

UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL

**CENTRO ESTADUAL DE PESQUISAS EM SENSORIAMENTO REMOTO E METEOROLOGIA
PROGRAMA DE PÓS-GRADUAÇÃO EM SENSORIAMENTO REMOTO**

**ESTUDO SOBRE O USO DE INFORMAÇÕES ESPECTRAIS E DE CONTEXTO
ESPACIAL NA PONDERAÇÃO DE AMOSTRAS SEMI-ROTULADAS**

Atilio Efrain Bica Grondona

Orientador: Prof. Vitor Haertel, PhD

Porto Alegre, março de 2011

UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL

CENTRO ESTADUAL DE PESQUISAS EM SENSORIAMENTO REMOTO E METEOROLOGIA
PROGRAMA DE PÓS-GRADUAÇÃO EM SENSORIAMENTO REMOTO

ESTUDO SOBRE O USO DE INFORMAÇÕES ESPECTRAIS E DE CONTEXTO
ESPACIAL NA PONDERAÇÃO DE AMOSTRAS SEMI-ROTULADAS

Atilio Efrain Bica Grondona

Lic. em Matemática (2004) - PUCRS

Orientador: Prof. Vitor Haertel, PhD

Banca Examinadora:

Prof. Dr. Claudio Jung – (Instituto de Informática - UFRGS)

Prof. Dr. David Fernandes – (Instituto Tecnológico de Aeronáutica - ITA)

Prof. Dr. Robin Clarke – (Instituto de pesquisas hidráulicas - UFRGS)

*Dissertação submetida ao
Programa de Pós-Graduação em
Sensoriamento Remoto da Universidade
Federal do Rio Grande do Sul como
requisito parcial para a obtenção do título
de Mestre em Sensoriamento Remoto*

Porto Alegre, março de 2010

APRESENTAÇÃO

Este trabalho foi desenvolvido no âmbito do Programa de Pós-Graduação em Sensoriamento Remoto do Centro Estadual de Pesquisas em Sensoriamento Remoto e Meteorologia (CEPSRM) da Universidade Federal do Rio Grande do Sul (UFRGS) durante o ano de 2009 e 2010, sob a orientação do Prof. Dr. Vitor Haertel.

AGRADECIMENTOS

Antes de qualquer coisa gostaria de agradecer ao professor Vitor Haertel, orientador, mas também de forma justa, co-autor desse trabalho, pela confiança depositada em mim, pelo profissionalismo e amizade dedicados o tempo todo juntamente com sua sapiência e entusiasmo contagiante pela ciência. A ele agradeço também pelo tópico desenvolvido nesse estudo, o qual se mostrou desafiador à medida que o trabalho se desenvolvia, ao mesmo tempo revelava crescente pertinência dentro do estado atual da arte;

A minha mãe Sônia Bica eu agradeço todo o apoio e incentivo dado, e ao meu pai Eduardo Grondona (in memoriam) eu dedico este trabalho, pois eles desempenharam um papel fundamental em minha vida, sendo meus primeiros professores, meus amigos, minha base;

Aos colegas do CEPSRM, pessoas que tiveram uma convivência extraordinária comigo, e aos quais eu devo muito por este e por outros trabalhos, e por tudo que aprendi com eles, em especial agradeço a Daniel Zanotta (turma anterior) e Juliana Costi (minha turma);

Aos meus mais que colegas do Bacharelado em Matemática Aplicada da UFRGS (turma de 2006), amigos para a vida, pessoas que compartilhei, e compartilho a construção do meu conhecimento, estes que junto comigo abdicamos dos finais de semana, das festas e de muitas outras coisas para, por muitas horas a fio, estudar, e em especial agradeço a Julio Lombardo, Everson Gomes, Victor Giacomello, Guilherme Monteiro e Guilherme Stocker por me brindarem sua amizade, pessoas estas que palavras não quantificam a minha gratidão;

Ao CNPq pelo suporte financeiro na forma de bolsa de estudos integral;

À UFRGS, meu segundo lar, que sempre proporcionou as mais favoráveis condições possíveis para a execução deste estudo. A professora Denise Fontana que abriu as portas do conhecimento para o Sensoriamento Remoto ao me selecionar como bolsista de graduação no CEP SRM há alguns anos atrás, e a prof^a. Silvia B. A. Rolim, por ser não só uma professora, mas sim por se tornar uma amiga acima de tudo.

“La science, mon garçon, est faite d’erreurs, mais d’erreurs qu’il est bon de commettre, car elles mènent peu à peu à la vérité.”

Julio Verne

ESTUDO SOBRE O USO DE INFORMAÇÕES ESPECTRAIS E DE CONTEXTO ESPACIAL NA PONDERAÇÃO DE AMOSTRAS SEMI-ROTULADAS

Atilio Efrain Bica Grondona

Orientador: Prof. Vitor Haertel, PhD

RESUMO

Esta dissertação aborda o problema da utilização de classificadores paramétricos em dados de alta dimensionalidade. As vantagens trazidas pelos dados em alta dimensionalidade são bem conhecidas. Classes que são muito semelhantes podem, não obstante, ser separadas com um alto grau de acurácia desde que a classificação dos dados seja realizada em um espaço de alta dimensionalidade e que as matrizes de covariância das classes difiram significativamente. Sistemas sensores capazes de adquirir dados de imagem em alta dimensionalidade (dados de imagens hiperespectrais) foram, em parte, desenvolvidos para tirar proveito dessa condição. Nas condições do mundo real, no entanto, temos de enfrentar o problema de estimar um grande número de parâmetros, geralmente, com um número limitado de amostras. Amostras de treinamento são geralmente caras e demoradas para adquirir. Diferentes abordagens para resolver ou, pelo menos, atenuar este problema tem sido um tópico de investigação por parte da comunidade internacional em sensoriamento remoto. Entre outras, uma possível abordagem que tem sido proposta na literatura consiste em aumentar o número de amostras pela adição de amostras semi-rotuladas ao processo de estimação dos parâmetros do classificador. A metodologia investigada nesta dissertação segue esta abordagem geral. O foco principal deste estudo consiste em investigar uma abordagem para estimar os pesos a serem associados às amostras semi-rotuladas. A abordagem proposta inclui duas etapas. Na primeira, as estimativas iniciais para os pesos são realizadas de forma interativa, por meio da utilização de informações espectrais somente. Em uma segunda etapa, os pesos estimados são refinados por meio de informações de contexto espacial. A metodologia proposta é avaliada através de experimentos que fazem uso de dados de imagens hiperespectrais AVIRIS. Os resultados são apresentados e discutidos. Sugestões para futuras pesquisas neste tópico também são apresentados.

Palavras chave: classificação, estimação dos parâmetros, contexto espacial, amostras semi-rotuladas.

STUDY ABOUT THE USE OF SPECTRAL AND SPATIAL CONTEXT INFORMATION IN THE WEIGHTHING OF SEMI-LABELED SAMPLES

Atilio Efrain Bica Grondona

Orientador: Prof. Vitor Haertel, PhD

ABSTRACT

This dissertation deals with the problem of using parametric classifiers in high dimensional data settings. The advantages brought by high dimensional data are well known. Classes that are very similar can nonetheless be separated with a high degree of accuracy provided that the classification is performed in high dimensional data settings and that the classes' covariance matrices differ significantly. Sensor system capable of acquiring high dimensional image data (hyperspectral image data) were in part developed to take advantage of this condition. In real world conditions, however, we have to face the problem of estimating a resulting large number of parameters with a generally limited number of samples. Training samples are usually expensive and time consuming to acquire. Different approaches to solve or at least mitigate this problem have been a topic of investigation by the international community in remote sensing. Among others, one possible approach that has been proposed in the literature consists in increasing the number of samples by adding semi-labeled samples to the process of estimating the classifier's parameters. The methodology investigated in this dissertation follows this general approach. The main focus in this study consists in investigating an approach to estimate the weights to be associated with the semi-labeled samples. The proposed approach includes two steps. In the first one, initial estimates for the weights are performed in an iterative way, by making use of spectral information only. In a second step, the estimated weights are further adjusted by means of spatial context information. The proposed methodology is evaluated by experiments making use of AVIRIS hyperspectral image data. The results are presented and discussed. Suggestions for further research in this topic are also presented.

Key-words: classification, parameter estimation, spatial context, semi-labeled samples.

SUMÁRIO

1 INTRODUÇÃO.....	10
1.1 SENSORES HIPERESPECTRAIS.....	10
1.2 SISTEMAS SENSORES HIPERESPECTRAIS.....	11
1.2.1 O SISTEMA HYPERION.....	11
1.2.2 O SISTEMA AVIRIS.....	11
1.3 RECONHECIMENTO DE PADRÕES E CLASSIFICAÇÃO DIGITAL DE IMAGENS.....	15
1.4 O PROBLEMA DA DIMENSIONALIDADE DOS DADOS.....	18
1.5 DADOS DE ALTA DIMENSIONALIDADE.....	19
1.6 O FENÔMENO DE HUGHES.....	20
1.7 OBJETIVOS.....	22
1.8 ESTRUTURA DA DISSERTAÇÃO.....	23
2 REVISÃO BIBLIOGRÁFICA.....	24
2.1 DADOS EM ALTA DIMENSIONALIDADE – VANTAGENS E PROBLEMAS.....	24
2.2 MÉTODOS DE REDUÇÃO DA DIMENSIONALIDADE DOS DADOS	25
2.2.1 SELEÇÃO DE VARIÁVEIS.....	25
2.2.2 EXTRAÇÃO DE VARIÁVEIS.....	27
2.2.2.1 TRANSFORMAÇÃO DE COMPONENTES PRINCIPAIS.....	28
2.2.2.2 ANÁLISE CANÔNICA.....	30
2.3 ANÁLISE DISCRIMINANTE REGULARIZADA (RDA).....	32
2.4 TÉCNICAS DE INCREMENTO DO NÚMERO DE AMOSTRAS DE TREINAMENTO SEMI-ROTULADAS.....	34
2.5 INSERÇÃO DE INFORMAÇÕES DE CONTEXTO ESPACIAL NO PROCESSO DE CLASSIFICAÇÃO DE IMAGENS DIGITAIS.....	37
2.5.1 RELAXAÇÃO.....	37
2.5.2 CAMPOS ALEATÓRIOS DE MARKOV.....	40
3 METODOLOGIA.....	43
3.1 INTRODUÇÃO.....	43
3.2 METODOLOGIA PROPOSTA.....	45
3.3 ESTIMAÇÃO INICIAL DE PESOS PARA AMOSTRAS SEMI- ROTULADAS.....	48
3.4 INTRODUÇÃO DE INFORMAÇÕES DE CONTEXTO ESPACIAL (VIZINHANÇA)	51
3.5 O COEFICIENTE DE COMPATIBILIDADE.....	52
3.6 O PROCESSO DE RELAXAÇÃO.....	53
3.7 LIMIAR E “CLASSE DE FUNDO.....	54
4 EXPERIMENTOS.....	57
4.1 INTRODUÇÃO.....	57
4.2 MATERIAIS.....	57

4.2.1 EQUIPAMENTOS.....	57
4.2.2 APLICATIVOS.....	57
4.2.3 DADOS.....	57
4.3 CENA DE ESTUDO.....	58
4.4 SELEÇÃO DE BANDAS ESPECTRAIS.....	61
4.5 SELEÇÃO DE CLASSES.....	63
4.6 SOFTWARE IMPLEMENTANDO A METODOLOGIA PROPOSTA.....	64
4.7 AMOSTRAS DE TREINAMENTO.....	66
4.8 O LIMIAR E A “CLASSE DE FUNDO”	68
4.9 ANÁLISE DOS EXPERIMENTOS.....	71
4.10 TEMPO COMPUTACIONAL.....	73
4.11 EXPERIMENTOS.....	74
4.11.1 EXPERIMENTO 1.....	74
4.11.2 EXPERIMENTO 2.....	76
4.11.3 EXPERIMENTO 3.....	78
4.11.4 EXPERIMENTO 4.....	82
4.11.5 EXPERIMENTO 5.....	84
4.11.6 EXPERIMENTO 6.....	88
4.11.7 EXPERIMENTO 7.....	91
4.12 ANÁLISE DOS EXPERIMENTOS	94
5 CONCLUSÕES E SUGESTÕES.....	96
5.1 CONCLUSÕES.....	96
5.2 SUGESTÕES.....	97
REFERÊNCIAS BIBLIOGRÁFICAS.....	99
APÊNDICE A: CÓDIGO FONTE (LINGUAGEM MATLAB)	105

1 INTRODUÇÃO

1.1 SENSORES HIPERESPECTRAIS

Desde a década de 1960 até hoje, um volume grande de recursos vem sendo investido em pesquisa, visando o desenvolvimento de novos sistemas sensores (hardware) bem como processos para a extração de informações, isto é, a utilização destes dados (desenvolvimento de novas metodologias e softwares). Um dos avanços possibilitados é fazer a coleta de dados multiespectrais em 200 ou mais bandas do espectro. Tais conjuntos de dados podem ser denominados como dados hiperespectrais [LAN 03]. Estes dados são adquiridos por sensores hiperespectrais, e consistem do conjunto de observações realizadas sobre um grande número de bandas do espectro eletromagnético simultaneamente (224 bandas espectrais no caso do sensor AVIRIS, e 220 no HYPERION). Estes dados têm como característica principal a alta dimensionalidade, o que exige, em muitos casos, uma abordagem distinta daquelas tradicionalmente empregadas em dados de baixa dimensionalidade (por exemplo, dados multiespectrais dos sensores LANDSAT, SPOT), tornando-se necessário o desenvolvimento de novas metodologias e técnicas para fins de classificação.

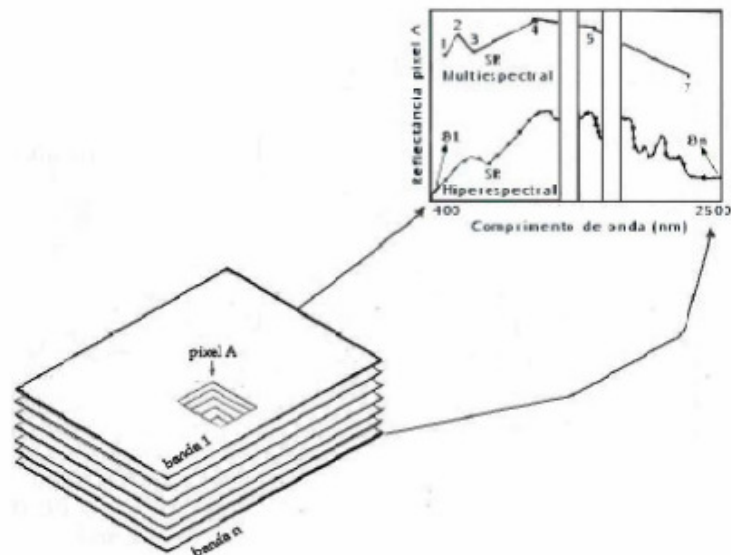


Figura 1 – Comparação entre baixa e alta resolução espectral

Na Figura 1, se observa 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. O grande poder discriminante dos dados de sensores hiperespectrais pode registrar o comportamento espectral de cada pixel com uma enorme riqueza de detalhes, este detalhamento fino das características do alvo oferece ainda a possibilidade de identificar e separar substâncias/alvos através de seus picos de absorção, por exemplo, em aplicações na geologia.

1.2 SISTEMAS SENSORES HIPERESPECTRAIS

1.2.1 O SISTEMA HYPERION

Como mencionado anteriormente, o sistema sensor HYPERION dispõe de um imageador de alta resolução capaz de adquirir dados em 220 bandas espectrais (de 0.4 a 2.5 μm) 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 (*Earth Observing-1*) lançado no final do ano 2000. Este sistema é o primeiro imageador hiperespectral da NASA operando em nível orbital [HYP 01].

1.2.2 O SISTEMA AVIRIS

O AVIRIS, acrônimo para *Airborne Visible InfraRed Imaging Spectrometer* foi desenvolvido pela NASA em 1983 e se trata de um sensor experimental, aerotransportado que já operou em quatro plataformas de aeronaves: ER-2 jet da NASA, dois outros turboélices internacionais, *Scaled Composites' Proteus* e o WB-57 também da NASA. O ER-2 voa a cerca de 20 km acima do nível do mar, a cerca de 730 km / h. A aeronave Twin Otter voa a 4 km acima do nível do solo em 130 km/h [AVI 01]. O sensor AVIRIS voou todo o EUA, além de Canadá e Europa. Este sensor coleta dados em 224 bandas espectrais contíguas, no intervalo de 0.4 μm a 2.5 μm , sendo o primeiro vôo desde sensor realizado em 1987 [AVI 02]. Estes dados são caracterizados, portanto, por uma alta dimensionalidade (alta resolução espectral), com esta característica a utilização destes dados requer uma abordagem diferente daquelas tradicionalmente utilizadas nos dados de baixa dimensionalidade (sensores multiespectrais, LANDSAT, SPOT, por exemplo), tornando-se necessário desta forma o desenvolvimento de novas técnicas e metodologias. Na Figura 2 observa-se a curva de resposta espectral de um pixel, gerada pelos sensores hiperespectrais. A curva de resposta

espectral do pixel é muito mais detalhada com as informações obtidas a partir de um sensor hiperespectral [AVI 04].

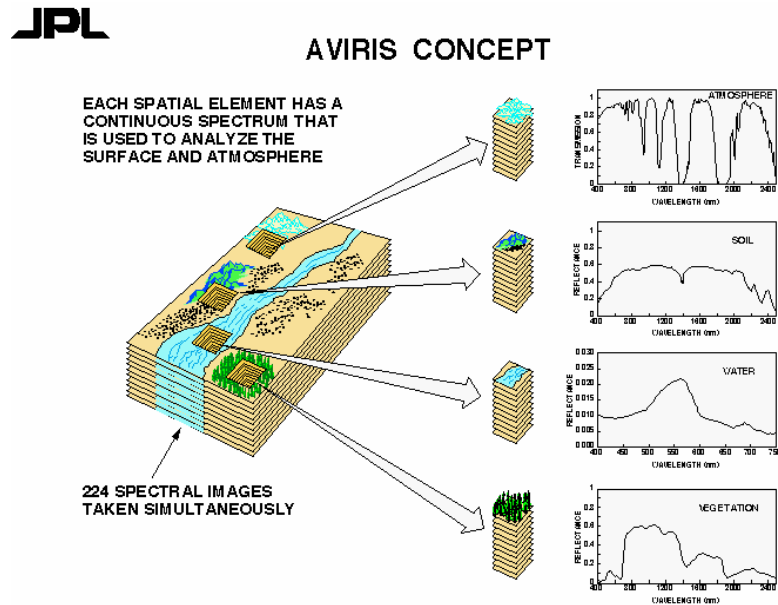


Figura 2 – Riqueza de detalhes em um pixel obtido por um sensor hiperespectral

O sistema óptico do sensor AVIRIS consiste de um scanner, 4 espectrômetros, e uma fonte de calibração acoplada por fibra óptica (Figura 3).

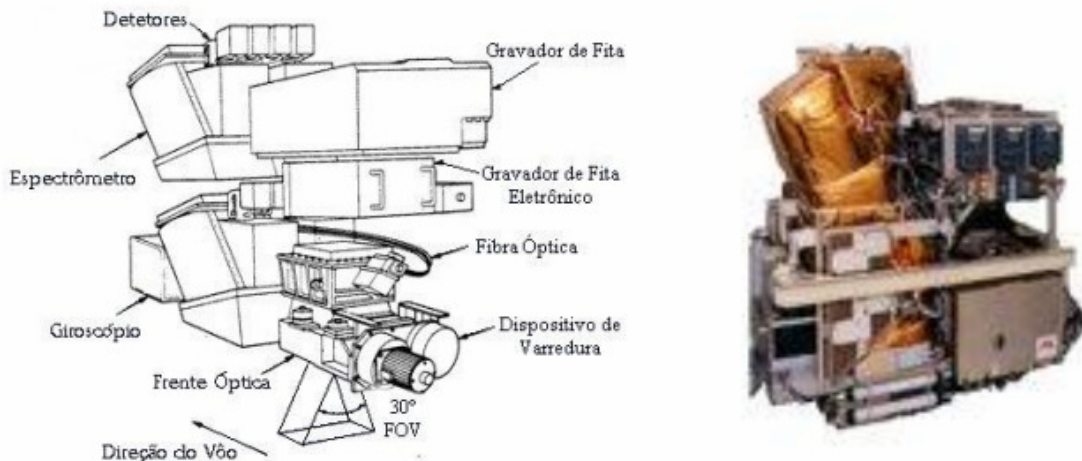


Figura 3 - Sensor AVIRIS

Fonte: JPL - Jet Propulsion Laboratory - NASA

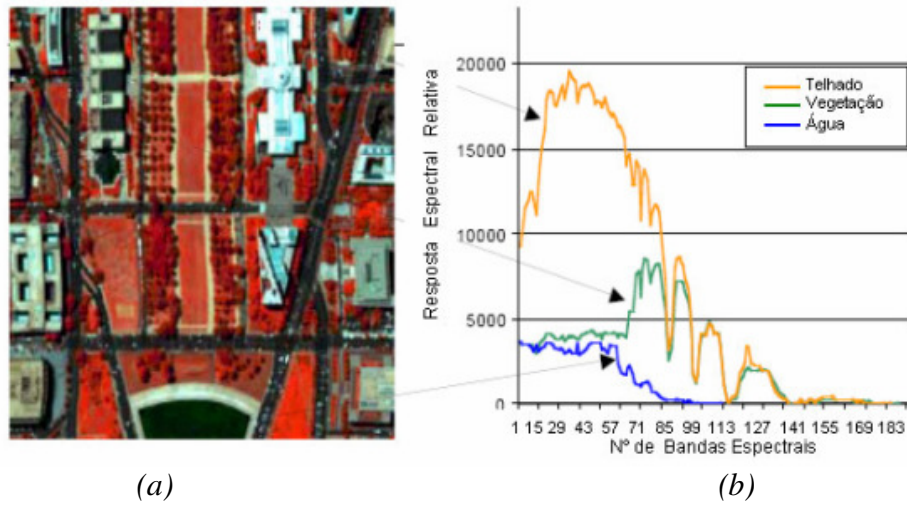
O sensor AVIRIS usa um espelho scanner que varre no sentido longitudinal, produzindo 614 pixels para cada varredura dos 224 detectores, cada um cobrindo um intervalo de comprimento de onda (banda espectral), de aproximadamente 10nm, permitindo uma cobertura completa entre 400nm e 2500nm. Cada pixel cobre, 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. A Tabela 1.1 mostra as características técnicas do sensor AVIRIS.

Tabela 1.1 – Ficha técnica do Sensor AVIRIS

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

Na Figura 4 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 4 - a) Espaço imagem b) Espaço espectral
Fonte: Adaptado de Landgrebe (2002).*

Outra forma interessante de visualizar a quantidade de informação, ou volume de dados, obtidos a partir deste tipo de sensor de alta resolução espectral é através da chamada imagem cubo ou também cubo hiperespectral. O exemplo ilustrado na Figura 5 é datado em 20 de agosto de 1992, quando foi feito um voo a uma altitude de 20.000 m (65.000 pés) sobre a região de Moffett Field, Califórnia, a parte final Sul da Baía de São Francisco. Descrevendo a imagem gerada, na parte superior do cubo foi usada uma imagem falsa-cor, feita para acentuar a estrutura da água e dos lagos. Também está visível na imagem o Aeroporto de Moffett Field.

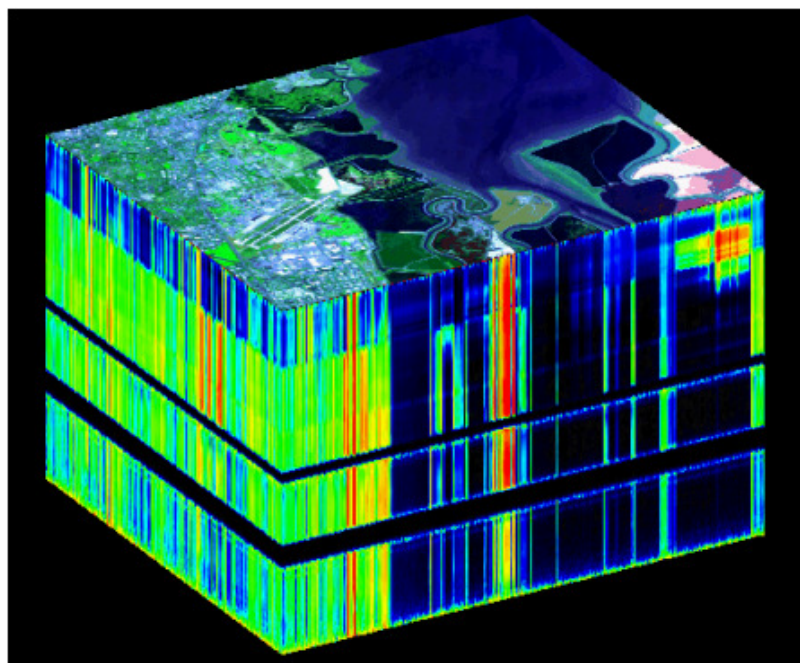


Figura 5 – Imagem Cubo (Moffet Field)

As laterais do cubo ilustram os limites das 224 bandas espectrais do sensor. A parte superior do cubo é a parte visível do espectro (comprimento de onda de 400nm), e a parte inferior pertence ao infravermelho médio (comprimento de onda de 2.500 nm). São usadas pseudo-cores, entre o preto e o azul (baixa resposta, isto é, alta absorção e baixa reflectância) até o vermelho (alta resposta, isto é, baixa absorção e alta reflectância). As regiões que apresentam duas falhas no cubo são as regiões de absorção da atmosfera [AVI 05].

1.3 RECONHECIMENTO DE PADRÕES E CLASSIFICAÇÃO DIGITAL DE IMAGENS

O emprego de dados de sensoriamento remoto usualmente requer o uso de técnicas de Reconhecimento de Padrões. Este estudo abrange um tópico de pesquisa que está inserido dentro desta área.

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

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

A classificação de imagens de Sensoriamento Remoto assistida por computador trata de atribuir a cada pixel da imagem um “rótulo” identificando uma determinada classe de cobertura do solo, dentre as classes pré-definidas pelo analista. Este processo 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. Cada pixel na imagem é representado por um vetor linha com dimensão igual ao número de bandas (NB) espectrais do sensor, e geralmente utiliza a forma $1 \times NB$. Cada elemento neste vetor fica definido pelo valor do contador digital (DN – Digital Number) na banda correspondente. Este valor de DN varia de 0 até $2^n - 1$, onde n é o número de bits utilizados pelo sensor na coleta dos dados. 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. Além da natureza espectral dos dados, as imagens de Sensoriamento Remoto podem fornecer ricas informações de contexto. O uso em conjunto, e de forma adequada, das informações de contexto e espectral pode melhorar a acurácia dos classificadores, quando em comparação com somente um dos tipos de atributos citados [JEO 92].

A relaxação é uma das técnicas mais conhecidas que tem sido propostas para explorar o contexto. A possibilidade de usar a relaxação, mais especificamente a relaxação probabilística, para a efetiva classificação contextual de imagens reais de Sensoriamento Remoto tem sido investigada por vários autores. Porém, vários problemas surgem no uso da relaxação em aplicações no Sensoriamento Remoto. Primeiro temos o problema de computar exaustivamente as probabilidades para um conjunto de pixels que servirão como estimativas iniciais do processo de relaxação. Outro problema está no cálculo do coeficiente de compatibilidade, que deve ser tendencioso para promover a estrutura da imagem [ZEN 87].

O desenvolvimento de sistemas de reconhecimento de padrões, visualizado na Figura 6, essencialmente envolve três aspectos: 1) aquisição de dados através de dispositivos de captação dos dados; 2) representação dos dados na forma de um vetor x ; 3) transformações necessárias e tomadas de decisão.

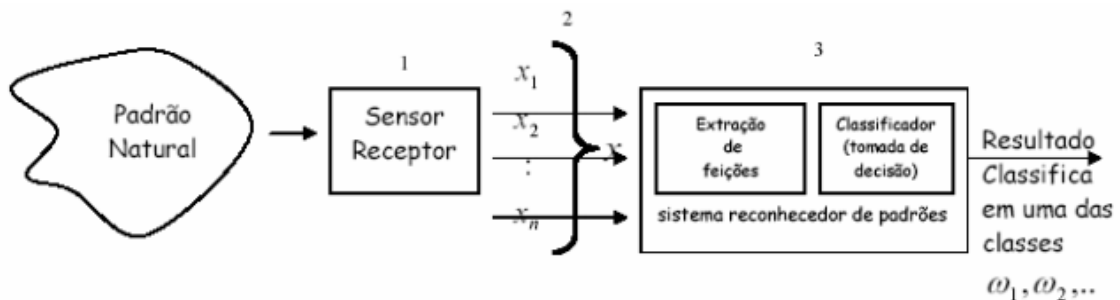


Figura 6 – Esquema básico para reconhecimento de padrões

O domínio do problema dita a escolha do sensor, técnica de pré-processamento, esquema de representação, e o modelo de tomada de decisão [JAI 00]. O aprendizado a partir de um conjunto de amostras (amostras de treinamento) é uma etapa inicial na maioria dos sistemas de reconhecimento de padrões. Os métodos desenvolvidos para fins de reconhecimento de padrões, isto é, classificação de dados em classes previamente estabelecidas, tem sido objeto de estudo por um número grande de pesquisadores. Um levantamento teórico dos vários métodos, hoje em utilização, pode ser encontrado em [JAI 00].

Diferentes técnicas de classificação são usadas dependendo do tipo do modelo escolhido [DUD 01]. Os métodos de classificação podem ser agrupados em duas grandes categorias: métodos supervisionados e métodos não-supervisionados. Nos métodos da classificação supervisionada, as classes são previamente definidas pelo analista, isto é, definidas ou caracterizadas através das amostras de treinamento. Cada classe pode então ser caracterizada por uma função decisão que pode ser: (1) de natureza probabilística ou (2) de natureza determinística.

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

- 1°. Identificar as classes existentes;
- 2°. Definir a forma geral da função decisão a ser adotada;
- 3°. Selecionar os pixels representativos de cada classe, que serão chamados de amostras de treinamento;
- 4°. Estimar os parâmetros ou calcular os coeficientes específicos da função decisão a ser usada;
- 5°. Usando o classificador treinado, se classificam 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 outro tipo de abordagem. Em alguns casos existem problemas na área de reconhecimento de padrões, nas quais a natureza (ou definição) das classes e mesmo o número de classes presentes são desconhecidos. Neste caso, o problema a ser tratado consiste não somente na classificação propriamente dita, mas também, na própria definição das classes. Ao contrário do método supervisionado, onde se tem um conhecimento prévio das classes, os métodos não-supervisionados atribuem à técnica ou ao algoritmo escolhido, a tarefa de identificar, por si só, as classes existentes num conjunto de dados.

Uma possível abordagem para resolver um problema desta natureza, freqüentemente utilizada em sensoriamento remoto, consiste em procurar aglomerados (clusters) de pontos

(pixels) no espaço dos atributos. 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 matemático e/ou lógico. Este processo é denominado de *clustering* ou análise de aglomerados. Os processos de classificação podem também ser agrupados em duas grandes categorias: métodos paramétricos e métodos não-paramétricos. Nos métodos paramétricos, a forma geral da função densidade de probabilidade que descreve o comportamento dos dados é suposta conhecida. Os parâmetros existentes são estimados a partir das amostras de treinamento disponíveis para cada classe. Ocorrem casos em que a forma geral da função densidade probabilidade associada aos dados não é conhecida. A alternativa nestes casos consiste na utilização dos chamados métodos não-paramétricos. Como exemplo de métodos não-paramétricos pode-se citar o método ou estimação de Parzen e o método do vizinho mais próximo [THE 89]. O método, dentro da abordagem estatística, mais utilizado em sensoriamento remoto, é o da Máxima Verossimilhança Gaussiana, que é um método supervisionado e paramétrico. É considerado paramétrico, pois o método envolve parâmetros (vetor média e matriz covariância) da distribuição gaussiana multivariada, e é supervisionado, pois estima estes parâmetros através das amostras de treinamento. Contudo, uma das maiores dificuldades ocorre na estimação dos parâmetros, em especial a matriz covariância. Em dados de alta dimensionalidade, esta estimação torna-se extremamente problemática, devido ao fato de que em situações reais o tamanho das amostras disponíveis nunca é adequado. A alta dimensionalidade dos dados se constitui, portanto, simultaneamente em um benefício e em um problema. Esta nova situação vem sendo objeto de intenso interesse por parte da comunidade científica internacional. O que se verifica na prática é que o aumento na dimensionalidade dos dados, isto é, do número de bandas espectrais, resulta inicialmente em um incremento na acurácia da resultante imagem classificada. Este fato seria de se esperar, pois com o aumento do número de bandas, ocorre um incremento na informação disponível. A partir de um certo ponto, entretanto, e utilizando-se as mesmas amostras de treinamento (para fins de estimação dos parâmetros do classificador), a acurácia começa a decrescer com o aumento da dimensionalidade dos dados. Este efeito, conhecido como o fenômeno de Hughes, se deve ao aumento do número de parâmetros a serem estimados, especialmente na matriz covariância. Isto serve como motivação para o desenvolvimento de novas pesquisas visando minimizar este efeito.

1.4 O PROBLEMA DA DIMENSIONALIDADE DOS DADOS

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

O aumento da dimensionalidade de tais dados presumivelmente aumenta o realce do conteúdo da informação dos dados, mas fornece um desafio a técnicas convencionais de analisar estes dados [LAN 03].

Outros dois inconvenientes de dados com alta dimensionalidade são [LAN 03]:

- Espaços de dimensão elevada são na maior parte virtualmente vazios;
- Dados normalmente distribuídos têm a tendência de se concentrar nas caudas da distribuição, fazendo a estimativa da densidade mais difícil.

1.5 DADOS DE ALTA DIMENSIONALIDADE

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

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

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

Um ponto a salientar aqui é que, para modelos em um espaço de características hiperespectral (N-dimensional), estatísticas de primeira ordem (por exemplo, vetor de médias) fornece informações sobre a localização da distribuição (classe) no espaço, enquanto que estatísticas de segunda ordem (por exemplo, variâncias e covariâncias) descrevem sua forma. No entanto, existem evidências que sugerem, de forma geral, que estatísticas de primeira ordem são freqüentemente mais importantes nos processos de discriminação, mas com o aumento da dimensionalidade, a informação de forma (fornecida pelas estatísticas de segunda ordem) podem em alguns casos começar a dominar, e a localização da distribuição começa a ser menos importante [LAN 03].

1.6 O FENÔMENO DE *HUGHES*

Dados em alta dimensionalidade oferecem novas possibilidades no processo de classificação de imagens. Classes com comportamento espectral muito semelhante, ou seja, que possuem os vetores de média muito próximos ou iguais, são praticamente indistinguíveis quando se emprega dados tradicionais em baixa dimensionalidade (LANDSAT, por exemplo), mas podem passar a ser distinguíveis ou separáveis com alto grau de acurácia quando se dispõe de dados em alta dimensionalidade, desde que sua matriz de covariância seja significativamente diferente [FUK 90].

Também se pode provar que o número necessário de amostras de treinamento é linearmente relacionado com a dimensionalidade dos dados, para classificadores lineares, enquanto que para classificadores quadráticos esta relação é quadrática também [LAN 03].

A alta dimensionalidade dos dados apresenta, entretanto, novos desafios ao processo de classificação. Possivelmente, o maior desafio neste caso consiste na estimação de parâmetros associados a classificadores estatísticos. O número de parâmetros a serem estimados, na matriz covariância particularmente, cresce rapidamente na medida em que a dimensionalidade dos dados aumenta, como ilustrado na Figura 7 e na Tabela 1.2. Se o número das amostras de treinamento não aumenta, a confiabilidade na estimação destes parâmetros decresce. Em outras palavras, quando o tamanho das amostras de treinamento permanece constante, o acréscimo de bandas espectrais adicionais resulta no aumento da variância, isto é, da incerteza da estimativa dos parâmetros.



