



UNIVERSIDADE TECNOLÓGICA FEDERAL DO PARANÁ  
CAMPUS CURITIBA

GERÊNCIA DE PESQUISA E PÓS-GRADUAÇÃO

PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA ELÉTRICA E  
INFORMÁTICA INDUSTRIAL — CPGEI

LUÍS ALBERTO LUCAS

**SISTEMA DE INFERÊNCIA *FUZZY* GERAL DO TIPO-2  
APLICADO À CLASSIFICAÇÃO**

TESE DE DOUTORADO

CURITIBA  
MARÇO DE 2009



**TESE DE DOUTORADO**

apresentada à UTFPR  
para obtenção do título de

**DOUTOR EM CIÊNCIAS**

por

**LUÍS ALBERTO LUCAS**

---

SISTEMA DE INFERÊNCIA *FUZZY* GERAL DO TIPO-2  
APLICADO À CLASSIFICAÇÃO

---

Banca Examinadora:

Presidente e co-orientadora:

Profª. Dra. MYRIAM REGATTIERI DELGADO UTFPR

Examinadores:

Profª. Dra. MARLEY MARIA BERNARDES REBUZZI VELLASCO PUC-Rio

Prof. Dr. JORGE ANTONIO SILVA CENTENO UFPR

Prof. Dr. JAQUES FACON PUCPR

Prof. Dr. CELSO ANTÔNIO ALVES KAESTNER UTFPR

Curitiba, março de 2009.



LUÍS ALBERTO LUCAS

SISTEMA DE INFERÊNCIA *FUZZY* GERAL DO TIPO-2  
APLICADO À CLASSIFICAÇÃO

Tese apresentada ao Programa de Pós-Graduação em Engenharia Elétrica e Informática Industrial da Universidade Tecnológica Federal do Paraná como requisito parcial à obtenção do título de “Doutor em Ciências” — Área de Concentração: Informática Industrial.

Orientadora: Profa. Dra. Tania Mezzadri Centeno.  
Co-orientadora: Profa. Dra. Myriam Regattieri Delgado

Curitiba 2009.

Ficha catalográfica elaborada pela Biblioteca da UTFPR – Campus Curitiba

L933s Lucas, Luís Alberto  
Sistema de inferência fuzzy geral do tipo-2 aplicado à classificação /  
Luís Alberto Lucas. – 2009.  
108 f. : il. ; 30 cm

Orientador: Tania Mezzadri Centeno  
Co-orientador: Myriam Regattieri Delgado  
Tese (Doutorado) – Universidade Tecnológica Federal do Paraná.  
Programa de Pós-Graduação em Engenharia Elétrica e Informática  
Industrial. Área de Concentração: Informática Industrial, Curitiba, 2009  
Bibliografia: f. 104-8

1. Conjuntos difusos. 2. Sensoriamento remoto. 3. Computação.  
4. Classificadores digitais. I. Centeno, Tania Mezzadri, orient.  
II. Delgado, Myriam Regattieri, co-orient. III. Universidade Tecnológica  
Federal do Paraná. Programa de Pós-Graduação em Engenharia Elétrica e  
Informática Industrial. Área de Concentração em Informática Industrial.  
IV. Título.

CDD 621.3

## AGRADECIMENTOS

Agradeço à Profa. Dra. Tania Mezzadri Centeno pela oportunidade concedida e pelas correções desta tese e à Profa. Dra. Myriam Regattieri Delgado pela inúmeras sugestões ao longo de todo o Doutorado.

Agradeço, também, à minha esposa Rosenei pelas palavras de estímulo e aos meus filhos, Rafael e João Vitor, por darem um propósito à minha vida.

Agradeço ao meu sogro Ruberval pela amizade e pela maneira alegre como encara a vida e supera as dificuldades que se apresentam.

Agradeço, finalmente, a meus pais, Francisco e Vilma, pelos anos de apoio, incentivo e paciência.





