Intervalo de colunas que sera retirado da imagem original
% - colINI   = coluna inicial a ser retirada da imagem total
% - colFINAL = coluna final a ser retirada da imagem total
% BANDAS = [bandINI : bandaFINAL]
% Intervalo de bandas que farao parte da amostra
% - colINI   = coluna inicial a ser retirada da imagem total
% - colFINAL = coluna final a ser retirada da imagem total
% RETbandas : inteiro
% Numero de bandas a serem extraidas da amostra
%=====
function retIMG = loadIMG(LINHA, COLUNA, BANDAS, RETbandas)
% CARREGA IMAGEM
[Reduzida_92av190,Nlin,Ncol,Nbandas] = ...
loadimage('92av190reduzida.lan',LINHA,COLUNA,BANDAS,'unsign');
% COMPOE AMOSTRAS COM BANDAS DISTRIBUIDAS EQUITATIVAMENTE POR TODA IMAGEM
switch RETbandas
    case 20,
        % << 20 BANDAS >>
        t = 9;
    case 40,
        % << 40 BANDAS >>
        t = 4;
    case 60,
        % << 60 BANDAS >>
        t = 3;
    case 80,
        % << 80 BANDAS >>
        t = 2;
    case 100,
        % << 100 BANDAS >>
        t = 1;
    otherwise
        % << >100 BANDAS >>
        t = 1;
end
Reduzida_92av190 = [Reduzida_92av190(:,1:t:190)];

```

```

Reduzida_92av190 = Reduzida_92av190(:,1:RETbandas);
save Reduzida_92av190 Reduzida_92av190;
retIMG = Reduzida_92av190;
return

% =====
% FUNÇÃO DE CARGA DA IMAGEM A SER CLASSIFICADA
% =====
function retIMG = CarregaImagem(bandas, teste)
% =====
% Estrutura da Imagem
% =====
% ATRIBUTOS
% -----
% Nome - Indica o nome da imagem.
%   -Tipo: Alfanumerico
% Imagem[] - Matriz da Imagem a classificar (imagem teste e/ou tematica)
%   -Tipo: Matriz
% Classe[] - Indica a classe a qual o Pixel pertence.
%   -Tipo: Matriz
% ATest = Imagem
% VamTest = IMG
limiteBANDAS = [];
IMG = struct( ...
    'Nome', '', ...
    'Imagem', [], ...
    'Classe', [] ...
);
% Estrutura da Imagem
NomeIMG = []; % Nome da Imagem
NomeARQ = []; % Nome do Arquivo a ser carregada a imagem
imgTEST = []; % Armazena a imagem a ser classificada
% INTERFACE PARA CARREGAR O ARQUIVO
NomeIMG = 'Imagem Geral';
switch bandas
    case 190,
        if teste
            % FORMA RESUMIDA - APENAS PARA TESTE
            % Carrega as imagens de teste
            load test_corn_notill_190; % w{1}
            load test_soy_notill_190; % w{4}
            load test_woods_190;      % w{7}
            % Concatena todas as imagens de teste
            imgTEST = [test_corn_notill_190; ... % w{1}
                test_soy_notill_190; ... % w{4}
                test_woods_190];          % w{7}
        else
            % FORMA COMPLETA
            % Carrega as imagens de teste
            load test_corn_notill_190; % w{1}
            load test_corn_min_190;   % w{2}
            load test_corn_190;       % w{3}
            load test_soy_notill_190; % w{4}
            load test_soy_min_190;    % w{5}
            load test_soy_clean_190;  % w{6}
            load test_woods_190;      % w{7}
        end
    end
end

```

```

    % Concatena todas as imagens de teste
    imgTEST=[test_corn_notill_190; ... % w{1}
            test_corn_min_190; ...   % w{2}
            test_corn_190; ...       % w{3}
            test_soy_notill_190; ... % w{4}
            test_soy_min_190; ...    % w{5}
            test_soy_clean_190; ...  % w{6}
            test_woods_190];        % w{7}
end
case 40,
if teste
    % FORMA RESUMIDA - APENAS PARA TESTE
    % Carrega as imagens de teste
    load test_corn_notill_300; % w{1}
    load test_corn_min_300;   % w{2}
    load test_corn_300;       % w{3}
    load test_soy_notill_300; % w{4}
    load test_soy_min_300;   % w{5}
    load test_soy_clean_300; % w{6}
    load test_woods_300;     % w{7}
    % Concatena todas as imagens de teste
    imgTEST=[test_corn_notill_300; ... % w{1}
            test_corn_min_300; ...   % w{2}
            test_corn_300; ...       % w{3}
            test_soy_notill_300; ... % w{4}
            test_soy_min_300; ...    % w{5}
            test_soy_clean_300; ...  % w{6}
            test_woods_300];        % w{7}
else
    % FORMA COMPLETA
    % Carrega as imagens de teste
    load test_corn_notill_300; % w{1}
    load test_corn_min_300;   % w{2}
    load test_corn_300;       % w{3}
    load test_soy_notill_300; % w{4}
    load test_soy_min_300;   % w{5}
    load test_soy_clean_300; % w{6}
    load test_woods_300;     % w{7}
    % Concatena todas as imagens de teste
    imgTEST = [test_corn_notill_300; ... % w{1}
              test_corn_min_300; ...   % w{2}
              test_corn_300; ...       % w{3}
              test_soy_notill_300; ... % w{4}
              test_soy_min_300; ...    % w{5}
              test_soy_clean_300; ...  % w{6}
              test_woods_300];        % w{7}
end
end
% Salva todas e um arquivo chamado imgTEST.mat
save imgTEST imgTEST;
loadingTEST;
if bandas > 110
    limiteBANDAS = 110;
else
    limiteBANDAS = bandas;
end
end

```

```

% Carrega para estrutura a imagem e zera a matriz Classes
IMG.Nome = NomeIMG;
IMG.Imagem = imgTEST(:,1:limiteBANDAS);
IMG.Classe = imgTEST(:,1) - imgTEST(:,1);
retIMG = IMG;