2 componentes = 3 parâmetros

3 componentes = 6 parâmetros

Figura 7 – Número de parâmetros a serem estimados na matriz de covariância

Aplicando a fórmula $np = \left(n + \frac{n^2 - n}{2} \right)$ 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, por exemplo, o sistema sensor AVIRIS com 185 bandas das 224, calcula-se a quantidade de parâmetros necessários para cada caso, conforme mostrado na Tabela 1.2.

Tabela 1.2 – Parâmetros necessários para o sensor LANDSAT – TM e o sensor AVIRIS

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

Nota-se que no uso do sistema AVIRIS é necessário estimar 820 vezes mais parâmetros do que a quantidade de parâmetros necessários no uso de sistemas convencionais como o LANDSAT-TM. O fato do aumento do número de parâmetros a serem estimados ocasionar a degradação da acurácia é responsável pelo efeito conhecido como fenômeno de Hughes ou a maldição da dimensionalidade [TAD 98]. O efeito, demonstrado na Figura 8, manifesta-se no comportamento da acurácia resultante de um processo de classificação estatística.

Inicialmente, o acréscimo de novas bandas espectrais faz com que a acurácia tenda a aumentar. Em um determinado ponto, um máximo para a acurácia é atingido. A partir daí, um

acréscimo adicional do número de bandas resulta em um decréscimo cada vez mais acelerado na acurácia fornecida pelo processo de classificação. Este problema, bem como possíveis alternativas para mitigá-lo, vem sendo objeto de estudo por parte de vários pesquisadores, como Shahshahani [SHA 94], Hsieh [HSI 98], Tadjudin [TAD 98] e Cortijo F.J. [COR 99].

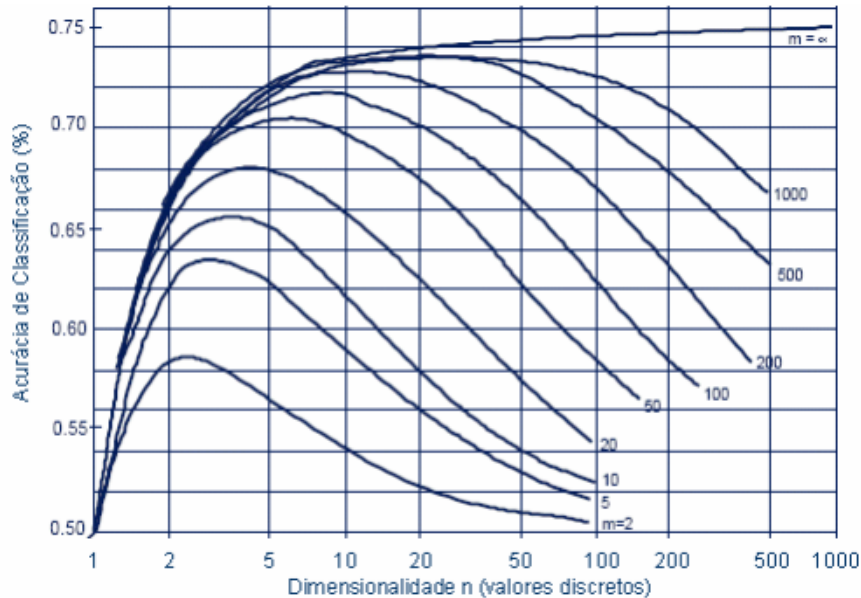


Figura 8 – Representação típica do fenômeno de Hughes
Fonte: Adaptado de Landgrebe (2002)

A solução para este problema está, basicamente, centrada em três abordagens: a) pela utilização de técnicas de análise discriminante regularizada [AEB 94], [FRI 89], [ERB 01]; b) redução da dimensionalidade dos dados, ou seja, a 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 [FUK 90], 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) com a introdução de amostras semi-rotuladas, amostras estas que são inseridas de forma iterativa e ponderadas durante o processo de classificação.

1.7 OBJETIVOS

Nesta dissertação se têm por objetivos:

1. Investigar uma metodologia que permita uma estimacão mais confiável dos parâmetros do classificador quando se utiliza dados em alta dimensionalidade;

2. Propor um método de ponderação das amostras semi-rotuladas que seja iterativo e eficiente;
3. Refinar esta ponderação através da relaxação probabilística, e com isso, melhorar a acurácia do classificador;
4. Desenvolver ferramentas computacionais para implementar e testar essa metodologia.

1.8 ESTRUTURA DA DISSERTAÇÃO

Esta dissertação está organizada em capítulos, onde:

- Primeiro capítulo - é feita uma introdução aos temas: sensoriamento remoto (radiação eletromagnética, espectro eletromagnético, bandas espectrais, assinatura espectral), É também descrito o funcionamento do sensor hiperespectral AVIRIS, os conceitos básicos de reconhecimento de padrões e da sua abordagem estatística, a dimensionalidade dos dados e seus problemas intrínsecos. E por último, estão sumarizados os objetivos desta dissertação.
- Segundo capítulo - consta de uma revisão bibliográfica sobre os métodos para mitigar os problemas associados à dimensionalidade dos dados, uma breve descrição das técnicas paramétricas citadas anteriormente e posteriormente neste texto.
- Terceiro capítulo - versa sobre a metodologia que será utilizada nos dados hiperespectrais disponíveis.
- Quarto capítulo - apresenta os experimentos do algoritmo em imagens hiperespectrais obtidas pelo sensor AVIRIS e análise dos mesmos.
- Quinto capítulo - são apresentadas as conclusões finais e sugestões.

2 REVISÃO BIBLIOGRÁFICA

2.1 DADOS EM ALTA DIMENSIONALIDADE – VANTAGENS E PROBLEMAS

Conforme mencionado na seção de Introdução, dados em alta dimensionalidade, como aqueles disponibilizados em imagens hiperespectrais, permitem uma separação mais acurada de classes espectrais, especialmente aquelas muito semelhantes entre si. Classes que apresentam estatísticas de primeira ordem (vetores de médias) muito similares entre si podem, não obstante, ser separadas com um alto grau de acurácia em espaços de alta dimensionalidade, sob a condição de que as matrizes de covariância difiram significativamente entre si ([FUK 90] e [LAN 03]).

A utilização de dados em alta dimensionalidade em classificadores paramétricos resulta, entretanto, em um número elevado de parâmetros a serem estimados a partir das amostras de treinamento disponíveis. Em situações práticas, o número de amostras de treinamento disponíveis é usualmente limitado, resultando em estimativas pouco confiáveis para os parâmetros do classificador e conseqüentemente em uma acurácia reduzida na imagem temática produzida. Esta dificuldade afeta particularmente as estimativas para as matrizes de covariância, resultando no comportamento conhecido como fenômeno de Hughes. A acurácia produzida por um classificador paramétrico tende inicialmente a crescer na medida em que informação adicional, isto é, novas variáveis são acrescentadas aos dados. Em um determinado ponto, a acurácia atinge um máximo para em seguida passar a decrescer, na medida em que a dimensionalidade dos dados, isto é, o número de bandas espectrais, continua a crescer. Este comportamento reflete a dificuldade em se obter estimativas confiáveis para um número crescente de parâmetros a partir de um número limitado de amostras de treinamento.

Métodos que visam reduzir este problema vêm 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 na dimensionalidade dos dados por seleção ou extração de variáveis, conforme Fukunaga (1990) e Serpico et alii. (2002), (2) métodos de análise discriminante regularizada, descrita por Friedman (1989) e Aeberhard et alii. (1994) e (3) utilização de amostras semi-rotuladas, de acordo com Shashahani e Landgrebe (1994) e Jackson e Landgrebe (2001).

2.2 MÉTODOS DE REDUÇÃO DA DIMENSIONALIDADE DOS DADOS

Em situações reais, a utilização simultânea de todas as bandas fornecidas pelo sensor (dimensionalidade total) ou de um número grande destas no processo de classificação, além de geralmente desnecessário pode se tornar 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 [LAN 03]. Geralmente é 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 do melhor subconjunto de variáveis para cada situação em particular. Este problema é conhecido como redução da dimensionalidade dos dados. Nas metodologias que visam reduzir a dimensionalidade dos dados, existem duas abordagens distintas: (1) seleção de variáveis e (2) extração de variáveis.

2.2.1 SELEÇÃO DE VARIÁVEIS

Seleção de variáveis consiste possivelmente na forma mais intuitiva de reduzir o problema da dimensionalidade dos dados. Consiste simplesmente em selecionar um subconjunto de bandas espectrais, contendo aquelas que oferecem melhores condições de separabilidade entre as classes em estudo. Este processo se torna ótimo somente após uma busca exaustiva, na qual todos os possíveis subconjuntos são testados. Para valores moderados da dimensionalidade dos dados, esta abordagem é viável. Para dados em alta dimensionalidade, entretanto, uma busca exaustiva pode se tornar computacionalmente inviável.

Conforme Webb (2002), a seleção de um subconjunto contendo d variáveis entre um total de p pode se tornar computacionalmente inviável. O número de possíveis subconjuntos com d variáveis (n_d) pode ser calculado pelo coeficiente binomial:

$$n_d = \binom{p}{d} = \frac{p!}{(p-d)!d!}$$

que pode se tornar muito grande, mesmo para valores moderados de p e d . Existem duas estratégias básicas para seleção de variáveis, sendo:

- Métodos ótimos: são métodos de busca exaustivos, viáveis somente para problemas de dimensionalidade pequena.
- Métodos sub-ótimos: buscam basicamente estimar a sobreposição entre as distribuições das classes com base em um determinado critério e escolher o subconjunto de variáveis para o qual a sobreposição entre as distribuições das classes é mínima (ou seja, maximiza a separabilidade).

Os algoritmos de busca sub-ótimos, embora não sendo capazes de examinar todas as combinações de variáveis, avaliam um conjunto potencialmente útil dessas combinações. Dentre as várias técnicas sub-ótimas temos:

“*Best Individual N*”: é o mais simples e talvez o menos eficiente. Consiste em selecionar as N variáveis que individualmente apresentam o maior poder discriminante para as classes em estudo. Em alguns casos este método pode produzir razoáveis conjuntos de variáveis, especialmente se as variáveis no conjunto original de dados não são correlacionadas.

“*Sequential Forward Selection*”: o algoritmo SFS identifica iterativamente o melhor subconjunto de variáveis que pode ser obtido pela adição de uma variável por vez ao subconjunto selecionado. Parte-se de um conjunto com d_1 variáveis originais, X_{d_1} . Para cada variável ξ_j que não pertence a d_1 , a função separabilidade $J_j = J(X_{d_1} + \xi_j)$ é calculada. A variável que tem o maior valor para a função J_j é escolhida como uma das variáveis a ser adicionada ao conjunto X_{d_1} . Assim, em cada etapa, a variável escolhida é aquela que, quando adicionada ao conjunto atual, maximiza o critério de seleção. O conjunto de variáveis é inicializado como sendo o conjunto vazio. Quando a melhor escolha torna o conjunto o conjunto de variáveis pior, ou quando o número máximo permitido de variáveis é atingido, o algoritmo termina. A principal desvantagem deste método é que ele não inclui um mecanismo de exclusão de variáveis do conjunto de variáveis, isto é, uma vez incluída, a variável permanece no subconjunto, mesmo que com a adição de novas variáveis seu poder discriminante fique negligível.

“*Generalised sequential forward selection*”: ao invés de adicionar uma variável por vez ao conjunto, neste algoritmo são adicionadas r variáveis por vez. Dado um conjunto com d_1 variáveis (X_{d_1}). Todos os possíveis conjuntos de tamanho r são gerados a partir das

restantes $n - d_1$ variáveis, isto dá $\binom{n - d_1}{r}$ conjuntos. Para cada conjunto de r variáveis a função custo (Y_r) é calculada para $X_{d_1} + Y_r$, e o conjunto que maximiza a função custo é

usado para incrementar o conjunto de variáveis. Isto é mais dispendioso que o algoritmo SFS, mas tem a vantagem de que a cada estágio é possível calcular algum grau de relação estatística entre as medidas disponíveis.

“Sequential Backward Selection (SBS)”: é o análogo inverso do SFS. Na busca de um subconjunto com d variáveis, inicia-se com todas as variáveis disponíveis. A cada etapa, uma variável, aquela que menos contribui ao poder de separação das classes é eliminada até que restem d variáveis. Este processo tem a desvantagem de ser mais dispendioso computacionalmente em comparação com o SFS, pois a função J precisa ser calculada sobre grandes conjuntos de variáveis.

“Generalised sequential backward selection (GSBS)”: por analogia, este método é essencialmente o mesmo que do SBS, porém, este método remove mais que uma variável por vez.

“Plus l – take away r selection”: neste método se inicia o processo com um conjunto de variáveis d , a este conjunto então são adicionadas l variáveis através do método SFS. A partir deste novo conjunto, composto pelas variáveis originais mais as adicionadas ($d \cup l$), r variáveis são descartadas utilizando o método SBS.

“Generalised Plus l – take away r selection”: esta é a versão generalizada do algoritmo descrito anteriormente. Este método, no entanto, utiliza a versão GSFS para adicionar l variáveis ao conjunto inicial e o GSBS para remover r variáveis do novo conjunto ($d \cup l$).

“Floating search methods”: as técnicas descritas anteriormente sofrem do chamado efeito *“nesting”*. Isto faz com que, uma vez a variável eliminada pelo método SBS ela não seja mais considerada para compor o conjunto de variáveis. O análogo ocorre para a adição de variáveis via SFS, uma vez a variável adicionada ela não poderá ser removida posteriormente. O *“Floating search methods”* proporciona um meio flexível para reconsiderar estas variáveis adicionadas/removidas.

2.2.2 EXTRAÇÃO DE VARIÁVEIS

Os métodos de extração de variáveis abordam o problema da redução da dimensionalidade dos dados de uma maneira distinta. São utilizadas transformações (lineares ou não-lineares) das variáveis originais, isto é, as variáveis são mapeadas do espaço original (X) para um novo espaço (Y). Estas transformações são selecionadas de forma que no novo

espaço o poder de separação fique concentrado em um número menor de variáveis, permitindo assim uma diminuição na dimensionalidade dos dados, com uma perda mínima de informações. Na literatura encontram-se citados vários métodos para extração de variáveis, sendo os mais conhecidos: transformação de componentes principais e análise canônica.

2.2.2.1 TRANSFORMAÇÃO DE COMPONENTES PRINCIPAIS

Em dados reais, as variáveis normalmente apresentam um grau de correlação, isto é, as variáveis apresentam um certo grau de redundância nos dados. Este fato resulta em que o aumento na dimensionalidade dos dados não é acompanhado por um correspondente ganho de novas informações. A transformação de componentes principais busca contornar este problema mapeando as variáveis para um novo espaço, no qual a correlação entre as variáveis é nula. Esta é uma transformação linear, e pode ser representada, convenientemente, na forma matricial:

$$Y = AX \tag{2.1}$$

onde, X representa o espaço original, Y o novo espaço e A a matriz que implementa a transformação.

Neste novo espaço (Figura 9), os dados não estão correlacionados. A primeira componente fica alinhada segundo a direção de maior dispersão dos dados (maior variância) e as demais formando um sistema ortogonal. Esta transformação está ilustrada na Figura 9, para um caso bi-dimensional, com a elipse ilustrando a dispersão dos dados e X uma amostra genérica, com coordenadas (x_1, x_2) no espaço original e (y_1, y_2) no novo espaço.

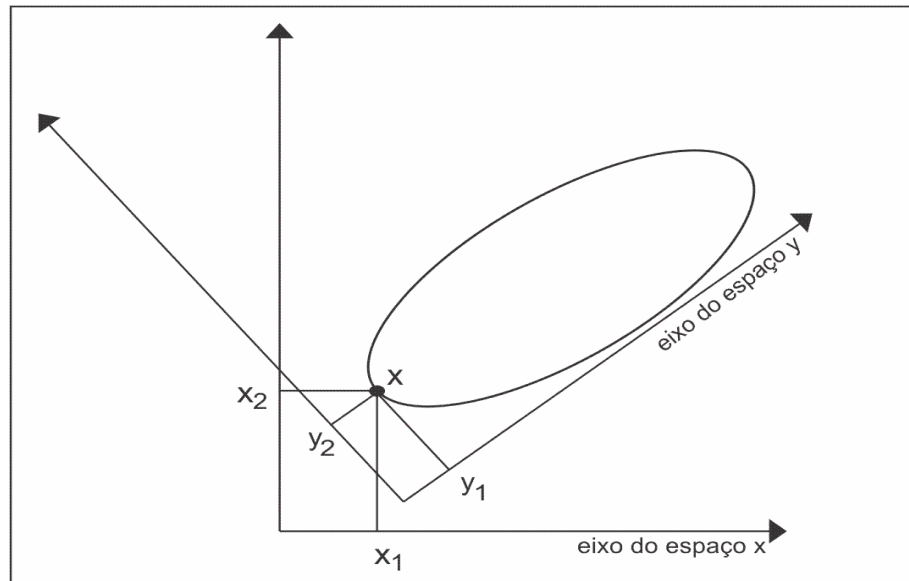


Figura 9 – Distribuição hipotética dos dados no espaço de atributos bi-dimensional
 Fonte: Adaptado de Landgrebe (2003)

Em transformações desta natureza, a matriz de covariância para as variáveis no novo espaço (Σ_y) e aquela no espaço original (Σ_x) estão relacionadas por:

$$\Sigma_y = A\Sigma_x A^T \quad (2.2)$$

Portanto, a matriz de transformação A deve ser tal que a matriz de covariância Σ_y resulte diagonal. Pode-se mostrar que este objetivo pode ser alcançado com uma matriz A cujas colunas sejam compostas pelos autovetores de Σ_x caso em que a matriz de covariância Σ_y resultará diagonal, com os elementos ao longo da diagonal principal dados pelos correspondentes autovalores de Σ_x . Os autovalores de Σ_x representam, portanto, as variâncias ao longo dos novos eixos no espaço Y e os novos eixos ficam alinhados segundo a direção dos autovetores de Σ_x . Neste novo espaço, pode-se então selecionar um subconjunto de variáveis que contenham as maiores variâncias, isto é, a maior quantidade de informação, descartando os demais. Este processo permite, assim, a redução na dimensionalidade dos dados com uma perda muito pequena de informação.

Para fins de classificação de dados, a transformação de componentes principais utilizada para redução na dimensionalidade dos dados, pode apresentar alguns problemas. Normalmente, a matriz de covariância no espaço original (Σ_x) é estimada para a totalidade dos dados na imagem. Este fato resulta que a transformação de componentes principais não

tem como critério otimizar a separação entre as classes. Em outras palavras, esta transformação maximiza o sinal, mas não a separação entre as classes. Este fato está ilustrado na Figura 10 [FUK 90], onde são apresentadas duas classes (homens, mulheres) em um espaço bi-dimensional (peso, altura). Pode-se observar, que na primeira componente (y_1), aquela que apresenta a maior variância, o sinal é maximizado, mas não a separação entre as duas classes. Neste exemplo, é a segunda componente (y_2) que apresenta maior poder discriminante e, portanto, a que deve ser utilizada para fins de classificação.

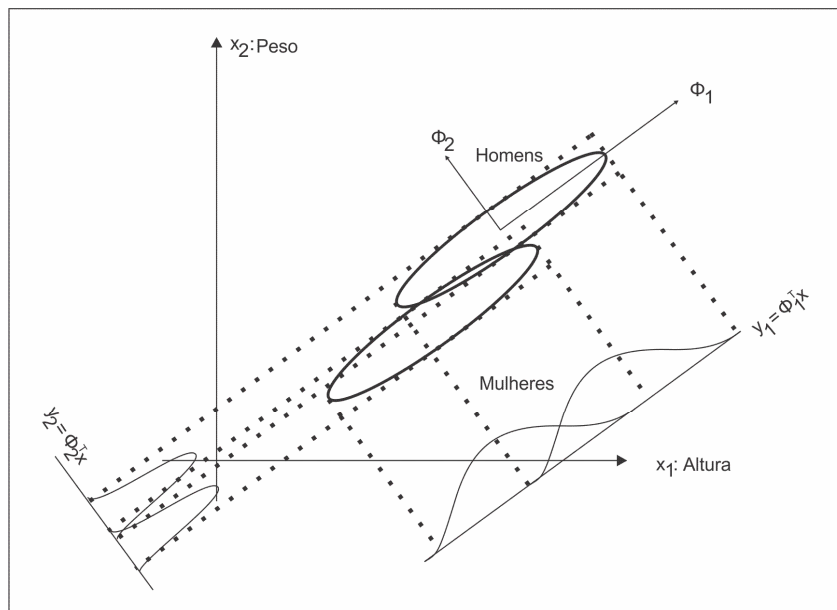
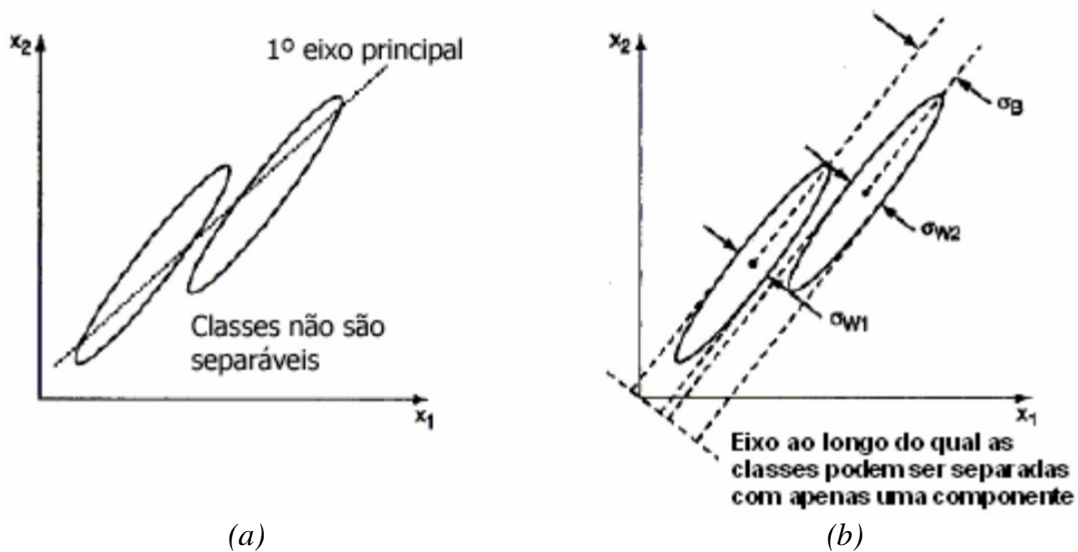


Figura 10 – Exemplo de extração de variáveis para classificação.
Fonte: Adaptado de Fukunaga (1990).

2.2.2.2 ANÁLISE CANÔNICA

Uma possível solução a este problema consiste na transformação de análise canônica. Neste caso, a matriz de transformação A implementa um critério que busca maximizar a separação entre as classes [LAN 03], conforme ilustrado na Figura 11.



(a) (b)
 Figura 11 – Distribuição hipotética de duas classes no espaço dos atributos bi-dimensional.
 Fonte: Adaptado de Landgrebe (2003).

Esta simples situação hipotética bi-dimensional será usada para ilustrar uma abordagem para reduzir a dimensionalidade de 2 para 1, enquanto realça a separabilidade entre as classes. Na Figura 11a se observa que as classes não são separáveis empregando-se uma única variável no espaço original, nem com as variáveis obtidas pela transformação de componentes principais. No entanto, na Figura 11b, se observa que as classes podem ser separadas se o eixo (em pontilhado) puder ser encontrado. O problema então é encontrar este eixo.

Intuitivamente pode-se observar que o critério a ser adotado aqui consiste em encontrar uma direção na qual (1) a distância entre as médias de cada classe seja maximizada, e (2) a dispersão (variância) seja minimizada. Este critério pode ser expresso pela maximização da expressão:

$$\frac{\Sigma_B^2}{\Sigma_W^2} = \frac{\text{variância entre classes}}{\text{variância média intra-classe}} \quad (2.3)$$

com Σ_W como a média de Σ_{W1} e Σ_{W2} . Na forma matricial a matriz de espalhamento intra-classes (Σ_W) e a matriz de espalhamento entre - classes (Σ_B) pode ser definida como:

$$\begin{aligned}\Sigma_W &= \sum_i P(\omega_i) \Sigma_i \\ \Sigma_B &= \sum_i P(\omega_i) (\mu_i - \mu_0) (\mu_i - \mu_0)^T \\ \mu_0 &= \sum_i P(\omega_i) \mu_i\end{aligned}\tag{2.4}$$

onde μ_i, Σ_i e $P(\omega_i)$ são o vetor de médias, a matriz de covariância e a probabilidade a priori da classe ω_i respectivamente. O critério de separabilidade (otimização) pode ser definido como

$$J = \text{tr}(\Sigma_W^{-1} \Sigma_B)\tag{2.5}$$

onde são selecionados os vetores de forma a maximizar o critério J .

Estudos relacionados com técnicas que envolvem redução da dimensionalidade obtiveram resultados notáveis, como as propostas por Jia & Richards (1999), Jimenez & Landgrebe (1999), Serpico & Bruzzone (2001), Wang (2008), Zhong & Wang (2008) e Zortea et alii. (2007), por exemplo.

2.3 ANÁLISE DISCRIMINANTE REGULARIZADA (RDA)

As técnicas de Análise Discriminante Regularizada se constituem em uma segunda possível via para fins de reduzir 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 se procura obter estimativas mais confiáveis para os parâmetros do classificador, introduzindo ou forçando uma tendenciosidade adequada em sua estimação, de forma que as estimativas sejam mais plausíveis fisicamente [FRI 89]. 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 (1996) mostram que as técnicas de regularização obtiveram alto sucesso na solução de

problemas de mal-condicionamento da matriz de covariância. Dentre as pesquisas recentes, destacam-se as feitas por Aeberhard et alii. (1994), Haste et alii. (1995), Hoffbeck & Landgrebe (1996), Cortijo & De La Blanca (1999), Kuo & Chang (2007) e Berge et alii. (2007).

Com um número limitado de amostras de treinamento, as estimativas da matriz de covariância das classes individuais, podem não refletir adequadamente a estrutura das classes. Nesses casos, uma combinação linear da matriz de covariância das classes e da matriz de covariância combinada pode fornecer uma melhor estimativa da subjacente estrutura da classe, e assim, conduzir a uma melhor classificação. Para isto, inicialmente, se substitui a estimativa amostral da matriz de covariância (Σ_k) da classe pela combinação linear ($\Sigma_k(\lambda)$) (equação 2.6) entre esta matriz e a matriz de covariância combinada (Σ).

$$\Sigma_k(\lambda) = \frac{(1-\lambda)Q_k + iQ}{N_k(i)}, \quad 0 \leq \lambda \leq 1 \quad (2.6)$$

com

$$N_k(\lambda) = (1-\lambda)N_k + \lambda N \quad (2.7)$$

e

$$Q = \sum_{k=1}^K Q_k, \quad Q_k = N_k \Sigma_k \quad (2.8)$$

aonde Σ_k é definido pela equação 2.9:

$$\Sigma_k = \frac{1}{N_k} \sum_{i=1}^{N_k} (X_i - \mu_k)(X_i - \mu_k)^T \quad (2.9)$$

com μ_k sendo a média amostral, N_k o número de amostras de treinamento na classe k , N é o número total de amostras de treinamento. A equação (2.7) ajusta o número de objetos associados com a matriz de covariância regularizada como uma combinação linear do número de amostras de treinamento na classe ω_k e o número total de amostras de treinamento. O

parâmetro $\lambda (0 \leq \lambda \leq 1)$ controla o grau de regularização. O valor $\lambda = 0$ resulta em uma discriminação quadrática, enquanto para $\lambda = 1$ a discriminação é linear. Valores intermediários resultam da combinação destes dois.

Para limitadas amostras de treinamento, é sabido, que as estimativas dos autovalores são tendenciosas, ou seja, com os menores autovalores sendo estimados muito pequenos, e os maiores autovalores muito grandes. A equação (2.10) tenta ajustar este viés através de um escalonamento da matriz de covariância com relação à matriz identidade multiplicada pela média do autovalor de $(\Sigma_k(\lambda))$.

$$\Sigma_k(\lambda, \gamma) = (1 - \gamma)\Sigma_k(\lambda) + \frac{\gamma}{d} \text{tr}[\Sigma_k(\lambda)]I, \quad 0 \leq \gamma \leq 1 \quad (2.10)$$

aonde $\text{tr}[\Sigma_k(\lambda)]$ é o traço de $\Sigma_k(\lambda)$, d a dimensionalidade e $\gamma (0 \leq \gamma \leq 1)$ um parâmetro controlando a segunda regularização. Após aplicar ambas as regularizações e usar $\Sigma_k(\lambda, \gamma)$ ao invés de Σ_k , se tem o valor classificativo c_k , dado por:

$$c_k(X_i) = (X_i - \mu_k)^T \Sigma_k^{-1}(\lambda, \gamma)(X_i - \mu_k) + \ln|\Sigma_k(\lambda, \gamma)| - 2\ln(P(\omega_k)) \quad (2.11)$$

Os valores ótimos para λ e γ não são conhecidos, e são escolhidos conjuntamente, de modo que a probabilidade de classificação correta seja maximizada. Isto é alcançado através da avaliação dos resultados deste classificador para um conjunto de pares (λ, γ) , e então escolhendo o par que resulta no melhor desempenho.

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. Amostras semi-rotuladas são pixels na imagem, cuja pertinência a uma das classes que ocorrem na cena pode ser inferida. Estas amostras são então adicionadas de uma forma ponderada ao conjunto das amostras de treinamento disponíveis, resultando em estimativas mais confiáveis para os parâmetros do classificador.

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 fazendo uso de amostras semi-rotuladas, 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.

Liczbinski e Haertel (2008) propõem um novo método de adquirir amostras semi-rotuladas para o processo de classificação de imagens digitais. A principal característica nesta nova proposta consiste na estimação simultânea dos pesos para amostras semi-rotuladas e das probabilidades *a priori* associadas às classes presentes na cena.

Possivelmente uma das propostas mais interessantes nesta linha que implementa o uso de amostras semi-rotuladas é aquela devida a Shahshahani e Landgrebe (1994) que propõem usar o algoritmo *Expectation Maximization* (EM) [DUD 01] para incluir mostras não rotuladas no processo de estimação dos parâmetros. A idéia básica consiste em assumir o conjunto de pixels em uma imagem como uma mistura de amostras das várias classes presentes na cena, todas elas normalmente distribuídas. O problema neste caso consiste em estimar os valores dos parâmetros para cada uma das classes que compõe esta mistura (vetores de médias, matrizes de covariância e probabilidades *a priori*). Neste algoritmo, parte-se de valores iniciais aproximados para os parâmetros de cada classe, estes são atualizados em um processo iterativo que emprega amostras disponíveis, isto é, pixels presentes na imagem (rotulados e não-rotulados). Esta abordagem permite, portanto, a inclusão de amostras não-rotuladas na estimação dos parâmetros do classificador, contribuindo desta forma para valores estimados mais acurados, e conseqüentemente em imagens temáticas mais confiáveis. O algoritmo EM consiste basicamente de dois passos principais, o *Expectation* seguido do *Maximization* [LUN 04]:

- passo E: encontram-se os valores esperados das estatísticas para o conjunto de dados completos, dado o conjunto de dados incompletos e as estimativas atuais dos parâmetros;
- passo M: utilizam-se estas estatísticas para fazer uma estimativa de máxima verossimilhança para os novos parâmetros.

O algoritmo EM pode ser sumarizado no seguinte conjunto de equações:

$$P(\omega_i) = \frac{1}{n} \sum_{k=1}^n P(\omega_i | \mathbf{X}_k, \boldsymbol{\theta}) \quad (2.12)$$

$$\boldsymbol{\mu}_i = \frac{\sum_{k=1}^n P(\omega_i | \mathbf{X}_k, \boldsymbol{\theta}) \mathbf{X}_k}{\sum_{k=1}^n P(\omega_i | \mathbf{X}_k, \boldsymbol{\theta})} \quad (2.13)$$

$$\boldsymbol{\Sigma}_i = \frac{\sum_{k=1}^n P(\omega_i | \mathbf{X}_k, \boldsymbol{\theta}) (\mathbf{X}_k - \boldsymbol{\mu}_i)(\mathbf{X}_k - \boldsymbol{\mu}_i)'}{\sum_{k=1}^n P(\omega_i | \mathbf{X}_k, \boldsymbol{\theta})} \quad (2.14)$$

$$P(\omega_i | \mathbf{X}_k, \boldsymbol{\theta}) = \frac{p(\mathbf{X}_k | \omega_i, \boldsymbol{\theta}_i) P(\omega_i)}{\sum_{j=1}^{cl} p(\mathbf{X}_k | \omega_j, \boldsymbol{\theta}_j) P(\omega_j)} \quad (2.15)$$

$$= \frac{|\boldsymbol{\Sigma}_i|^{-1/2} \exp\left[-\frac{1}{2}(\mathbf{X} - \boldsymbol{\mu}_i)' \boldsymbol{\Sigma}_i^{-1} (\mathbf{X} - \boldsymbol{\mu}_i)\right] P(\omega_i)}{\sum_{j=1}^{cl} |\boldsymbol{\Sigma}_j|^{-1/2} \exp\left[-\frac{1}{2}(\mathbf{X} - \boldsymbol{\mu}_j)' \boldsymbol{\Sigma}_j^{-1} (\mathbf{X} - \boldsymbol{\mu}_j)\right] P(\omega_j)}$$

com $\boldsymbol{\mu}_i$, $\boldsymbol{\Sigma}_i$, $P(\omega_i)$ representando respectivamente o vetor de médias, a matriz de covariância e a probabilidade *a priori* associados à classe ω_i , o vetor $\boldsymbol{\theta}_i$ representando os parâmetros da classe ω_i , cl o número de classes presentes na mistura, e n o número de amostras obtidas da mistura.

O processo iterativo inicia com a estimação dos parâmetros do classificador a partir das amostras rotuladas disponíveis. Estas estimativas iniciais são inseridas na equação (2.15) que é por sua vez é inserida nas equações (2.12), (2.13) e (2.14), produzindo estimativas mais confiáveis para os parâmetros, devido à inclusão de amostras adicionais, não-rotuladas. Estes novos valores são inseridos novamente na equação (2.15) e o processo iterativo segue, produzindo a cada etapa valores mais confiáveis para os parâmetros. Esta abordagem foi proposta por Shashahani e Landgrebe (1994) com resultados bastante positivos.

2.5 INSERÇÃO DE INFORMAÇÕES DE CONTEXTO ESPACIAL NO PROCESSO DE CLASSIFICAÇÃO DE IMAGENS DIGITAIS

As informações que podem ser extraídas em imagens digitais podem ser agrupadas em duas categorias: (1) informações espectrais, (2) informações espaciais (textura e contexto espacial). Nas imagens multiespectrais empregadas em sensoriamento remoto, os atributos espectrais têm sido os mais freqüentemente empregados. Esta preferência deve-se entre outros fatores, a facilidade com que este tipo de informação pode ser quantificado numericamente para implementação em ambiente de computador.

Conforme Chen (2008), os atributos do tipo espacial são inerentes em dados de sensoriamento remoto. Basicamente o efeito do contexto no problema de classificação de imagens é que quando um pixel é considerado isoladamente, este pode fornecer informações incompletas sobre as características desejadas, mas ao considerar o pixel em seu contexto, informações mais completas podem ser obtidas. No entanto, este tipo de atributo (espacial) é de quantificação mais complexa, quando comparado com atributos puramente espectrais.