# *Lista de Figuras*

2.1	Exemplo de resposta espectral de uma plantação de milho, obtida a partir do <i>Landsat</i> , com 7 bandas espectrais. . . . .	p. 19
2.2	Estrutura típica de um sistema de inferência <i>fuzzy</i> . . . . .	p. 25
2.3	Conjunto unitário ou <i>singleton</i> . . . . .	p. 26
2.4	Exemplo da inferência escalonada para um sistema <i>fuzzy</i> com duas regras e dois antecedentes. . . . .	p. 29
2.5	Estrutura típica de um sistema de inferência <i>fuzzy</i> do tipo-2. . . . .	p. 30
2.6	Um conjunto <i>fuzzy</i> discreto do tipo-2. . . . .	p. 34
2.7	Um conjunto <i>fuzzy</i> discreto do tipo-2 rotacionado. . . . .	p. 34
3.1	GEFOU de um conjunto <i>fuzzy</i> geral do tipo-2 com PMF triangular. . . . .	p. 42
3.2	União de dois conjuntos <i>fuzzy</i> gerais do tipo-2. GEFOU de: a) $\tilde{A}$ ; b) $\tilde{B}$ ; c) $\tilde{A} \cup \tilde{B}$ . . . . .	p. 43
3.3	Interseção de dois conjuntos <i>fuzzy</i> gerais do tipo-2. GEFOU de: a) $\tilde{A}$ ; b) $\tilde{B}$ ; c) $\tilde{A} \cap \tilde{B}$ . . . . .	p. 43
3.4	Um <i>singleton</i> do tipo-2 definido em $x = x'$ . . . . .	p. 44
3.5	GEFOU de: a) Conjunto do tipo-2 $\tilde{A}$ ; b) Supremo $\sup[\tilde{A}]$ ; c) Extensão cilíndrica $\text{cyl}_{x \in X} [\mu_{\tilde{A}}(x_1 = 0)]$ . . . . .	p. 45
3.6	Conjunto <i>fuzzy</i> do tipo-2. . . . .	p. 46
3.7	Exemplo de obtenção da CTR. . . . .	p. 47
3.8	Exemplo de obtenção da VSCTR. . . . .	p. 47
3.9	Redução de tipo de um conjunto do tipo-2. a) GEFOU de um conjunto do tipo-2 $\tilde{B}$ . b) Conjunto reduzido, do tipo-1, $C_{\tilde{B}}$ . . . . .	p. 48
3.10	Exemplo de obtenção do <i>matching</i> $\tilde{M}_i^l$ (em c) a partir da interseção $\tilde{I}$ (em b) calculada a partir da entrada $\tilde{F}_i$ com o antecedente $\tilde{F}_i^l$ (em a). . . . .	p. 50

3.11	A $l$ -ésima regra de um sistema de inferência do tipo-2. O GEFOU de: a) Extensão cilíndrica do grau de ativação ao longo do eixo $y$ ; b) Consequente do tipo-2; c) Saída obtida da $l$ -ésima regra. . . . .	p. 51
3.12	Exemplo da inferência escalonada <i>fuzzy</i> geral do tipo-2 para um sistema <i>fuzzy</i> com duas regras, com dois antecedentes cada. . . . .	p. 53
4.1	Aspectos gerais do classificador do tipo-2 proposto. . . . .	p. 55
4.2	Amostras de treinamento obtidas de uma banda espectral. . . . .	p. 56
4.3	Construção de um conjunto <i>fuzzy</i> do tipo-2 a partir das amostras de treinamento. . . . .	p. 58
4.4	Consequentes, $C_s$ , $C_m$ e $C_a$ , para uma base de regras com três classes, soja, milho e alfafa, respectivamente. . . . .	p. 59
4.5	Amostras de teste para a banda espectral usada no treinamento. . . . .	p. 61
4.6	Mecanismo de inferência do classificador <i>fuzzy</i> do tipo-2 proposto. . . . .	p. 62
4.7	Exemplo da construção da entrada 7-dimensional do tipo-1. . . . .	p. 63
4.8	Um exemplo de <i>matching</i> entre um antecedente do tipo-2 e uma entrada <i>fuzzy</i> convencional. . . . .	p. 64
4.9	Obtenção do grau de ativação, FS, de uma regra com sete antecedentes, onde, $M_1$ a $M_7$ são os <i>matchings</i> destes antecedentes com o <i>pixel</i> 7-dimensional de entrada. . . . .	p. 65
4.10	Exemplo de saída inferida para a regra $l$ . . . . .	p. 66
4.11	Exemplo de <i>defuzzificação</i> da saída de uma das regras do classificador geral do tipo-2. . . . .	p. 67
4.12	Exemplo de tomada de decisão: classe = milho. . . . .	p. 68
5.1	O primeiro conjunto de dados com sete bandas espectrais obtidas pelo satélite <i>Landsat TM</i> em 17 de julho de 1986. . . . .	p. 73
5.2	Verdade de campo para <i>Tippecanoe County</i> . . . . .	p. 73
5.3	O segundo conjunto de dados com nove bandas espectrais obtido pelo espectrômetro AVIRIS em junho de 1992. . . . .	p. 74
5.4	Verdade de campo para o segundo conjunto de testes. . . . .	p. 75