%=====
% FUNÇÃO QUE PREPARA AS AMOSTRAS (TREINAMENTO E TESTE)
% COM INTERVALO UNIFORME DE BANDAS SOBRE TODO ESPECTRO
%=====
function loadAmostras190(bandas, NumPixelsTR, NumPixelsTE)
T_BANDAS = 190;

% Carrega Amostras de Teste
load test_corn_min_190;
load test_corn_notill_190;
load test_soy_clean_190;
load test_soy_min_190;
load test_soy_notill_190;
load test_woods_190;

% Carrega Amostras de Treinamento
load train_corn_min_190;
load train_corn_notill_190;
load train_soy_clean_190;
load train_soy_min_190;
load train_soy_notill_190;
load train_woods_190;

% Padroniza a Nomenclatura
test_corn_min      = test_corn_min_190;
test_corn_notill   = test_corn_notill_190;
test_soy_clean     = test_soy_clean_190;
test_soy_min       = test_soy_min_190;
test_soy_notill    = test_soy_notill_190;
test_woods         = test_woods_190;
train_corn_min     = train_corn_min_190;
train_corn_notill  = train_corn_notill_190;
train_soy_clean    = train_soy_clean_190;
train_soy_min      = train_soy_min_190;
train_soy_notill   = train_soy_notill_190;
train_woods        = train_woods_190;

% Verifica o Tamanho Total de cada amostra
ttr_cm = size(train_corn_min,1);
ttr_cn = size(train_corn_notill,1);
ttr_s  = size(train_soy_clean,1);
ttr_sm = size(train_soy_min,1);
ttr_sn = size(train_soy_notill,1);
ttr_w  = size(train_woods,1);
tmte_cm = size(test_corn_min,1);
tmte_cn = size(test_corn_notill,1);
tmte_s  = size(test_soy_clean,1);
tmte_sm = size(test_soy_min,1);
tmte_sn = size(test_soy_notill,1);
tmte_w  = size(test_woods,1);

```

```

% Seleção alternada de pixels para gerar as amostras de TREINAMENTO
sttr_cm = div(ttr_cm, NumPixelsTR);
sttr_cn = div(ttr_cn, NumPixelsTR);
sttr_s  = div(ttr_s, NumPixelsTR);
sttr_sm = div(ttr_sm, NumPixelsTR);
sttr_sn = div(ttr_sn, NumPixelsTR);
sttr_w  = div(ttr_w, NumPixelsTR);

% Seleção alternada de pixels para gerar as amostras de TESTE
stte_cm = div(tmte_cm, NumPixelsTE);
stte_cn = div(tmte_cn, NumPixelsTE);
stte_s  = div(tmte_s, NumPixelsTE);
stte_sm = div(tmte_sm, NumPixelsTE);
stte_sn = div(tmte_sn, NumPixelsTE);
stte_w  = div(tmte_w, NumPixelsTE);

% Ajusta o tamanho das amostras de treinamento para
% o numero de pixels solicitado
if NumPixelsTR == 0
    nptr_cm = sttr_cm;
    nptr_cn = sttr_cn;
    nptr_s  = sttr_s;
    nptr_sm = sttr_sm;
    nptr_sn = sttr_sn;
    nptr_w  = sttr_w;
else
    nptr_cm = NumPixelsTR;
    nptr_cn = NumPixelsTR;
    nptr_s  = NumPixelsTR;
    nptr_sm = NumPixelsTR;
    nptr_sn = NumPixelsTR;
    nptr_w  = NumPixelsTR;
end
if NumPixelsTE == 0
    npte_cm = tmte_cm;
    npte_cn = tmte_cn;
    npte_s  = tmte_s;
    npte_sm = tmte_sm;
    npte_sn = tmte_sn;
    npte_w  = tmte_w;
else
    npte_cm = NumPixelsTE;
    npte_cn = NumPixelsTE;
    npte_s  = NumPixelsTE;
    npte_sm = NumPixelsTE;
    npte_sn = NumPixelsTE;
    npte_w  = NumPixelsTE;
end
% Amostras de treinamento
if (sttr_cm == 0) | (sttr_cm == Inf)
    sttr_cm = 1;
    nptr_cm = ttr_cm;
end
if (sttr_cn == 0) | (sttr_cn == Inf)
    sttr_cn = 1;
    nptr_cn = ttr_cn;
end

```



```

if (sttr_s == 0) | (sttr_s == Inf)
    sttr_s = 1;
    nptr_s = ttr_s;
end
if (sttr_sm == 0) | (sttr_sm == Inf)
    sttr_sm = 1;
    nptr_sm = ttr_sm;
end
if (sttr_sn == 0) | (sttr_sn == Inf)
    sttr_sn = 1;
    nptr_sn = ttr_sn;
end
if (sttr_w == 0) | (sttr_w == Inf)
    sttr_w = 1;
    nptr_w = ttr_w;
end
% Amostras de teste
if (stte_cm == 0) | (stte_cm == Inf)
    stte_cm = 1;
    npte_cm = tmte_cm;
end
if (stte_cn == 0) | (stte_cn == Inf)
    stte_cn = 1;
    npte_cn = tmte_cn;
end
if (stte_s == 0) | (stte_s == Inf)
    stte_s = 1;
    npte_s = tmte_s;
end
if (stte_sm == 0) | (stte_sm == Inf)
    stte_sm = 1;
    npte_sm = tmte_sm;
end
if (stte_sn == 0) | (stte_sn == Inf)
    stte_sn = 1;
    npte_sn = tmte_sn;
end
if (stte_w == 0) | (stte_w == Inf)
    stte_w = 1;
    npte_w = tmte_w;
end

% Define o intervalo entre bandas
b = div(T_BANDAS,bandas);