Informações do tipo espacial, ou de contexto espacial, fazem uso da correlação que ocorre entre as classes em cenas naturais. A classe à qual um pixel pertence e as classes de pixels vizinhos apresentam uma dependência que pode ser explorada para fins de incrementar a acurácia do processo de classificação. Em outras palavras, o classificador ao rotular um pixel deve levar em conta as decisões tomadas ao rotular pixels vizinhos.

O uso de atributos de contexto espacial vem sendo explorado desde meados de 1970, mas só mais recentemente é que tem recebido uma atenção maior. Como este tipo de atributo apresenta um grande potencial para a classificação, e/ou refinamento deste processo, a principal questão é como utilizar esta informação da melhor maneira possível. Se essa dependência puder ser estimada de alguma maneira para um dado conjunto de dados esta informação poderá incrementar a acurácia do processo de classificação. Métodos para inserir esse tipo de informação em imagens classificadas foram discutidos por alguns autores como Therrien (1989). Os dois processos mais freqüentemente empregados na extração de informações de contexto espacial são os processos de relaxação e os campos randômicos de Markov.

2.5.1 RELAXAÇÃO

Conforme Richards (2006), o processo de relaxação começa assumindo que uma classificação, baseada somente em informações espectrais, já foi realizada. Desta forma, está disponível um conjunto de probabilidades, para cada pixel, que descreve as chances de um pixel pertencer a cada um das classes em consideração. Este conjunto de probabilidades pode ser inicialmente computado utilizando o classificador MVG, por exemplo, e é representado, para um determinado pixel (m), por:

$$P_m(\omega_i), \quad i = 1, \dots, J \quad (2.16)$$

onde J é o número total de classes, ω_i é a classe em questão, e $P_m(\omega_i)$ é a probabilidade de que ω_i seja a classe correta para o pixel m .

Para iniciar o processo de relaxação deve-se então determinar uma vizinhança em torno de um pixel (m), que pode ser de qualquer tamanho, e a princípio, deva ser grande o suficiente para assegurar que todos os pixels considerados (vizinhos) tenham qualquer correlação espacial com o pixel m em questão. Também deve ser possível encontrar uma função vizinhança ($Q_m(\omega_i)$) que permita aos pixels vizinhos influenciarem a possível classificação do pixel m .

Esta função vizinhança deve levar em conta uma medida de compatibilidade entre a classificação atual do pixel m e a classificação dos pixels adjacentes. Seja esta medida de compatibilidade chamada $r_{mm}(\omega_i, \omega_j)$, que descreve quão compatível é ter o pixel m classificado na classe ω_i , enquanto que, ao mesmo tempo se tem o pixel n classificado na classe ω_j ([HSI 89] e [EKL 80]). Portanto, é de se esperar que, esta medida seja alta se pixels adjacentes são rotulados na mesma classe, ao passo que se estes pixels forem rotulados em classes diferentes esta medida deva ser baixa.

Existem vários métodos de definir esta medida de compatibilidade ([HSI 89] e [ZEN 89]), porém, intuitivamente a mais atraente é aquela dada pelas probabilidades condicionais. Assim o coeficiente de compatibilidade passa a medir a probabilidade do pixel m ser atribuído corretamente a classe ω_i ao mesmo tempo em que o pixel n é atribuído a classe ω_j . Uma pequena indicação do pixel m pertencer à classe ω_i é dada por $P_m(\omega_i/\omega_j)P_n(\omega_j)$, enquanto que a evidência total do pixel n pertencer à classe ω_j é dada por:

$$\sum_j P_m(\omega_i/\omega_j) P_n(\omega_j) \quad (2.17)$$

Assim, de forma a todos os vizinhos contribuírem para a classificação do pixel m na classe ω_i , basta adicionar as contribuições de forma ponderada, conforme:

$$Q_m^k(\omega_i) = \sum_n d_n \sum_j P_m(\omega_i/\omega_j) P_n^k(\omega_j) \quad (2.18)$$

aonde k é a iteração, e d_n o coeficiente ponderador da vizinhança. Este coeficiente reconhece que alguns pixels vizinhos podem ser mais influentes que outros. Desta forma está definida a função vizinhança ($Q_m(\omega_i)$).

A influência da vizinhança é então, finalmente, calculada multiplicando as probabilidades da classe pela função vizinhança. No entanto, esse novo conjunto deve ser normalizado para que as probabilidades somem um, logo esses novos valores são divididos pela sua soma (equação 2.19):

$$P_m^i(\omega_i) = \frac{P_m(\omega_i) Q_m(\omega_i)}{\sum_j P_m(\omega_j) Q_m(\omega_j)} \quad (2.19)$$

tal cálculo é realizado sobre todos os pixels da imagem. Conseqüentemente, a equação (2.19) é aplicada quantas vezes forem necessárias para que a $P_m^i(\omega_i)$ tenha estabilidade, ou seja, atualizando os valores até que não mude com mais iterações. Assume-se então que $P_m^i(\omega_i)$ representa o conjunto correto de probabilidades para o pixel m , tendo tomado em conta ambos os tipos de informações, espectral para a determinação da probabilidade inicial e espacial via função vizinhança. Uma vez que o processo é iterativo, a equação (2.19) pode ser escrita na forma:

$$P_m^{k+1}(\omega_i) = \frac{P_m^k(\omega_i) Q_m^k(\omega_i)}{\sum_j P_m^k(\omega_j) Q_m^k(\omega_j)} \quad (2.20)$$

Dependendo do tamanho da imagem e de sua complexidade espacial o número de iterações necessárias para a convergência do processo pode ser demasiadamente grande, para contornar este problema se deve adotar um critério de convergência adicional.

A relaxação probabilística foi investigada por Di Zeno et al. (1987) para incrementar a acurácia em classificações temáticas, outros autores como Rosenfeld et alii (1976), Zucker et alii (1978), Eklundh et alii (1980), Hummel e Zucker (1983), Kalayeh e Landgrebe (1984), Haertel e Moraes (2009) também investigaram o uso da relaxação pra fins de inclusão de informação de contexto espacial no processo de classificação de imagens digitais.

Kalayeh e Landgrebe (1984) propõem o uso do coeficiente de compatibilidade não fixo (NSCC, onde o coeficiente de compatibilidade é calculado sobre uma vizinhança do pixel em questão) na implementação da relaxação probabilística para efeito de classificação. Esta proposta se deve ao fato de que o uso do coeficiente de compatibilidade fixo (SCC, onde o coeficiente de compatibilidade é calculado sobre todos os pixels da imagem) freqüentemente diverge na classificação a partir de um erro mínimo de classificação alcançado, enquanto que, ao se usar o NSCC a classificação estabiliza a partir de um erro mínimo, geralmente alcançado nas primeiras iterações.

Haertel e Moraes (2009) propõem um passo adicional à metodologia proposta por Jackson e Landgrebe (2001) para selecionar e ponderar amostras semi-rotuladas fazendo uso de informações de contexto espacial via relaxação probabilística, para estimação mais confiável dos parâmetros do classificador paramétrico MVG.

Outro tipo de abordagem freqüentemente empregada para a inserção de informação de contexto espacial em processos de classificação refere-se aos campos aleatórios de Markov (Markov Random Fields, MRF). Os MRF são uma útil ferramenta para caracterizar a informação de contexto, sendo ainda amplamente empregado em segmentação e restauração de imagens.

2.5.2 CAMPOS ALEATÓRIOS DE MARKOV

Conforme Tso e Mather (2001), seja um conjunto de variáveis aleatórias $D = \{D_1, D_2, \dots, D_m\}$ definidas sobre o conjunto S de todos os pixels da imagem, onde cada variável aleatória D_i assume um valor d_i sobre o conjunto das classes $L = \{\omega_1, \dots, \omega_k\}$. O valor d_i é o evento da variável aleatória D_i ser classificada como pertencente algumas das

classes em L , e o conjunto $D = \{D_1 = d_1, D_2 = d_2, \dots, D_m = d_m\}$ é o evento conjunto de tais variáveis aleatórias, e define a configuração de um campo aleatório. Desta forma o conjunto D é chamado de campo aleatório de Markov, em relação a uma vizinhança, se satisfaz as seguintes três propriedades:

1. Positividade: $P(D) > 0$ para todas as configurações possíveis de D ;
2. Markovianidade: $P(d_r | d_{S-\{r\}}) = P(d_r | d_{N_r})$, e;
3. Homogeneidade: $P(d_r | d_{S-\{r\}})$ para todos os lugares de r .

com $S - \{r\}$ sendo o conjunto diferença (ou seja, todos os pixels no conjunto S excluindo r), $d_{S-\{r\}}$ denota o conjunto de classes dos pixels de $S - \{r\}$, e N_r a vizinhança do pixel r .

A primeira propriedade (positividade), na prática geralmente é satisfeita, e a probabilidade conjunta $P(D)$ pode ser determinada pelas propriedades condicionais locais. A Markovianidade indica que a classificação do pixel na posição r é unicamente dependente de sua vizinhança, enquanto a última propriedade (homogeneidade) especifica a probabilidade condicional da classificação de um pixel na posição r , dada a classificação dos pixels de sua vizinhança, independentemente da posição do pixel r dentro da imagem (S).

Autores como Jhung e Swain (1996), Solberg et alii. (1996) e Solberg (1999), Tso e Mather (1999), Jackson e Landgrebe (2002), Melgani e Serpico (2003) e Jia e Richards (2008) tem desenvolvido pesquisas nesta área.

Solberg et alii. (1996) propõem um modelo geral para classificação multi-fonte de imagens de sensoriamento remoto baseado em campos aleatórios de Markov, bem como também é apresentado um modelo específico para fusão de dados provenientes de imagens ópticas, imagens SAR e GIS. O MRF explora a dependência espacial da classe entre pixels vizinhos (informação de contexto espacial), e a dependência temporal de classes entre diferentes imagens da mesma área. Os resultados dos experimentos deste trabalho foram promissores para o modelo MRF, indicando que o modelo proposto é útil para a classificação multi-fonte de imagens de satélite.

Jackson e Landgrebe (2002) desenvolveram um classificador Bayesiano contextual, que utiliza tanto a informação espectral como a dependência interpixel espacial na estimação das estatísticas e na classificação. Neste classificador, as probabilidades conjuntas das classes de cada pixel e seus vizinhos são modelados por campos aleatórios de Markov.

Jia e Richards (2008) apresentam um método simples para classificação de imagens digitais que consiste em ponderar as contribuições provenientes de informações espectrais e de contexto espacial, conjuntas, através de campos randômicos de Markov.

3 METODOLOGIA

3.1 INTRODUÇÃO

O processo de classificação de dados em imagens de cenas naturais consiste em rotular cada pixel, individualmente, conforme definido pelo seu vetor de observações, em uma das classes existentes. Sistemas sensores produzindo um número limitado de bandas espectrais como, por exemplo, os sistemas LANDSAT-TM e SPOT vêm sendo amplamente utilizados para fins de obtenção de informações em áreas diversas tais como florestas, agricultura, geologia, além de várias outras. Na maioria destes casos, as classes de interesse apresentam aproximadamente uma distribuição normal multivariada e o processo de classificação pode ser feito utilizando o bem conhecido classificador quadrático Máxima Verossimilhança Gaussiana (MVG). Este classificador estatístico pode ser entendido como um caso particular do classificador de Bayes (3.1).

$$G_i(X) = p(X / \omega_i) P(\omega_i) \quad (3.1)$$

sendo X o pixel a ser rotulado, $p(X/\omega_i)$ a função densidade de probabilidade condicional à classe ω_i , e $P(\omega_i)$ a probabilidade *a priori* associada à mesma classe.

O processo de classificação consiste em rotular o pixel na classe ω_j conforme a regra de classificação:

$$X \in \omega_j \quad \text{se} \quad G_j(X) > G_i(X) \quad \forall i \neq j \quad (3.2)$$

Em situações práticas a estimação da probabilidade *a priori* $P(\omega_i)$ pode se constituir em um problema difícil. Por esta razão, é comum se supor valores idênticos para esta probabilidade para todas as classes em consideração. Esta simplificação permite, portanto eliminar o fator $P(\omega_i)$ em (3.1) resultando na função de decisão denominada de Máxima Verossimilhança (*maximum likelihood*). Supondo-se ainda que os dados apresentem uma distribuição Normal Multivariada, esta função de decisão passa a ser expressa por:

$$G_i(X) = \frac{1}{(2\pi)^{p/2}} \frac{1}{|\Sigma_i|^{1/2}} \exp \left[-\frac{1}{2} (X - \mu_i)^T \Sigma_i^{-1} (X - \mu_i) \right] \quad (3.3)$$

$$G_i(X) \propto \frac{1}{|\Sigma_i|^{p/2}} \exp\left[-\frac{1}{2}(X - \mu_i)^T \Sigma_i^{-1} (X - \mu_i)\right] \quad (3.4)$$

sendo μ_i e Σ_i respectivamente o vetor de médias e a matriz de covariância associados à classe ω_i , p a dimensionalidade dos dados. Dado p fixo, o termo $1/(2\pi)^{p/2}$ se torna comum para todas as classes na equação (3.3) e pode ser cancelado, obtendo assim a equação (3.4).

Uma versão computacionalmente mais adequada de (3.4) pode ser obtida empregando-se o logaritmo natural da mesma:

$$G_i(X) = -\frac{1}{2}(X - \mu_i)^T \Sigma_i^{-1} (X - \mu_i) - \frac{1}{2} \ln |\Sigma_i| \quad (3.5)$$

Deve-se aqui observar que a função logaritmo é uma função monotonicamente crescente. Desta forma, o resultado da regra de classificação (3.3) empregando a função de decisão (3.4) não fica alterado.

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). Pode-se mostrar que classes que apresentam características muito semelhantes, isto é, que compartilham do mesmo vetor de médias podem, não obstante, ser separadas com um alto grau de acurácia em um espaço de dimensionalidade mais alta, desde que suas matrizes de covariância difiram entre si, substancialmente [FUK 90].

O sistema AVIRIS, por exemplo, coleta dados de cenas naturais em 224 bandas espectrais, na região do espectro eletromagnético compreendida entre 0.4 μm e 2.4 μm .

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 dimensionalidade elevada. O número de amostras de treinamento, normalmente disponíveis em situações práticas é insuficiente para uma estimativa confiável deste número crescente de

parâmetros, resultando em estimativas pouco confiáveis para o vetor de médias, e principalmente, para a matriz de covariância, o que por sua vez, compromete o desempenho do classificador e, conseqüentemente, a acurácia da imagem temática produzida.

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 certo grau de incerteza, o que se traduz pelo peso reduzido a elas atribuído no processo de estimação dos parâmetros do classificador. A proposta investigada nesta dissertação se insere na linha do uso de amostras semi-rotuladas (seção 2.4).

3.2 METODOLOGIA PROPOSTA

A metodologia proposta nesta dissertação tem por objetivo investigar o uso de amostras semi-rotuladas no processo de classificação de imagens digitais, em particular na classificação de imagens em alta dimensionalidade para as quais o número de amostras de treinamento disponíveis (amostras rotuladas) é geralmente insuficiente para uma estimativa confiável dos parâmetros do classificador. A metodologia proposta consiste de duas etapas. Em uma primeira etapa amostras semi-rotuladas são iterativamente obtidas durante o próprio processo de classificação e adicionadas de uma forma ponderada às amostras rotuladas, para fins de estimação dos parâmetros do classificador. Nesta etapa, o peso atribuído a cada amostra semi-rotulada deve expressar o grau de confiança de que este pixel efetivamente pertença à classe em questão. A cada iteração os pesos para as amostras semi-rotuladas são re-estimados a partir dos resultados da iteração anterior, produzindo uma nova estimativa para os parâmetros do classificador e uma nova imagem temática. O processo iterativo nesta etapa é encerrado quando o número de pixels que troca de classe entre duas iterações sucessivas é menor do que um limiar pré-fixado ou é atingido um número máximo de iterações.

A segunda etapa consiste no refinamento das estimativas para os pesos das amostras semi-rotuladas com base em informações de contexto espacial. A motivação básica nesta etapa consiste no fato de que em cenas naturais as diferentes classes não ocorrem de uma forma aleatória, isto é, apresentam uma correlação espacial. Se esta correlação espacial puder ser estimada, então o grau de pertinência de cada pixel às classes presentes, isto é, o peso, pode ser avaliado em função dos correspondentes graus de pertinência de seus vizinhos. Este

procedimento permite um refinamento nos pesos das amostras semi-rotuladas e conseqüentemente em estimativas mais confiáveis para os parâmetros do classificador.

O processo de classificação final pode então ser realizado, empregando valores estimados para os parâmetros do classificador, com base em um conjunto de amostras rotuladas e semi-rotuladas, estas últimas ponderadas conforme o resultado final da segunda etapa.

As diferentes etapas da metodologia proposta encontram-se ilustradas na Figura 12. A contribuição original nesta dissertação está presente nas duas etapas do processo, sendo: (1) na metodologia empregada na estimação dos pesos em amostras semi-rotuladas, e (2) na aplicação de processos de relaxação, não no refinamento de imagens temáticas conforme já proposto anteriormente por outros autores, mas sim aplicado ao conjunto dos pesos das amostras semi-rotuladas.

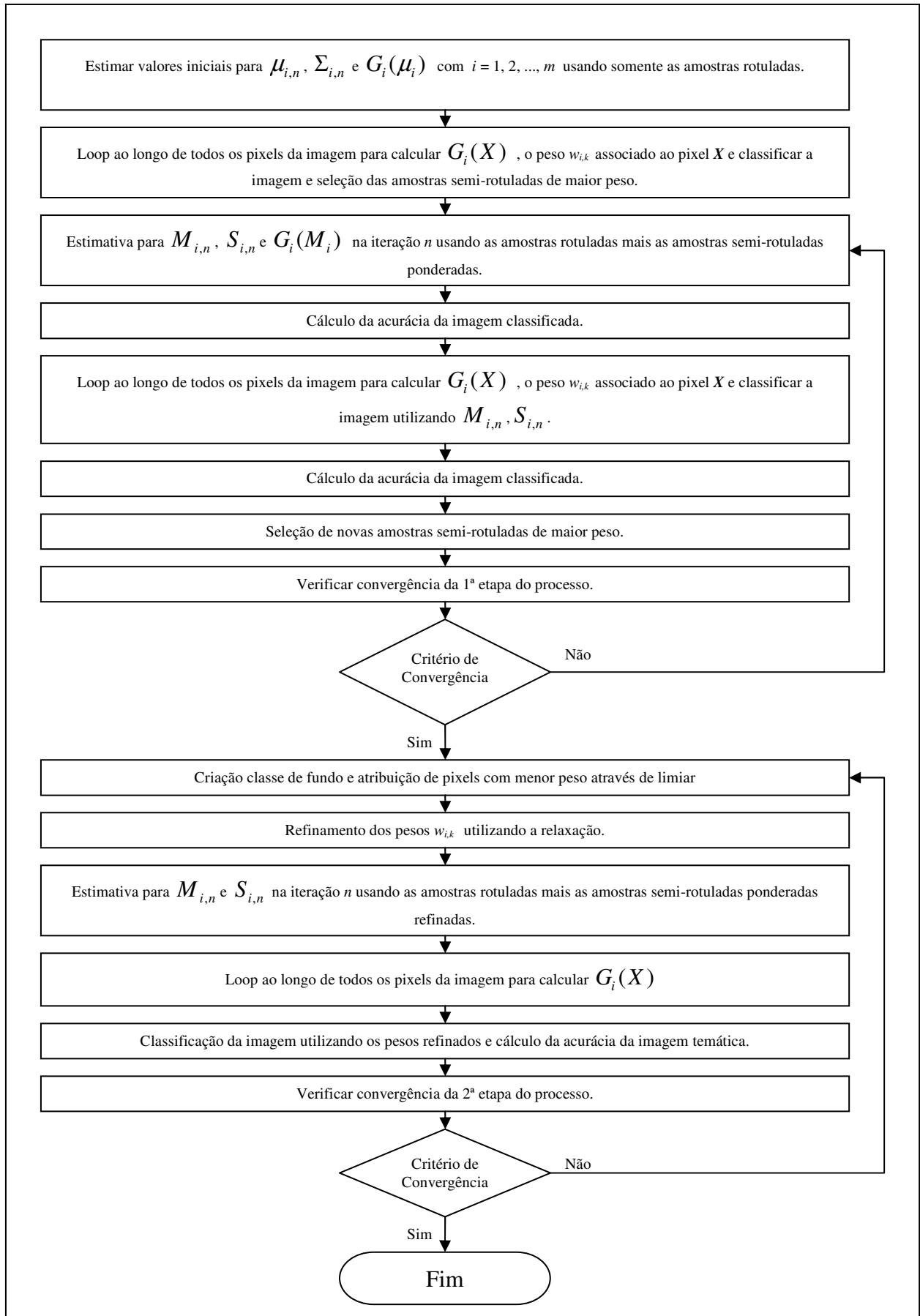


Figura 12 – Fluxograma do processo iterativo.

3.3 ESTIMAÇÃO INICIAL DE PESOS PARA AMOSTRAS SEMI-ROTULADAS

O passo inicial no processo de seleção de amostras semi-rotuladas consiste na estimação do peso de cada amostra à respectiva classe. Nesta dissertação se propõe estimar o peso pela probabilidade *a posteriori*, seguindo uma abordagem similar à proposta por [JAC 01] e [LIC 08]:

$$w_{i,k} = p(\omega_k / \mathbf{X}_i) = \frac{p(\mathbf{X}_i / \omega_k) P(\omega_k)}{\sum_{j=1}^m p(\mathbf{X}_i / \omega_j) P(\omega_j)} \quad (3.6)$$

onde $w_{i,k}$ representa o peso associado ao pixel \mathbf{X}_i com relação à classe ω_k e m o número de classes presentes na cena. Esta equação pode ser também expressa em termos das funções de decisão $G_i(X)$:

$$w_{i,k} = \frac{G_i(\mathbf{X}_k)}{\sum_{j=1}^m G_j(\mathbf{X}_k)} \quad (3.7)$$

Como citado anteriormente, na prática, esta abordagem quando implementada diretamente pode apresentar dificuldades numéricas, devido aos valores muito pequenos resultantes da parte exponencial em (3.4), gerando estimativas não confiáveis para os pesos em (3.7). Este problema é comumente evitado empregando-se a função logaritmo natural (3.5). Embora numericamente mais adequada, esta alternativa apresenta a peculiaridade de que os valores obtidos pelas funções de decisão resultam sempre negativos. Para fins exclusivos de classificação dos pixels em uma imagem, este fato não apresenta nenhum inconveniente, pois a regra de classificação (3.2) permanece inalterada: cada pixel é rotulado conforme a classe que produz o valor mais alto para a função de decisão. Como para todas as classes os valores produzidos por (3.5) serão negativos, a classe vencedora será aquela que apresentar o menor valor absoluto. Esta peculiaridade inviabiliza, entretanto, o emprego direto da equação (3.7) para fins de estimação dos pesos, isto é, os valores obtidos para os pesos $w_{i,k}$ conforme (3.7) não podem ser empregados diretamente na estimação dos graus de pertinência às classes em consideração, e portanto como pesos, para fins de ponderação de amostras semi-rotuladas. Nesta dissertação é proposta uma abordagem nova para solução deste problema: Inicialmente valores são calculados para cada pixel, em cada uma das m classes, aplicando a

expressão logarítmica da função de decisão $G_i(\mathbf{X})$ (3.5) em (3.7). Após, toma-se o complemento destes valores (Equação 3.8):

$$w_{i,k}^* = 1 - w_{i,k} \quad (3.8)$$

Embora estes novos valores para os pesos estejam corretamente seqüenciados de acordo com os graus de pertinência de cada pixel a cada uma das m classes, torna-se ainda necessário efetuar a normalização, isto é, para qualquer pixel a soma dos pesos sobre as m classes deva ser igual à unidade (Equação 3.9):

$$w_{i,k}^{**} = \frac{\left(w_{i,k}^* / \sum_{k=1}^m w_{i,k}^* \right)}{m-1} \quad (3.9)$$

O denominador $(m - 1)$ em (3.9) serve para normalizar os valores obtidos para os pesos, isto é, para um mesmo pixel, o somatório dos pesos referentes às m classes deve ser igual a 1.

$$(1 - w_{i,1}) + (1 - w_{i,2}) + \dots + (1 - w_{i,m}) = m - \underbrace{(w_{i,1} + w_{i,2} + \dots + w_{i,m})}_{1} = m - 1$$

Um problema final a ser ainda considerado se refere ao valor assumido por cada pixel individual com relação à função densidade de probabilidade Normal Multivariada (3.4). Observa-se pela expressão (3.7) que os pesos são estimados relativamente aos valores nas m funções de decisão $G_i(\mathbf{X})$. Nesta abordagem podem ocorrer situações em que o valor de $G_i(\mathbf{X})$ para a classe vencedora seja reduzido, mas aqueles para as demais $(m-1)$ classes sejam ainda menores, resultando em um valor alto para o peso da classe vencedora. Torna-se, portanto, necessário escalonar o peso da classe vencedora conforme a posição efetiva do pixel \mathbf{X}_i na distribuição daquela classe, isto é, do valor de $G(\mathbf{X})$ em termos do valor máximo desta função, o que ocorre para $\mathbf{X}_i = \mu_i$.

$$w_{i,k}^{***} = w_{i,k}^{**} \frac{G_i(\mathbf{X})}{G_i(\mu_i)} \quad (3.10)$$

sendo M_i a estimativa amostral para o vetor de médias μ_i da classe vencedora (ω_i). A equação (3.10) exprime a forma final proposta nesta dissertação para estimação dos pesos em amostras semi-rotuladas. Para fins de simplificação da notação, no restante desta dissertação os pesos $w_{i,k}^{***}$ passam a ser representados simplesmente por $w_{i,k}$.

Os parâmetros do classificador passam, portanto, a serem estimados por um conjunto de amostras de treinamento, que inclui amostras rotuladas (peso unitário) e um determinado número de amostras semi-rotuladas, com peso estimado pelo procedimento acima descrito. As expressões empregadas para as estimativas amostrais do vetor de médias e da matriz de covariância (M_i , S_i) são dadas por [LIC 08] (Equação 3.11):

$$M_i = \frac{1}{\sum_{j=1}^{N_i} w_{i,j}} \sum_{j=1}^{N_i} w_{i,j} X_j$$

$$S_i = \frac{1}{\left(\sum_{j=1}^{N_i} w_{i,j} \right) - 1} \sum_{j=1}^{N_i} w_{i,j} (X_j - M_i)(X_j - M_i)^T$$
(3.11)

onde N_i representa o número total de amostras de treinamento empregadas (rotuladas mais semi-rotuladas) para a classe ω_i . Deve-se notar aqui, que os pesos referentes às amostras rotuladas são iguais à unidade, refletindo o grau de confiança na pertinência da amostra à classe em questão.

Na metodologia proposta estes pesos estimados são atualizados a cada iteração. A partir da segunda iteração, os parâmetros (M_i , S_i) passam a ser estimados por um conjunto composto pelas amostras rotuladas disponíveis para cada classe, acrescidas de um determinado número de amostras semi-rotuladas, estas últimas ponderadas de acordo com os correspondentes pesos obtidos na iteração anterior. Este processo se repete até que um número pré-definido de iterações seja atingido ou o número de pixels que muda de classe entre duas iterações consecutivas seja inferior a um limiar também pré-definido.

O produto final desta etapa inicial consiste, portanto, em um conjunto de estimativas para os pesos associados às possíveis amostras semi rotuladas, isto é, pixels que não fazem parte do conjunto das amostras rotuladas disponíveis para treinamento do classificador. Nesta etapa, estes pesos foram estimados individualmente para cada pixel, isto é, não são levados em consideração possíveis contribuições de vizinhança.

3.4 INTRODUÇÃO DE INFORMAÇÕES DE CONTEXTO ESPACIAL (VIZINHANÇA)

O processo iterativo desenvolvido na etapa anterior está baseado nos atributos de cada pixel considerados isoladamente. Em Reconhecimento de Padrões, existem entretanto outras fontes de informação que podem ser utilizadas para fins de classificação de pixels em uma imagem, como atributos espaciais (forma, textura) e atributos de vizinhança ou contexto espacial. Nesta dissertação é proposto o uso de informações de contexto espacial para refinar as estimativas para os pesos de possíveis amostras semi-rotuladas conforme desenvolvido na etapa anterior.

Experimentos diversos mostram a existência de uma correlação espacial entre as classes presentes na grande maioria das cenas. Este fato pode ser utilizado na estimação dos diferentes graus de probabilidade de um pixel pertencer a uma determinada classe, em função de sua vizinhança. Na literatura são citadas as duas abordagens mais freqüentemente empregadas na quantificação de informações de contexto espacial: técnicas empregando Campos Randômicos de Markov e abordagens implementando técnicas de relaxação [THE 89], [SOL 99], [ZEN 87] e [CEN 91]. Nesta Dissertação é empregado o segundo processo.

No processo geral de relaxação, parte-se de um conjunto inicial de valores, no caso presente os pesos de cada pixel com relação a cada uma das classes presentes na cena. Estes valores são iterativamente ajustados a partir de estimativas para a correlação espacial entre um pixel e sua vizinhança. Nesta abordagem, uma etapa inicial importante refere-se à estimativa das probabilidades de co-ocorrência entre as várias classes presentes na imagem em uma determinada vizinhança. A vizinhança utilizada para coletar as informações de contexto neste trabalho está ilustrada na Figura 13(b).

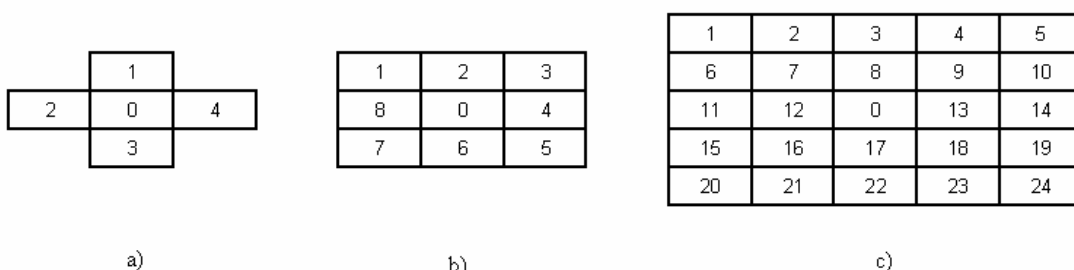


Figura 13 – Exemplos de vizinhanças mais freqüentemente empregadas (0 corresponde ao pixel em análise)

Conforme Landgrebe (1980), observa-se que pixels espacialmente próximos entre si apresentam um grau mais elevado de correlação que decresce na medida em que aumenta a

distância entre eles. De uma forma geral, a correlação cai lentamente com a distância, e uma separação da ordem de vários pixels é necessária para que a correlação atinja valores não significativos. Desta forma, os maiores valores para a correlação são encontrados em pixels adjacentes, isto justifica o uso de vizinhanças menores, como na Figura 13(b). Comumente, vizinhanças com um número ímpar de pixels são utilizadas, isto é devido ao fato que vizinhanças deste tipo são invariantes frente à rotação, além de terem um pixel central bem definido, e com isso poder atribuir a mesma importância a pixels vizinhos que se encontrem a iguais distâncias do pixel central. Assim a informação obtida a partir deste tipo de vizinhança pode ser mais apropriadamente atribuída ao pixel central.

Nas subseções seguintes são examinadas as propostas para a estimação dos graus de correlação espacial por meio do coeficiente de compatibilidade e o emprego desta informação no refinamento dos valores estimados para os pesos, em um processo iterativo de relaxação.

3.5 O COEFICIENTE DE COMPATIBILIDADE

O processo de relaxação utilizado neste trabalho tem como base o coeficiente de compatibilidade de contexto.

O coeficiente de compatibilidade, representado por $r(i,j,h,k)$, que por facilidade de notação passa aqui a ser representado por $r_{i,j,h,k}$, estima a probabilidade de que o pixel central na posição central i pertença à classe ω_h ($X_i \in \omega_h$) ao mesmo tempo em que o seu vizinho na posição j pertence à classe ω_k ($X_j \in \omega_k$).

Peleg (1978) sugere que o coeficiente de compatibilidade para os processos de relaxação contextual devam ter as seguintes propriedades:

- $r(i,j,h,k) > 0$ se (ω_h, ω_k) são compatíveis com (X_h, X_k)
- $r(i,j,h,k) < 0$ se (ω_h, ω_k) são incompatíveis com (X_h, X_k)
- $r(i,j,h,k) \approx 0$ se as classes i e j são independentes

Di Zenzo et alii. (1986) propõem a seguinte expressão para o coeficiente de compatibilidade:

$$r(i, j, h, k) = E \left[\ln \frac{p(i \in h, j \in k)}{p(i \in h)p(j \in k)} \right] \quad (3.12)$$

Nesta dissertação, a expressão utilizada para o coeficiente de compatibilidade é [HSI 89]:

$$r_{i,j,h,k} = \frac{1}{5} \ln \left(\frac{E [p(i \in h, j \in k)]}{E [p(i \in h)] E [p(j \in k)]} \right) \quad (3.13)$$

Uma estimativa amostral para os termos em (3.12) pode ser obtida a partir da imagem classificada produzida na primeira etapa desta metodologia (subseção 3.3): empregando os parâmetros estimados a partir de amostras rotuladas e semi-rotuladas:

$$r_{i,j,h,k} = \frac{1}{5} \ln \left(\frac{N(i, j, h, k) / N}{\left(\sum_{h=1}^m N(i, j, h, k) / N \right) \left(\sum_{k=1}^m N(i, j, h, k) / N \right)} \right) \quad (3.14)$$

onde $N(i, j, h, k)$ representa o número de vezes em que um pixel na posição i pertence à classe ω_h , ao mesmo tempo em que seu vizinho na posição j pertence à classe ω_k , e N o número de pixels na imagem. O termo $N(i, j, h, k) / N$ estima portanto a probabilidade conjunta $p(i \in \omega_h, j \in \omega_k)$, enquanto $\left(\sum_{h=1}^m N(i, j, h, k) / N \right)$ e $\left(\sum_{k=1}^m N(i, j, h, k) / N \right)$ estimam respectivamente as duas probabilidades marginais no denominador de (14). Em termos probabilísticos, o valor encontrado em (3.14) estima a probabilidade de ocorrência dos dois eventos simultaneamente (pixel central $i \in \omega_h$ e seu vizinho na posição $j \in \omega_k$).

Na abordagem investigada nesta dissertação assume-se que o valor estimado para o coeficiente de compatibilidade é invariante com relação à posição do pixel central X_i .

3.6 O PROCESSO DE RELAXAÇÃO

O processo iterativo de relaxação pode então ser utilizado para atualizar o valor dos pesos, em função da vizinhança selecionada [THE 89] e [ZEN 87]:

$$w_{i,h}^{t+1} = w_{i,h}^t \frac{1}{L} \left(1 + (1/n) \sum_{j=1}^n \left(\sum_{k=1}^m r_{i,j,h,k} w_{j,k} \right) \right) \quad (3.15)$$

onde t representa o número da iteração, n o número de pixels na vizinhança, m o número de classes e L um fator de normalização, dado por (Equação 3.16):

$$L = \sum_{h=1}^m \left(1 + (1/n) \sum_{j=1}^n \sum_{k=1}^m r_{i,j,h,k} w_{j,k}^t \right) \quad (16)$$

Assim sendo, é interessante notar que valores altos da probabilidade $w_{j,k}$, associada ao pixel vizinho, combinado com valores elevados, positivos ou negativos, de $r_{i,j,h,k}$, possuem alta influência na alteração de $w_{i,h}$ na equação (3.15). Os efeitos da contribuição desses dois fatores são representados no quadro abaixo:

Tabela 3.1: Efeito do coeficiente de compatibilidade no valor estimado para o peso do pixel central.