5.5	O terceiro conjunto de dados com doze bandas espectrais obtido por um <i>scanner</i> aero-embarcado em junho de 1966. . . . .	p. 76
5.6	Verdade de campo para o terceiro conjunto de testes. . . . .	p. 77
5.7	O quarto conjunto de dados com cinco bandas espectrais obtidas pelo satélite <i>Landsat TM</i> . . . . .	p. 78
5.8	Verdade de campo para o quarto conjunto de testes. . . . .	p. 79
5.9	O quinto conjunto de dados com três bandas espectrais obtidas pelo satélite <i>Ikonos</i> com resolução espacial de 1m. . . . .	p. 79
5.10	Verdade de campo para o quinto conjunto de testes. . . . .	p. 80
6.1	Amostras de avaliação para o primeiro conjunto de dados. . . . .	p. 83
6.2	Amostras de avaliação para o segundo conjunto de dados. . . . .	p. 84
6.3	Amostras de avaliação para o terceiro conjunto de dados. . . . .	p. 86
6.4	Amostras de avaliação para o quarto conjunto de dados. . . . .	p. 87
6.5	Amostras de avaliação para o quinto conjunto de dados. . . . .	p. 88
6.6	Desempenho e pontos não classificados de um classificador BRE-E (“Experimento 1”). . . . .	p. 91
6.7	Os dois tipos de pontos não classificados apresentados por um classificador BRE-E (“Experimento 1”). . . . .	p. 92
6.8	Desempenho e pontos não classificados de um classificador do tipo-2 BRE-OU (“Experimento 2”). . . . .	p. 93
6.9	Desempenho e pontos não classificados de um classificador do tipo-2 com base de regras específica e operador E seguido de OU, BRE-E/OU (“Experimento 3”). . . . .	p. 94
6.10	Desempenho e pontos não classificados de um classificador do tipo-2 com base de regras genérica e operador E, BRG-E (“Experimento 4”). . . . .	p. 95
6.11	Classificadores ordenados pelos rankings médios e agrupados pela distância crítica. . . . .	p. 98
6.12	Interpretação 1: BRG-E, BRE-E/OU e <i>MaxVer</i> teriam o mesmo desempenho e os três seriam superiores ao T1. . . . .	p. 99
6.13	Interpretação 2: BRG-E e BRE-E/OU seriam superiores a <i>MaxVer</i> e a T1. . . . .	p. 99
6.14	Interpretação 3: BRG-E seria superior a BRE-E/OU, <i>MaxVer</i> e T1. . . . .	p. 99

## *Lista de Tabelas*

2.1	Exemplo de matriz de confusão. . . . .	p. 22
4.1	<i>Pixel</i> de entrada e seus 4-vizinhos. . . . .	p. 63
6.1	Matriz de confusão para o classificador tipo-2 com base de regras específica e operador E para agregação dos antecedentes (BRE-E) utilizando apenas a primeira banda espectral. . . . .	p. 89
6.2	Matriz de confusão para o classificador tipo-2 BRE-E utilizando duas bandas espectrais. . . . .	p. 89
6.3	Matriz de confusão para o classificador tipo-2 BRE-E utilizando três bandas espectrais. . . . .	p. 89
6.4	Matriz de confusão para o classificador tipo-2 BRE-E utilizando quatro bandas espectrais. . . . .	p. 90
6.5	Matriz de confusão para o classificador tipo-2 BRE-E utilizando cinco bandas espectrais. . . . .	p. 90
6.6	Matriz de confusão para o classificador tipo-2 BRE-E utilizando seis bandas espectrais. . . . .	p. 90
6.7	Matriz de confusão para o classificador tipo-2 BRE-E utilizando todas as bandas disponíveis (7 bandas). . . . .	p. 91
6.8	Comparação entre o número de <i>pixels</i> não classificados para os quatro classificadores testados: <i>MaxVer</i> , <i>fuzzy</i> convencional (T1), BRE-E/OU e BRG-E. . . . .	p. 95
6.9	Comparação entre as precisões (P) dos classificadores: <i>MaxVer</i> , <i>fuzzy</i> convencional (T1), BRE-E/OU e BRG-E. . . . .	p. 96
6.10	Comparação entre os rankings médios dos classificadores: <i>MaxVer</i> , <i>fuzzy</i> convencional (T1), BRE-E/OU e BRG-E. . . . .	p. 97