%***** ORGANIZA AMOSTRAS *****
% Seleciona Bandas(intervalos iguais),cobrindo todo espectro eletromagnetico
test_corn_notill = [test_corn_notill(1:stte_cn:tmte_cn,1:b:T_BANDAS)];
test_corn_min = [test_corn_min(1:stte_cm:tmte_cm,1:b:T_BANDAS)];
test_soy_notill = [test_soy_notill(1:stte_sn:tmte_sn,1:b:T_BANDAS)];
test_soy_min = [test_soy_min(1:stte_sm:tmte_sm,1:b:T_BANDAS)];
test_soy_clean = [test_soy_clean(1:stte_s:tmte_s,1:b:T_BANDAS)];
test_woods = [test_woods(1:stte_w:tmte_w,1:b:T_BANDAS)];
train_corn_notill = [train_corn_notill(1:sttr_cn:ttr_cn,1:b:T_BANDAS)];
train_corn_min = [train_corn_min(1:sttr_cm:ttr_cm,1:b:T_BANDAS)];
train_soy_notill = [train_soy_notill(1:sttr_sn:ttr_sn,1:b:T_BANDAS)];
train_soy_min = [train_soy_min(1:sttr_sm:ttr_sm,1:b:T_BANDAS)];

```

```

train_soy_clean = [train_soy_clean(1:sttr_s:ttr_s,1:b:T_BANDAS)];
train_woods     = [train_woods(1:sttr_w:ttr_w,1:b:T_BANDAS)];

% Seleciona Amostras alternadas de acordo com o
% n° total de pixels pre-definidos para cada Classe
test_corn_notill = test_corn_notill(1:npte_cn,1:bandas);
test_corn_min    = test_corn_min(1:npte_cm,1:bandas);
test_soy_notill  = test_soy_notill(1:npte_sn,1:bandas);
test_soy_min     = test_soy_min(1:npte_sm,1:bandas);
test_soy_clean   = test_soy_clean(1:npte_s,1:bandas);
test_woods       = test_woods(1:npte_w,1:bandas);
train_corn_notill = train_corn_notill(1:nptr_cn,1:bandas);
train_corn_min    = train_corn_min(1:nptr_cm,1:bandas);
train_soy_notill  = train_soy_notill(1:nptr_sn,1:bandas);
train_soy_min     = train_soy_min(1:nptr_sm,1:bandas);
train_soy_clean   = train_soy_clean(1:nptr_s,1:bandas);
train_woods       = train_woods(1:nptr_w,1:bandas);

% Salva em Arquivo as Amostras de Teste
save test_corn_notill test_corn_notill;
save test_corn_min test_corn_min;
save test_soy_notill test_soy_notill;
save test_soy_min test_soy_min;
save test_soy_clean test_soy_clean;
save test_woods test_woods;

% Salva em Arquivo as Amostras de treinamento
save train_corn_notill train_corn_notill;
save train_corn_min train_corn_min;
save train_soy_notill train_soy_notill;
save train_soy_min train_soy_min;
save train_soy_clean train_soy_clean;
save train_woods train_woods;
return

% =====
% FUNÇÃO QUE CALCULA OS PARAMETROS MK, SK
% E PROB. A PRIORI PARA TODAS AS CLASSES
% =====
function retW = calculoEst(Imagem, w)
for i=1 : prod(size(w))
    NLinRot = size(w{i}.Rotulados,1);
    NLinPERT = size(w{i}.Pertence,1);

    % Junção de Amostras Rotuladas c/ Semi-Rotuladas, e dos Pesos
    RotSR = [w{i}.Rotulados; ...
             w{i}.Pertence]; % Junção dos Pixels R e SR
    PesoRotSR = [ones(NLinRot,1); ...
                w{i}.Peso(1:NLinPERT)]; % Junção dos Pesos
    % Calculo da Mk, Sk e PROB. PRIORI
    w{i}.Mk = mediaEst(RotSR, PesoRotSR);
    w{i}.Sk = covEst(RotSR, w{i}.Mk, PesoRotSR);
    w{i}.PPriori = ProbPriori(Imagem, RotSR);
end
retW = w;

```

```

% =====
% FUNÇÃO QUE ESTIMA A MÉDIA ( $M_k$ ) DA CLASSE  $w\{i\}$ 
% =====
% SINTAXE: mediaEst(Amostra, Peso);
% -----
% Amostra(X)=Amostra de Rotulados e Semi-Rotulados
% Peso      =Pesos dos Pixels da classe  $W\{i\}$ 
% =====
function retMEDIA = mediaEst(X,Peso)
peso_X      = [];
pesoMEDIA   = [];
NLin = size(X,1);
for i = 1 : NLin
    peso_X = [peso_X; Peso(i,1) * X(i,:)];
end
pesoMEDIA = sum(peso_X) / sum(Peso);
retMEDIA = pesoMEDIA;
return

% =====
% FUNÇÃO QUE CALCULA A MATRIZ COVARIÂNCIA ( $S_k$ )
% DA CLASSE  $w\{i\}$ 
% =====
% SINTAXE: covEst(x, Mk, Peso);
% -----
% PARAMETROS
% -----
% x      = Amostra de Rotulados e Semi-Rotulados
% Mk     = Media da Amostra.
% Peso= Pesos dos Pixels da classe  $W\{i\}$ 
% =====
function xy = covEst(x, varargin)
if nargin==0, error('Not enough input arguments.');
```

```

Mk = varargin{1};
Peso = varargin{2};
[m,n] = size(x);
if m==1, % Handle special case
    xy = 0;
else
    auxMk = repmat(Mk,m,1); %sum(x)/m,m,1); % Remove mean
    xc = x - auxMk;
    auxXC = [];
    for i = 1 : m
        auxXC = [auxXC; Peso(i) * xc(i,:)];
    end
    xc = auxXC;
    xy = xc' * xc / (sum(Peso)-1);
end

% =====
% FUNÇÃO PARA CÁLCULO DO LIMIAR - Threshold (Ti)
% =====
% SINTAXE: Ti = funcaoLimiar(QuiQuadrado, Sk, PPriori);
% -----
% Qui-quadrado- Grau Confiança=95% / Nivel significancia=5%
% Sk          - Matriz Covariância da Classe W(i)
% PPriori     - Prob. Priori da Classe W(i)
% =====
function retTi = funcaoLimiar(QuiQuadrado, Sk, PPriori)
if size(PPriori,1) == 0
    retTi = - 2*QuiQuadrado - (log(det(Sk)));
else
    retTi = - 2*QuiQuadrado - (log(det(Sk))) + 2*log(PPriori);
end
return