		Coeficiente de compatibilidade	
		Alto	Baixo
$w_{j,k}$	Alta	+	-
	Baixa	0	0

A combinação de valores altos na probabilidade $w_{j,k}$, do pixel vizinho na posição j pertencer à classe k , com valores altos do coeficiente de compatibilidade, deverá ter uma alta contribuição positiva ao valor de $w_{i,h}$. Enquanto valores muito baixos para o coeficiente de compatibilidade associado a valores altos de $w_{j,k}$ deverão ter efeito negativo elevado, pois se o vizinho j no contexto de i pertence à classe k e esta classe é pouco compatível com a classe h , então a classe h dificilmente apareceria na posição i .

O processo de relaxação proposto nesta dissertação atualiza os valores estimados para os pesos $w_{i,h}$ iterativamente, até que um critério de convergência seja atendido.

3.7 LIMIAR E “CLASSE DE FUNDO”

Finalmente deve-se aqui mencionar um cuidado a ser observado em todo o processo que visa extrair informações de contexto espacial em uma imagem. Esse cuidado refere-se ao caráter exaustivo que as classes devem apresentar. Em processos de classificação de imagens cobrindo cenas naturais, o número de classes presentes normalmente é elevado, com algumas classes sendo representadas por um número reduzido de pixels. Por esta razão, em situações

reais, apenas um subconjunto das classes presentes é efetivamente considerado. Em processos de classificação de imagens, o expediente mais comumente empregado para evitar a ocorrência de pixels erroneamente classificados, consiste na introdução de uma condição adicional na regra de classificação

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

$$\text{e } G_j(X) > T_j$$

sendo T_j um limiar previamente selecionado. A função do limiar neste caso consiste em recusar pixels para os quais a probabilidade de pertencerem efetivamente à classe vencedora é muito baixa. Com esta segunda condição elimina-se o número de erros de inclusão na imagem classificada. Pixels que não são aceitos em função do critério do limiar provavelmente pertencem a alguma das classes não amostradas, e por esta razão, não são considerados no processo de classificação. Para os fins de extração de informações de contexto espacial, estes pixels devem ser agrupados em uma classe adicional, denominada de classe de fundo. Deve-se aqui enfatizar, que a classe de fundo, assim definida, consiste na realidade de uma mistura de várias classes, mas que para fins de extração de informações de contexto espacial podem ser agrupadas em um conjunto único. Tradicionalmente os critérios empregados na seleção de um valor para o limiar T_j buscam excluir pixels situados nas caudas das distribuições, isto é, pixels com baixa probabilidade de pertinência à classe vencedora [RIC 06]. Esta abordagem apresenta, entretanto, problemas quando se trata de dados em dimensionalidade mais elevada. Sabe-se que na medida em que a dimensionalidade dos dados cresce, ocorre uma tendência das amostras se deslocarem em direção à cauda das curvas de distribuição [LAN 03] e [FUK 90]. Este fenômeno resulta na necessidade de se definir valores muito extremos para o limiar, inviabilizando este procedimento.

Nesta dissertação é proposto um outro critério para identificação de pixels com baixa probabilidade de pertinência às m classes. A abordagem aqui proposta consiste em definir um limiar de corte aplicado não ao valor da função de decisão da classe vencedora, mas aos pesos estimados conforme a metodologia proposta em 3.3. O escalonamento dos pesos conforme proposto na equação (3.10), garante que pixels com baixa probabilidade de pertinência as classes em consideração, apresentarão um valor baixo para o peso em todas as m classes. Se o maior destes m pesos for inferior a um determinado limiar, este pixel é rotulado como membro da classe de fundo.

No processo de estimação dos coeficientes de compatibilidade ($r_{i,j,h,k}$), a classe de fundo é portanto levada em consideração, isto é, na etapa de extração de informações de contexto espacial são consideradas as m classes originais, mais a classe de fundo, totalizando portanto $(m+1)$ classes. No Capítulo 4 –Experimentos– são fornecidos os detalhes operacionais deste procedimento.

4 EXPERIMENTOS

4.1 INTRODUÇÃO

Neste capítulo são apresentados e discutidos os experimentos realizados com a finalidade de ilustrar e validar a metodologia proposta de incremento do número de amostras de treinamento, através da utilização de amostras semi-rotuladas, ponderadas por meio de informações de natureza espectral e de contexto espacial (vizinhança), no processo de classificação de imagens em alta dimensionalidade (imagens hiperespectrais). Na metodologia proposta é estudado um processo iterativo, em duas etapas, para fins de estimação dos pesos na ponderação de amostras semi-rotuladas. Na primeira etapa, os pesos são estimados dentro de um processo iterativo que leva em conta somente atributos de natureza espectral. Na segunda etapa, estes valores são ajustados por meio da introdução de informações de contexto espacial (vizinhança) em um processo também iterativo de relaxação. Os parâmetros do classificador são então estimados por um conjunto de amostras de treinamento que inclui a mostras rotuladas e amostras semi-rotuladas, estas últimas ponderadas por meio dos valores finais estimados para os pesos. Este capítulo apresenta e analisa os experimentos desenvolvidos.

4.2 MATERIAIS

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

4.2.1 EQUIPAMENTOS

Microcomputador Pentium Intel Core 2 Duo 3 GHz, RAM 3 Gb, HD 250Gb e sistema operacional Microsoft Windows XP;

4.2.2 APLICATIVOS

MATLAB R2008b;

MultiSpecW32.

4.2.3 DADOS

Imagem AVIRIS cedida para estudo, pelo prof. David Landgrebe - Purdue University, para fins de validação [IPD 09];

4.3 CENA DE ESTUDO

Os dados hiperespectrais usados neste experimento são um segmento (145 x 118), num total de 17110 pixels (Figura 17), obtido a partir de outro segmento (435 x 435), num total de 189.225 pixels (Figura 15), de uma cena de dados AVIRIS, tomada em junho de 1992, cobrindo uma região agrícola (região experimental de testes) chamada *Indian Pine*, localizada a noroeste no estado de Indiana – EUA (Figura 14). A cena observada nas figuras 15 e 17 possui 190 bandas espectrais (das 224 bandas originais do sensor AVIRIS), cobrindo intervalos de comprimentos de onda distanciados uniformemente de modo a cobrir todos os comprimentos de onda do espectro eletromagnético entre 400 nm e 2.500 nm. A imagem (Figura 15) cobre uma área agrícola, com culturas de milho e soja, além de outras. A imagem foi obtida no início da época de crescimento, ocorrendo por isso grande quantidade de solo exposto. Foram ainda utilizadas técnicas de plantio distintas (plantio tradicional e plantio direto). Esta situação resulta em classes espectralmente muito semelhantes, constituindo-se num grande desafio para o classificador, 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.

A verdade terrestre para a cena completa está ilustrada na Figura 16, enquanto a Figura 18 mostra a verdade terrestre da cena utilizada nos experimentos desta metodologia (Figura 17).

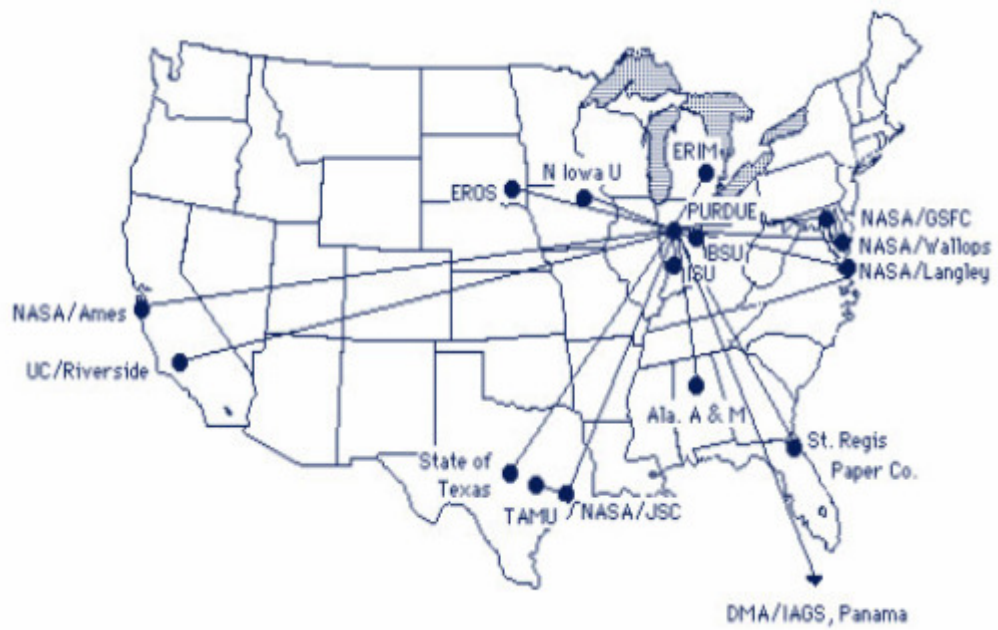


Figura 14 - Localização da Universidade Purdue, Estado de Indiana – EUA.
Fonte: Landgrebe (2002).



Figura 15 - AVIRIS: Composição RGB com as bandas 50, 27 e 17, da área experimental de teste, (NW Indiana's Indian Pine) - 220 bandas espectrais

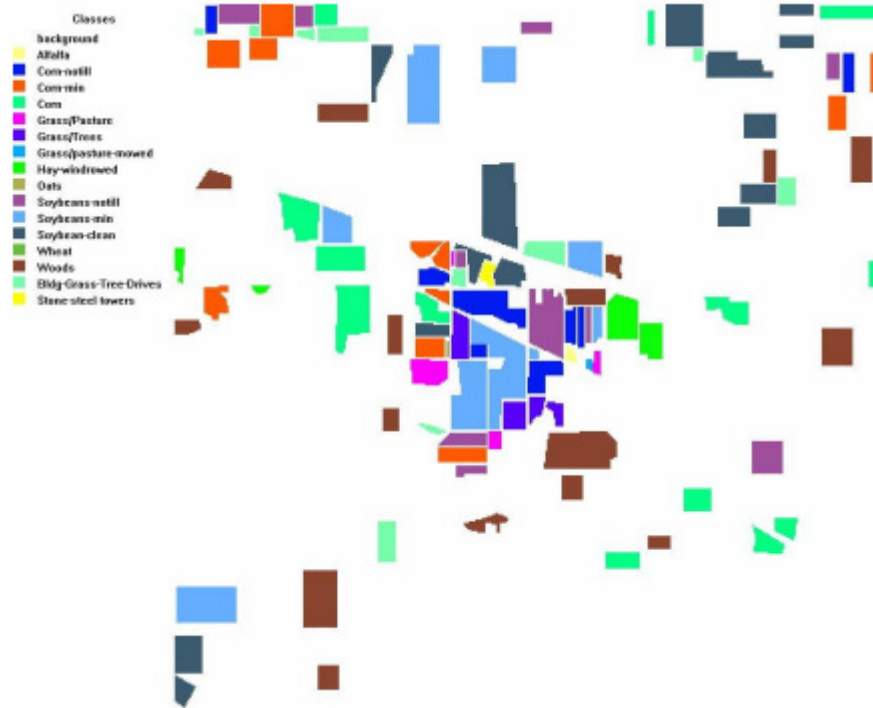


Figura 16 - Imagem temática da verdade terrestre (ground truth) da Figura 15.

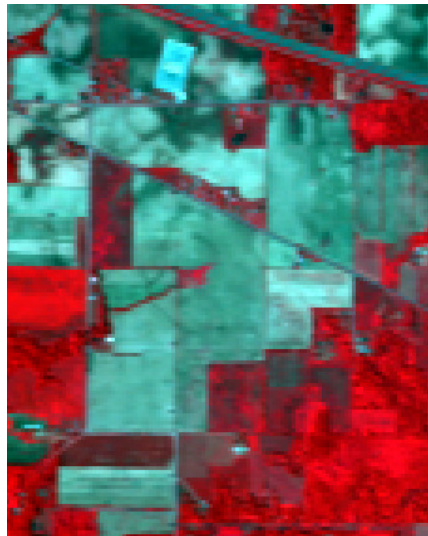


Figura 17 - AVIRIS: Composição RGB com as bandas 50, 27 e 17, do segmento utilizado nos experimentos, (NW Indiana's Indian Pine) - 220 bandas espectrais.

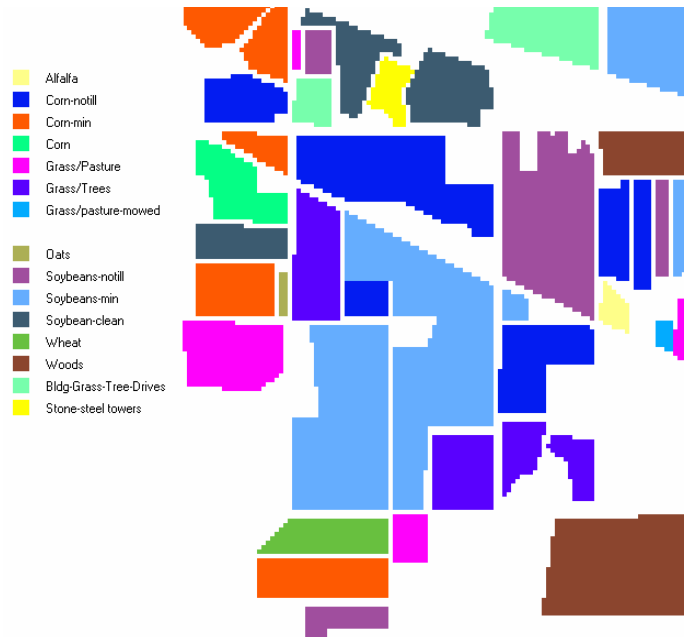


Figura 18 - Imagem temática da verdade terrestre (ground truth) para o segmento utilizado nos experimentos (Figura 17).

4.4 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 (H_2O), dióxido de carbono (CO_2) e ozônio (O_3), que resultam em imagens com um nível muito alto de ruído.

Foram excluídas as seguintes bandas espectrais, conforme está indicado (Tabela 4.1).

TABELA 4.1 - 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.

Diversos experimentos foram desenvolvidos empregando diferentes valores para a dimensionalidade dos dados, isto é, para o número de bandas espectrais. Em cada experimento, uma vez definida a dimensionalidade dos dados, o critério adotado para a seleção das bandas espectrais consiste em selecionar um subconjunto das 190 bandas espectrais de tal forma que as bandas selecionadas estejam distribuídas da maneira mais uniformemente possível ao longo de todo o conjunto original contendo a totalidade das bandas. O procedimento adotado está ilustrado no pseudocódigo abaixo:

```

*****
*****
% total_bands: dimensionalidade dos dados originais (total de bandas)
% n_bands: numero de bandas desejado

p = total_bands/n_bands;

% forma um vetor com as posições uniformemente distribuídas (numeros reais)

de 1 até numero total de bandas, com passo p
    adiciona mais um elemento ao vetor posições;
fim
%
% arredonda os elementos no vetor posições para o inteiro mais proximo;
*****
*****

```

4.5 SELEÇÃO DE CLASSES

No segmento de imagem empregado nos experimentos (Figura 17) estão presentes várias classes (diferentes tipos de culturas agrícolas). Dentre estas, cinco classes apresentam um desafio ao processo de classificação: milho com preparo mínimo do solo (*corn_min*), milho plantio direto (*corn_notill*), soja com preparo mínimo do solo (*soy_min*), soja plantio direto (*soy_notill*) e soja plantio tradicional (*soy_clean*). Estas cinco classes são espectralmente muito semelhantes. Ainda, a imagem foi tomada no início da época de crescimento das culturas de soja e milho, onde apenas aproximadamente 5% da área estão efetivamente cobertas pela vegetação, sendo os restantes 95% composto por solo exposto e resíduos de colheitas anteriores (Figura 19). Todos estes fatores contribuem para tornar a classificação destas cinco classes em um desafio ao classificador. Nestas condições, o processo de classificação empregando dados tradicionais em média ou baixa dimensionalidade (dados LANDSAT-TM ou SPOT, por exemplo), resultam em imagens temáticas com baixa acurácia [LAN 03].

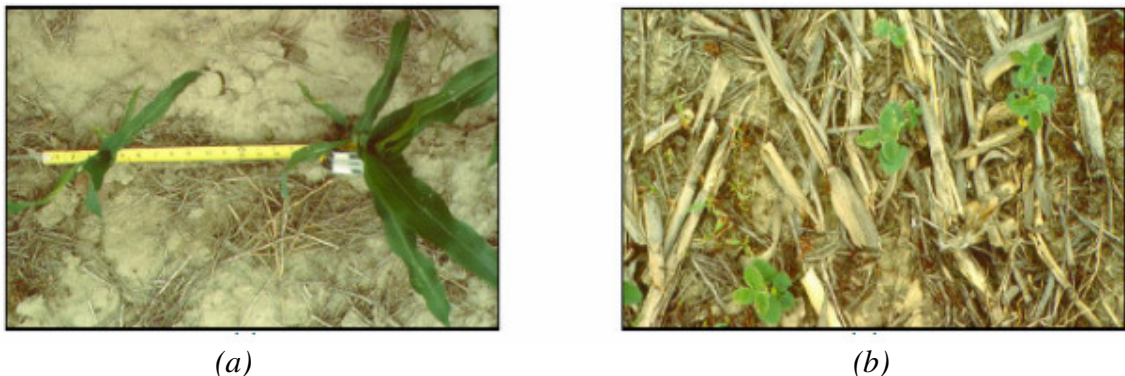


Figura 19 - Vista de áreas contendo as culturas de milho (a) e soja (b), mostrando a baixa porcentagem que representa frente à área emaciada no momento da coleta dos dados.

Fonte: Adaptado de Landgrebe (2003).

Utilizando os dados disponíveis de verdade terrestre, que foram coletados para cada uma das cinco classes, são criados dois conjuntos de pixels: o primeiro para fins de treinamento e o segundo para fins de teste (processo *holdout* [FUK 90]). Como as áreas de plantio ocupadas por cada uma destas classes de informação não são iguais, o tamanho das amostras é variável de uma classe para outra. A Tabela 4.2 ilustra o número total de pixels disponíveis.

TABELA 4.2 - Total de pixels extraídos da imagem AVIRIS

Nome da classe	Identificador Numérico	Total de pixels da amostra de treinamento	Total de pixels da amostra de teste
corn_min	1	417	417
corn_notill	2	717	717
soy_min	3	1234	1234
soy_notill	4	484	484
soy_clean	5	307	307

A similaridade entre o comportamento espectral destas cinco classes pode ser ilustrada pelas curvas de resposta espectral média, conforme ilustrado na Figura 20.

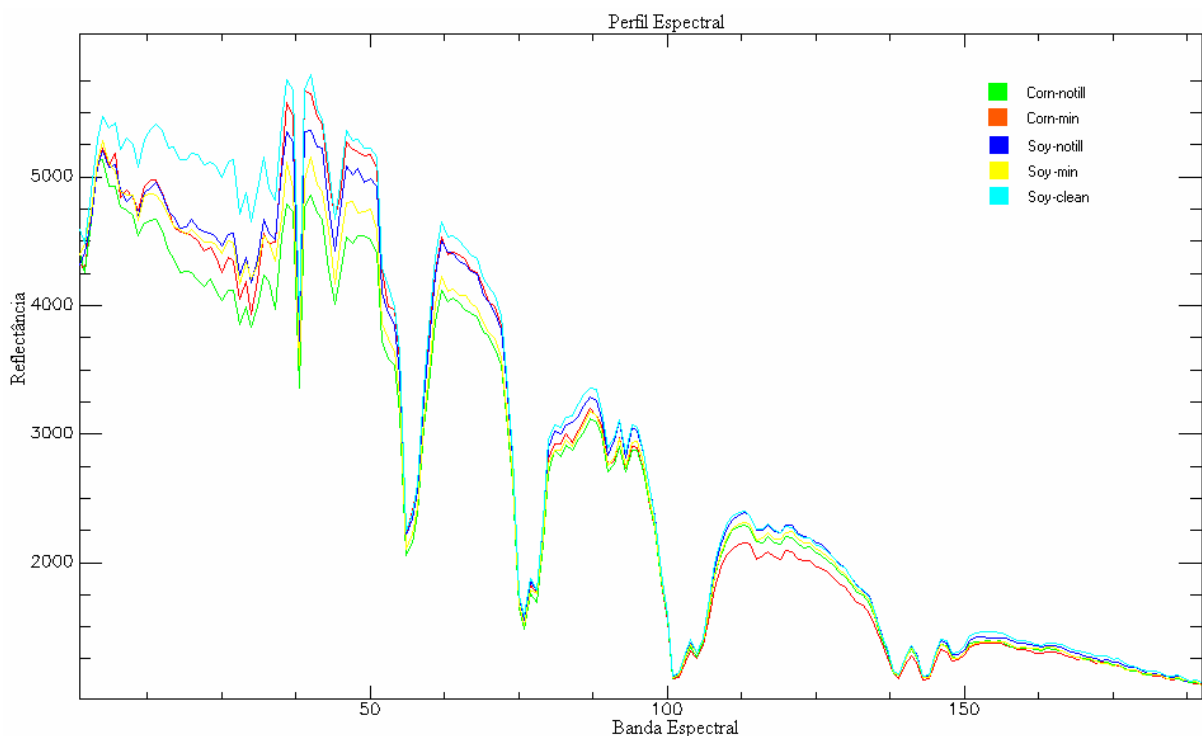


Figura 20 - Curvas de Resposta Espectral média para as classes Milho notill (verde), Milho min (vermelho), Soja notill (azul), Soja Min (amarelo), Soja clean (ciano).

4.6 SOFTWARE IMPLEMENTANDO A METODOLOGIA PROPOSTA

Com o objetivo de testar a metodologia proposta, foi desenvolvido um software em ambiente MATLAB. O software implementa a metodologia proposta (Capítulo 3), além de executar o pré-processamento dos dados, como por exemplo, a seleção das bandas a serem empregadas em função da dimensionalidade escolhida, a seleção das amostras de treinamento e de teste, classe de fundo, critério de convergência e número de amostras semi-rotuladas a ser inserido no processo.

O software é composto por uma função principal:

- context_20_Oct_B.m

e por seis funções auxiliares:

-select_samples_ctx_Oct_06.m

-GML.m

-accuracies_05_class_June_08.m

-weighted_mean_sl_Sept_06.m

-prob_relax_Oct_10.m

-weighted_mean_pr_Jun_03_B.m

As operações implementadas em cada uma das funções são resumidas a seguir:

1- context_20_Oct_B.m

Recebe os dados de entrada (imagem com a totalidade das bandas espectrais, imagem com os dados de verdade terrestre, dimensionalidade dos dados a ser empregada, amostras rotuladas disponíveis para treinamento e para teste, número de amostras rotuladas desejado, número de amostras semi-rotuladas desejado, número máximo de iterações e correspondentes limiares permitido em ambas as etapas da metodologia). Empregando as funções auxiliares, esta função principal implementa a metodologia proposta, fornecendo como dados de saída as acurácias do usuário, do produtor e a global obtidas no processo de classificação.

As funções auxiliares realizam as seguintes tarefas:

2- select_samples_ctx_Oct_06.m:

Seleciona as amostras de treinamento, de teste, e as bandas utilizadas da forma mais uniformemente possível distribuída, conforme citado no tópico seleção de bandas.

3- GML.m

Calcula o valor para a versão logarítmica da função de decisão de Máxima Verossimilhança Gaussiana (equação 3.5), e requer como argumentos de entrada os parâmetros do classificador (vetor de médias e matriz de covariância).

4- accuracies_05_class_June_08.m

Calcula as acurácias do produtor, usuário e global, e também gera a matriz de contingência do processo.

5- weighted_mean_sl_Sept_06.m

Estima os parâmetros (vetor de médias e matriz de covariância) empregando ponderadamente amostras rotuladas (peso unitário) acrescidas de amostras semi-rotuladas (peso estimado por meio de atributos espectrais somente).

6- prob_relax_Oct_10.m

Esta função realiza o processo de relaxação para o refinamento dos pesos para as amostras semi-rotuladas.

7- **weighted_mean_pr_Jun_03_B.m**

Estima os parâmetros (vetor de médias e matriz de covariância) empregando ponderadamente amostras rotuladas (peso unitário) acrescidas de amostras semi-rotuladas (peso estimado por meio de atributos espectrais e informações de contexto espacial).

4.7 AMOSTRAS DE TREINAMENTO

Do conjunto das amostras disponíveis para cada classe individual foram extraídos dois subconjuntos: um subconjunto para fins de treinamento do classificador e um segundo para fins de estimação da acurácia dos resultados. Esta abordagem é conhecida na literatura como *holdout* ([FUK 90], [LAN 03]). Com o objetivo de se obter dois conjuntos com boa representatividade dos dados, as amostras em ambos os conjuntos são extraídas alternadamente do conjunto total das amostras disponíveis. Também, à semelhança do procedimento adotado na seleção das bandas espectrais, estas amostras são uniformemente distribuídas ao longo do conjunto 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. Conforme a metodologia proposta, e em cada uma das etapas deste processo, a cada iteração o conjunto original com as amostras de treinamento rotuladas é incrementado por um número adicional de amostras extraídas pelo próprio processo, denominadas de amostras semi-rotuladas, com cada pixel recebendo um peso conforme as equações descritas na seção 3.3. O número de amostras semi-rotuladas agregadas ao conjunto das amostras de treinamento, a cada etapa do processo, é definido pelo analista.

Com a finalidade de testar a eficácia da metodologia proposta neste estudo, uma série de testes foram realizados para diferentes combinações de valores das variáveis (dimensionalidade dos dados, número de amostras rotuladas, número de amostras semi-rotuladas e número de iterações). A avaliação dos resultados é feita tomando-se a dimensionalidade dos dados como variável independente e a resultante acurácia na classificação como variável dependente. Em cada experimento a dimensionalidade dos dados variou no intervalo [1, (número de amostras de treinamento-1)]. Para valores mais altos da

dimensionalidade dos dados a matriz de covariância resulta singular. Para fins de análise da metodologia proposta, valores mais altos para a dimensionalidade dos dados, isto é, próximos ao limite superior do intervalo, apresentam maior interesse. Um número maior de bandas espectrais contribui com uma maior quantidade de informação, mas acarreta também em um número maior de parâmetros a serem estimados, especialmente na matriz de covariância. Nesta situação, a inclusão de amostras semi-rotuladas no treinamento do classificador pode contribuir significativamente no incremento da acurácia da classificação.

Deve-se aqui observar um cuidado ao comparar os resultados de experimentos empregando diferentes valores para a dimensionalidade dos dados. No processo aqui proposto, em cada caso, as bandas espectrais são selecionadas uniformemente ao longo de todo o conjunto das 190 bandas espectrais disponíveis. O objetivo aqui consiste em minimizar o problema da alta correlação existente entre bandas espectrais próximas. Esta abordagem resulta, entretanto, em que a cada incremento na dimensionalidade dos dados o novo conjunto de bandas espectrais não necessariamente inclui aquelas selecionadas nos experimentos anteriores (com dimensionalidade menor). Este fato explica as irregularidades que se pode notar nas curvas que ilustram a acurácia obtida em função da dimensionalidade dos dados. Com o objetivo de reduzir esta variabilidade, que apresenta características de um ruído, propõe-se aqui expressar a acurácia para cada valor da dimensionalidade dos dados, como a média das acurácias estimadas para um conjunto de cinco valores: a dimensionalidade em questão, os dois valores imediatamente mais altos e os dois imediatamente mais baixos. Esta abordagem apresenta, portanto, as características de uma filtragem passa-baixas, reduzindo o ruído nos 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 empregando a metodologia proposta, presente na seção 4.6 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 refinadas, ao final do processo iterativo), é ilustrado através das imagens temáticas produzidas, dos gráficos da acurácia média obtida (acurácias do usuário e do produtor), e das matrizes de erro onde pode ser observado o número total de pixels que foram classificados em cada uma das classes selecionadas para o experimento. Para evitar o viés no cálculo da matriz de contingência se optou por tomar o número de pixels igual para todas as classes de teste, sendo este número igual ao da menor

das amostras de teste, ou seja, neste caso, 307 pixels (Tabela 4.2) que é referente à classe *soy_clean*. Isto permite uma posterior comparação direta dos resultados obtidos.

No classificador utilizado nestes experimentos, foi incluído um limiar, que 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, ou neste caso, como pertencentes à classe de fundo.

4.8 O LIMIAR E A “CLASSE DE FUNDO”

Conforme detalhado anteriormente no Capítulo 3 (Metodologia), em imagens de cenas naturais, o número de classes existentes é geralmente elevado. Algumas das classes presentes na cena freqüentemente são representadas por um número reduzido de pixels. Esta situação resulta em que freqüentemente nem todas as classes são amostradas. Nestas condições, pixels pertencentes a classes não amostradas (ou simplesmente pixels ruidosos) acabam sendo erroneamente rotulados em uma das classes para quais amostras foram coletadas. Pixels nestas condições podem usualmente serem identificados por uma baixa probabilidade de pertinência à classe vencedora. Este problema é normalmente contornado por meio da adição de uma restrição adicional na regra de classificação (equação 3.2). No contexto desta dissertação, estes pixels duvidosos são identificados por apresentarem um baixo grau de pertinência a todas as classes as classes na imagem.

Este problema se reveste de importância aqui. Para fins de extração de informações de contexto espacial é necessário que a imagem classificada, que servirá para a estimação do coeficiente de compatibilidade (equação 3.14), esteja completa, isto é, todos os pixels devem estar rotulados em uma das classes. Nesta dissertação propõe-se resolver este problema pela criação de uma classe adicional, denominada de “classe de fundo” à qual são rotulados todos os pixels cujo grau de pertinência w_{ik} à classe vencedora encontre-se abaixo de um determinado limiar. Nesta dissertação, a seleção deste limiar é feita experimentalmente com base no histograma dos graus de pertinência máximo para o conjunto dos pixels na imagem. A Figura 21 ilustra um destes histogramas. Resultados obtidos aplicando-se diferentes valores para o limiar (aplicado à cauda esquerda do histograma) estão ilustrados na Figura 23. Uma avaliação visual da adequação desta abordagem pode ser feita comparando a área definida para a classe de fundo na Figura 23, com as áreas ocupadas pelas demais classes conforme ilustradas na imagem de verdade terrestre (Figura 22).

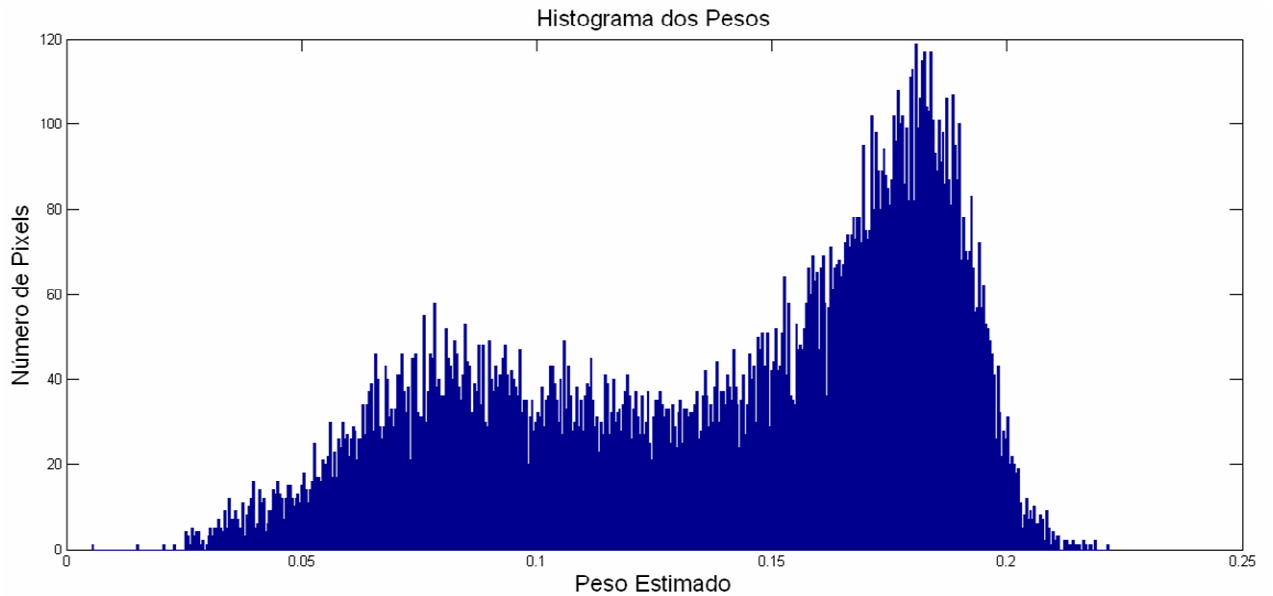


Figura 21 – Exemplo de histograma para o valor médio dos pesos máximos.

O conjunto dos pixels limiarizados por este critério é agrupado na chamada “classe de fundo”. Deve-se aqui observar que a “classe de fundo” não se constitui propriamente em uma classe, mas sim em um agrupamento de várias classes. Este fato pode ser facilmente verificado pelo caráter multimodal de seu histograma. Entretanto, para fins de extração de contexto espacial, esta abordagem é aceitável. Para cada pixel rotulado nesta classe, atribui-se um peso igual ao complemento da soma dos pesos associados as demais classes.