## *Lista de abreviaturas e siglas*

<i>MaxVer</i>	Classificador estatístico da máxima verossimilhança,	p. 23
FIS	<i>Fuzzy Inference System</i> (Sistema de Inferência <i>Fuzzy</i> ),	p. 24
BR	Base de Regras,	p. 27
BRE	Base de Regras Específica,	p. 27
BRG	Base de Regras Genérica,	p. 27
T2 FIS	<i>Type-2 Fuzzy Inference System</i> (Sistema de Inferência <i>Fuzzy</i> do Tipo-2),	p. 30
PMF	<i>Principal Membership Function</i> (Função de Pertinência Principal),	p. 33
<i>join</i>	Operador que permite calcular a união entre dois conjuntos <i>fuzzy</i> gerais do tipo-2,	p. 36
<i>meet</i>	Operador que permite calcular a interseção entre dois conjuntos <i>fuzzy</i> gerais do tipo-2,	p. 36
CTR	Redução de Tipo Baseada no Centróide ( <i>Centroid Type Reduction</i> ),	p. 37
FOU	<i>Footprint Of Uncertainty</i> (“Mancha” de Incerteza),	p. 42
GEFOU	<i>General Footprint of Uncertainty</i> (“Mancha” de Incerteza Geral),	p. 42
VSCTR	<i>Vertical Slice Centroid Type Reduction</i> (Redução de Tipo Através dos Centróides das Fatias Verticais),	p. 46
NC	quantidade de <i>pixels</i> não classificados obtida por um classificador em uma tarefa de classificação de pontos de teste,	p. 67
BRE-E	Base de Regras Específica com agregação dos antecedentes dada pelo operador E,	p. 67
BRE-OU	Base de Regras Específica com agregação dos antecedentes dada pelo operador OU,	p. 67
BRE-E/OU	Base de Regras Específica com agregação dos antecedentes dada pelo operador E seguido do OU,	p. 67
BRG-E	Base de Regras Genérica com agregação dos antecedentes dada pelo operador E,	p. 67
CD	<i>Critical Distance</i> : distância crítica utilizada para comparar dois classificadores através do teste de Nemenyi,	p. 98

# *Sumário*

## **Resumo**

## **Abstract**

<b>1</b>	<b>Introdução</b>	p. 15
1.1	Considerações iniciais . . . . .	p. 15
1.2	Objetivos . . . . .	p. 16
1.3	Contribuições do trabalho . . . . .	p. 17
1.4	Organização do texto . . . . .	p. 17
<b>2</b>	<b>Fundamentação teórica</b>	p. 18
2.1	Introdução . . . . .	p. 18
2.2	Sensoriamento remoto . . . . .	p. 18
2.2.1	Reflectância . . . . .	p. 18
2.2.2	Assinatura espectral . . . . .	p. 19
2.2.3	Valores do contador digital . . . . .	p. 19
2.2.4	Resolução espacial . . . . .	p. 19
2.3	Classificadores digitais . . . . .	p. 20
2.3.1	Treinamento . . . . .	p. 21
2.3.2	Classificação . . . . .	p. 21
2.3.3	Validação de um classificador . . . . .	p. 21

2.3.4	Classificador da Máxima Verossimilhança . . . . .	p. 23
2.4	Sistemas <i>fuzzy</i> . . . . .	p. 24
2.4.1	Sistema de inferência <i>fuzzy</i> . . . . .	p. 24
2.4.1.1	Regras <i>fuzzy</i> . . . . .	p. 26
2.4.1.2	Base de regras . . . . .	p. 27
2.4.1.3	Inferência escalonada . . . . .	p. 27
2.4.1.4	<i>Defuzzificação</i> . . . . .	p. 28
2.4.2	Sistema de inferência <i>fuzzy</i> do tipo-2 . . . . .	p. 29
2.4.3	Conjunto <i>fuzzy</i> do tipo-2 . . . . .	p. 31
2.4.4	Fatia vertical . . . . .	p. 32
2.4.5	Função de pertinência principal . . . . .	p. 33
2.4.6	Operações sobre conjuntos do tipo-2 . . . . .	p. 35
2.4.6.1	União de dois conjuntos <i>fuzzy</i> do tipo-2 . . . . .	p. 35
2.4.6.2	Interseção de dois conjuntos <i>fuzzy</i> do tipo-2 . . . . .	p. 36
2.4.6.3	Redução de tipo . . . . .	p. 37
2.5	Trabalhos relacionados . . . . .	p. 38
2.6	Conclusão . . . . .	p. 40
<b>3</b>	<b>Proposta de Formalização do Sistema de Inferência Escalonada Fuzzy Geral do Tipo-2</b>	<b>p. 41</b>
3.1	Introdução . . . . .	p. 41
3.2	“Mancha” de Incerteza Geral . . . . .	p. 41
3.3	<i>Singleton</i> do tipo-2 . . . . .	p. 43
3.4	Translação de fatia vertical . . . . .	p. 44
3.5	Supremo de um conjunto <i>fuzzy</i> geral do tipo-2 . . . . .	p. 44