% =====
% FUNÇÃO CONTROLA PIXELS EXCLUÍDOS PELO LIMIAR
% =====
% Retorna o Nr de pixels excluidos pelo Limiar,
% relativos a cada uma das classes consideradas
% =====
function retPixels = verificaTi(ImagemClass, Ti, CFG)
NumIteracoes = [];
NumPixels = [];
Pixels.QtdIT = [];
Pixels.Posicao = [];
Pixels.Classe = [];
NumPixels = size(ImagemClass,1);
NumIteracoes = size(ImagemClass,2);
Pixels.QtdIT(1:NumIteracoes) = 0;
for Xpos=1 : NumPixels
    disp(sprintf('%d',Xpos));
    for ITERACAO = 1 : NumIteracoes
        Xclasse = ImagemClass(Xpos, ITERACAO);
        ACHEI = false;
        i = 1;

```

```

while (~ACHEI) & (i < size(Ti.Posicao,1))
    if Ti.Posicao(i) == Xpos
        Pixels.QtdIT(1,ITERACAO) = Pixels.QtdIT(1,ITERACAO) + 1;
        Pixels.Posicao = [Pixels.Posicao; Xpos];
        Pixels.Classe = [Pixels.Classe; Xclasse];
        ACHEI=true;
        disp('Sim!');
    else
        i=i+1;
        if i >= size(Ti.Posicao,1)
            i;
        end
    end
end
end
end
retPixels = Pixels;
return

%=====
% FUNÇÃO DECISÃO DO CLASSIFICADOR MAXVER ADAPTATIVO
%=====
% X          = Pixel para calcular do valor de MAXVER
% ATrain    = Matriz da classe candidata
% PPriori   = Prob. a Priori da classe candidata
%=====
% CALCULA O MAXVER
% -----
function retMaxVer = fdecisao(X, Mk, Sk, PPriori)

    MaxVer = [];
    MaxVer = -logdet(Sk) - (X - Mk) * inv(Sk) * (X - Mk)' + 2 * log(PPriori);
    retMaxVer = MaxVer;

return

% =====
% FUNÇÃO CLASSIFICADOR
% =====
%FLUXO DE CLASSIFICAÇÃO(DINÂMICO)
function [NumClassificados,imgResult,wResult]=...
    Classificador(IMG,w,ITERACAO,opPPriori)
matSR      = []; % Matriz auxiliar para formar a matriz de SR
matMaxVer  = []; % Matriz auxiliar para formar a matriz da MaxVer
matSomeMV  = []; % Matriz auxiliar para formar a matriz do SomaMV
Classe     = []; % Salva a nova classe do pixel X
Xmaxver    = []; % MaxVer de X na classe vencedora
Xmaxver_2  = []; % Auxiliar
XsomaMV    = []; % Soma do MaxVer de X em todas as classes

% Para controle do numero de Pixels que trocaram de classes
NumP=0;
NumCLASSIF = []; % Numero de pixels classificados
salvaclasse = [];
auxITERACAO = [];

```

```

img.NLin      = 0;
img.NCol      = 0;
for i = 1 : prod(size(w))
    if ITERACAO ~= 0
        w{i}.qtdCLASSIF(ITERACAO) = 0;
        w{i}.qtdTROCA(ITERACAO)   = 0;
    end
    w{i}.SemiRot   = []; % OBS: Remove todos os elementos dos campos da
    w{i}.MaxVer    = []; % estrutura W{i} que serao atualizados, pois esses
    w{i}.Pertence  = []; % campos devem ser preenchidos de acordo com a nova
    w{i}.SomaMV    = []; % classificaçao
    w{i}.Posicao    = [];
    w{i}.Peso      = [];
end
[img.NLin, img.NCol] = size(IMG.Imagem);
qtdCLASSIF = 0; % Numero de pixels classificados
for k=1 : img.NLin
    X = IMG.Imagem(k,:);
    Xmaxver = 0;
    Xmaxver_2 = 0;
    XsomaMV = 0;
    Classe = 0; % Posiçao anterior ao "i". Calcula e compara a MaxVer
    [w, Classe, Xmaxver, XsomaMV, Xmaxver_2] = DecideClasse(X, w, opPPriori);

    % Decide a classe vencedora do pixel "IMG.Imagem(k,m)"
    % Conta quantos Pixels mudaram de Classe
    if ITERACAO == 1
        auxITERACAO = 1;
    else
        auxITERACAO = ITERACAO - 1;
    end
    if Classe ~= IMG.Classe(k,auxITERACAO)
        NumP=NumP+1;
        % Contabiliza os Pixels que mudaram de classe
        if IMG.Classe(k,auxITERACAO) > 0
            IND = IMG.Classe(k,auxITERACAO); % IND = Classe anterior
            w{IND}.qtdTROCA(ITERACAO) = w{IND}.qtdTROCA(ITERACAO) + 1;
        end
    end
end

% Salva a nova classe do Pixel na matriz de classes
IMG.Classe(k) = Classe;
salvaclasse(k) = Classe;

% Salva Pixel e seu MAXVER da Classe vencedora
w{Classe}.SemiRot = [w{Classe}.SemiRot; X];
w{Classe}.MaxVer = [w{Classe}.MaxVer; [Xmaxver Xmaxver_2]];
w{Classe}.Posicao = [w{Classe}.Posicao; k]; % Salva a posiçao do Pixel
                                           % na Imagem Geral na classe à
                                           % qual ele pertence

w{Classe}.SomaMV = XsomaMV;
w{Classe}.Peso = [w{Classe}.Peso; Xmaxver/XsomaMV]; % Atribui Peso ao
                                                    % Pixel X

if ITERACAO ~= 0
    w{Classe}.qtdCLASSIF(ITERACAO) = w{Classe}.qtdCLASSIF(ITERACAO) + 1;
end

```

```

    % As matrizes W(Classe).SemiRot e W(Classe).MaxVer ficam c/ dimensao 1xN
    % O nr de colunas deve ser igual aquele da matriz W(Classe).Rotulados,
    % pois sera necessario que o Peso tambem possua o mesmo numero de colunas
    % para o calculo da Mk e Sk
end
IMG.Classe(:,ITERACAO) = salvaclasse;
NumClassificados = NumP;
imgResult = IMG;
wResult = w;