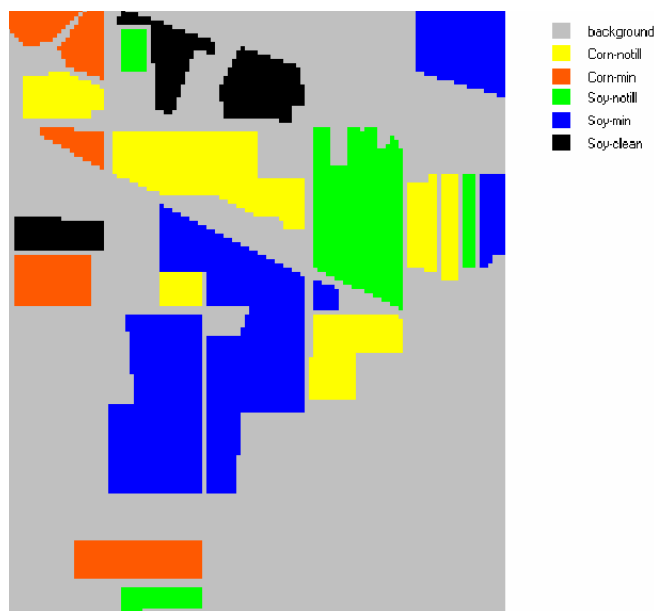
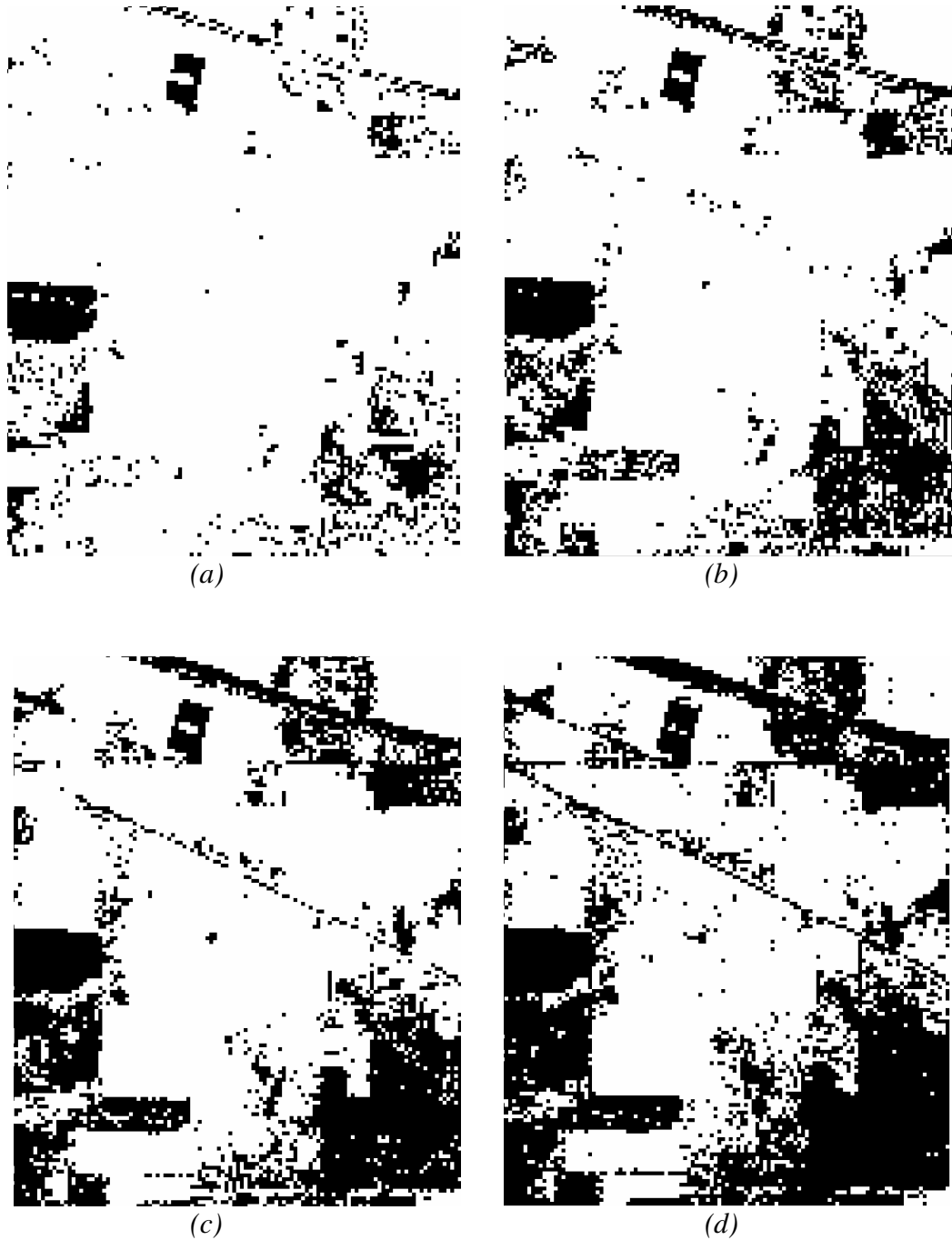


Figura 22 – Imagem ideal para a classificação da cena utilizada neste trabalho

Na Figura 22 temos em cinza a classe de fundo, e nas demais cores a verdade de campo das 5 classes utilizadas neste trabalho, conforme a legenda.

O limiar de corte para este trabalho foi estabelecido através de experimentação com vários valores, estes variando de 0% - 60%, e seus respectivos resultados podem ser vistos na Figura 23.



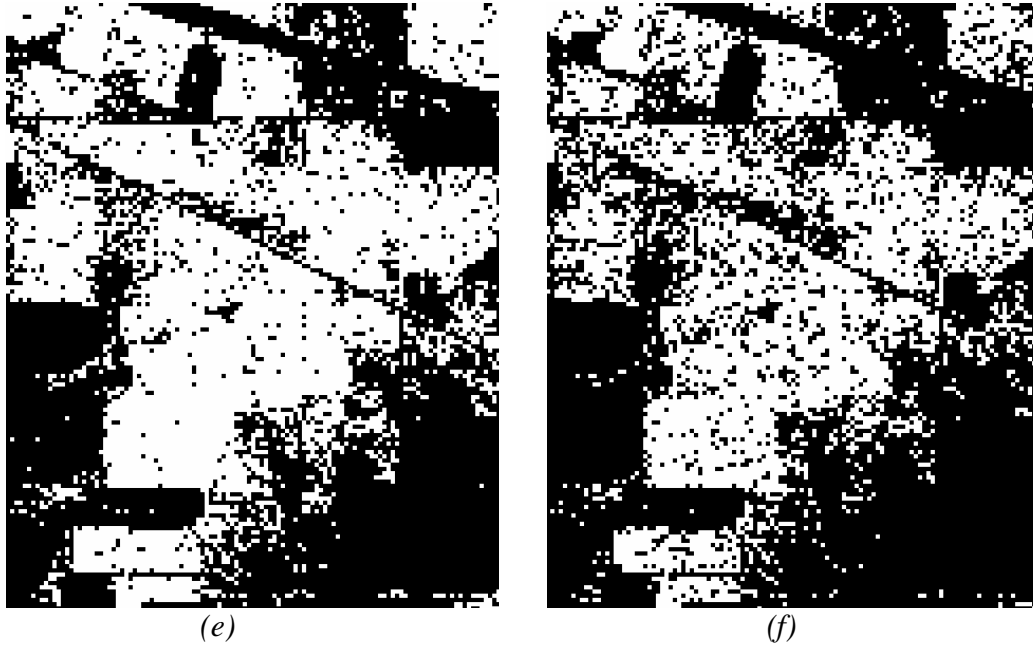


Figura 23 – a) limiar de 10%, b) limiar de 20%, c) limiar de 30%, d) limiar de 40%, e) limiar de 50% e f) limiar de 60%. A “classe de fundo” está representada pela cor branca(preto).

Nas Figuras 23a, 23b e 23c se pode observar que ainda existem diversos pixels atribuídos indevidamente as classes, já nas Figuras 23d e 23e é possível ver uma classificação mais semelhante a aquela esperada na Figura 22, porém na última Figura 23f, se observa um excesso de atribuição de pixels à classe de fundo. Embora as Figuras 23d e 23e, com limiares respectivamente 40% e 50%, sejam visualmente muito semelhantes, elas diferem na acurácia da classificação, onde a classificação utilizando o limiar de 40% se mostrou mais acurada que a classificação utilizando o limiar de 50%. Por este motivo neste trabalho se optou por utilizar o limiar de 40%.

4.9 ANÁLISE DOS EXPERIMENTOS

Uma avaliação prévia mostrou ser aceitável a hipótese de uma distribuição normal multivariada para os dados. Desta forma, o classificador empregado nos experimentos foi o da Máxima Verossimilhança Gaussiana (MVG).

Para uma completa avaliação da metodologia proposta, os experimentos foram organizados em três grupos:

1- Estimação dos parâmetros do classificador MVG empregando apenas um conjunto composto por amostras rotuladas (processo tradicional);

2- Estimação dos parâmetros do classificador MVG adicionando ao conjunto das amostras de treinamento um certo número de amostras semi-rotuladas ponderadas por meio de informações espectrais somente (seção 3.10);

3- Estimação dos parâmetros do classificador MVG adicionando ao conjunto das amostras de treinamento um certo número de amostras semi-rotuladas ponderadas por meio de informações espectrais (seção 3.10) e de informações de contexto espacial (seção 3.15).

As variáveis testadas nos experimentos foram:

1- dimensionalidade dos dados (número de bandas espectrais);

2- número de amostras rotuladas;

3- número de amostras semi-rotuladas;

4- número de iterações em cada um dos dois processos iterativos.

Vários experimentos foram realizados empregando valores diferentes para estas variáveis, estimando-se, em cada caso, o valor obtido para a acurácia no processo de classificação. Nesta dissertação são apresentados e discutidos os experimentos considerados mais representativos, conforme sumarizados na Tabela 4.3.

Tabela 4.3 - Dados dos Experimentos

Experimento	Número de Bandas	Número de Amostras Rotuladas	Número de Amostras Semi-rotuladas
1	15	16	30
2	15	16	60
3	15	16	90
4	30	31	30
5	30	31	60
6	30	31	90
7	100	101	80

A avaliação da metodologia proposta é realizada comparando os resultados fornecidos pelo conjunto de experimentos supracitados, e o critério para avaliação destes é a acurácia da imagem temática produzida em casa caso, mas também são mostradas as matrizes de erro. Com a finalidade de tornar os resultados obtidos para as várias classes comparáveis entre si, foram utilizados subconjuntos de amostras de treinamento de mesmo tamanho para todas as classes em estudo. Estes subconjuntos foram extraídos do conjunto de amostras disponíveis (Tabela 4.2).

Na fase de experimentos, utilizaram-se quantidades de amostras de treinamento de mesmo tamanho para todas as classes obtidas conforme descrito na seção 4.6. Note-se que o número mínimo admissível para as amostras de treinamento é igual à dimensionalidade dos

dados mais um. Um valor inferior resultará que a estimativa para a matriz de covariância resulta singular e, portanto, não utilizável no classificador MVG [LAN 03]. O número de amostras de treinamento foi escolhido intencionalmente pequeno quando comparado com a dimensionalidade dos dados, simulando desta forma situações encontradas em problemas reais e ilustrando a utilidade do uso de amostras semi-rotuladas.

O processo iterativo é encerrado quando é atingido algum dos critérios de convergência descritos abaixo:

- Quando o número de pixels que trocam de classe entre duas iterações sucessivas é inferior a 2% do total de pixels na imagem temática;
- 1ª etapa – ponderação das amostras semi-rotuladas: 2 iterações;
- 2ª etapa – refinamento dos pesos das amostras semi-rotuladas com a introdução de informações de contexto espacial via relaxação: 1 iteração.

As acurácias obtidas em cada experimento foram estimadas com o auxílio da matriz de contingência [RIC 06]. Esta matriz é construída empregando o conjunto das amostras de teste disponíveis, com as seguintes acurácias sendo estimadas:

a) Acurácia do Produtor (producer's accuracy): – é a razão das amostras classificadas corretamente em cada classe, pelo total de amostras de teste da respectiva classe;

b) Acurácia do Usuário (user's accuracy): – é a razão das amostras classificadas corretamente em cada classe, pelo total de amostras classificadas na respectiva classe;

c) Acurácia Geral (global accuracy): – traduz a razão do total de amostras classificadas corretamente em cada classe, ou seja, a soma dos valores ao longo da diagonal principal, pelo total das amostras de teste.

Aqui se deve ressaltar que devido ao fato de se ter amostras de teste de igual tamanho para todas as classes, a Acurácia do Produtor resulta igual à Acurácia Geral.

As imagens temáticas que serão apresentadas a seguir têm como referência a Figura 22.

4.10 TEMPO COMPUTACIONAL

Outro ponto interessante é avaliar o desempenho computacional da metodologia proposta, já que técnicas demasiadamente demoradas se tornam computacionalmente inaplicáveis. O tempo de processamento de cada experimento foi medido em segundos, e é contado a partir do início da função principal (context_20_Oct_B.m) até o fim do processo. Abaixo na Tabela 4.4 pode ser visto os tempos de execução de cada experimento.

Tabela 4.4 – Custo computacional dos experimentos

Experimento	Tempo (s)
1	456.6
2	472.4
3	466.0
4	1404.1
5	1326.6
6	1394.6
7	27396,9

4.11 EXPERIMENTOS

4.11.1 EXPERIMENTO 1

Para este experimento foram utilizados dados com dimensionalidade igual a 15, 16 amostras rotuladas e 30 amostras semi-rotuladas. As imagens temáticas resultantes estão ilustradas na Figura 24.

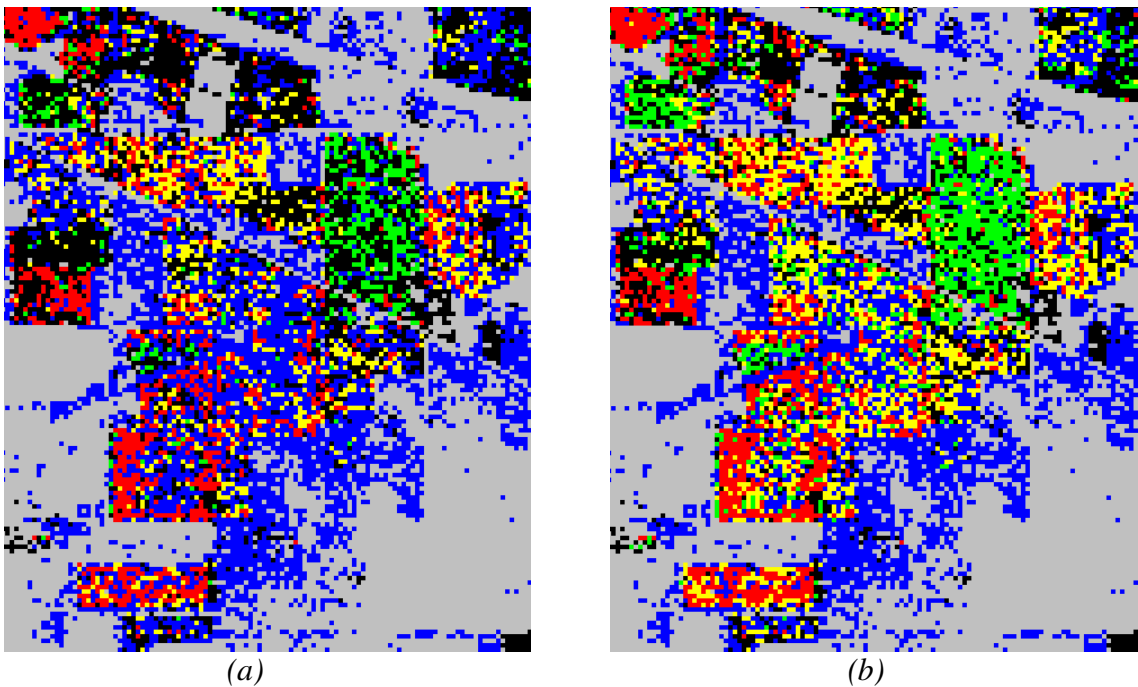


Figura 24 – (a) Imagem temática sem a inclusão de amostras semi-rotuladas (início do processo) (b) Imagem temática com a inclusão de amostras semi-rotuladas refinadas (final do processo). Legenda de cores conforme Figura 22(seção4.8).

A contribuição das amostras semi-rotuladas para o incremento no valor estimado da acurácia pode ser observada na Figura 24. Este incremento é especialmente visível na classe *soy_notill* (em cor verde).

Os valores estimados para as acurácias estão ilustrados nas Figuras 25 e 26 para a acurácia média do usuário e do produtor, respectivamente, ilustrando o aumento da acurácia média com a inclusão das amostras semi rotuladas. Os resultados, na forma de matriz de erro, estão apresentados na Tabela 4.5.

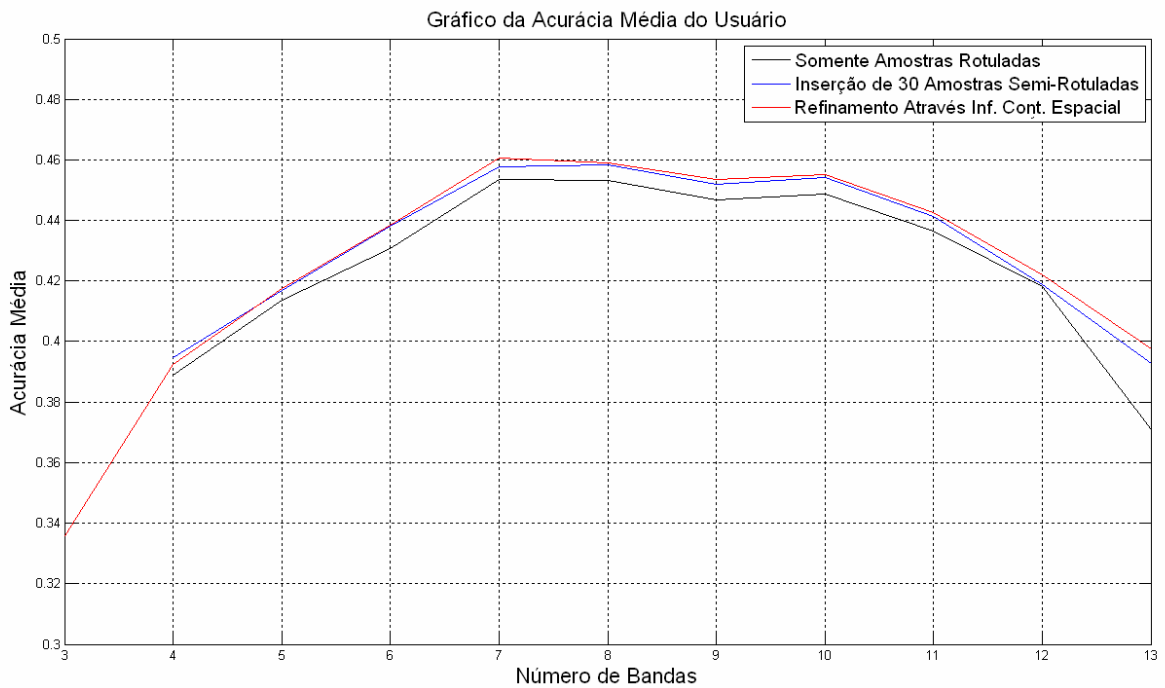


Figura 25 - Acurácia da classificação (usuário), utilizando 15 bandas, 16 amostras de treinamento rotuladas e incremento de 30 amostras semi-rotuladas.

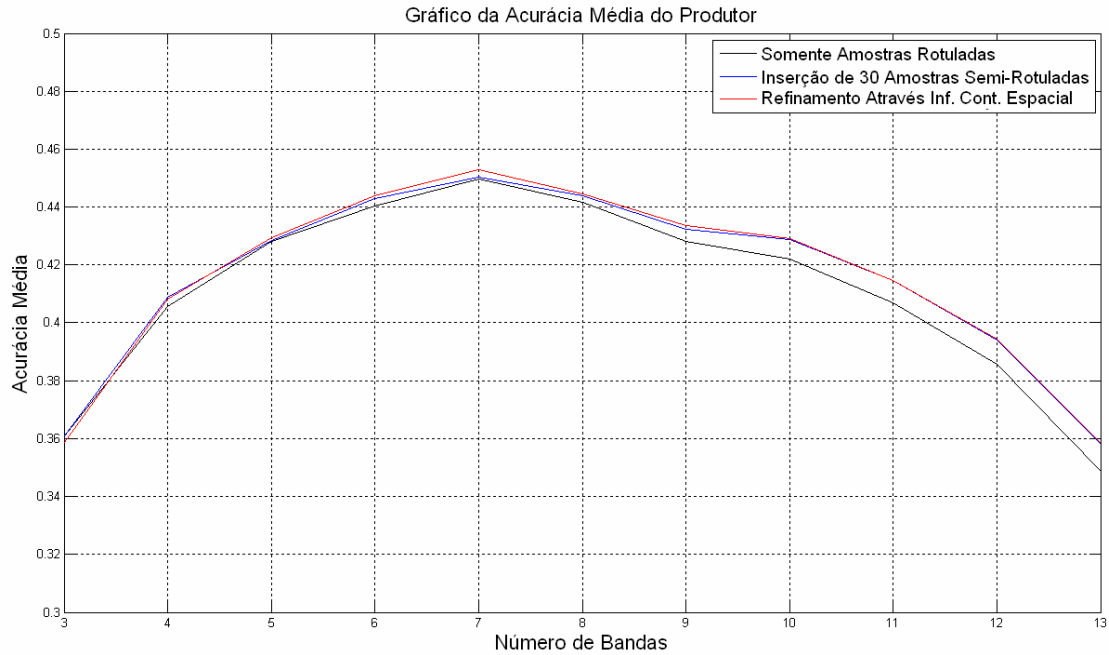


Figura 26 - Acurácia da classificação (produtor), utilizando 15 bandas, 16 amostras de treinamento rotuladas e incremento de 30 amostras semi-rotuladas.

TABELA 4.5 – Matriz de contingência resultante da metodologia proposta neste estudo, empregando dados com dimensionalidade igual a 15, 16 amostras rotuladas e 30 amostras semi-rotuladas.

Classes da imagem classificada	Classes da Imagem de Referência					Total de Amostras	Acurácia do Usuário (%)
	corn_min	corn_notill	soy_clean	soy_min	soy_notill		
corn_min	130	59	70	16	26	301	0,432
corn_notill	42	112	79	29	44	306	0,366
soy_clean	50	50	87	46	36	269	0,323
soy_min	25	21	29	127	21	223	0,570
soy_notill	60	65	42	89	180	436	0,413
Total de Amostras	307	307	307	307	307		
Acurácia do Produtor (%)	0,423	0,365	0,283	0,414	0,586		

Na Tabela 4.5 pode ser visualizado o comportamento da acurácia para as classes empregadas no experimento.

4.11.2 EXPERIMENTO 2

Para a execução deste experimento foram utilizados dados com dimensionalidade igual a 15, 16 amostras rotuladas e 60 amostras semi-rotuladas. As imagens temáticas

mostradas na Figura 27 ilustram os resultados da classificação ao final do processo. Os resultados, em forma de matriz de erro, estão apresentados na Tabela 4.6.

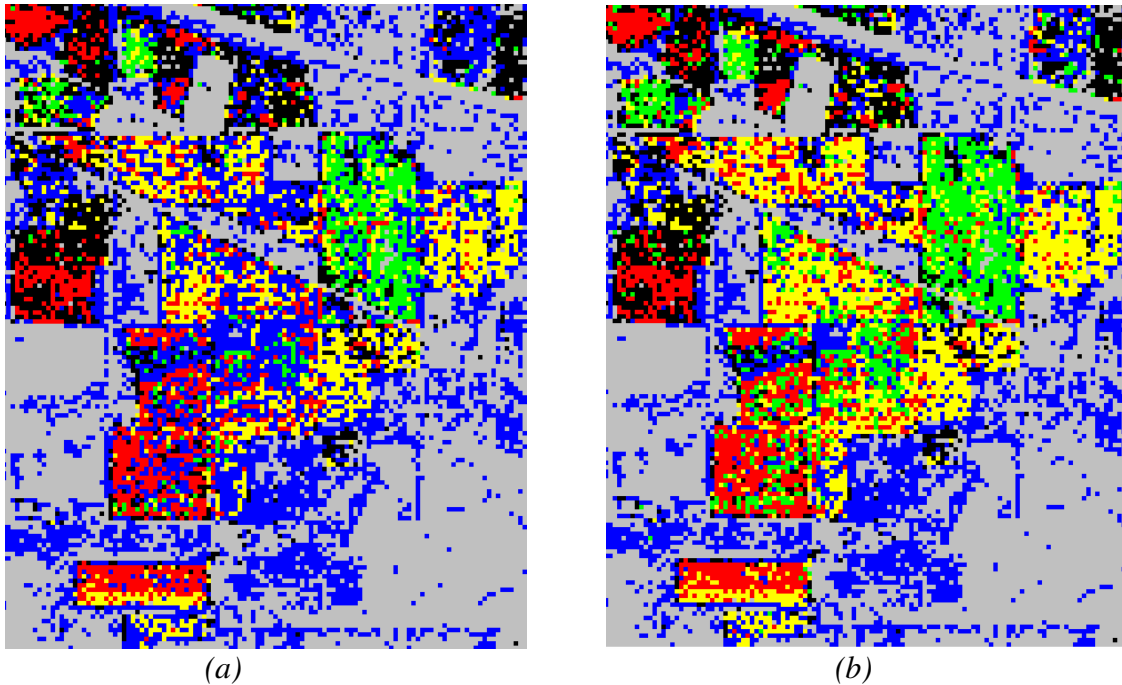


Figura 27 – (a) Imagem temática sem a inclusão de amostras semi-rotuladas (início do processo) (b) Imagem temática com a inclusão de amostras semi-rotuladas refinadas (final do processo). Legenda de cores conforme Figura 22(seção4.8) .

Na Figura 27 se pode observar que após a inclusão das amostras semi-rotuladas refinadas por meio da informação de contexto houve um incremento na discriminação da classe *corn_notill* (amarelo) e *soy_notill* (verde), no entanto, ainda há uma confusão entre as classes *soy_min* (azul) e *corn_min* (vermelho).

Os valores estimados para as acurácias estão ilustrados nas Figuras 28 e 29 e representam a acurácia média do usuário e do produtor, respectivamente. Nas figuras fica evidente o aumento da acurácia média dos experimentos nas duas etapas do processo, sendo a maior contribuição ao aumento da acurácia devido à introdução de amostras semi-rotuladas ponderadas em função apenas de informações espectrais (em cor azul).

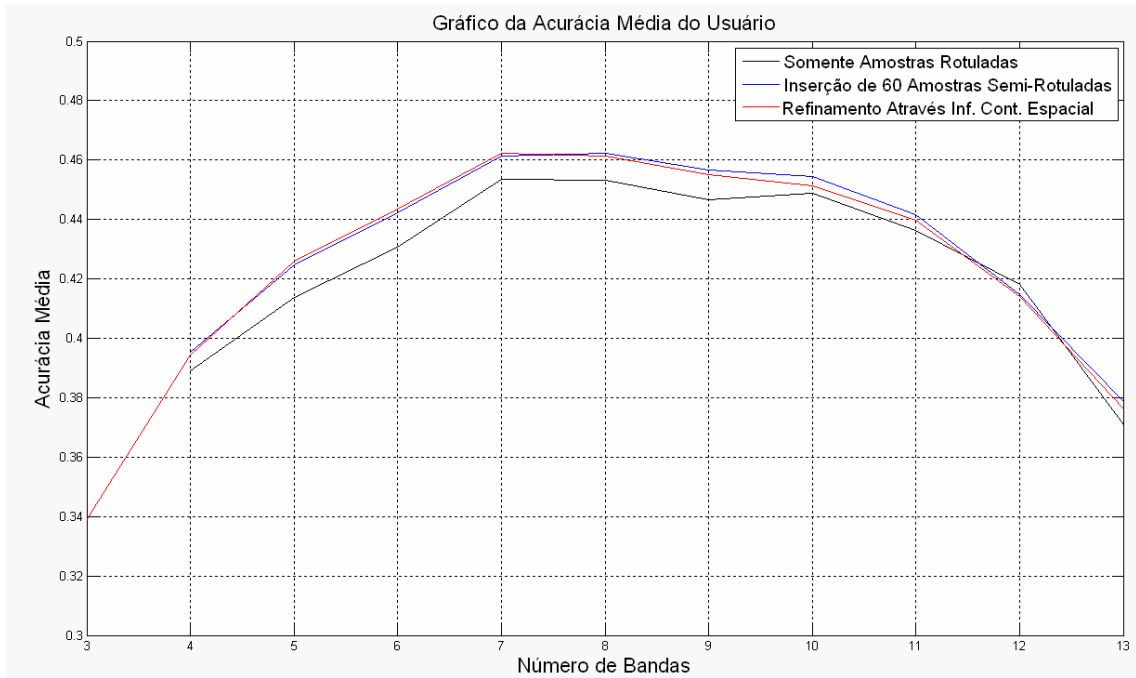


Figura 28 - Acurácia da classificação (usuário), utilizando 15 bandas, 16 amostras de treinamento rotuladas e incremento de 60 amostras semi-rotuladas.

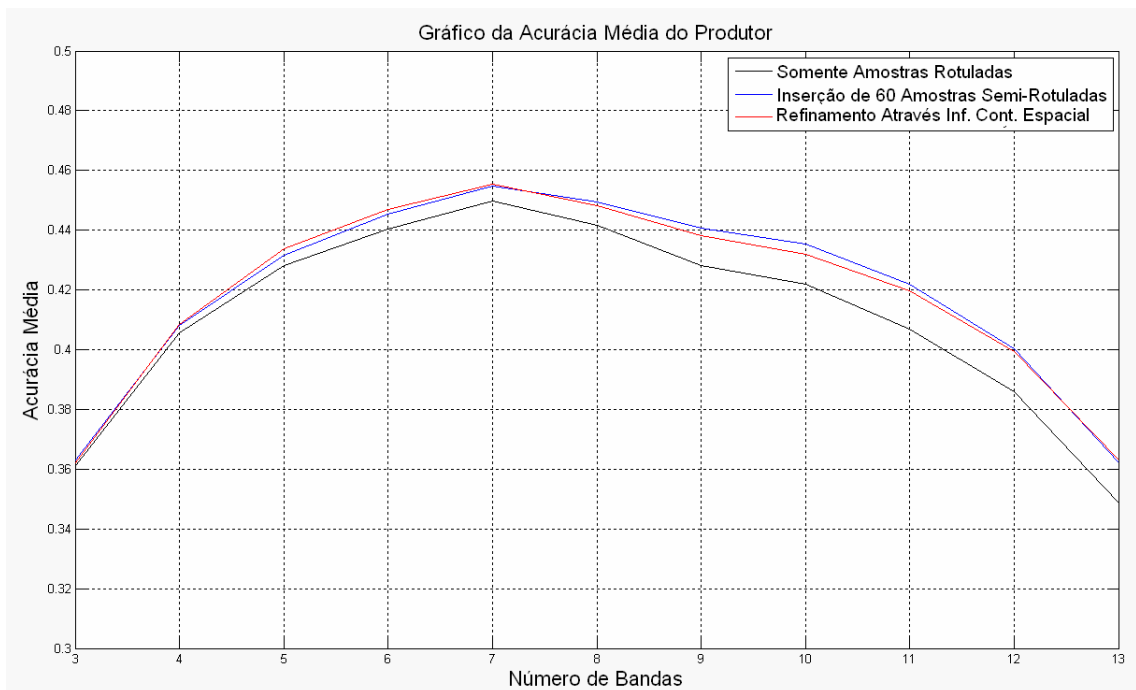


Figura 29 - Acurácia da classificação (produtor), utilizando 15 bandas, 16 amostras de treinamento rotuladas e incremento de 60 amostras semi-rotuladas.

TABELA 4.6 – Matriz de contingência resultante da metodologia proposta neste estudo, empregando dados com dimensionalidade igual a 15, 16 amostras rotuladas e 60 amostras semi-rotuladas.

Classes da imagem classificada	Classes da Imagem de Referência					Total de Amostras	Acurácia do Usuário (%)
	corn_min	corn_notill	soy_clean	soy_min	soy_notill		
corn_min	135	37	83	11	49	315	0,429
corn_notill	47	162	68	58	19	354	0,458
soy_clean	12	55	65	57	44	233	0,279
soy_min	13	21	60	163	22	279	0,584
soy_notill	100	32	31	18	173	354	0,489
Total de Amostras	307	307	307	307	307		
Acurácia do Produtor (%)	0,440	0,528	0,212	0,531	0,564		

Na tabela acima pode ser visto que a classe *soy_clean* obteve menos da metade da acurácia na classificação, em comparação com as demais classes, e que as amostras de teste foram distribuídas quase que igualmente entre as demais classes.

4.11.3 EXPERIMENTO 3

Para a execução deste experimento foram utilizados dados com dimensionalidade igual a 15, 16 amostras rotuladas e 90 amostras semi-rotuladas. As imagens temáticas mostradas na Figura 30 ilustram os resultados finais do processo de classificação. Os resultados para a matriz de contingência estão apresentados na Tabela 4.7.

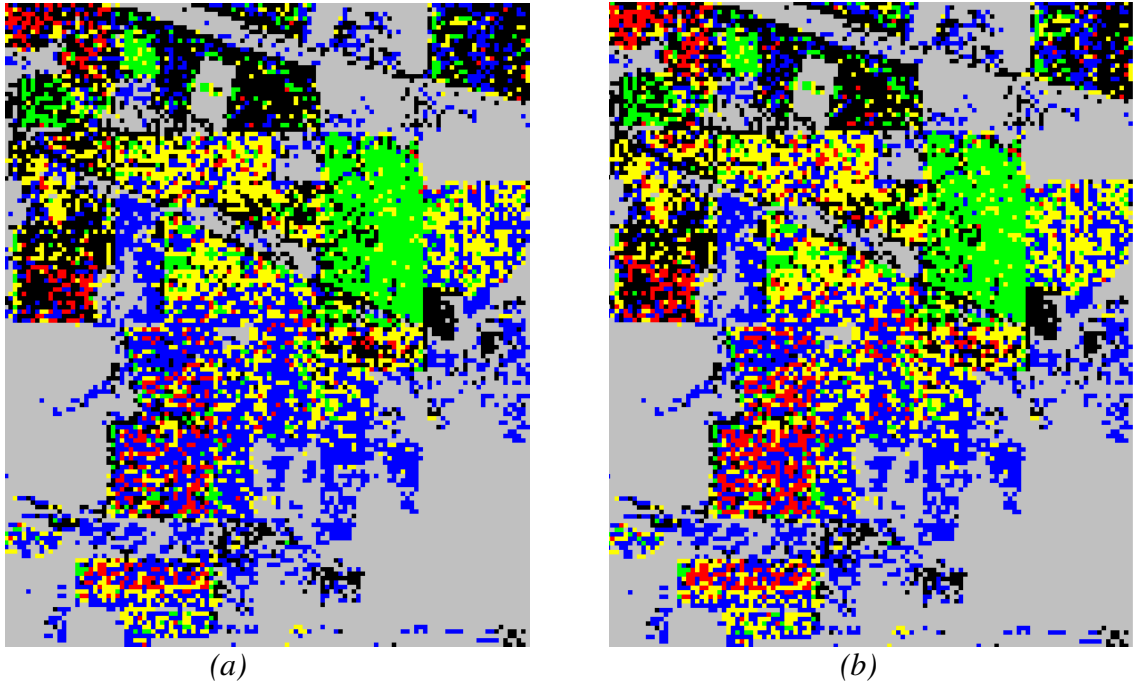


Figura 30 – (a) Imagem temática sem a inclusão de amostras semi-rotuladas (início do processo) (b) Imagem temática com a inclusão de amostras semi-rotuladas refinadas (final do processo). Legenda de cores conforme Figura 22(seção4.8).

Na imagem temática deste experimento houve um incremento na acurácia da classificação, e isto pode ser visto na classe *corn_min* (vermelho) no canto superior esquerdo, na classe *corn_notill* (amarelo) e também na classe *soy_clean* (preto).

Nas Figuras 31 e 32 estão representados os gráficos da acurácia média do usuário e do produtor, respectivamente, onde é possível observar um considerável aumento na acurácia entre as etapas do processo e na acurácia final, quando comparado com os experimentos anteriores.

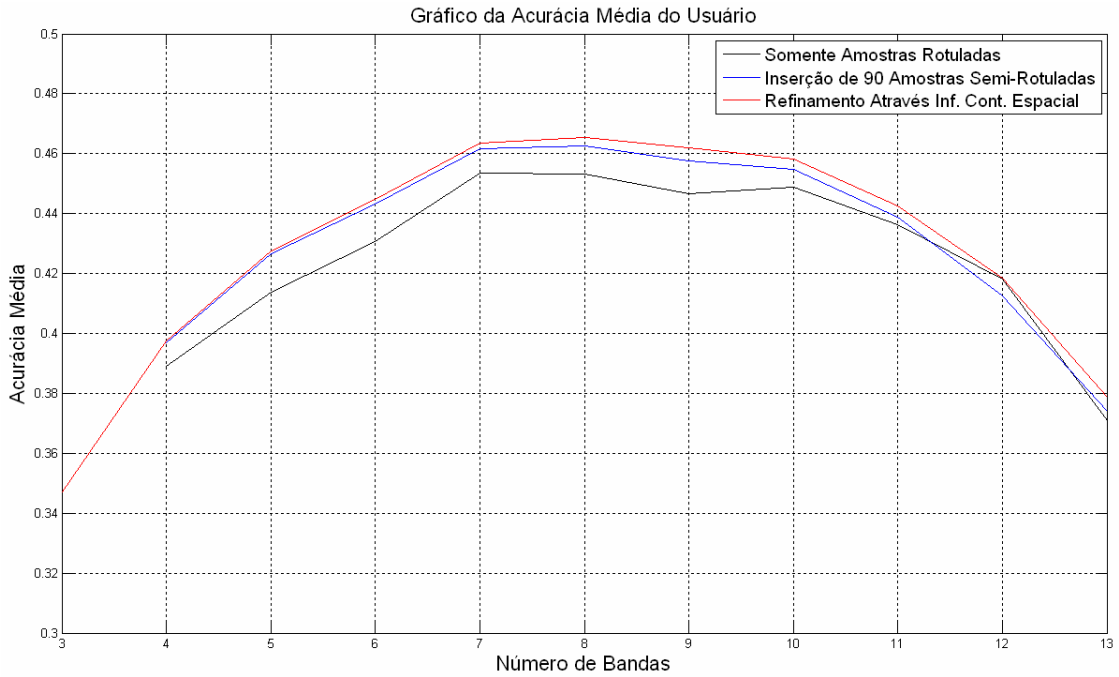


Figura 31 - Acurácia da classificação (usuário), utilizando 15 bandas, 16 amostras de treinamento rotuladas e incremento de 90 amostras semi-rotuladas.