3.6	Extensão cilíndrica de uma fatia vertical . . . . .	p. 45
3.7	Redução de tipo baseada no centróide das fatias verticais . . . . .	p. 45
3.8	Um sistema de inferência <i>fuzzy</i> geral do tipo-2 . . . . .	p. 48
3.9	Conclusão . . . . .	p. 52
<b>4</b>	<b>Metodologia de Desenvolvimento do Classificador <i>Fuzzy</i> do Tipo-2 Geral</b>	<b>p. 54</b>
4.1	Introdução . . . . .	p. 54
4.2	O classificador <i>fuzzy</i> do tipo-2 geral . . . . .	p. 54
4.2.1	Fase de projeto . . . . .	p. 54
4.2.1.1	Obtenção dos antecedentes das regras . . . . .	p. 54
4.2.1.2	Definição dos consequentes . . . . .	p. 59
4.2.1.3	Obtenção da base de regras . . . . .	p. 59
4.2.2	Processo de avaliação . . . . .	p. 61
4.2.3	Base de regras específica com operador de agregação E (BRE-E) . . . . .	p. 68
4.2.4	Base de regras específica com operador de agregação OU (BRE-OU) . . . . .	p. 68
4.2.5	Base de regras específica com operador de agregação E seguido de OU (BRE-E/OU) . . . . .	p. 69
4.2.6	Base de regras genérica com operador de agregação E (BRG-E) . . . . .	p. 70
4.3	Conclusão . . . . .	p. 71
<b>5</b>	<b>Estudo de casos</b>	<b>p. 72</b>
5.1	Introdução . . . . .	p. 72
5.2	Caso 1 . . . . .	p. 72
5.3	Caso 2 . . . . .	p. 72
5.4	Caso 3 . . . . .	p. 75
5.5	Caso 4 . . . . .	p. 75
5.6	Caso 5 . . . . .	p. 78
5.7	Conclusão . . . . .	p. 80



<b>6</b>	<b>Resultados e discussão</b>	p. 81
6.1	Introdução . . . . .	p. 81
6.2	Descrição dos experimentos . . . . .	p. 81
6.3	Resultados e Discussões . . . . .	p. 85
6.3.1	Experimento 1 . . . . .	p. 85
6.3.2	Experimento 2 . . . . .	p. 92
6.3.3	Experimento 3 . . . . .	p. 93
6.3.4	Experimento 4 . . . . .	p. 94
6.3.5	Experimento 5 . . . . .	p. 95
6.3.5.1	Avaliação estatística . . . . .	p. 96
6.4	Conclusão . . . . .	p. 100
<b>7</b>	<b>Conclusões e trabalhos futuros</b>	p. 101
	<b>Referências Bibliográficas</b>	p. 104



# *Resumo*

Propõe-se, nesta tese, o desenvolvimento de uma nova ferramenta baseada em conjuntos *fuzzy* gerais do tipo-2 para aplicação em processos de classificação digital de dados. O problema de classificação a ser considerado está relacionado à identificação de regiões de floresta em imagens de satélite com o objetivo de auxiliar em tarefas de monitoramento florestal. O classificador digital desenvolvido utiliza um mecanismo de inferência denominado de “inferência escalonada *fuzzy* geral do tipo-2” para classificar os *pixels* das imagens de satélite de acordo com sua cobertura vegetal. Tal classificador é inovador pois, além de utilizar conjuntos *fuzzy* tipo-2 gerais, pode utilizar tanto uma base de regras específica quanto uma base genérica (ambas de forma hierárquica) para reclassificar pontos que, do contrário, permaneceriam sem classificação. Isto permite a obtenção de uma base de regras compacta (composta de poucas regras). A justificativa para o uso de sistemas de inferência do tipo-2 é que estes, apesar do custo computacional maior, apresentam desempenho superior aos sistemas do tipo-1 equivalentes. Os testes realizados mostram que, de fato, o sistema proposto é melhor do que o classificador *fuzzy* convencional usualmente empregado em aplicações semelhantes e possui desempenho comparável ao classificador estatístico da máxima verossimilhança, sendo uma alternativa viável ao último.

# *Abstract*

This work proposes the development of a new tool based on general type-2 fuzzy sets to be applied to digital classification of data. The classification problem considered here regards the identification of areas of forest in satellite images. The goal is to assist users in tasks related to monitoring forest. The developed digital classifier employs an inference mechanism called “general type-2 scaled inference” to classify pixels in images according to their vegetation cover. Such classifier is innovative because, besides using general type-2 fuzzy sets, it can use specific and generic rules base (both in a hierarchical way) to reclassify pixels that remain unclassified. Such hierarchical reclassification leads to a compact rule base (with few rules). The reason why one should use type-2 fuzzy inference is that they present better performance than their type-1 counterparts, in spite of their bigger computational cost. The carried out tests showed, for sure, that the proposed system is better than the conventional fuzzy classifier usually employed in similar applications and its performance is comparable to the statistical likelihood classifier, proving to be an alternative choice to this last one.