%=====
% FUNÇÃO CLASSIFICADOR_2
%=====
%FLUXO DE CLASSIFICAÇÃO(ESTÁTICO)
function [retCLASSIF, retW, retVETPeso, retV, retNumTROCA]=
    Classificador2(Imagem,ImagemClass, SemiRot, Peso, ITERACAO, CTRL,
    NumTROCA, v)
vetClassif = [];
vetPeso    = {};
Mk         = {};
Sk         = {};
Rotulados  = {};
Xpeso      = [];
v{ITERACAO}= {};

% Reinicializando o vetor w{i}
for i=1 : size(SemiRot,2)
    w{i}=[];
    vetPeso{i} = [];
    v{ITERACAO}{i} = struVerifica;
end
% Carrega amostras treinamento
load train_corn_notill;
load train_corn_min;
load train_corn;
load train_soy_notill;
load train_soy_min;
load train_soy_clean;
load train_woods;
if ITERACAO == 1 % NA PRIMEIRA ITERACAO CALCULA A Mk e Sk Sem Ponderação

    % CALCULO DO VETOR DE MEDIAS C/ VALORES PESOS IDENTICOS
    Mk_cn = mean(train_corn_notill);
    Mk_cm = mean(train_corn_min );
    Mk_c  = mean(train_corn      );
    Mk_sn = mean(train_soy_notill );
    Mk_sm = mean(train_soy_min   );
    Mk_s  = mean(train_soy_clean );
    Mk_w  = mean(train_woods     );

    % CALCULO DA MATRIZ COVARIANCIA C/ VALORES IDENTICOS
    Sk_cn = cov(train_corn_notill);
    Sk_cm = cov(train_corn_min   );
    Sk_c  = cov(train_corn       );
    Sk_sn = cov(train_soy_notill );
    Sk_sm = cov(train_soy_min    );

```

```

Sk_s = cov(train_soy_clean );
Sk_w = cov(train_woods );
else
% JUNCAO DOS PIXELS ROTULADOS COM OS SEMI-ROTULADOS
RotSR_cn = [train_corn_notill; SemiRot{1}];
RotSR_cm = [train_corn_min ; SemiRot{2}];
RotSR_c = [train_corn ; SemiRot{3}];
RotSR_sn = [train_soy_notill ; SemiRot{4}];
RotSR_sm = [train_soy_min ; SemiRot{5}];
RotSR_s = [train_soy_clean ; SemiRot{6}];
RotSR_w = [train_woods ; SemiRot{7}];

% JUNCAO DOS PESOS DOS PIXELS ROTULADOS COM OS PESOS DOS SEMI-ROTULADOS
PesoRotSR_cn = [ones(size(train_corn_notill,1),1); Peso{1}];
PesoRotSR_cm = [ones(size(train_corn_min,1),1) ; Peso{2}];
PesoRotSR_sn = [ones(size(train_soy_notill,1),1) ; Peso{4}];
PesoRotSR_sm = [ones(size(train_soy_min,1),1) ; Peso{5}];
PesoRotSR_s = [ones(size(train_soy_clean,1),1) ; Peso{6}];
PesoRotSR_w = [ones(size(train_woods,1),1) ; Peso{7}];

% CALCULO DO VETOR DE MEDIAS COM VALORES PESOS ESTIMADOS
Mk_cn = mediaEst(train_corn_notill, PesoRotSR_cn);
Mk_cm = mediaEst(train_corn_min, PesoRotSR_cm);
Mk_c = mediaEst(train_corn, PesoRotSR_c );
Mk_sn = mediaEst(train_soy_notill, PesoRotSR_sn);
Mk_sm = mediaEst(train_soy_min, PesoRotSR_sm);
Mk_s = mediaEst(train_soy_clean, PesoRotSR_s );
Mk_w = mediaEst(train_woods, PesoRotSR_w );

% CALCULO DA MATRIZ COVARIANCIA COM VALORES ESTIMADOS
Sk_cn = covEst(RotSR_cn, Mk_cn, PesoRotSR_cn);
Sk_cm = covEst(RotSR_cm, Mk_cm, PesoRotSR_cm);
Sk_c = covEst(RotSR_c, Mk_c, PesoRotSR_c );
Sk_sn = covEst(RotSR_sn, Mk_sn, PesoRotSR_sn);
Sk_sm = covEst(RotSR_sm, Mk_sm, PesoRotSR_sm);
Sk_s = covEst(RotSR_s, Mk_s, PesoRotSR_s );
Sk_w = covEst(RotSR_w, Mk_w , PesoRotSR_w );
end
NumTROCA{ITERACAO} = 0; % Verifica o numero de trocas na iteracao

for Xpos=1 : size(Imagem,1)
X = Imagem(Xpos,:); % Pega o Pixel X na posicao Xpos na imagem geral
% FUNCOES DECISAO (Calcula Max. Ver. Classes)
MV_cn = fdecisao(X, Mk_cn, Sk_cn);
MV_cm = fdecisao(X, Mk_cm, Sk_cm);
MV_c = fdecisao(X, Mk_c, Sk_c );
MV_sn = fdecisao(X, Mk_sn, Sk_sn);
MV_sm = fdecisao(X, Mk_sm, Sk_sm);
MV_s = fdecisao(X, Mk_s, Sk_s );
MV_w = fdecisao(X, Mk_w, Sk_w );
% DEFINE CLASSIFICACAO(Classse vencedora = pixel > Max Ver)
XmaxverClasses = [MV_cn MV_cm MV_c MV_sn MV_sm MV_s MV_w];
[Xmaxver, Xclasse] = max(XmaxverClasses);
% DEFINE O PESO DO PIXEL X
Xpeso = PesoEst(XmaxverClasses, Xclasse);
% SALVANDO OS DADOS
vetClassif(Xpos,1)= Xclasse; % Salva classe vencedora

```



```

vetPeso{Xclasse} = [vetPeso{Xclasse}; Xpeso]; % Salva o peso de X
w{Xclasse}       = [w{Xclasse}; X]; % Salva o pixel X na classe vencedora