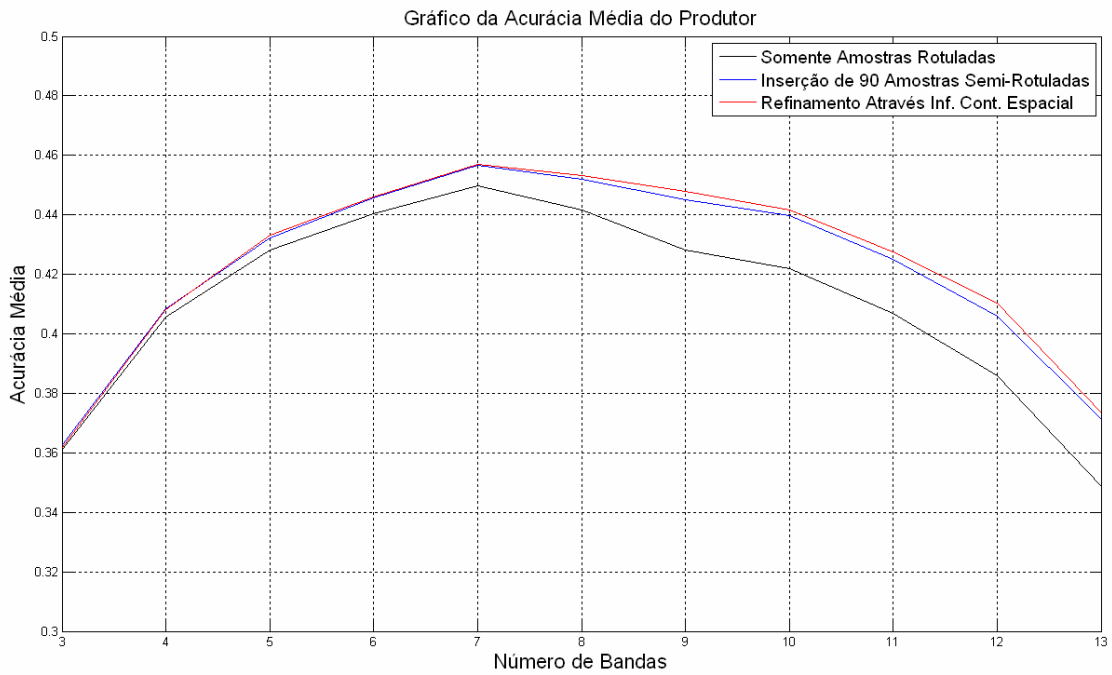


Figura 32 - Acurácia da classificação (produtor), utilizando 15 bandas, 16 amostras de treinamento rotuladas e incremento de 90 amostras semi-rotuladas.

TABELA 4.7 – Matriz de contingência resultante da metodologia proposta neste estudo, empregando dados com dimensionalidade igual a 15, 16 amostras rotuladas e 90 amostras semi-rotuladas.

Classes da imagem classificada	Classes da Imagem de Referência					Total de Amostras	Acurácia do Usuário (%)
	corn_min	corn_notill	soy_clean	soy_min	soy_notill		
corn_min	102	27	42	4	11	186	0,548
corn_notill	64	129	80	41	26	340	0,379
soy_clean	56	73	104	42	36	311	0,334
soy_min	6	30	29	194	25	284	0,683
soy_notill	79	48	52	26	209	414	0,505
Total de Amostras	307	307	307	307	307		
Acurácia do Produtor (%)	0,332	0,420	0,339	0,632	0,681		

Neste experimento pode ser visto (Tabela 4.7) que as acurácias das classes obtiveram um ganho considerável em comparação ao Experimento 3, no entanto, ainda ocorrem erros de inclusão em todas as classes.

4.11.4 EXPERIMENTO 4

Para a execução deste experimento foram utilizados dados com dimensionalidade igual a 30, 31 amostras rotuladas e 30 amostras semi-rotuladas. As imagens temáticas mostradas na Figura 33 ilustram os resultados finais obtidos, enquanto a Tabela 4.8 apresenta os resultados para a matriz de contingência.

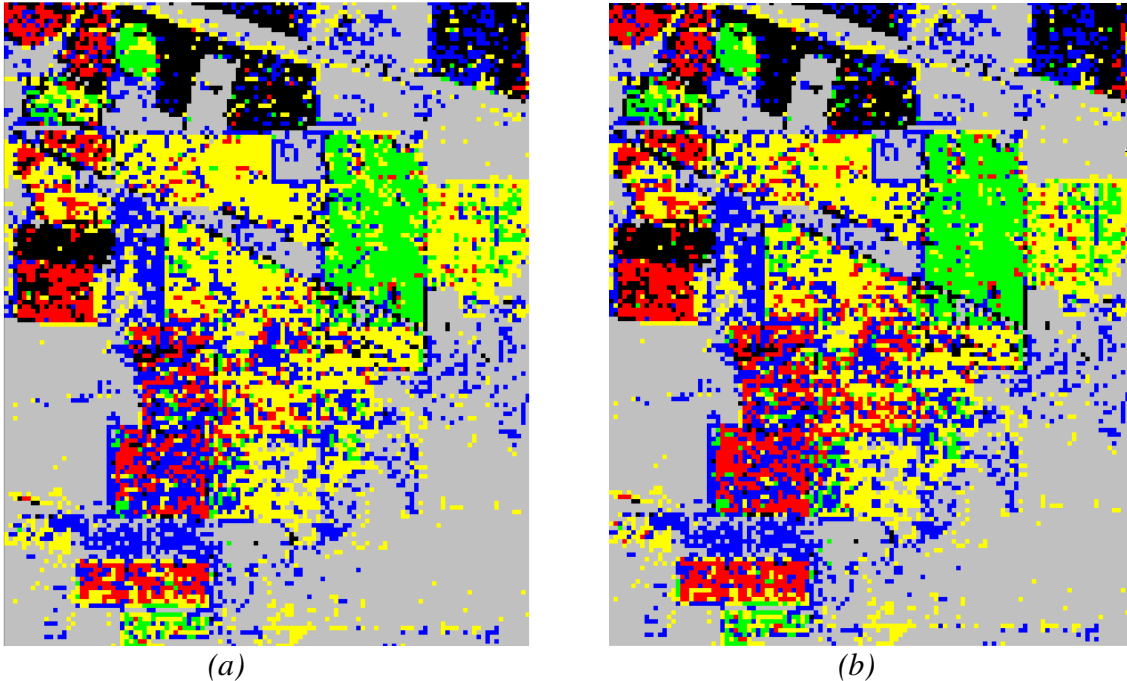


Figura 33 – (a) Imagem temática sem a inclusão de amostras semi-rotuladas (início do processo) (b) Imagem temática com a inclusão de amostras semi-rotuladas (final do processo). Legenda de cores conforme Figura 22(seção4.8).

Na Figura 33 os pixels que não pertencem a nenhuma das classes começam a ser atribuídos, de forma mais precisa, à classe de fundo, enquanto que as demais classes sofrem um pequeno incremento na classificação, como pode ser visto nas classes *corn_min* (vermelho) e *soy_clean* (preto),

As acurácias, média, do usuário e do produtor estão representadas nas Figuras 34 e 35, respectivamente. Neste estágio a contribuição para o aumento da acurácia com a inclusão de informações de contexto espacial começa a ser mais significativa do que a contribuição que ocorre quando a ponderação das amostras semi-rotuladas é feita empregando somente informações espectrais.

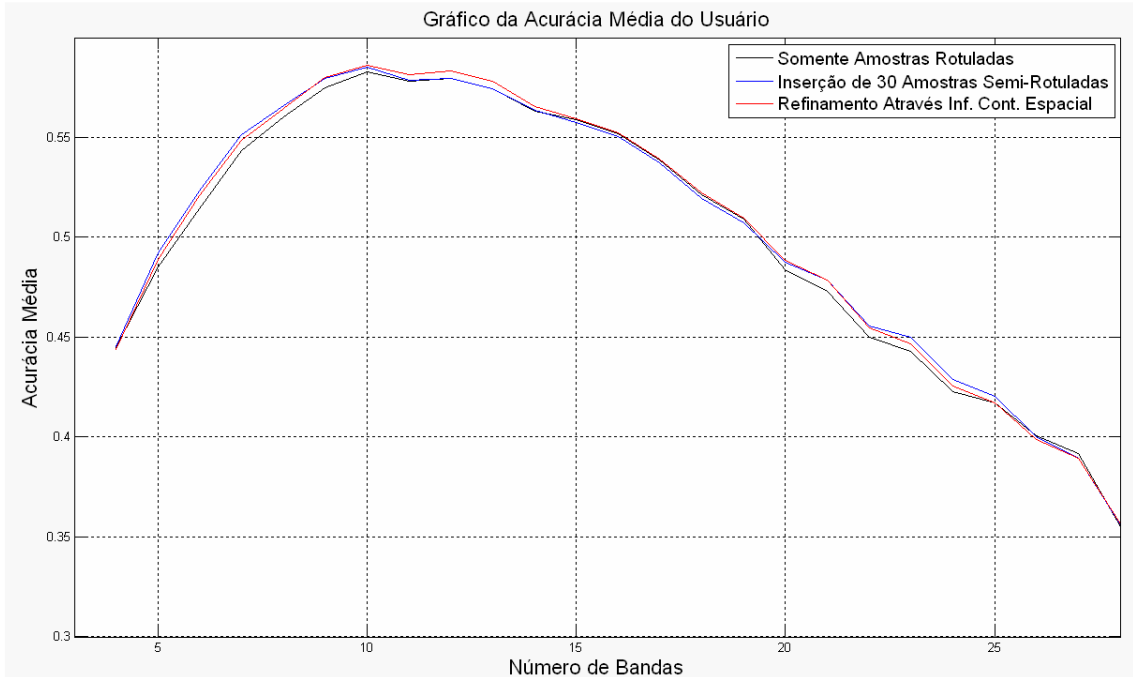


Figura 34 - Acurácia da classificação (usuário), utilizando 30 bandas, 31 amostras de treinamento rotuladas e incremento de 30 amostras semi-rotuladas.

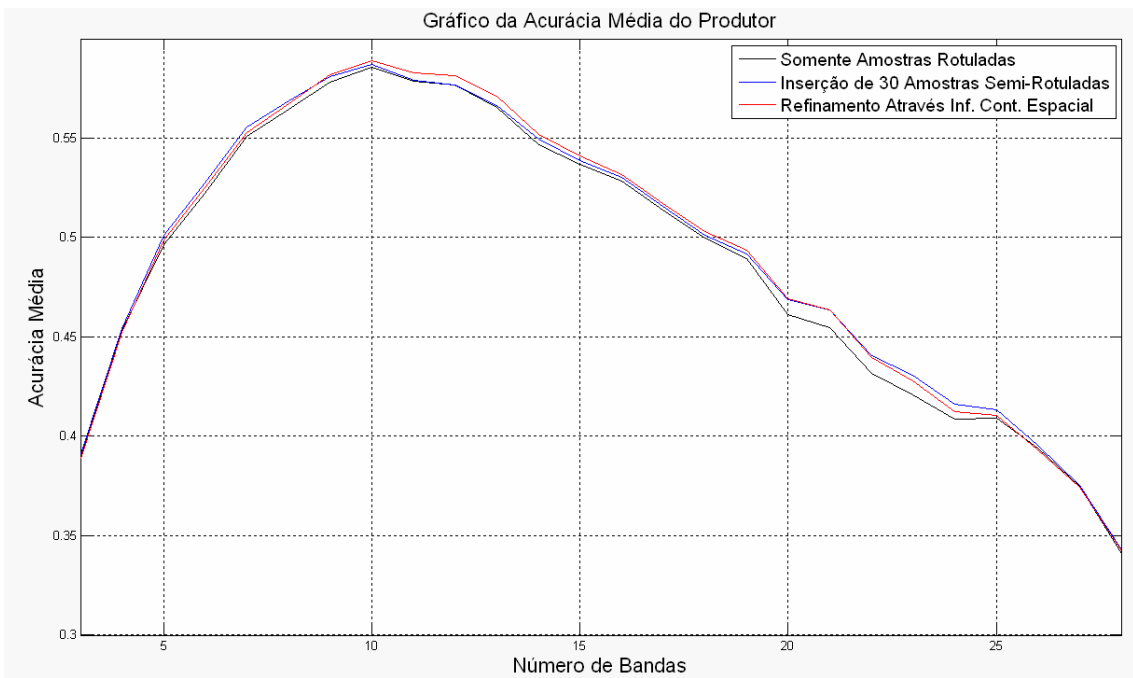


Figura 35 - Acurácia da classificação (produtor), utilizando 30 bandas, 31 amostras de treinamento rotuladas e incremento de 30 amostras semi-rotuladas.

TABELA 4.8 – Matriz de contingência resultante da metodologia proposta neste estudo, empregando dados com dimensionalidade igual a 30, 31 amostras rotuladas e 30 amostras semi-rotuladas.

Classes da imagem classificada	Classes da Imagem de Referência					Total de Amostras	Acurácia do Usuário (%)
	corn_min	corn_notill	soy_clean	soy_min	soy_notill		
corn_min	185	27	73	12	11	308	0,601
corn_notill	32	179	75	52	9	347	0,516
soy_clean	39	53	88	22	33	235	0,374
soy_min	3	32	32	215	7	289	0,744
soy_notill	48	16	39	6	247	356	0,694
Total de Amostras	307	307	307	307	307		
Acurácia do Produtor (%)	0,603	0,583	0,287	0,700	0,805		

Na tabela 4.8 é possível ver o aumento na acurácia da classificação na maioria das classes, embora a classe *soy_clean* ainda seja a de menor acurácia, tanto para o usuário como para o produtor.

4.11.5 EXPERIMENTO 5

Para a execução deste experimento foram utilizados dados com dimensionalidade igual a 30, 31 amostras rotuladas e 60 amostras semi-rotuladas. A Figura 36 ilustra os resultados obtidos representados na forma de imagens temáticas.

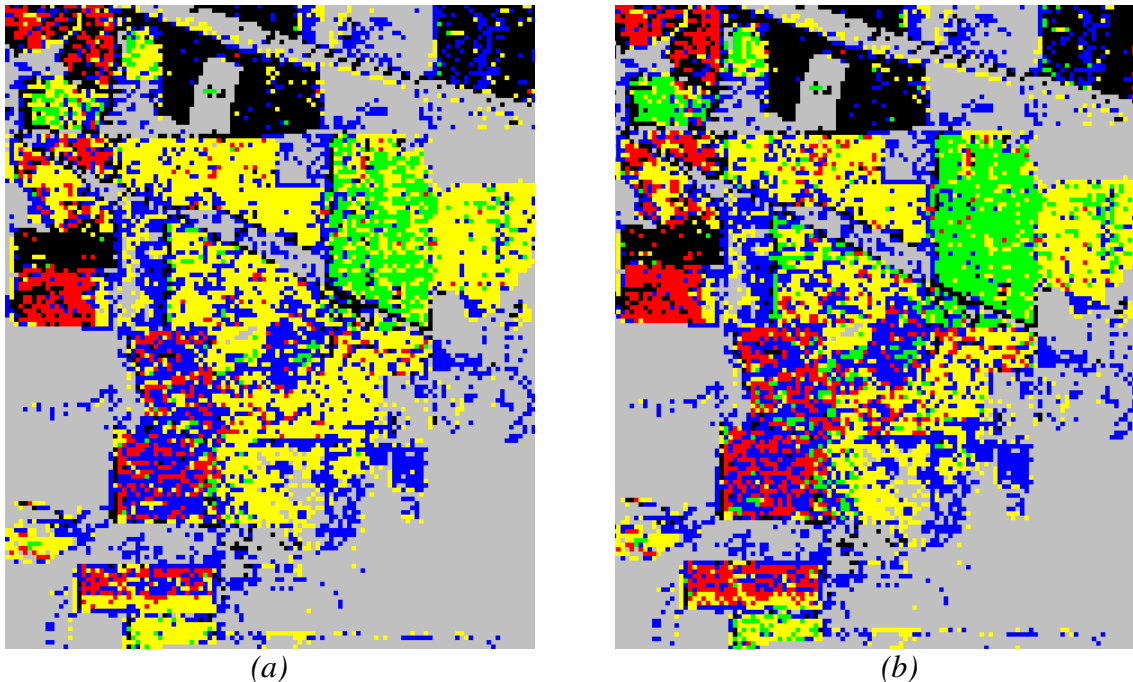


Figura 36 – (a) Imagem temática sem a inclusão de amostras semi-rotuladas (início do processo) (b) Imagem temática com a inclusão de amostras semi-rotuladas (final do processo). Legenda de cores conforme Figura 22(seção4.8).

Na Figura 36 a classe *soy-clean* (preto), *soy_notill* (verde) e *corn_min* (vermelho) no canto superior esquerdo, têm uma melhora considerável na classificação, embora a classe *soy_min* (azul) ainda tenha confusão com as classes *corn_min* (vermelho) e *corn_notill* (amarelo) no centro da imagem.

Os resultados para a matriz de contingência estão apresentados na Tabela 4.9. As Figuras 37 e 38 representam na forma de gráficos a acurácia média do usuário e do produtor, respectivamente.

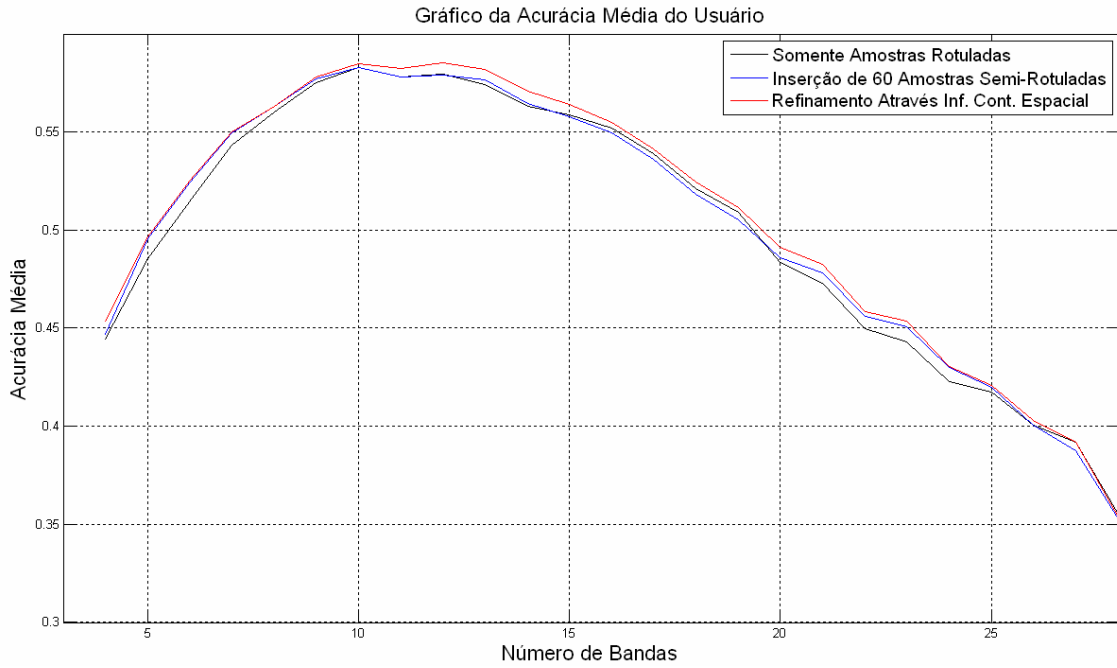


Figura 37 - Acurácia da classificação (usuário), utilizando 30 bandas, 31 amostras de treinamento rotuladas e incremento de 60 amostras semi-rotuladas.

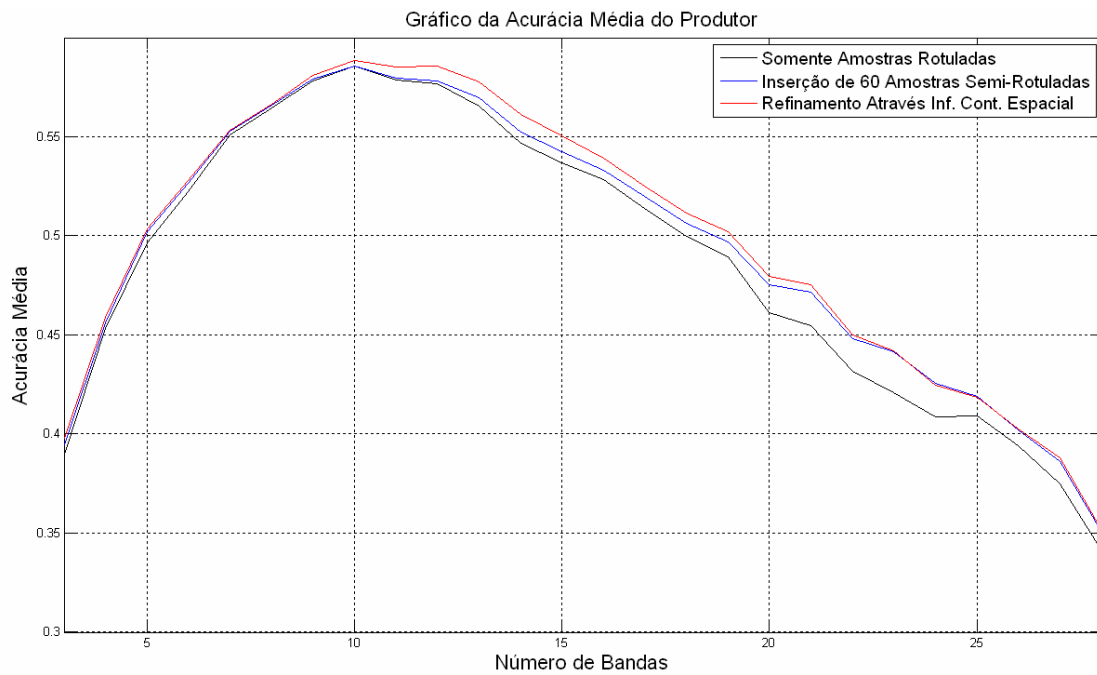


Figura 38 - Acurácia da classificação (produtor), utilizando 30 bandas, 31 amostras de treinamento rotuladas e incremento de 60 amostras semi-rotuladas.

TABELA 4.9 – Matriz de contingência resultante da metodologia proposta neste estudo, empregando dados com dimensionalidade igual a 30, 31 amostras rotuladas e 60 amostras semi-rotuladas.

Classes da imagem classificada	Classes da Imagem de Referência					Total de Amostras	Acurácia do Usuário (%)
	corn_min	corn_notill	soy_clean	soy_min	soy_notill		
corn_min	172	20	58	5	5	260	0,662
corn_notill	35	190	65	82	11	383	0,496
soy_clean	50	49	107	18	20	244	0,439
soy_min	1	36	35	199	2	273	0,729
soy_notill	49	12	42	3	269	375	0,717
Total de Amostras	307	307	307	307	307		
Acurácia do Produtor (%)	0,560	0,619	0,349	0,648	0,876		

A tabela 4.9 mostra que ainda existe uma confusão entre as classes, mesmo que no final do processo a acurácia média tenha aumentado.

4.11.6 EXPERIMENTO 6

Para a execução deste experimento foram utilizados dados com dimensionalidade igual a 30, 31 amostras rotuladas e 90 amostras semi-rotuladas. As imagens temáticas mostradas na Figura 39 ilustram os resultados finais obtidos utilizando somente as amostras rotuladas, e com a inclusão das amostras semi-rotuladas respectivamente.

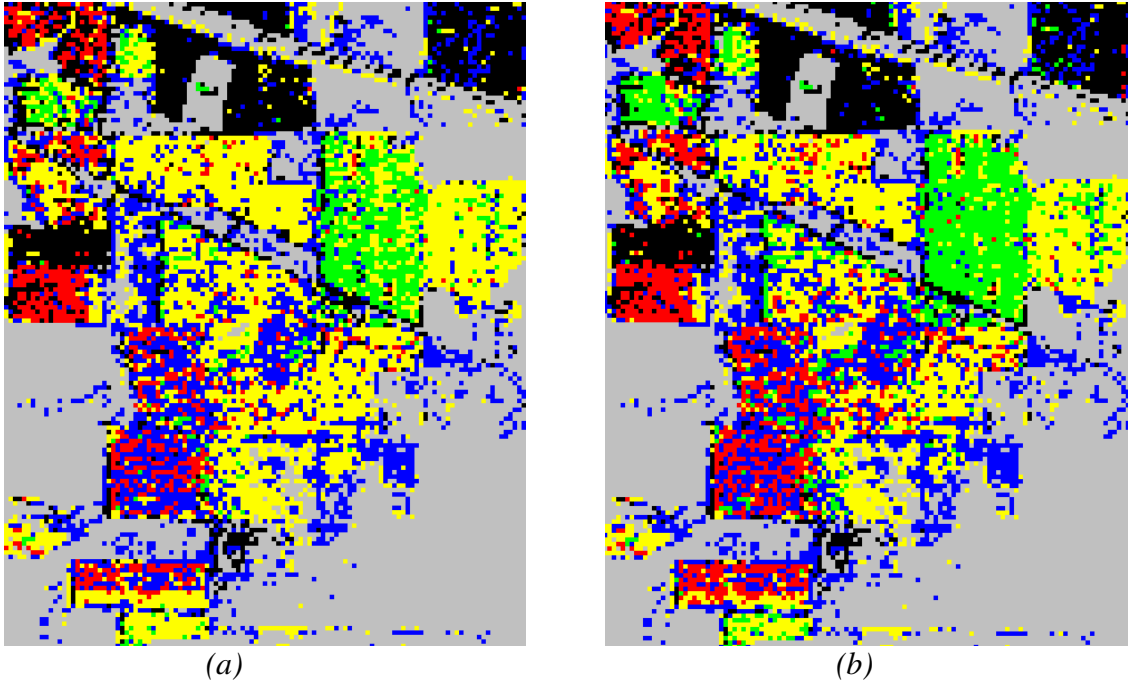


Figura 39 – (a) Imagem temática sem a inclusão de amostras semi-rotuladas (início do processo) (b) Imagem temática com a inclusão de amostras semi-rotuladas (final do processo). Legenda de cores conforme Figura 22(seção4.8).

A Figura 39, por inspeção visual, não mostra significativa melhora no processo de classificação após a inclusão das amostras semi-rotuladas refinadas utilizando informações de contexto espacial. No entanto, os gráficos da acurácia média do usuário e do produtor, que são vistos nas Figuras 40 e 41, respectivamente, mostram que o processo obteve um modesto aumento na acurácia quando se utilizou as amostras semi-rotuladas ponderadas por meio da informação espectral, enquanto que, com as amostras semi-rotuladas refinadas por meio da informação de contexto espacial este aumento se torna significativamente maior. Os resultados podem ser avaliados também na Tabela 4.10.

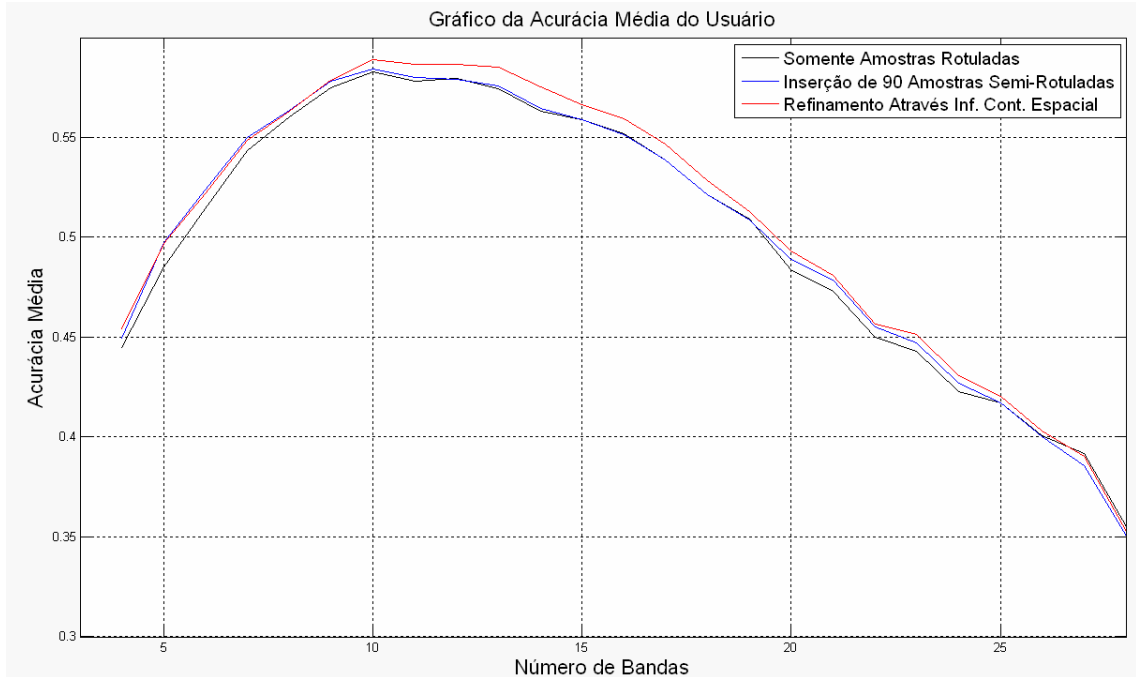


Figura 40 - Acurácia da classificação (usuário), utilizando 30 bandas, 31 amostras de treinamento rotuladas e incremento de 90 amostras semi-rotuladas.

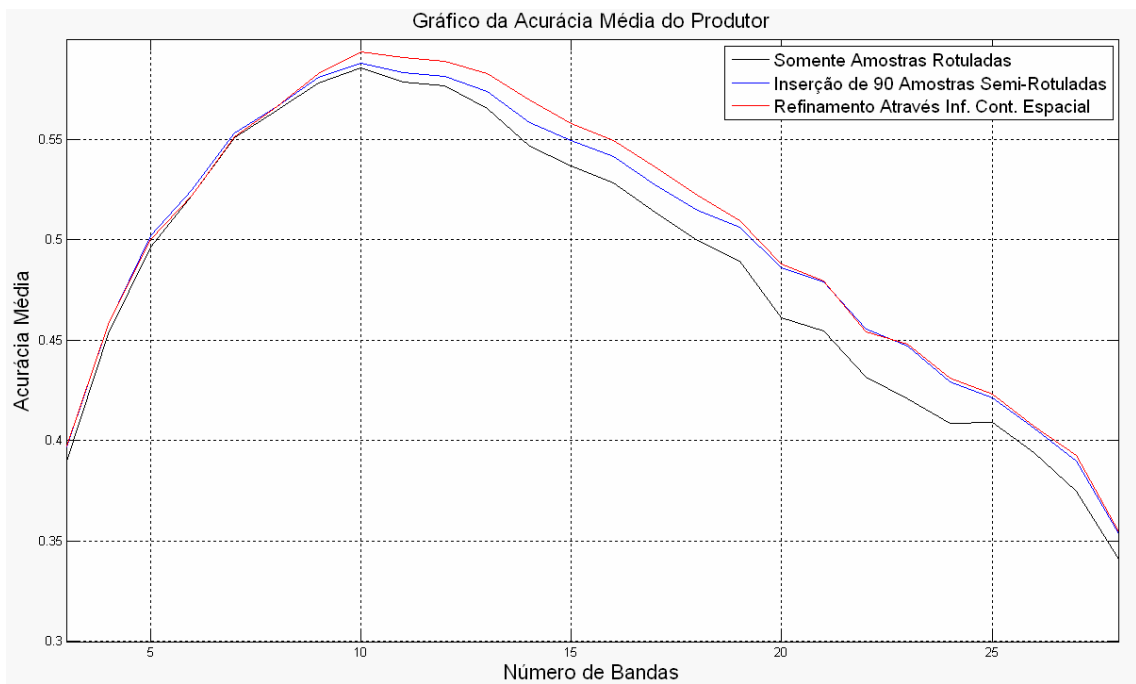


Figura 41 - Acurácia da classificação (produtor), utilizando 30 bandas, 31 amostras de treinamento rotuladas e incremento de 90 amostras semi-rotuladas.

TABELA 4.10 – Matriz de contingência resultante da metodologia proposta neste estudo, empregando dados com dimensionalidade igual a 30, 31 amostras rotuladas e 90 amostras semi-rotuladas.

Classes da imagem classificada	Classes da Imagem de Referência					Total de Amostras	Acurácia do Usuário (%)
	corn_min	corn_notill	soy_clean	soy_min	soy_notill		
corn_min	173	21	58	5	3	260	0,665
corn_notill	32	191	61	80	11	375	0,509
soy_clean	50	48	110	18	20	246	0,447
soy_min	1	35	35	201	2	274	0,734
soy_notill	51	12	43	3	271	380	0,713
Total de Amostras	307	307	307	307	307		
Acurácia do Produtor (%)	0,564	0,622	0,358	0,655	0,883		

Na Tabela 4.10, é possível observar que houve uma boa diferenciação entre algumas classes, assim como no Experimento 5. Novamente, como ocorrido anteriormente, a classe *soy_clean* apresentou problemas na classificação, tendo mais da metade dos pixels distribuídos entre as demais classes.

4.11.7 EXPERIMENTO 7

Para a execução deste experimento foram utilizados dados com dimensionalidade igual a 100, 101 amostras rotuladas e 80 amostras semi-rotuladas. As imagens temáticas mostradas na Figura 42 ilustram os resultados finais obtidos utilizando somente as amostras rotuladas, e com a inclusão das amostras semi-rotuladas respectivamente.

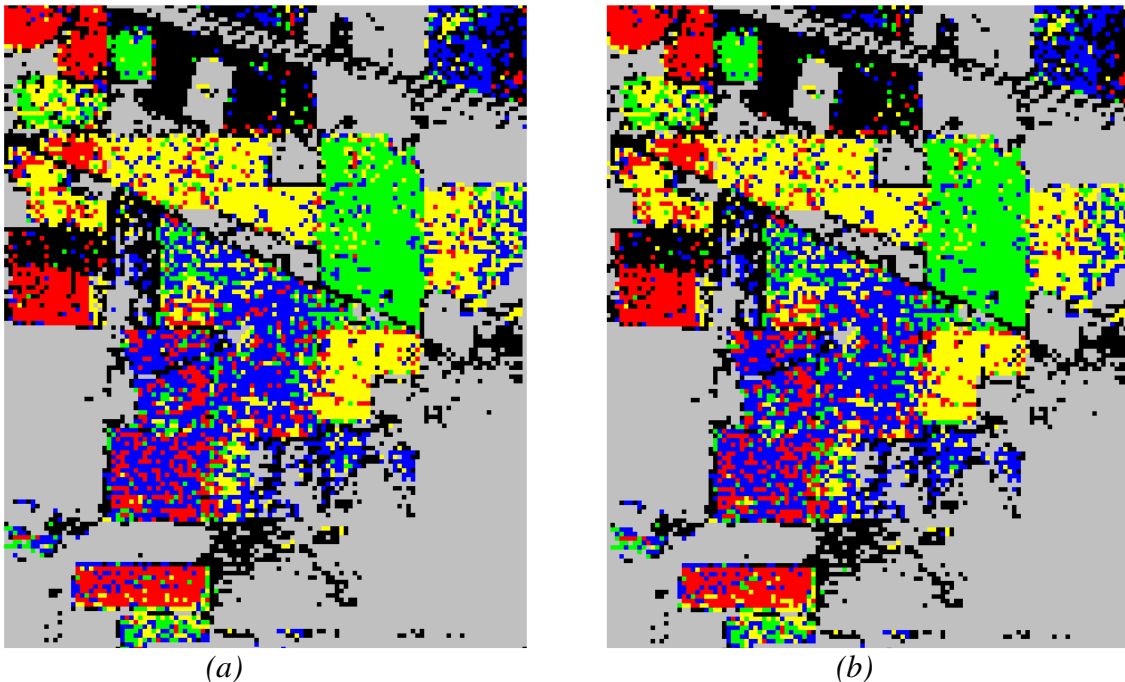


Figura 42 – (a) Imagem temática sem a inclusão de amostras semi-rotuladas (início do processo) (b) Imagem temática com a inclusão de amostras semi-rotuladas (final do processo). Legenda de cores conforme Figura 22(seção4.8).