# 1 *Introdução*

## 1.1 **Considerações iniciais**

Os conjuntos do tipo-2 foram apresentados pela primeira vez por Zadeh (1975) como uma evolução da sua teoria dos conjuntos *fuzzy*. Posteriormente, tais conjuntos foram estudados com mais profundidade e sua base teórica foi adequadamente estabelecida (KARNIK; MENDEL, 2001). A principal definição aceita atualmente é a de que os conjuntos do tipo-2 são aqueles cujos graus de pertinência são, eles próprios, conjuntos do tipo-1. Os conjuntos do tipo-2 que atendem estritamente a definição apresentada são chamados de conjuntos gerais do tipo-2 e aqueles que utilizam funções de pertinência constantes (unitárias) são conhecidos como conjuntos intervalares do tipo-2.

Recentemente, começaram a surgir diversas aplicações envolvendo sistemas de inferência *fuzzy* do tipo-2 os quais podem ser considerados evoluções dos sistemas *fuzzy* tradicionais, pois utilizam conjuntos do tipo-2. Os sistemas de inferência do tipo-2 são interessantes, uma vez que permitem tratar os dois tipos de incerteza usualmente encontrados quando se trabalha com sistemas *fuzzy*: a incerteza presente na medição do sinal (ruído) e a incerteza presente no nível semântico (o significado dos termos linguísticos pode variar de um especialista para outro) (ZENG; LIU, 2007).

Uma área bastante comum de aplicação de sistemas *fuzzy* é a classificação digital. Classificadores digitais são sistemas capazes de agrupar dados de acordo com algum critério de similaridade. Quando se utilizam estes classificadores para separar regiões de uma imagem em classes de cobertura do solo, por exemplo, costuma-se chamá-los de classificadores digitais de cobertura vegetal.

Os classificadores de cobertura vegetal são muito empregados em sensoriamento remoto para monitorar a degradação das áreas de preservação como florestas, mangues e matas ciliares. Além disso, é importante acompanhar o avanço da fronteira agrícola (e.g.: plantação de cana-de-açúcar, soja, trigo, arroz e milho) sobre as áreas de preservação. Particularmente, as florestas são o ecossistema terrestre mais importante pois sustentam a vida através da biodiversidade, da conservação da água e do solo, entre outros. Há estudos que mostram que a área florestal cobre 30% da superfície terrestre e abriga 80% da biodiversidade e mais, 25% da população mundial necessitam, em diferentes proporções, dos recursos florestais para sobreviverem (WFC... , 2003). Apesar da sua

importância, as florestas têm sido ameaçadas tanto pela atividade humana como a agricultura, as queimadas, a urbanização, quanto pelo mau uso dos recursos naturais.

Já existem, em uso, bons classificadores estatísticos, como o *MaxVer* (baseado no método da máxima verossimilhança), e os classificadores *fuzzy* convencionais (classificadores que utilizam conjuntos fuzzy do tipo-1), ambos com bons índices de acerto. O *MaxVer* é um classificador bastante utilizado, porém exige conhecimento de um especialista para que se possa fazer o melhor uso do mesmo. Em contrapartida, o classificador *fuzzy* é normalmente obtido a partir da adaptação de um sistema de inferência *fuzzy* e, como tal, procura capturar o conhecimento do especialista (ou obter conhecimento via dados de treinamento), de forma que possa ser bem utilizado pelo usuário comum, com pouca interação deste com o sistema. Os sistemas de inferência baseados em conjuntos do tipo-2 têm grande potencial para serem aplicados na classificação digital mas a maior parte das aplicações do tipo-2 vem utilizando conjuntos intervalares, que são formas bastante simplificadas dos conjuntos gerais do tipo-2, mas apresentam, como vantagem, o baixo custo computacional.

Propõe-se, nesta tese, o desenvolvimento de uma nova ferramenta baseada em conjuntos *fuzzy* gerais do tipo-2 para aplicação em processos de classificação digital de dados. Para isso, foi preciso desenvolver e formalizar um sistema de inferência *fuzzy* geral do tipo-2. O classificador proposto será validado em problemas de classificação de cobertura vegetal, utilizando dados provenientes de plataformas de sensoriamento remoto. O desempenho deste classificador será comparado com os de referência (*MaxVer* e *fuzzy* convencional) utilizando ferramentas de significância estatística. Convém mencionar que se pretende obter um classificador aplicável em quaisquer problemas de classificação digital e, portanto, o uso de dados de sensoriamento remoto visa apenas à validação do sistema, não estando este limitado aos problemas de classificação de cobertura vegetal. Ainda mais, a inferência a ser formalizada não se restringe à aplicação em classificação, podendo ser utilizada em outras áreas onde os sistemas *fuzzy* vêm sendo utilizados com sucesso, tais como, o controle de plantas industriais (LEE, 1990; WAN; KAMAL, 2006; PAREEK; KAR, 2006) e a aproximação de funções não-lineares (ISHIBUCHI; YAMAMOTO; NAKASHIMA, 2006).