% VERIFICA QUANTOS PIXELS TROCARAM DE CLASSE
if ITERACAO >= 2
    if ImagemClass(Xpos, ITERACAO-1) ~= Xclasse
        NumTROCA{ITERACAO} = NumTROCA{ITERACAO} + 1;
    end
end
v= verificaMatErro(Xpos, Xclasse, v, CTRL, ITERACAO); % Verificar se
% a classificacao esta ocorrendo corretamente

end
% ATUALIZACAO DOS DADOS
% -----
ImagemClass(:,ITERACAO) = vetClassif;
Peso                    = vetPeso;

% RETORNO DA FUNCAO
% -----
retCLASSIF = ImagemClass;
retW       = w;
retVETPeso = Peso;
retV

%=====
% FUNÇÃO MONITORA A TROCA DE PIXELS ENTRE CLASSES
%=====
function [retPORC, retCLASSE] = TrocaClasse(ImagemClass, CFG )
NClasses = CFG.NumClass;
ITERACAO = CFG.ITERACAO;
PORC     = [];
porcCLASSE = [];
ANT      = [];
ATUAL    = [];
vetPORC  = [];
mediaIT  = [];
t.qtdANT = {};
t.qtdATUAL = {};
t.Entrada = {};
t.Saida = {};
t.Fica = {};
for i=1 : NClasses
    t.qtdANT{i} = 0;
    t.qtdATUAL{i}= 0;
    t.Entrada{i} = 0;
    t.Saida{i} = 0;
    t.Fica{i} = 0;
end
if ITERACAO < 2
    PORC = 1;
Else

% Define a iteraçao Anterior e a Atual
ANT = ITERACAO - 1;
ATUAL = ITERACAO;

```

```

% Contabiliza os Pixels que trocaram de Classe
for p=1 : size(ImagemClass,1)
    classeANT = ImagemClass(p,ANT);
    classeATUAL = ImagemClass(p,ATUAL);
    if (classeANT ~= 0)&(classeANT~=NClasses+1)
        if classeANT ~= classeATUAL
            t.Saida{classeANT} = t.Saida{classeANT} + 1;
            t.Entrada{classeANT} = t.Entrada{classeANT} + 1;
        else
            t.Fica{classeANT} = t.Fica{classeANT} + 1;
        end
        t.qtdANT{classeANT} = t.qtdANT{classeANT} + 1;
        t.qtdATUAL{classeATUAL} = t.qtdATUAL{classeATUAL} + 1;
    end
end
% Porcentagem de Trocas na Iteração Corrente
for i=1 : NClasses
    if t.qtdANT{i} == 0
        vetPORC = [vetPORC 0];
    else
        vetPORC = [vetPORC t.Saida{i}/t.qtdANT{i}];
    end
    mediaIT = [mediaIT; ];
end
[PORC, porcCLASSE] = max(vetPORC);
retPORC = PORC;
retCLASSE = porcCLASSE;
return

% =====
% FUNÇÃO MONITORA A CLASSIFICAÇÃO DE CADA PIXEL
% =====
% Utilizada na CERTIFICAÇÃO da Classificação
% =====
function [retV, retVetClassERRO] = verificaMatErro(Xpos, Xclasse, v, CTRL,
vetClassERRO, CFG)
NumClass = CFG.NumClass;
ITERACAO = CFG.ITERACAO;
% Verifica a qual classe o pixels X deveria pertencer
if Xclasse == 0
    vetClassERRO(Xpos,1) = Xclasse;
elseif (Xpos >= CTRL.Teste.Inicio{Xclasse})&(Xpos <=CTRL.Teste.Fim{Xclasse})
    % Grava o numero do pixel e a classe para informar que esta correto
    v{ITERACAO}{Xclasse}.Certo = [v{ITERACAO}{Xclasse}.Certo; Xpos ];
    v{ITERACAO}{Xclasse}.Classe = [v{ITERACAO}{Xclasse}.Classe; Xclasse];
    vetClassERRO(Xpos,1) = Xclasse;
else
    % Grava o numero do pixel e a classe para informar que esta correto
    v{ITERACAO}{Xclasse}.Errado = [v{ITERACAO}{Xclasse}.Errado; Xpos];
    vetClassERRO(Xpos,1)=Xclasse+NumClass; %Indica pixel cadastrado errado.

    % Verifica à qual classe X deveria pertencer
    for m = 1 : size(CTRL.Teste.Inicio,2) % "m" percorre as classes
        if (Xpos >= CTRL.Teste.Inicio{m}) & (Xpos <= CTRL.Teste.Fim{m})
            v{ITERACAO}{Xclasse}.ClasseERRO =

```

```

        [v{ITERACAO}{Xclasse}.ClasseERRO; m];
    end
end
end
retV = v;
retVetClassERRO = vetClassERRO;
return

%=====
% FUNÇÃO P/ GERAR A MATRIZ DE ERRO EM CADA ITERAÇÃO
%=====
function retMatERRO = montaMatErro(ImagemClass, CFG, CTRL, Ti)
NumClass = CFG.NumClass;
TamTi = {};
ImagemREF = []; % Imagem de referencia
MatERRO = [];
MatERRO_1 = [];
MatERRO_2 = [];
Xpos = [];
clasTEMA = [];
classREF = [];
if size(CTRL,1) == 0 % Se CTRL == Vazio
    MatERRO = [];
else
    % Constroi parcialmente a Matriz de Erro
    MatERRO_2(1:NumClass,1:NumClass) = 0;
    ImagemREF = montaORI(CTRL,'n');
    for Xpos=1 : size(ImagemClass,1)
        % Confirma a Classificação do Pixel na Classe devida
        classREF = ImagemREF(Xpos);
        if
            ImagemClass(Xpos,1) <= NumClass clasTEMA = ImagemClass(Xpos,1);
        else
            clasTEMA = ImagemClass(Xpos,1)-NumClass;
        end
        % Constroi a Matriz de Erro Final
        % Pixels que sai de classREF e vai para clasTEMA
        MatERRO_2(classREF, clasTEMA) = MatERRO_2(classREF, clasTEMA) + 1;
    end
    % -----
    % MatERRO_1
    % -----
    % Coluna 1 = Acuracia do Produtor(%)
    % Coluna 2 = Erro de Omissao (%)
    % Coluna 3 = Nr de Pixels Referencia(nr de amostras teste p/cada classe)
    MatERRO_1(1:NumClass,1:3) = 0;
    for Yclasse=1 : NumClass
        TamTi{Yclasse}=0;
    end
    if size(Ti.Posicao,1)~= 0
        for i=1 : size(Ti.Posicao,1)
            for Yclasse=1 : NumClass
                if
                    (Ti.Posicao(i)>=CTRL.Teste.Inicio{Yclasse})&(Ti.Posicao(i)<=
                    CTRL.Teste.Fim{Yclasse})
                        TamTi{Yclasse} = TamTi{Yclasse} + 1;
                end
            end
        end
    end
end

```