A Figura 42, por inspeção visual, não mostra significativa melhora no processo de classificação após a inclusão das amostras semi-rotuladas refinadas utilizando informações de contexto espacial, assim como ocorre nas imagens temáticas do experimento 6 (Figura 39). Um pequeno incremento pode, no entanto, ser observado nos gráficos da acurácia média do usuário e do produtor, que são mostrados nas Figuras 43 e 44, respectivamente,

Com o aumento na dimensionalidade dos dados, os efeitos do fenômeno de Hughes tornam-se mais visíveis. O pico da curva desloca-se para valores mais altos da dimensionalidade dos dados, com acurácias mais elevadas. Em ambos os experimentos pode-se observar que na região do pico da curva (em torno da dimensão 30), a contribuição das amostras semi-rotuladas tende a exibir o comportamento esperado, com as amostras semi-rotuladas ponderadas por meio de informações espectrais e de contexto resultando nos valores mais altos para a acurácia. A contribuição devido à adição de informações de contexto espacial tende, entretanto, a cair mais rapidamente do que aquela que utiliza somente informações espectrais, para valores crescentes na dimensionalidade dos dados. Este comportamento deve-se ao fato de que as informações de contexto espacial são estimadas a partir da imagem temática. Na medida em que a acurácia da imagem temática diminui, a contribuição do contexto espacial também tende a decrescer. Os resultados podem ser avaliados também na Tabela 4.11.

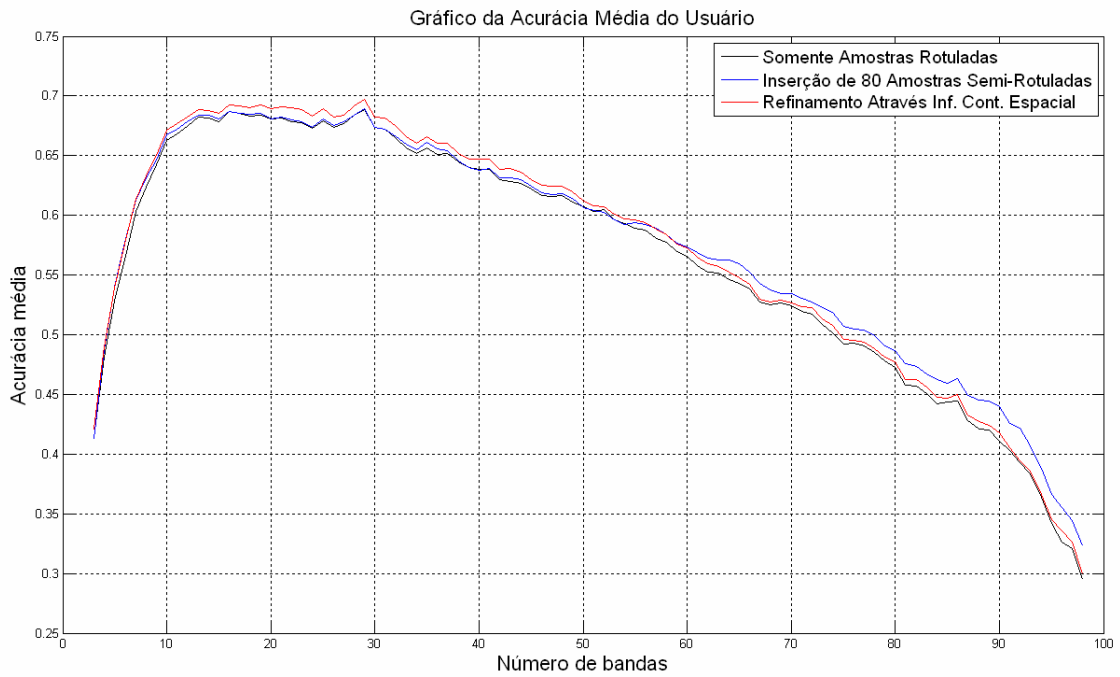


Figura 43 - Acurácia da classificação (usuário), utilizando 100 bandas, 101 amostras de treinamento rotuladas e incremento de 80 amostras semi-rotuladas.

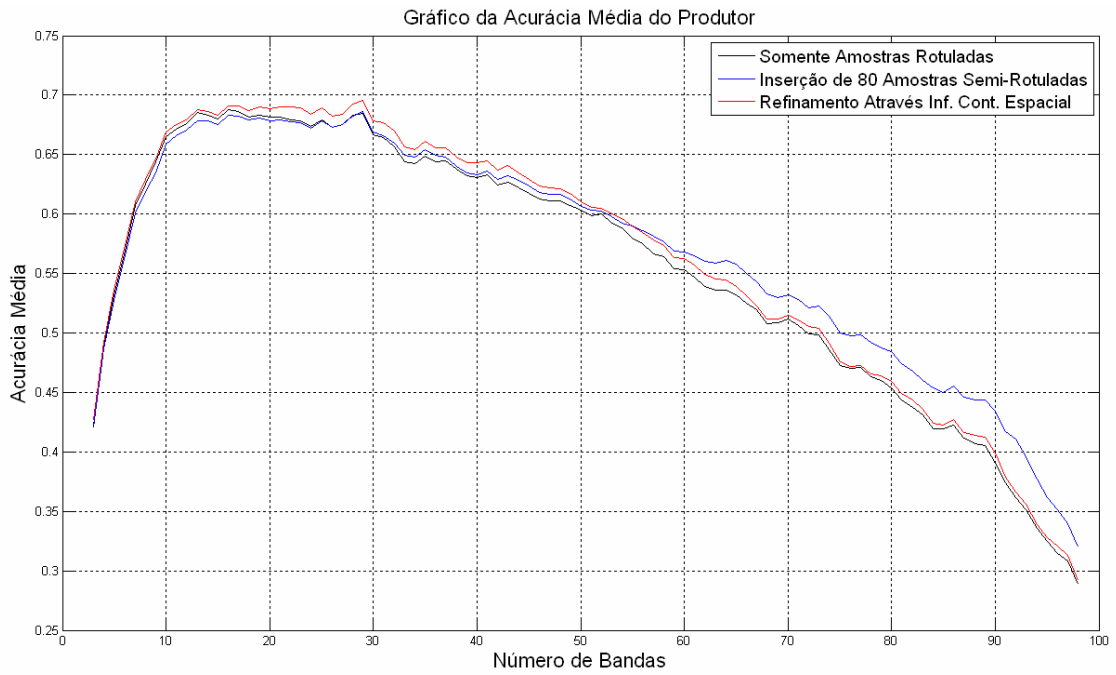


Figura 44 - Acurácia da classificação (produtor), utilizando 100 bandas, 101 amostras de treinamento rotuladas e incremento de 80 amostras semi-rotuladas.

TABELA 4.11 – Matriz de contingência resultante da metodologia proposta neste estudo, empregando dados com dimensionalidade igual a 100, 101 amostras rotuladas e 80 amostras semi-rotuladas.

Classes da imagem classificada	Classes da Imagem de Referência					Total de Amostras	Acurácia do Usuário (%)
	corn_min	corn_notill	soy_clean	soy_min	soy_notill		
corn_min	221	29	48	4	7	309	0,715
corn_notill	30	214	30	47	2	323	0,663
soy_clean	15	37	160	29	8	249	0,643
soy_min	8	18	41	221	7	295	0,749
soy_notill	33	9	28	6	283	359	0,788
Total de Amostras	307	307	307	307	307		
Acurácia do Produtor (%)	0,720	0,697	0,521	0,720	0,922		

Na Tabela 4.11, é possível observar que houve um aumento na acurácia de mais de 12% em todas as classes, quando comparado com os demais experimentos. No entanto, como ocorrido anteriormente, a classe *soy_clean* apresentou a mais baixa acurácia, com problemas na classificação, tendo quase metade dos pixels distribuídos entre as demais classes.

Uma observação que se deve fazer neste experimento é a utilização de um limiar diferenciado, o qual usa um limiar de corte sobre a média dos pesos máximos obtidos durante o processo. Este limiar, diferentemente do sugerido anteriormente, tem o número de pixels atribuído a classe de fundo atrelado à média dos pesos, desta forma se torna variável ao longo do processo, e com o aumento da dimensionalidade, este limiar tende a atribuir um número maior de pixels a classe de fundo, quando comparado com dimensionalidades mais baixas. No entanto, este limiar não foi utilizado nos experimentos anteriores devido ao fato, já mencionado, de que com o aumento da dimensionalidade a atribuição de pixels à classe de fundo se torna cada vez maior.

4.12 ANÁLISE DOS EXPERIMENTOS

Os experimentos aqui apresentados são apenas um pequeno grupo de todos os possíveis experimentos que podem ser realizados com as variáveis mencionadas. No entanto, estes experimentos representam um conjunto elucidativo que serve para ilustrar o uso de amostras semi-rotuladas na classificação de dados em alta dimensionalidade. Na metodologia proposta nesta dissertação as amostras semi-rotuladas são ponderadas por meio de informações espectrais e de contexto espacial. No caso de informações espectrais a

metodologia empregada na estimação dos pesos representa uma contribuição original desta dissertação.

Os experimentos servem para ilustrar a contribuição destas amostras no incremento da acurácia dos resultados produzidos pelo classificador paramétrico MVG. De um modo geral, pode-se observar nestes experimentos o comportamento que seria de esperar. A inclusão de amostras semi-rotuladas ponderadas somente em função de informações espectrais contribui para um incremento, embora modesto, na acurácia dos resultados, num intervalo extenso de valores para a dimensionalidade dos dados. O refinamento dos pesos por meio da adição de informações de contexto espacial contribui para um incremento adicional na acurácia. Pode-se aqui também observar que a contribuição das informações de contexto espacial tende em alguns casos a decrescer, para valores crescentes da dimensionalidade dos dados. Este comportamento pode ser facilmente explicado pelo fato de que as informações de contexto espacial são extraídas da imagem temática produzida para a mesma dimensionalidade dos dados. Devido aos efeitos do fenômeno de Hughes, a acurácia da imagem temática tende a cair após um certo valor da dimensionalidade dos dados. Como consequência, a contribuição proveniente das informações de contexto espacial tende também a diminuir. Uma variável adicional que também influencia a contribuição das informações de contexto espacial refere-se à questão da classe de fundo. A definição desta classe depende do critério adotado e contribui significativamente para uma maior ou menor contribuição dos dados de contexto espacial.

5 CONCLUSÕES E SUGESTÕES

5.1 CONCLUSÕES

Nesta dissertação foi investigado o uso de amostras semi-rotuladas com o objetivo de incrementar o número de amostras de treinamento, e assim aumentar a confiabilidade das estimativas dos parâmetros em classificadores paramétricos. Como já é conhecido, o uso de variáveis em alta dimensionalidade apresentam um grande potencial na separação de classes muito semelhantes entre si, que por esta razão são de difícil discriminação sempre que se empregada dados mais tradicionais em media/baixa dimensionalidade. O problema, neste caso, resulta em um maior número de parâmetros a serem estimados, principalmente na matriz de covariância.

Na literatura, estão relatadas possíveis metodologias que objetivam minimizar o problema que surge na estimação de um número grande de parâmetros a partir de um número geralmente limitado de amostras de treinamento. Entre as possíveis alternativas propostas, figura a opção de incrementar o número de amostras de treinamento pela inclusão das chamadas amostras semi-rotuladas. Neste caso, um ponto chave constitui a estimação dos pesos para estas amostras, isto é, a estimação do grau de confiança de que cada amostra efetivamente pertença à classe em questão.

Nesta dissertação investigou-se uma metodologia visando estimar os pesos das amostras semi-rotuladas em um processo a duas etapas. Na primeira são utilizadas informações espectrais somente, e a metodologia proposta apresenta uma contribuição nova na estimação dos pesos. Na segunda etapa, os pesos estimados anteriormente são refinados por meio de informações de contexto espacial, em um processo de relaxação.

Alguns experimentos foram realizados com o objetivo de avaliar a eficácia da metodologia proposta. A acurácia produzida por um classificador paramétrico (Máxima Verossimilhança Gaussiana) aplicado a dados em alta dimensionalidade (190 bandas), foi estimada em função de algumas variáveis: a dimensionalidade dos dados, o número de amostras rotuladas, o número de amostras semi-rotuladas. O número de iterações na estimação dos pesos das amostras semi-rotuladas e o valor para o limiar é que definem a “classe de fundo” no processo. Com base neste limiar são identificados pixels com baixo grau de pertinência a qualquer uma das classes em estudo, e assim, estes pixels são agrupados em uma única classe, denominada “classe de fundo”.

Nesta dissertação foram apresentados e discutidos alguns destes experimentos. Pode-se observar, que de modo geral, a inclusão das amostras semi-rotuladas contribuiu com um pequeno incremento na acurácia média final dos resultados produzidos pelo classificador. Conforme era esperado, os maiores ganhos na acurácia foram obtidos com a inclusão de amostras semi-rotuladas ponderadas por informações espectrais com os pesos re-estimados (refinados) pelas informações de contexto espacial.

Embora os resultados obtidos tenham confirmado a adequação da metodologia proposta, fica claro que pesquisas adicionais neste tópico podem resultar em ganhos maiores na acurácia. Na etapa que estima os pesos por meio de informações espectrais, outras abordagens na ponderação das amostras semi-rotuladas podem resultar em valores estimados mais confiáveis para os pesos. Na etapa que inclui informações de contexto espacial, também, outras abordagens de inserção de informações de contexto espacial podem retornar melhores resultados, no entanto, o item que merece estudos adicionais é o do critério para definição da classe de fundo. O conjunto total dos experimentos realizados serviu para ilustrar a importância que a classe de fundo apresenta na extração de informações de contexto espacial.

5.2 SUGESTÕES

Conforme mencionado anteriormente, embora alguns resultados positivos tenham sido obtidos, outros itens da metodologia proposta nesta dissertação podem ser mais bem compreendidos e se constituem em tópicos interessantes para futuras pesquisas.

Na primeira etapa do processo de estimação dos pesos das amostras semi-rotuladas, se inicia estimando o peso de cada amostra de uma forma relativa, em termos do grau de pertinência de cada amostra às várias classes presentes na cena. Em um segundo momento o peso da classe vencedora é colocado em uma escala de pertinência intra-classe por meio da comparação deste com o máximo valor possível para aquela classe, isto é, o valor correspondente à média da classe. Embora esta abordagem tenha resultado em valores aceitáveis sugere-se aqui que seja investigada em maior profundidade, com o objetivo de que estes pesos representem mais fidedignamente o grau de pertinência de cada amostra às classes presentes na cena.

Na etapa que adiciona informações de contexto espacial, um item adequado para futuras investigações refere-se à questão da classe de fundo. O processo de extração de informações de contexto espacial baseia-se nos resultados previamente obtidos pelo classificador. Isto significa que a confiabilidade das informações de contexto espacial é

fortemente influenciada pela acurácia da imagem temática produzida na etapa anterior. Em imagens de cenas naturais, o número de classes presentes é geralmente muito grande, com algumas classes sendo representadas por um número reduzido de pixels. Deste fato resulta que nem todas as classes são amostradas e, portanto, não dispõem de funções de decisão no classificador. Um cuidado que se deve ter consiste, portanto, em não permitir que estes pixels pertencentes às classes não amostradas sejam erroneamente atribuídos a uma das classes em consideração. Normalmente este procedimento pode ser feito através da aplicação de um limiar na própria função de decisão, rejeitando pixels que apresentam uma baixa probabilidade de pertinência mesmo à classe vencedora. A estimação de limiares para dados em média ou baixa dimensionalidade se constitui em uma tarefa simples, no entanto, para dados em alta dimensionalidade, surgem alguns problemas devido ao fato das amostras tenderem a se deslocar para a cauda das curvas das distribuições. Nesta dissertação se propõe uma alternativa a ser pesquisada, com o limiar sendo definido a partir do histograma dos pesos e de suas respectivas médias. Sugere-se, entretanto, que investigações adicionais sejam desenvolvidas no sentido de se obter critérios mais seguros e objetivos na escolha deste limiar e, portanto, na formação da “classe de fundo”. Um critério mais confiável certamente contribuirá decisivamente na melhora da contribuição da informação de contexto espacial.

REFERÊNCIAS BIBLIOGRÁFICAS

- [AEB 94] AEBERHARD, S., COOMANS, D., De Vel, O., Comparative Analysis of Statistical Pattern Recognition Methods in High Dimensional Settings, **Pattern Recognition**, vol. 27, no. 8, pp. 1065-1077, 1994.
- [BER 07] BERGE, A., JENSEN, A.C., SOLBERG, A.S., Sparse inverse covariance estimates for hyperspectral image classification, **IEEE Transactions on Geoscience and Remote Sensing**, vol. 45, no. 5, pp. 1399–1407, May de 2007.
- [BYE 92] BYEUNGWOO, J., LANDGREBE, D., Classification with Spatio-Temporal Interpixel Class Dependency Contexts, **IEEE Transactions on Geoscience and Remote Sensing**, vol. 30, no. 4, pp. 663-672, 1992.
- [CEN 91] CENTENO, J. A. S., **Exploração de atributos de contexto em imagens digitais no estudo da cobertura vegetal em bacias hidrográficas**. 1991. Dissertação (Mestrado em Sensoriamento Remoto) – Programa de Pós-Graduação em Sensoriamento Remoto e Meteorologia, Centro Estadual de Pesquisas em Sensoriamento Remoto e Meteorologia, UFRGS, Porto Alegre, RS, 1991.
- [CHE 08] CHEN, C. H., *Image Processing for Remote Sensing*, Taylor & Francis, 2008.
- [COR 99] 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**, vol.20, no. 17, pp. 3345-3365, 1999.
- [DUD 01] DUDA, R. O., HART P.E.,STORK D.G., *Pattern Classification*, 2ª ed., John Wiley & Sons Inc. EUA, 2001.
- [ERB 01] ERBERT, M., **Uso da Análise discriminante regularizada no reconhecimento de padrões em imagens digitais hiperespectral de sensoriamento remoto**. 2001. Dissertação (Mestrado em Sensoriamento Remoto) – Programa de Pós-Graduação em Sensoriamento Remoto e Meteorologia, Centro Estadual de Pesquisas em Sensoriamento Remoto e Meteorologia, UFRGS, Porto Alegre, RS, 2001.
- [EKL 80] EKLUNDH, J. O., YAMAMOTO, H., ROSENFELD A., A Relaxation Method for Multispectral Pixel Classification, **IEEE Transactions on Pattern Analysis and Machine Intelligence**, vol. PAMI-2, no. 1, Jan. 1980.
- [FRI 89] FRIEDMAN, J. H. Regularized Discrimination Analysis. **Journal of the American Statistical Association**, Theory and Methods, vol. 84, no. 405, 1989.
- [FUK 90] FUKUNAGA, K., *Introduction to Statistical Pattern Recognition*, 2ª ed., School of Electrical Engineering, Purdue University, West Lafayette, Academic Press Inc. 1990.

- [FUK 89] FUKUNAGA, K., HAYES R.R., Effects of Sample Size in Classifier Design, **IEEE Transaction Pattern Analysis and Machine Intelligence**, vol. 11, no. 8, pp. 873-885, Aug. 1989.
- [HAR 09] HAERTEL, V., MORAES, D. A. O., Contexto Espacial na Ponderação de Amostras Semi-Rotuladas, **Anais: XIV Simpósio Brasileiro de Sensoriamento Remoto**, Natal, Brasil, 25-30 abril 2009, INPE, pp. 7805-7812.
- [HAS 95] HASTIE, T., BUJA, A., TIBSHIRANI, R., Penalized Discriminant Analysis, **Annals of Statistics**, vol. 23, pp. 73-102, 1995.
- [HSI 89] HSIAO, J. Y., SAWCHUK, A. A., Supervised Texture Image Segmentation Using Feature Smoothing and Probabilistic Relaxation, **IEEE Transaction on Pattern Analysis and Machine Intelligence**, vol. 11, no. 12, Dez. 1989.
- [HOF 96] HOFFBECK J. P., LANDGREBE D. A., Covariance Matrix Estimation and Classification With Limited Training Data, **IEEE Transactions on Pattern Analysis and Machine Intelligence**, vol. 18, no 7, July 1996.
- [HSI 98] HSIEH, P., LANDGREBE, D.A., Classification of High Dimensional Data, **TR-ECE 98-4**, School of Elec. and Comp. Eng., Purdue University, May 1998.
- [HUM 83] HUMEEL, R. A., ZUCKER, S. W., On the Foundations of Relaxation Labeling Processes, **IEEE Transactions on Pattern Analysis and Machine Intelligence**, vol. PAMI-5, no. 3, May 1983.
- [INP 01] INPE - INSTITUTO DE PESQUISAS ESPACIAIS - **Introdução ao Sensoriamento Remoto**, disponível em: <http://www.dpi.inpe.br/spring/usuario/sensorr.htm#espectro> Acesso: 10 Novembro 2010.
- [JAC 01] JACKSON, Q., AND LANDGREBE, D., An Adaptive Classifier Design for High-Dimensional Data Analysis With a Limited Training Data Set, **IEEE Transactions on Geoscience and Remote Sensing**, vol. 39, no. 12, pp. 2664-2679, 2001.
- [JAC 02] JACKSON, Q., LANDGREBE, L., Adaptive Bayesian Contextual Classification Based on Markov Random Fields, **IEEE Transactions on Geoscience and Remote Sensing**, vol. 14, no. 11, pp. 2454-2463, 2002.
- [JAI 00] JAIN, A., DUIN, R.P.W., MAO, J., Statistical Pattern Recognition: A Review, **IEEE Transactions on Pattern Analysis and Machine Intelligence**, vol. 22, no. 1, pp. 4-37, 2000.
- [JEO 92] JEON, B., LANDGREBE, D. A., Classification with Spatio-Temporal Interpixel Class Dependency Contexts, **IEEE Transactions on Geoscience and Remote Sensing**, vol. 30, no. 4, pp. 663-672, July 1992.
- [AVI 01] JET PROPULSION LABORATORY - **AVIRIS** - disponível em: <http://aviris.jpl.nasa.gov/>. Acesso: 10 Novembro 2010.

- [AVI 02] JET PROPULSION LABORATORY - **AVIRIS** - disponível em:
<<http://aviris.jpl.nasa.gov/html/aviris.overview.html>>. Acesso: 10 Novembro 2010.
- [AVI 03] JET PROPULSION LABORATORY - **AVIRIS** - disponível em:
<<http://aviris.jpl.nasa.gov/html/aviris.instrument.html>>. Acesso: 10 Novembro 2010.
- [AVI 04] JET PROPULSION LABORATORY - **AVIRIS** - disponível em:
<<http://aviris.jpl.nasa.gov/html/aviris.concept.html>>. Acesso: 10 Novembro 2010.
- [AVI 05] JET PROPULSION LABORATORY - **AVIRIS** - disponível em:
<<http://aviris.jpl.nasa.gov/html/aviris.cube.html>>. Acesso: 10 Novembro 2010.
- [JHU 96] JHUNG, Y., SWAIN, P. H., Bayesian contextual classification based on modified M-estimates and Markov random fields, **IEEE Trans. Geosci. Remote Sensing**, vol.34, no.1, pp. 68-75, Jan. 1996.
- [JIA 99] JIA, X., RICHARDS, J., Segmented Principal Components Transformation for Efficient Hyperspectral Remote-Sensing Image Display and Classification, **IEEE Transactions on Geoscience and Remote Sensing**, vol. 37, no1, pp. 538-542, January 1999.
- [JIA 08] JIA, X., RICHARDS, J., Managing the Spectral-Spatial Mix in Context Classification Using Markov Random Field, **IEEE Geoscience and Remote Sensing Letters**, vol. 5, no. 2, pp. 311-314, 2008.
- [JIM 99] JIMENEZ, L.O., LANDGREBE, D.A., Hyperspectral Data Analysis and Supervised Feature Reduction Via Projection Pursuit. **IEEE Transactions on Geoscience and Remote Sensing**, vol.37, no. 6, pp. 2653 – 2667, novembro 1999.
- [KAL 84] KALAYEH, H. M., LANDGREBE, D. A., Adaptive Relaxation Labeling, **IEEE Transactions on Pattern Analysis and Machine Intelligence**, vol. PAMI-6, no. 3, May 1984.
- [KUO 07] KUO, B.C., CHANG, K.Y., Feature Extractions for Small Sample Size Classification Problem, **IEEE Transaction on Geoscience and Remote Sensing**, vol. 45, no. 3, pp. 756–764, Mar. 2007.
- [LAN 02] LANDGREBE, D. A., Hyperspectral Image Data Analysis as a High Dimensional Signal Processing Problem, **Reprinted from Special Issue of the IEEE Signal Processing Magazine**, vol. 19, no. 1 pp. 17-28, January, 2002.
- [LAN 03] LANDGREBE. D., *Signal Theory Methods in Multispectral Remote Sensing*, Wiley-Interscience, 2003.
- [LAN 80] LANDGREBE. D., The Development of a Spectral-Spatial Classifier for Earth Observation Data, **Pattern Recognition**, 1980, vol. 12, pp. 165-175.
- [LEM 03] 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.

- [LEW 94] LEWOTSKY, K.; Hyperspectral Imaging: Evolution of Imaging Spectrometry. **OE Reports**. november 1994 issue, Disponível em <http://www.spie.org/web/oer/november/image_spectro.html>.
- [LIS 09] LI, S. Z., *Markov Random Field Modeling in Image Analysis*, Terceira Edição, Springer, 2009.
- [LIC 08] LICZBINSKI, C., HAERTEL, V., A New Approach to Estimate A priori Probabilities in Remote Sensing Digital Image Classification, **Canadian Journal of Remote Sensing**, vol. 34, no. 2, pp. 135-142, 2008.
- [LUN 04] LUNA, J. E. O., **Algoritmos EM para Aprendizagem de Redes Bayesianas a partir de Dados Incompletos**, 2004. 132 f. Dissertação (Mestrado em Ciência da Computação) – Programa de Pós-Graduação Computação e Estatística, Centro de Ciências Exatas e Tecnologia, Universidade Federal de Mato Grosso do Sul, MS, 2004.
- [MEL 03] MELGANI, F. E., SERPICO, S. B., A Markov Random Field Approach to Spatio-Temporal Contextual Image Classification, **IEEE Transactions on Geoscience and Remote Sensing**, vol. 41, no. 11, pp. 2478-2487, 2003.
- [MOO 96] MOON, T., The Expectation Maximization Algorithm, **Signal Processing Magazine**, vol. 13(6), pp. 47-60, 1996.
- [OSU 86] O'SULLIVAN, F., A Statistical Perspective on Ill-Posed Inverse Problems, **Statistical Science**, vol. 1, pp. 502-527, 1986.
- [PEL 78] PELEG, S., ROSENFELD, A., Determining Compatibility Coefficients for Curve Enhancement Relaxation Processes, **IEEE Transactions on Systems, Man and Cybernetics**, vol. SAMC-8, no. 7, pp. 548-555, Jul. 1978.
- [IPD 09] PURDUE UNIVERSITY , West Lafayette, Indiana. Disponível em: <<ftp://ftp.ecn.purdue.edu/biehl/MultiSpec/92AV3C>>.
- [RIC 06] RICHARDS, J. A.; JIA, X., *Remote Sensing Digital Image Analysis*, Springer, 2006. Fourth Edition.
- [ROS 76] ROSENFELD, A., HUMMEL, R. A., ZUCKER, S. W., Scene Labeling by Relaxation Operations, **IEEE Transactions on Systems, Man, and Cybernetics**, vol. SMC-6, no. 6, Jun. 1976.
- [SER 02] SERPICO, S. B., D'INCA, M., MELGANI, F., MOSER, G., A Comparison of Feature Reduction Techniques for Classification of Hyperspectral Remote-Sensing Data, **In: Proceedings of SPIE - The International Society for Optical Engineering**, 2002. Agia Pelagia, Greece. vol. 4885, pp. 347-358.

- [SER 01] SERPICO, S. B., BRUZZONE, L., A New Search Algorithm for Feature Selection in Hiperespectral Remote Sensing Images, **IEEE Transaction on Geoscience and Remote Sensing**; Especial Issue on Analysis of Hiperspectral Image Data, vol.39, no7, pp. 1360-1367, julho de 2001.
- [SHA 94] 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**, vol. 32, no 5, 1087-1095, sept. 1994
- [SOL 99] SOLBERG, A. H. S., Contextual Data Fusion Applied to Forest Map Revision, **IEEE Transactions on Geoscience and Remote Sensing**, vol. 37, no. 3, pp. 1234-1243, 1999.
- [SOL 96] SOLBERG, A. H. S., TAXT T., JAIN A. K., A Markov Random Field Model for Classification of Multisource Satellite Imagery, **IEEE Trans. Geosci. Remote Sensing**, vol. 34, pp. 100–113, Jan. 1996.
- [TAD 98] TADJUDIN, 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.
- [THE 89] THERRIEN, C. W., *Decision Estimation and Classification – An Introduction to Pattern Recongnition and Related Topics*, Naval Postgraduate School, John Wiley & Sons Inc.EUA, 1989
- [TIT 85] TITTERINGTON, D. M., Common Structure of Smoothing Techniques in Statistics, **International Statistical Review**, vol. 53,pp. 141-170, 1985.
- [TSO 99] TSO, B. C. K., MATHER, P. M., Classification of Multisource Remote Sensing Imagery Using a Genetic Algorithm and Markov Random Fields, **IEEE Trans. Geosci. Remote Sensing**, vol.37, no.3, pp. 1255-1260, May 1999.
- [TSO 01] TSO, B., MATHER, P. M., *Classification Methods for Remotely Sensed Data*, Taylor & Francis, 2001.
- [HYP 01] U.S. GEOLOGICAL SURVEY – HYPERION – disponível em: <<http://edcsns17.cr.usgs.gov/eo1/sensors/hyperion>>. Acesso: 10 Novembro 2010.
- [WAN 08] WANG, L., Feature Selection with Kernel Class Separability, **IEEE Transactions on Pattern Analysis and Machine Intelligence**, vol.30, no9, pp1534-1546, Setembro de 2008.
- [WEB 02] WEBB, A., *Statistical Pattern Recognition*, 2nd edition, John Wiley, 2002.
- [ZEN 87] ZENZO, S., BERNSTEIN, R., DEGLORIA, S. D., KOLSKY, H. G., Gaussian Maximun Likelihood and Contextual Classification Algorithms for Multicrop Classification, **IEEE Transactions on Geoscience and Remote Sensing**, vol. GE-25, no 6, pg. 805-814, november 1987.

- [ZHO 08] ZHONG, P., WANG, R., Learning Sparse CRFs for Feature Selection and Classification of Hyperspectral Imagery. **IEEE Transactions on Geoscience and Remote Sensing**, vol.46, no.12, pp. 4186–4197, dezembro de 2008.
- [ZOR 07] ZORTEA, M., HAERTEL, V., CLARKE, R., Feature Extraction in Remote Sensing High-Dimensional Image Data, **IEEE Geoscience and Remote Sensing Letters**, vol.4, no. 1, pp. 107-111, janeiro de 2007.
- [ZUC 78] ZUCKER, S. W., KRISHNAMURTHY E. V., HAAR, R. L., Relaxation Processes for Scene Labeling: Convergence, Speed, and Stability, **IEEE Transaction on Systems, Man, and Cybernetics**, vol. SMC-8, no. 1, Jan. 1978.

APÊNDICE A: CÓDIGO FONTE (LINGUAGEM MATLAB)

1) Função principal que implementa a metodologia proposta.

```

function x = context_20_Oct_B(img_sa,img_no_train_sa,nlin,ncol, \
nclass,total_nbands,nsamples_train,nsamples_test, \
img_bin_5_class,train_corn_min_sa,train_corn_notill_sa,... \
train_soy_clean_sa,train_soy_min_sa,train_soy_notill_sa, \
threshold_sl,maxiter_sl,number_semi_labeled,threshold_pr, \
maxiter_pr,... \
priori, test_corn_min_sa,test_corn_notill_sa, \
test_soy_clean_sa,test_soy_min_sa,test_soy_notill_sa, \
pos_corn_min,pos_corn_notill,... \
pos_soy_min,pos_soy_notill,pos_soy_clean,thrsltd)
tic
%
%
% INPUT:
% total_nbands:      upper limit in the loop
% nbands:           number of bands at each particular step \
in the loop
% nsamples_train:   desired number of labeled samples
% The data with extention "sa":
% img_sa:           full image data set, including all \
spectral bands (ERDAS-74 format)
% img_no_train_sa:  sub-set of img, excluding all training \
samples (labeled samples)
% img_bin_5:        file the same size as img, the lines \
corresponding to training samples (labeled samples)
%                   have 1 in the column corresponding to its \
class. Otherwise it is zeros everywhere
% train_XX :        training samples (same dimension as in \
"img_sa")
% threshold_sl:     criterion to stop iterations in the fase \
one of the program (no context info)
%                   number of pixels that do not change \
classes in two successive iterations
% maxiter_sl:       second criterion to stop iterations: \
maximum number of iterations allowed
% number_semi_labeled: desired number of semi-labeled samples
% threshold_pr:     criterion to stop iterations in the fase \
two of the program (relaxation)
%                   number of pixels that do not change \
classes in two successive iterations
% maxiter_pr:       second criterion to stop iterations in \
relaxation: maximum number of iterations allowed
% priori:           if 0, equal a priori probabilities \
assumed, if not zero, these probabilities are estimated.
% test_XX:          test samples (same dimension as in \

```

```

"img_sa")
%
% OUTPUT:
% img_thematic_context.gis: thematic image updated by semi-
labeled samples (spectral plus spatial context)
% M1, S1, M2, S2, M3, S3, M4, S4, M5, S5: parameters estimated
by labeled plus semi-labeled samples
% accuracies_no_context: user's, producer's and overall accuracy
at every step in the iterative process (no context)
% accuracies_context: user's, producer's and overall accuracy at
every step in the iterative process (including context)
%
%
%*****
%*****
%
% select the sub-set with the desired number of spectral bands
"nbands" and labeled samples "nsamples_train"
% these selected samples have an extension "ctx"
%
% produce an initial classification (img_class) to estimate the
weights of non-labeled pixels
%
%
% accuracy
%
overall_accuracy=[];
%
pixels_change_all=[];
%
for nbands = 1:100
    nbands
    % initialize arrays for this bands
    w=[];
    class=[];
    img_class=[];
    img_class_new=[];
    img_thematic=[];
    %
    select_samples_ctx_Oct_06(img_sa, img_no_train_sa,
nsamples_train, nsamples_test, nbands, train_corn_min_sa,
train_corn_notill_sa, train_soy_clean_sa, train_soy_min_sa, ...
    train_soy_notill_sa, test_corn_min_sa,
test_corn_notill_sa, test_soy_clean_sa, test_soy_min_sa,
test_soy_notill_sa, pos_corn_min, ...