## 1.2 Objetivos

Os principais objetivos desta tese são:

- desenvolver um sistema de inferência geral que trabalhe com conjuntos *fuzzy* gerais do tipo-2;
- definir e formalizar alguns termos e, também, desenvolver algoritmos para dar suporte ao mecanismo de inferência associado aos conjuntos *fuzzy* gerais do tipo-2;

- adaptar este mecanismo de inferência à classificação digital;
- testar o classificador produzido em diferentes problemas de classificação digital;
- validar o novo classificador comparando-o com classificadores convencionais como o *MaxVer* e o *fuzzy* tradicional, utilizando técnicas estatísticas adequadas.

### 1.3 Contribuições do trabalho

A principal contribuição desta tese é a formalização de um mecanismo de inferência para sistemas *fuzzy* baseados em conjuntos do tipo-2 denominado “sistema de inferência escalonada *fuzzy* geral do tipo-2”. Para dar suporte a este mecanismo, foram propostas: uma ferramenta de visualização para conjuntos do tipo-2 gerais (a “mancha” de incerteza geral), a translação de uma fatia vertical, o supremo de um conjunto *fuzzy* geral do tipo-2, a extensão cilíndrica de uma fatia vertical e o *singleton* do tipo-2. Além disso, para que a inferência geral fosse possível era necessário diminuir seu custo computacional e, para isso, propôs-se também uma técnica de *defuzzificação* de conjuntos gerais baseada nos centróides das fatias verticais (*VSCTR: Vertical Slice Centroid Type Reduction*).

Também foram propostos o “método de classificação hierárquica” e a adaptação da inferência escalonada *fuzzy* geral do tipo-2 aos classificadores de cobertura vegetal. O “método de classificação hierárquica” permite utilizar base de regras específica, com operador E seguido de OU, ou base genérica, com operador E, onde as regras possuem variáveis irrelevantes. A “inferência escalonada *fuzzy* geral do tipo-2” opera sobre conjuntos *fuzzy* gerais do tipo-2 e pode ser adaptada aos classificadores de cobertura vegetal.

### 1.4 Organização do texto

O capítulo 2 apresenta a base teórica necessária ao entendimento da metodologia proposta, bem como trabalhos relacionados ao tema estudado. O capítulo 3 apresenta a proposta de formalização do sistema de inferência escalonada *fuzzy* geral do tipo-2. O capítulo 4 mostra o desenvolvimento de um classificador *fuzzy* do tipo-2 geral a partir da adaptação do mecanismo de inferência geral. O capítulo 5 descreve os conjuntos de dados (estudos de caso) que serão utilizados para validar os classificadores desenvolvidos. No capítulo 6, são apresentados os resultados relativos à comparação dos classificadores desenvolvidos com os classificadores *MaxVer* e *fuzzy* tradicional e é mostrada a técnica estatística utilizada para avaliar estes resultados. As conclusões e as sugestões para trabalhos futuros são feitas no capítulo 7.

## 2 *Fundamentação teórica*

### 2.1 Introdução

Neste capítulo, são apresentados conceitos importantes para o sensoriamento remoto, tais como, assinatura espectral, resolução espacial e classificação digital. Os classificadores digitais são analisados e são vistas as principais métricas utilizadas para validar tais sistemas. Também são apresentados os conceitos básicos dos sistemas *fuzzy* e são feitas algumas considerações sobre a adaptação da etapa de *defuzzificação* de um sistema de inferência *fuzzy* convencional para que este funcione como classificador digital. O capítulo traz ainda os conjuntos *fuzzy* do tipo-2 e os sistemas de inferência utilizados por alguns autores para trabalhar com esses conjuntos. Finalmente, são apresentados os trabalhos relacionados à área de classificação *fuzzy*.

### 2.2 Sensoriamento remoto

#### 2.2.1 Reflectância

A reflectância é o nome dado à relação entre o fluxo de energia refletido e o fluxo de energia incidente (e.g.: luz solar) nos alvos terrestres. O valor da reflectância depende do tipo de alvo que está sendo imageado, sendo diferente para cada região do espectro eletromagnético ou comprimento de onda considerado. Quando se analisa um vegetal, nota-se que a reflectância é dominada pela presença dos pigmentos das folhas (clorofila, caroteno, xantofilas e antocianina) na região visível do espectro eletromagnético, em comprimentos de onda que variam de 0,4 a 0,7  $\mu\text{m}$ . Nota-se, também, uma região de transição, em torno de 0,7  $\mu\text{m}$ , conhecida como “borda vermelha” (*red edge*) (USTIN et al., 2004), onde a reflectância aumenta significativamente e, finalmente, uma faixa do espectro compreendida entre 1,3 a 3,0  $\mu\text{m}$ , onde a reflectância é dominada pela presença de água, principalmente devido à existência de três picos de absorção, em 1,4, 1,9 e 2,7  $\mu\text{m}$  (CENTENO, 2004).