```

        end
    end
end
for i=1 : size(CTRL.Teste.Inicio,2)
    MatERRO_1(i,3)=(CTRL.Teste.Fim{i}-(CTRL.Teste.Inicio{i}-1))-TamTi{i};
    MatERRO_1(i,2)=(sum(MatERRO_2(i,:))-MatERRO_2(i,i))/MatERRO_1(i,3))*100;
    MatERRO_1(i,1) = (MatERRO_2(i,i) / MatERRO_1(i,3)) * 100;
End
% -----
% Junçao das Partes
% -----
    MatERRO = [MatERRO_1 MatERRO_2];
% MatERRO
% -----
% Linha 1+NumClass = Numero de pixels da imagem classificada
% Linha 2+NumClass = Erro de Inclusao/Concessao
% Linha 3+NumClass = Acuracia do Usuario
% Linha 4+NumClass = Acuracia Geral
MatERRO(1+NumClass,:) = 0;
MatERRO(2+NumClass,:) = 0;
MatERRO(3+NumClass,:) = 0;
MatERRO(4+NumClass,:) = 0;
% Linha 1+NumClass
% -----
MatERRO(1+NumClass,3) = sum(MatERRO_1(:,3));
MatERRO(1+NumClass,4:size(MatERRO,2)) = sum(MatERRO_2);
% Linha 2+NumClass
% -----
acDIAGONAL = 0;
for i = 1 : size(MatERRO_2,2)
    acDIAGONAL = acDIAGONAL + MatERRO_2(i,i);
    MatERRO(2+NumClass,i+3) = ((MatERRO(1+NumClass, i+3) -
    MatERRO_2(i,i)) / MatERRO_1(i,3)) * 100;
    MatERRO(3+NumClass,3+i)=(MatERRO_2(i,i)/MatERRO(1+NumClass,3+i))*100;
end
MatERRO(4+NumClass,1) = (acDIAGONAL / MatERRO(1+NumClass,3)) * 100;
end
retMatERRO = MatERRO;
return

%=====
% FUNÇÃO PARA SALVAR A IMAGEM ORIGINAL
%=====
% Chamada na Construcao da Matriz de Erro
%=====
function retORI = montaORI(CTRL, Gravar)
IMGOoriginal = [];
INICIO = [];
for i=1 : size(CTRL.Teste.Inicio,2)
    INICIO = CTRL.Teste.Inicio{i};
    for j=INICIO : CTRL.Teste.Fim{i}
        IMGOoriginal = [IMGOoriginal; i];
    end
end
if (Gravar == 'S') | (Gravar == 's')

```

```

    ImagemORI = IMGOriginal;
    save ImagemORI ImagemORI;
    NPixels = size(ImagemORI,1);
    [L, C] = temaResol(NPixels);
    NomeARQ = 'Original.lan';
    salvarImagem(ImagemORI, NomeARQ, L, C);
end
retORI = IMGOriginal;
return

% =====
% FUNÇÃO PARA ESTIMAR O PESO DE CADA PIXEL CLASSIFICADO
% =====
% X = Pixel que esta sendo atribuido o peso estimado
% Xmaxver = Valor da função decisao da classe vencedora de X
% XsomaMV = Valor do MaxVer do Pixel X em todas as classes
% P = Retorna o peso estimado para o Pixel X
% VetTrain= Vetor contem a matriz de amostras das classes(Wi)
% SomaF_dec = Somatorio da função decisao de X em cada classe
% -----
% Ponderação: Rotuladas = 1;    Semi-rotuladas = [0,1]
% -----
% X = Pixel
% W = Classe
% G = Total de pixels em cada classe
% -----
%      W1      W2      W3 . . . Wi . . . Wm
%      |      |      |      |      |
%      G1(X)  G2(X)  G3(X)  Gi(X)  Gm(X)
% -----
%
%                               1/Gi(X)
% Peso =  $\frac{1/G_i(X)}{1/G_1(X) + 1/G_2(X) + \dots + 1/G_i(X) + \dots + 1/G_m(X)}$ 
%
% OBS.: O inverso, devido valores das Gi(x) serem negativos
% -----
%
%      MaxVer do Pixel X na Classe Vencedora
% Peso=  $\frac{\text{MaxVer do Pixel X na Classe Vencedora}}{\text{Soma MaxVer do Pixel X em todas as Classes}}$ 
% -----
% Acrescentar cada pixel à respectiva matriz de Wi
% =====
function RetW=EstimaPeso(w) % Calcula Peso de todos os Pixels SR de 1 classe
% Variaveis
X      = []; % Pixels a ser atribuido o peso
Xmaxver = []; % MaxVer do Pixel X
XsomaMV = []; % Soma dos MaxVer's do X em todas as classes
Xpeso   = []; % Peso do Pixel X
MATpeso = []; % Matriz auxiliar para armazenamento de pesos
% Contadores
i=[];
k=[];
m=[];
for i = 1 : prod(size(w))% Percorre cada classe w(i)
    w{i}.Peso = []; % OBS: Remove todos os elementos de Peso, os valores
                    % devem ficar de acordo com a nova classificação

```