```

```

        pos_corn_notill, pos_soy_min, pos_soy_notill, \
pos_soy_clean)
    %
    load samples_ctx
    %
    % initial estimation of the parameters (using labeled \
samples -training samples- only)
    %
    M1 = mean(train_corn_min_ctx);
    M2 = mean(train_corn_notill_ctx);
    M3 = mean(train_soy_min_ctx);
    M4 = mean(train_soy_notill_ctx);
    M5 = mean(train_soy_clean_ctx);
    %
    S1 = cov(train_corn_min_ctx);
    S2 = cov(train_corn_notill_ctx);
    S3 = cov(train_soy_min_ctx);
    S4 = cov(train_soy_notill_ctx);
    S5 = cov(train_soy_clean_ctx);
    %

    P = (1/5)*ones(5,1);
    %
    %
    % estimate accuracy using labeled samples only
    %
    accuracies_05_class_June_08(M1,M2,M3,M4,M5,S1,S2,S3,S4,S5,P, \
nclass, test_corn_min_ctx, test_corn_notill_ctx, ...
        test_soy_min_ctx, test_soy_notill_ctx, test_soy_clean_ctx, \
priori)
    %
    load accuracies
    accuracy_simple=accuracies;
    %
    \
    ##### \
    ##### \
    % this part adds "number_semi_labeled" semi-labeled samples \
to
    % estimate the parameters in "maxiter_sl" iterations
    \
    ##### \
    ##### \
    %
    % start by an initial estimation of weights for pixels in \

```

```

"img_no_train_ctx"
%
% maximum value of G(X), to scale the weights
max_1 = GML(M1,M1,S1,P(1),priori);
max_2 = GML(M2,M2,S2,P(2),priori);
max_3 = GML(M3,M3,S3,P(3),priori);
max_4 = GML(M4,M4,S4,P(4),priori);
max_5 = GML(M5,M5,S5,P(5),priori);
%
%
-----
%
*****
%
-----
% Selects only the image pixel that are not used to test
(fixed) and
% training (variable)
load pos_test
pos = [pos_test; pos_train_corn_min_ctx;
pos_train_corn_notill_ctx; ...
pos_train_soy_min_ctx; pos_train_soy_notill_ctx; ...
pos_train_soy_clean_ctx];

pos_no_train = sort(pos,'ascend');

pos_img_no_train_ctx = comple(pos_no_train,size(img_ctx,
1)');
img_no_train_ctx = img_ctx(pos_img_no_train_ctx,:);
%
-----
%
*****
%
-----
%
%
for pixel_no_GT = 1:size(img_no_train_ctx,1)
class(1)=GML(img_no_train_ctx(pixel_no_GT,:),M1,S1,P(1),

```

```

priori);
    class(2)=GML(img_no_train_ctx(pixel_no_GT,:),M2,S2,P(2),
priori);
    class(3)=GML(img_no_train_ctx(pixel_no_GT,:),M3,S3,P(3),
priori);
    class(4)=GML(img_no_train_ctx(pixel_no_GT,:),M4,S4,P(4),
priori);
    class(5)=GML(img_no_train_ctx(pixel_no_GT,:),M5,S5,P(5),
priori);
    %
    den_w = class(1) + class(2) + class(3) + class(4) + class
(5);
    %
    % initial weights for all unlabeled samples in image
data:
    %
    w1 = (1-(class(1)/den_w))/(nclass-2);
    w2 = (1-(class(2)/den_w))/(nclass-2);
    w3 = (1-(class(3)/den_w))/(nclass-2);
    w4 = (1-(class(4)/den_w))/(nclass-2);
    w5 = (1-(class(5)/den_w))/(nclass-2);
    %
    % scaling the weights
    %
    w1 = w1 *(max_1/class(1));
    w2 = w2 *(max_2/class(2));
    w3 = w3 *(max_3/class(3));
    w4 = w4 *(max_4/class(4));
    w5 = w5 *(max_5/class(5));
    %
    w(pixel_no_GT,:) = [w1 w2 w3 w4 w5];
    %
    [Y,C]=max(class);
    img_class(pixel_no_GT) = C;
end
%
% start iterative procedure introducing semi-labeled samples
and a priori probabilities
% using the weights stored in vector "w"
%
% the criteria to stop iterations are:
% (1) pixel_changes: number of pixels changing classes
% (2) iter: iteration number

```

```

↙
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%% ↙
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% initial values:
pixel_changes = 100;
iter = 0;
%
while (pixel_changes >= threshold_sl && iter < maxiter_sl)
    iter = iter + 1
    %
    % estimate the mean vector and covariance matrix and a ↙
priori probabilities (equation 4, CJRS)
    %
    [M1,S1,P(1)] = weighted_mean_sl_Sept_06 (1, ↙
train_corn_min_ctx,img_no_train_ctx,number_semi_labeled,w);
    [M2,S2,P(2)] = weighted_mean_sl_Sept_06 (2, ↙
train_corn_notill_ctx,img_no_train_ctx,number_semi_labeled,w);
    [M3,S3,P(3)] = weighted_mean_sl_Sept_06 (3, ↙
train_soy_min_ctx,img_no_train_ctx,number_semi_labeled,w);
    [M4,S4,P(4)] = weighted_mean_sl_Sept_06 (4, ↙
train_soy_notill_ctx,img_no_train_ctx,number_semi_labeled,w);
    [M5,S5,P(5)] = weighted_mean_sl_Sept_06 (5, ↙
train_soy_clean_ctx,img_no_train_ctx,number_semi_labeled,w);
    %
    % estimate accuracies no context information
    %
    accuracies_05_class_June_08(M1,M2,M3,M4,M5,S1,S2,S3,S4, ↙
S5,P,nclass,test_corn_min_ctx,test_corn_notill_ctx,...
test_soy_min_ctx,test_soy_notill_ctx,test_soy_clean_ctx, ↙
priori)
    %
    load accuracies
    accuracy_sl = accuracies;
    %
    % update the estimates for the weights and a priori ↙
probabilities
    % by re-classifying the image data using the updated ↙
weighted mean vector
    % and covariance matrix
    %
    % maximum value of G(X), to scale the weights
    max_1 = GML(M1,M1,S1,P(1),priori);
    max_2 = GML(M2,M2,S2,P(2),priori);
    max_3 = GML(M3,M3,S3,P(3),priori);
    max_4 = GML(M4,M4,S4,P(4),priori);

```

```

max_5 = GML(M5,M5,S5,P(5),priori);
%
w=[];
for pixel_no_GT = 1:size(img_no_train_ctx,1)
    class(1)=GML(img_no_train_ctx(pixel_no_GT,:),M1,S1,P(1),priori);
    class(2)=GML(img_no_train_ctx(pixel_no_GT,:),M2,S2,P(2),priori);
    class(3)=GML(img_no_train_ctx(pixel_no_GT,:),M3,S3,P(3),priori);
    class(4)=GML(img_no_train_ctx(pixel_no_GT,:),M4,S4,P(4),priori);
    class(5)=GML(img_no_train_ctx(pixel_no_GT,:),M5,S5,P(5),priori);
    %
    den_w = class(1) + class(2) + class(3) + class(4) +
class(5);
    %
    % initial weights for all unlabeled samples in image
data:
    %
    den_w = class(1) + class(2) + class(3) + class(4) +
class(5);
    %
    % initial weights for all unlabeled samples in image
data:
    %
    w1 = (1-(class(1)/den_w))/(nclass-2);
    w2 = (1-(class(2)/den_w))/(nclass-2);
    w3 = (1-(class(3)/den_w))/(nclass-2);
    w4 = (1-(class(4)/den_w))/(nclass-2);
    w5 = (1-(class(5)/den_w))/(nclass-2);
    %
    % scaling the weights
    %
    w1 = w1 *(max_1/class(1));
    w2 = w2 *(max_2/class(2));
    w3 = w3 *(max_3/class(3));
    w4 = w4 *(max_4/class(4));
    w5 = w5 *(max_5/class(5));
    %
    w(pixel_no_GT,:) = [w1 w2 w3 w4 w5];
    %
    [Y,C]=max(class);
    img_class_new(pixel_no_GT) = C;

```



```

    w5 = w5 *(max_5/class(5));
    %
    w(pixel,:) = [w1 w2 w3 w4 w5];
    [Y,C]=max(class);
    img_thematic(pixel) = C;
end
%
% set the weights for training samples equal to one and zero
%
w(pos_train_corn_min_ctx,1)=1;
w(pos_train_corn_min_ctx,[2 3 4 5])=0;
%
w(pos_train_corn_notill_ctx,2)=1;
w(pos_train_corn_notill_ctx,[1 3 4 5])=0;
%
w(pos_train_soy_min_ctx,3)=1;
w(pos_train_soy_min_ctx,[1 2 4 5])=0;
%
w(pos_train_soy_notill_ctx,4)=1;
w(pos_train_soy_notill_ctx,[1 2 3 5])=0;
%
w(pos_train_soy_clean_ctx,5)=1;
w(pos_train_soy_clean_ctx,[1 2 3 4])=0;
%
% save thematic image and weights, no context information
%
img_thematic=img_thematic';
save img_thematic img_thematic
saveimage(img_thematic,'img_thematic.gis',145,118)%%%
<=====
w_no_context=w;
save w_no_context w_no_context
%
% this section re-estimates the weight matrix , rescaling to
zero
% entries lower than a selected threshold based on the mean
weight, and
% introduces the background class (class=6)into the thematic
image
%
%
-----
-----
%
-----

```

```

-----
peso_maximo = max(w_no_context, [], 2); ##### teste #####
save w_max peso_maximo

[Y,C]=max(w_no_context, [], 2);
VEC = [Y C];
[val_VEC, pos_VEC]=sort(VEC(:,1), 'descend');

ini = round(size(VEC,1) - size(VEC,1)*thrsld);
pos = pos_VEC(ini:end,1);
VEC(pos,1) = 0;
VEC(pos,2) = 6;
img_thematic = VEC(:,2);

masc=[];
    for pixel=1:size(VEC,1)
        if VEC(pixel,1)==0
            masc(pixel)=0;
        else
            masc(pixel)=1;
        end;
    end;
%<
-----<
-----<
%<
-----<
%
masc=masc';
save masc masc
save img_thematic img_thematic
saveimage(img_thematic, 'img_no_context.gis', 145, 118)
saveimage(masc, 'masc.gis', 145, 118)
%
<
#####<
#####<
##### THE FOLLOWING SECTION IMPLEMENTS<
CONTEXT INFORMATION #####<
<
#####<
#####<
%

```

```

    % refine the weights "w" by implementing context information
    % to skip this section (no context information), set
max_iter_pr = 0
    %
    % re-arrange the weight matrix to include the full image
adding the
    % class background (6th class)
    %
    aux_0 = [masc masc masc masc masc];
    aux_1 = aux_0 .* w_no_context;
    aux_2 = ones(size(aux_1,1),1);
    aux_3 = aux_2 - sum(aux_1,2);
    w_full =[aux_1 aux_3];
    save w_full w_full
    %
    if maxiter_pr ~= 0
        pixel_changes=100;
        iter_pr=0;
        while (iter_pr < maxiter_pr) && (pixel_changes >
threshold_pr)
            %
            iter_pr = iter_pr+1
            w_pr = prob_relax_Oct_10(img_thematic,w_full,nlin,
ncol,nclass,threshold_pr,maxiter_pr);
            %
            % set the weights for training samples equal to one
and zero
            %
            w_pr(pos_train_corn_min_ctx,1)=1;
            w_pr(pos_train_corn_min_ctx,[2 3 4 5])=0;
            %
            w_pr(pos_train_corn_notill_ctx,2)=1;
            w_pr(pos_train_corn_notill_ctx,[1 3 4 5])=0;
            %
            w_pr(pos_train_soy_min_ctx,3)=1;
            w_pr(pos_train_soy_min_ctx,[1 2 4 5])=0;
            %
            w_pr(pos_train_soy_notill_ctx,4)=1;
            w_pr(pos_train_soy_notill_ctx,[1 2 3 5])=0;
            %
            w_pr(pos_train_soy_clean_ctx,5)=1;
            w_pr(pos_train_soy_clean_ctx,[1 2 3 4])=0;
            %
            %
            save w_pr w_pr

```

```

        % re-estimated the parameters after including
spatial context information
        %
        [M1,S1,P(1)] = weighted_mean_pr_Jun_03_B (1,
train_corn_min_ctx,img_ctx,number_semi_labeled,img_bin_5_class,
w_pr);
        [M2,S2,P(2)] = weighted_mean_pr_Jun_03_B (2,
train_corn_notill_ctx,img_ctx,number_semi_labeled,
img_bin_5_class,w_pr);
        [M3,S3,P(3)] = weighted_mean_pr_Jun_03_B (3,
train_soy_min_ctx,img_ctx,number_semi_labeled,img_bin_5_class,
w_pr);
        [M4,S4,P(4)] = weighted_mean_pr_Jun_03_B (4,
train_soy_notill_ctx,img_ctx,number_semi_labeled,
img_bin_5_class,w_pr);
        [M5,S5,P(5)] = weighted_mean_pr_Jun_03_B (5,
train_soy_clean_ctx,img_ctx,number_semi_labeled,img_bin_5_class,
w_pr);
        %
        % estimate accuracies (user's, producer's and
overall accuracy)
        %
        accuracies_05_class_June_08(M1,M2,M3,M4,M5,S1,S2,S3,
S4,S5,P,nclass,test_corn_min_ctx,test_corn_notill_ctx,...
test_soy_min_ctx,test_soy_notill_ctx,
test_soy_clean_ctx,priori)
        %
        load accuracies
        accuracy_pr = accuracies;
        %
        % classify image data after including spatial
context information
        %
        img_thematic_new=[];
        for pixel = 1:size(img_ctx,1)
            class(1)=GML(img_ctx(pixel,:),M1,S1,P(1),
priori);
            class(2)=GML(img_ctx(pixel,:),M2,S2,P(2),
priori);
            class(3)=GML(img_ctx(pixel,:),M3,S3,P(3),
priori);
            class(4)=GML(img_ctx(pixel,:),M4,S4,P(4),
priori);
            class(5)=GML(img_ctx(pixel,:),M5,S5,P(5),
priori);

```

```

        %
        [Y,I]=max(class);
        img_thematic_new(pixel) = I;
    end
    img_thematic_new=img_thematic_new';
    %
    % assign to class background (class 6) pixels
    %
    img_tem_before=img_thematic_new;
    img_thematic_new(masc==0,:)=6;
    img_tem_after=img_thematic_new;
    %
    pixel_changes=sum(sum(img_thematic ~=img_thematic_new))
    %
    save w_pr w_pr
    % go to the next iteration
    %
    img_thematic=img_thematic_new;
end
end
%
%
if maxiter_sl==0
    accuracy_sl=[0 0 0];
end
%
if maxiter_pr==0
    accuracy_pr=[0 0 0];
end
%
overall_accuracy=[overall_accuracy;[nbands 0 0]];
accuracy_simple;accuracy_sl;accuracy_pr]
%
end
tempo=toc
save overall_accuracy overall_accuracy tempo
saveimage(img_thematic,'img_context.gis',145,118)

```

2) Função de seleção das amostras de treinamento, amostras de teste e bandas.

```

function x = select_samples_ctx_Oct_06(img_sa, img_no_train_sa, \
nsamples_train, nsamples_test, nbands, train_corn_min_sa, \
train_corn_notill_sa, train_soy_clean_sa, train_soy_min_sa, ...
    train_soy_notill_sa, test_corn_min_sa, test_corn_notill_sa, \
test_soy_clean_sa, test_soy_min_sa, test_soy_notill_sa, \
pos_corn_min, pos_corn_notill, ...
    pos_soy_min, pos_soy_notill, pos_soy_clean)
%
% this function selects nsamples training samples, with nbands \
each.
%
% INPUT:
% nsamples: number of training samples to be selected
% nbands: data dimensionality (number of spectral bands)
% train_XX: full set of available training samples
% test_XX: full set of available test samples
%
% OUTPUT:
% samples_ctx: training and test samples to be used in \
"context_05_Dec_22.m"
% img_ctx: sub_set of "img_sa", including the selected \
bands only
% img_no_train_ctx: sub_set of "img_ctx", excluding training \
samples
%
% ===== \
% ===== \
%
% select bands to be used
%
total_bands=size(img_sa,2);
step_bands = total_bands/nbands;
bands=[1:step_bands:total_bands];
bands=round(bands);
%
% image with the selected bands
%
img_ctx=img_sa(:, [bands]);
%
% img_no_train
%
img_no_train_ctx=img_no_train_sa(:, [bands]);
%
% select training and test samples
%

```

```

% corn_min
step_corn_min_train = size(train_corn_min_sa,1)/nsamples_train;
step_corn_min_test = size(test_corn_min_sa,1)/nsamples_test;
samples_train = 1:step_corn_min_train:size(train_corn_min_sa,1);
samples_test = 1:step_corn_min_test:size(test_corn_min_sa,1);
samples_train=round(samples_train);
samples_test=round(samples_test);
train_corn_min_ctx= train_corn_min_sa(samples_train,bands);
test_corn_min_ctx= test_corn_min_sa(samples_test,bands);
pos_train_corn_min_ctx=pos_corn_min(samples_train);
%
% corn_notill
step_corn_notill_train = size(train_corn_notill_sa,1) \
/nsamples_train;
step_corn_notill_test = size(test_corn_notill_sa,1) \
/nsamples_test;
samples_train = 1:step_corn_notill_train:size \
(train_corn_notill_sa,1);
samples_test = 1:step_corn_notill_test:size(test_corn_notill_sa, \
1);
samples_train=round(samples_train);
samples_test=round(samples_test);
train_corn_notill_ctx= train_corn_notill_sa(samples_train, \
bands);
test_corn_notill_ctx= test_corn_notill_sa(samples_test,bands);
pos_train_corn_notill_ctx=pos_corn_notill(samples_train);
%
% soy_min
step_soy_min_train = size(train_soy_min_sa,1)/nsamples_train;
step_soy_min_test = size(test_soy_min_sa,1)/nsamples_test;
samples_train = 1:step_soy_min_train:size(train_soy_min_sa,1);
samples_test = 1:step_soy_min_test:size(test_soy_min_sa,1);
samples_train=round(samples_train);
samples_test=round(samples_test);
train_soy_min_ctx= train_soy_min_sa(samples_train,bands);
test_soy_min_ctx= test_soy_min_sa(samples_test,bands);
pos_train_soy_min_ctx=pos_soy_min(samples_train);
%
% soy_notill
step_soy_notill_train = size(train_soy_notill_sa,1) \
/nsamples_train;
step_soy_notill_test = size(test_soy_notill_sa,1)/nsamples_test;
samples_train = 1:step_soy_notill_train:size \
(train_soy_notill_sa,1);
samples_test = 1:step_soy_notill_test:size(test_soy_notill_sa, \

```

```

1);
samples_train=round(samples_train);
samples_test=round(samples_test);
train_soy_notill_ctx= train_soy_notill_sa(samples_train,bands);
test_soy_notill_ctx= test_soy_notill_sa(samples_test,bands);
pos_train_soy_notill_ctx=pos_soy_notill(samples_train);
%
% soy_clean
step_soy_clean_train = size(train_soy_clean_sa,1) \
/nsamples_train;
step_soy_clean_test = size(test_soy_clean_sa,1)/nsamples_test;
samples_train = 1:step_soy_clean_train:size(train_soy_clean_sa, \
1);
samples_test = 1:step_soy_clean_test:size(test_soy_clean_sa,1);
samples_train=round(samples_train);
samples_test=round(samples_test);
train_soy_clean_ctx= train_soy_clean_sa(samples_train,bands);
test_soy_clean_ctx= test_soy_clean_sa(samples_test,bands);
pos_train_soy_clean_ctx=pos_soy_clean(samples_train);
%
% save selected samples
%
save samples_ctx img_ctx img_no_train_ctx train_corn_min_ctx \
test_corn_min_ctx train_corn_notill_ctx test_corn_notill_ctx ...
train_soy_min_ctx test_soy_min_ctx train_soy_notill_ctx \
test_soy_notill_ctx train_soy_clean_ctx test_soy_clean_ctx...
pos_train_corn_min_ctx pos_train_corn_notill_ctx \
pos_train_soy_min_ctx pos_train_soy_notill_ctx \
pos_train_soy_clean_ctx

```

3) Algoritmo que calcula a função de decisão MVG.

```

function z = GML(X,M,S,P,priori)
%
% this function implements the GML decision function (natural \
logarith)
%
if priori ~= 0
    z = -(X-M)*inv(S)*(X-M)' - logdet(S) + 2*log(P);
else
    z = -(X-M)*inv(S)*(X-M)' - logdet(S);
end

```


4) Função que calcula as acurácias, e também gera a matriz de erro do processo.

```

function x = accuracies_05_class_June_08(M1,M2,M3,M4,M5,S1,S2,
S3,S4,S5,P,nclass,test_corn_min,test_corn_notill,...
test_soy_clean,test_soy_min,test_soy_notill,priori)
%
% this function estimates the user and producer accuracy using
parameters estimated by "class_semi_label.m"
% INPUT:
% mean vectors (M) and covariance matrices (S) as estimated by
"class_semi_label.m"
% all available test samples for the classes under analysis
%
% build the confusion matrix: ground truth along the columns;
classified data along the rows
%
conf_matrix =zeros(nclass,nclass);
user_accuracy=[];
producer_accuracy=[];
c=[];
%
% build matrix along the columns
%
% column # 1 (corn_min)
for pixel = 1:size(test_corn_min,1)
    c(1) = GML(test_corn_min(pixel,:),M1,S1,P(1),priori);
    c(2) = GML(test_corn_min(pixel,:),M2,S2,P(2),priori);
    c(3) = GML(test_corn_min(pixel,:),M3,S3,P(3),priori);
    c(4) = GML(test_corn_min(pixel,:),M4,S4,P(4),priori);
    c(5) = GML(test_corn_min(pixel,:),M5,S5,P(5),priori);
    [Y,I] = max(c);
    %
    conf_matrix(I,1)=conf_matrix(I,1)+1;
end
% column # 2 (corn_notill)
for pixel = 1:size(test_corn_notill,1)
    c(1) = GML(test_corn_notill(pixel,:),M1,S1,P(1),priori);
    c(2) = GML(test_corn_notill(pixel,:),M2,S2,P(2),priori);
    c(3) = GML(test_corn_notill(pixel,:),M3,S3,P(3),priori);
    c(4) = GML(test_corn_notill(pixel,:),M4,S4,P(4),priori);
    c(5) = GML(test_corn_notill(pixel,:),M5,S5,P(5),priori);
    [Y,I] = max(c);
    %
    conf_matrix(I,2)=conf_matrix(I,2)+1;
end
% column # 3 (soy_clean)
for pixel = 1:size(test_soy_clean,1)

```

```

c(1) = GML(test_soy_clean(pixel, :), M1, S1, P(1), priori);
c(2) = GML(test_soy_clean(pixel, :), M2, S2, P(2), priori);
c(3) = GML(test_soy_clean(pixel, :), M3, S3, P(3), priori);
c(4) = GML(test_soy_clean(pixel, :), M4, S4, P(4), priori);
c(5) = GML(test_soy_clean(pixel, :), M5, S5, P(5), priori);
[Y, I] = max(c);
%
conf_matrix(I, 3)=conf_matrix(I, 3)+1;
end
% column # 4 (soy_min)
for pixel = 1:size(test_soy_min, 1)
c(1) = GML(test_soy_min(pixel, :), M1, S1, P(1), priori);
c(2) = GML(test_soy_min(pixel, :), M2, S2, P(2), priori);
c(3) = GML(test_soy_min(pixel, :), M3, S3, P(3), priori);
c(4) = GML(test_soy_min(pixel, :), M4, S4, P(4), priori);
c(5) = GML(test_soy_min(pixel, :), M5, S5, P(5), priori);
[Y, I] = max(c);
%
conf_matrix(I, 4)=conf_matrix(I, 4)+1;
end
% column # 5 (soy_notill)
for pixel = 1:size(test_soy_notill, 1)
c(1) = GML(test_soy_notill(pixel, :), M1, S1, P(1), priori);
c(2) = GML(test_soy_notill(pixel, :), M2, S2, P(2), priori);
c(3) = GML(test_soy_notill(pixel, :), M3, S3, P(3), priori);
c(4) = GML(test_soy_notill(pixel, :), M4, S4, P(4), priori);
c(5) = GML(test_soy_notill(pixel, :), M5, S5, P(5), priori);
[Y, I] = max(c);
%
conf_matrix(I, 5)=conf_matrix(I, 5)+1;
end
%
% user's accuracy
%
user_acc_corn_min = conf_matrix(1, 1)/sum(conf_matrix(1, :));
user_acc_corn_notill = conf_matrix(2, 2)/sum(conf_matrix(2, :));
user_acc_soy_clean = conf_matrix(3, 3)/sum(conf_matrix(3, :));
user_acc_soy_min = conf_matrix(4, 4)/sum(conf_matrix(4, :));
user_acc_soy_notill = conf_matrix(5, 5)/sum(conf_matrix(5, :));
%
user_accuracy=[user_acc_corn_min user_acc_corn_notill\
user_acc_soy_clean user_acc_soy_min user_acc_soy_notill];
mean_user=mean(user_accuracy);
%
% producer's accuracy

```

```
%  
prod_acc_corn_min = conf_matrix(1,1)/sum(conf_matrix(:,1));  
prod_acc_corn_notill = conf_matrix(2,2)/sum(conf_matrix(:,2));  
prod_acc_soy_clean = conf_matrix(3,3)/sum(conf_matrix(:,3));  
prod_acc_soy_min = conf_matrix(4,4)/sum(conf_matrix(:,4));  
prod_acc_soy_notill = conf_matrix(5,5)/sum(conf_matrix(:,5));  
%  
producer_accuracy=[prod_acc_corn_min prod_acc_corn_notill  
prod_acc_soy_clean prod_acc_soy_min prod_acc_soy_notill];  
mean_producer=mean(producer_accuracy);  
%  
% overall accuracy  
%  
overall_accuracy = trace(conf_matrix)/sum(sum(conf_matrix));  
%  
accuracies = [mean_user mean_producer overall_accuracy];  
save accuracies accuracies  
%  
save conf_matrix conf_matrix
```

5) Estima os parâmetros do conjunto de treinamento (amostras rotuladas e semi-rotuladas) utilizando somente atributos espectrais.

```

function [M,S,P] = weighted_mean_sl_Sept_06(class, \
labeled_samples, img_no_train_ctx, number_semi_labeled, w)
%
% ===== \
% semi-labeled samples
% ===== \
%
% select the weights for this class
%
w_class = w(:,class);
%
% sort the weights in descending order
%
[Y,I] = sort(w_class,1,'descend');
w_semi = Y(1:number_semi_labeled);
%
% pick the corresponding pixels in "image_no_train_ctx"
%
for pixel=1:number_semi_labeled
    samples_sl(pixel,:) = img_no_train_ctx(I(pixel),:);
end
%
% ===== \
% labeled samples
% ===== \
%
% set the weight vector ww_labeled for the labeled samples \
(weights are set equal to one)
%
w_labeled = ones(size(labeled_samples,1),1);
%
% merge labeled and semi-labeled samples and the weights
%
samples=[labeled_samples;samples_sl];
ww=[w_labeled;w_semi];
%
% estimate the weighted mean vector (equation 4, CJRS)
%
M=0;
for count = 1:size(samples,1)
    M = M+samples(count,:)*ww(count);

```

```
end
M = (1/sum(w)) * M;
%
% estimate the weighted covariance matrix (equation 4, CJRS)
%
S=0;
for count = 1: size(samples,1)
    S = S+ww(count)*(samples(count,:)-M)'*(samples(count,:)-M);
end
S = (1/(sum(ww)-1)) * S;
%
% estimate the a priori probabilities (equation 5, CJRS)
%
P=sum(w(:,class))/size(w,1);
save P P
save ww ww
```

6) **Função que realiza o processo de relaxação.**

```

function w_pr = prob_relax_Oct_10(img_thematic,w_full,nlin,ncol,
nclass,threshold,max_iter)
%
% this function adjusts the weight associated with every pixel
in the image
% by using context information
%
% for more details see Di Zenzo et ali. TGARS, vol GE-25, No. 6,
pp. 805-814, Nov. 1987.
%
% INPUT:
% img_thematic: thematic image produced in the current iteration
in main program
% w_full: full weight matrix
% nlin, ncol: number of lines and columns in img
% nclass: number of classes (05 plus the background class)
% threshold, maxiter: criterion to stop iterations
%
% OUTPUT:
% w_pr: weight matrix updated by adding spatial context
information
%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%
% estimate the compatibility coefficient  $r_{ij}(h,k)$  - DiZenzo eq.
(11)
%
N=zeros(8,nclass,nclass);
diff=[];
%
% loop across pixels not on the border of the image
% compute the number of pairs of pixels N(i,j,h,k): pixel in the
central position "i"
% belongs to class "h" and its neighbor in position "j" (j=1,
2,...,8)
%
for lin = 2:nlin-1
    for col = 2:ncol-1
        %
        pos = (lin-1)*ncol + col;
        pos1 = (lin-2)*ncol + col-1;
        pos2 = (lin-2)*ncol + col;
        pos3 = (lin-2)*ncol + col+1;
        %
    
```

```

    pos4 = (lin-1)*ncol + col-1;
    pos5 = (lin-1)*ncol + col+1;
    %
    pos6 = (lin)*ncol + col-1;
    pos7 = (lin)*ncol + col;
    pos8 = (lin)*ncol + col+1;
    %
    N(1,img_thematic(pos),img_thematic(pos1)) = N(1, \
img_thematic(pos),img_thematic(pos1))+1;
    N(2,img_thematic(pos),img_thematic(pos2)) = N(2, \
img_thematic(pos),img_thematic(pos2))+1;
    N(3,img_thematic(pos),img_thematic(pos3)) = N(3, \
img_thematic(pos),img_thematic(pos3))+1;
    N(4,img_thematic(pos),img_thematic(pos4)) = N(4, \
img_thematic(pos),img_thematic(pos4))+1;
    N(5,img_thematic(pos),img_thematic(pos5)) = N(5, \
img_thematic(pos),img_thematic(pos5))+1;
    N(6,img_thematic(pos),img_thematic(pos6)) = N(6, \
img_thematic(pos),img_thematic(pos6))+1;
    N(7,img_thematic(pos),img_thematic(pos7)) = N(7, \
img_thematic(pos),img_thematic(pos7))+1;
    N(8,img_thematic(pos),img_thematic(pos8)) = N(8, \
img_thematic(pos),img_thematic(pos8))+1;
    end
end
save N N
Ntotal = sum(sum(sum(N)));
save Ntotal Ntotal
%
% compute the compatibility coefficients r(8, nclass, nclass)
%
% loop around the 8 neighbors
for neighbor = 1:8
    for class_i = 1:nclass
        for class_j = 1:nclass
            num = N(neighbor,class_i,class_j)*Ntotal;
            %
            % denominator
            % loop around class in central pixel "h"
            den1=0;
            for h = 1:nclass
                den1 = den1 + N(neighbor,h,class_j);
            end
            den2=0;
            for k = 1:nclass

```

```

        den2 = den2 + N(neighbor, class_i, k);
    end
    %
    if (num == 0)
        r(neighbor, class_i, class_j)=-1;
    elseif (den1*den2==0)
        r(neighbor, class_i, class_j)=1;
    else
        r(neighbor, class_i, class_j)=(1/5)*log(num/
(den1*den2));
    end
    %
end
end
end
save r r
%
% update iteratively the weights (Zenko eq (8) p. 808)
%
w_new=w_full;
iter=0;
dif=100000;
while(iter<max_iter)&(dif>threshold)
    %
    iter=iter+1;
    dif;
    for lin = 2:nlin-1
        for col = 2:ncol-1
            pos = (lin-1)*ncol + col;
            pos1 = (lin-2)*ncol + col-1;
            pos2 = (lin-2)*ncol + col;
            pos3 = (lin-2)*ncol + col+1;
            %
            pos4 = (lin-1)*ncol + col-1;
            pos5 = (lin-1)*ncol + col+1;
            %
            pos6 = (lin)*ncol + col-1;
            pos7 = (lin)*ncol + col;
            pos8 = (lin)*ncol + col+1;
            %
            % normalization factor
            %
            LL=0;
            for class_h = 1:nclass
                L=0;

```



```

        for class_k = 1:nclass
            L= L +...
                r(1,class_h,class_k)*w_full(pos1,class_k) ←
+...
                r(2,class_h,class_k)*w_full(pos2,class_k) ←
+...
                r(3,class_h,class_k)*w_full(pos3,class_k) ←
+...
                r(4,class_h,class_k)*w_full(pos4,class_k) ←
+...
                r(5,class_h,class_k)*w_full(pos5,class_k) ←
+...
                r(6,class_h,class_k)*w_full(pos6,class_k) ←
+...
                r(7,class_h,class_k)*w_full(pos7,class_k) ←
+...
                r(8,class_h,class_k)*w_full(pos8,class_k);
        end
        LL = LL + w_full(pos,class_h)*(1 + L/8);
    end
    %
    % updated estimated value for the weights (w_new)
    %
    for class_h = 1:nclass
        ww=0;
        for class_k = 1:nclass
            ww = ww +...
                r(1,class_h,class_k)*w_full(pos1,class_k) ←
+...
                r(2,class_h,class_k)*w_full(pos2,class_k) ←
+...
                r(3,class_h,class_k)*w_full(pos3,class_k) ←
+...
                r(4,class_h,class_k)*w_full(pos4,class_k) ←
+...
                r(5,class_h,class_k)*w_full(pos5,class_k) ←
+...
                r(6,class_h,class_k)*w_full(pos6,class_k) ←
+...
                r(7,class_h,class_k)*w_full(pos7,class_k) ←
+...
                r(8,class_h,class_k)*w_full(pos8,class_k);
        end
        w_new(pos,class_h)=(1/LL)*w_full(pos,class_h)*←
(1+ ww/8);

```

```
        end
    end
end
%
% check for convergence
%
dif=max(max(abs(w_new-w_full)));
diff=[diff;dif];
save diff diff
w_full=w_new;
end
w_pr=w_new;
```

7) **Função utilizada para estimar os parâmetros inserindo informações de contexto espacial.**

```

function [M,S,P] = weighted_mean_pr_Jun_03_B(class, \
labeled_samples,img,number_semi_labeled,img_bin_5_class,w_pr)
%
% select semi labeled samples (number_semi_labeled)
%
%
[Y,I] = sort(w_pr(:,class),'descend');
ini = size(labeled_samples,1) + 1;
fim = size(labeled_samples,1) + number_semi_labeled;
samples=labeled_samples;
pos_samples = I(ini:fim);
samples_semi = img(pos_samples,:);
samples=[samples;samples_semi];
%
%
ww = ones(size(labeled_samples,1),1);
ww_semi = Y(ini:fim,1);
ww=[ww;ww_semi];
save TESTE ww ww_semi
%
% estimate the weighted mean vector (equation 4, CJRS)
%
M=0;
for count = 1:size(samples,1)
    M = M+samples(count,:)*ww(count);
end
M = (1/sum(ww)) * M;
%
% estimate the weighted covariance matrix (equation 4, CJRS)
%
S=0;
for count = 1: size(samples,1)
    S = S+ww(count)*(samples(count,:)-M)'*(samples(count,:)-M);
end
S = (1/(sum(ww)-1)) * S;
%
% estimate the a priori probabilities (equation 5, CJRS)
%
P=sum(w_pr(:,class))/size(w_pr,1);
save P P
save ww ww

```