```

[NLin,NCol] = size(w{i}.SemiRot);
for k = 1 : NLin      % insere o peso para o pixel SR
    X = w{i}.SemiRot(k,:);
    Xmaxver = w{i}.MaxVer(k);
    XsomaMV = XsomaMV + (1/w{i}.SomaMV(k));
    % Gerando o Novo Peso
    Xpeso=(1/Xmaxver(Classe))/XSomaMV; % Xpeso=Peso(X,Xmaxver,w);
    w{i}.Peso = [w{i}.Peso; Xpeso];    % Insere no vetor de pesos
end
end
RetW = w;

% =====
% FUNÇÃO PARA INSERÇÃO DE AMOSTRAS SEMI-ROTULADAS (SR) DE MAIOR PESO
% =====
function [retW, retPesoSR, retSemiRot]= IncluiSemiRot(w, Peso, CFG)
ITERACAO = CFG.ITERACAO;
vetPESO = []; % Vetor auxiliar de Pesos
vetSR = []; % Vetor auxiliar de Semi-Rotulados
PesoORD = []; % Peso Ordenado
LIMITE = []; % Numero maximo de SR incluidos na classe w{i}
IND = []; % Pega a posição do Pixel na classe w{i}
NumPixels = []; % Numero de Pixels
IndiceDEC = []; % Vetor de indices em ordem decrescente
SemiRot = {}; % Array de pixels incluidos na classe w{i}
PesoSR = {}; % Array de pesos de pixels SR
pxlANTES = []; % Variaveis para monitoração de pixels Antes e
pxlDEPOIS = []; % Depois da iteração
porcINC = CFG.PorcetagemINC; % Porcentagem de inclusao de Amostras SR
for i=1 : size(w,2)      % Zerando o conteudo das variaveis auxiliares
    vetPESO = [];
    vetSR = [];
    PesoORD = [];
    IND = []; % Recupera a posição do Pixel na classe w{i}
    NumPixels = size(Peso{i},1); % Numero de Pixels
    IndiceDEC = [NumPixels:-1:1]'; % Cria vetor de indices, ordem decrescente
    switch CFG.Inclusao
        case 'Fixo', %Inclui quantidade fixa de pixels SR na classe
            LIMITE=round(NumPixels * porcINC); % N° maximo de SR incluidos na
                % classe w{i}
        case 'Gradual' %Inclui pixels SR gradativamente na classe
            if (porcINC * ITERACAO) < 1
                LIMITE = round(NumPixels * (porcINC * ITERACAO));
            else
                LIMITE = NumPixels;
            end
        case 'Total', %Inclui todos os pixels SR disponíveis na classe
            LIMITE = NumPixels;
    end
End

% Ordenação dos pesos
PesoORD = [Peso{i} w{i}];
pxlANTES = PesoORD;
PesoORD = sortrows(PesoORD);
PesoORD = [IndiceDEC PesoORD];
PesoORD = sortrows(PesoORD);

```

```

pxlDEPOIS = PesoORD;
% Atualização das amostras com base no maior Peso
if size(PesoORD,1) == 0 % Se PesoORD estiver vazio
    Peso{i} = [];
    w{i}     = [];
else
    Peso{i} = PesoORD(:,2);
    w{i}     = PesoORD(:,3:size(PesoORD,2));
end
% Inserindo as Amostras SR na estrutura(Vetor):SemiRot
SemiRot{i} = w{i}(1:LIMITE,:);
PesoSR{i}  = Peso{i}(1:LIMITE,:);
for Ipos=1 : % Monitoração dos Pixels
    pxlANTES = PesoORD;
    pxlDEPOIS = PesoORD;
end
disp('');
end
retW      = w;
retPesoSR = PesoSR;
retSemiRot = SemiRot;
return

% =====
% FUNÇÃO PARA ESTIMAR O VALOR DA PROBABILIDADE A PRIORI
% =====
function PPriori = ProbPriori(ImagemT, Classe)
%=====
% Incluir Probabilidade à Priori P(Wi)
%=====
%
%          N° total de pixels classificados em W1
% ProbPriori(W1)= -----
%          .          N° total de pixel na imagem temática
%          .
%          .
%          N° total de pixels classificados em Wn
% ProbPriori(Wn)= -----
%          .          N° total de pixel na imagem temática
%=====

% Definição de variaveis
TPIT = 0;      % Total de Pixels da Imagem Tematica
TPClasse = 0; % Total de Pixels da Classe
CalcPP=0;     % Armazena a prob. a priori da classe

% Atribuições e Calculo
TPIT      = size(ImagemT,1); % Recebe o n° de pixels da Imagem tematica

TPClasse = size(Classe,1);   % Recebe o n° de pixels da Classe

CalcPP    = (TPClasse / TPIT); % Calcula a probabilidade a priori da classe

PPriori   = CalcPP;
Return

```

```

%=====
% FUNÇÃO PARA ARMAZENAR AS IMAGENS GERADAS
%=====
% SALVA EM ARQ UMA IMAGEM A CADA ITERAÇÃO
%=====
% IMG = Estrutura do tipo IMAGEM a ser gravada
% w{i} = Estrutura das classes
%=====
function GravaImagem(ImagemOriginal, ImagemClass, CFG, vetClassERRO, CTRL)
Bandas = CFG.Bandas;
ITERACAO = CFG.ITERACAO;
NlinIMG = [];
NColIMG = [];
NovaImagem = [];

NovaImagem = ImagemClass;
save NovaImagem NovaImagem;
NPixels = size(NovaImagem,1);
[L, C] = temaResol(NPixels);
NomeARQ = sprintf('imgT00%d.lan', ITERACAO);
salvarImagem(NovaImagem, NomeARQ, L, C);
if size(CTRL,1)~= 0 % Se CTRL == Vazio

    % Gravar Imagem separando os pixels errados
    NovaImagem2 = vetClassERRO;
    save NovaImagem2 NovaImagem2;
    NPixels = size(NovaImagem2,1);
    [L, C] = temaResol(NPixels);
    NomeARQ = sprintf('ErroIMGT00%d.lan', ITERACAO);
    salvarImagem(NovaImagem2, NomeARQ, L, C);
end
return

```