## 2.2.2 Assinatura espectral

Analisando-se a reflectância como função do comprimento de onda, percebeu-se que cada tipo de alvo apresentava uma curva com formato diferente e, por esta razão, a curva foi chamada de “assinatura espectral” do alvo (LANDGREBE, 2002). Assim, a assinatura espectral de um vegetal é diferente da assinatura de um solo, por exemplo. Na figura 2.1, podem ser vistas as assinaturas espectrais para alguns *pixels* de uma imagem de área cultivada com milho, feitas a partir de 7 bandas espectrais do satélite *Landsat*. Alguns autores questionam o termo “assinatura espectral” já que, como pode ser visto na figura, as curvas referentes aos *pixels* de uma mesma cultura, embora semelhantes, não são iguais.

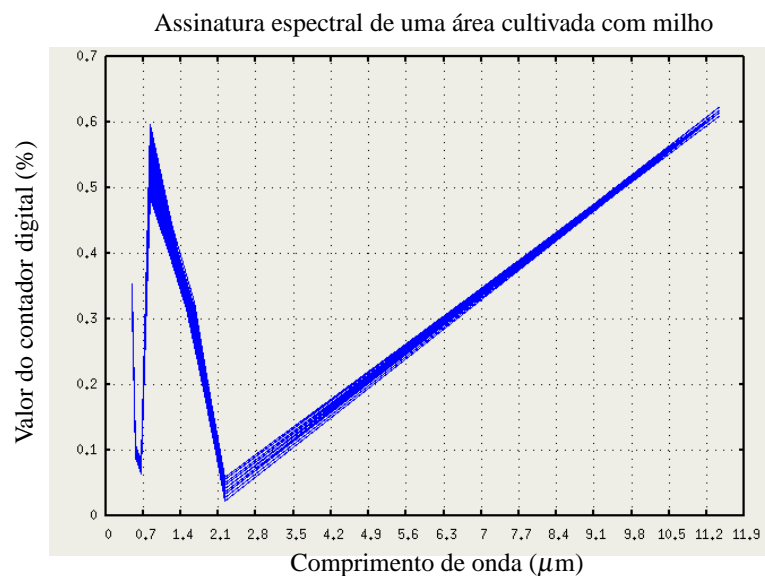


Figura 2.1: Exemplo de resposta espectral de uma plantação de milho, obtida a partir do *Landsat*, com 7 bandas espectrais.

## 2.2.3 Valores do contador digital

Os dados obtidos a partir dos satélites de sensoriamento remoto, como o *Landsat*, o *CBERS* e o *Ikonos*, costumam ser apresentados em “valores do contador digital”, que são números inteiros proporcionais ao valor da reflectância dos alvos terrestres imageados. Os valores do contador digital dependem das características do detector e da quantidade de energia recebida e podem ser utilizados para calcular a reflectância, através da curva de resposta do detector, do ângulo de iluminação solar, do ângulo de visada e das diferenças nas condições atmosféricas (USTIN et al., 2004).

## 2.2.4 Resolução espacial

A resolução espacial de uma imagem de satélite é a área unitária de terreno representada por um pixel (RICHARDS; JIA, 1999). Em sensoriamento remoto, tem havido uma evolução significativa

na resolução espacial: as primeiras versões da série *Landsat*, nos anos 1970, adquiriam imagens com resolução espacial de 80m (HSIEH; LEE; CHEN, 2001), as versões mais atuais do *Landsat* possuem resolução de 30m (BARATA; PINA, 2002) e satélites mais modernos, como o *Ikonos* e o *Quickbird*, lançados em 1999, têm resolução de 70cm (ROSENFELD, 2001; HSIEH; LEE; CHEN, 2001).

## 2.3 Classificadores digitais

Classificadores digitais de cobertura vegetal são sistemas capazes de separar regiões de uma imagem em classes de cobertura do solo, ou alvos terrestres, partindo da hipótese que regiões que cobrem um mesmo alvo devem ter características espectrais e espaciais similares (SCHOWENGERDT, 1983; RICHARDS; JIA, 1999; MAULIK; BANDYOPADHYAY, 2003). O resultado da classificação é uma imagem conhecida como mapa temático ou mapa de classes que é útil para mapeamento topográfico, monitoramento de produção agrícola, proteção ambiental, entre outros (SOLAIMAN; PIERCE; ULABY, 1999).

Existem dois tipos de classificadores digitais, os supervisionados e os não supervisionados. Os últimos, também conhecidos como algoritmos de agrupamento (KLIR; FOLGER, 1988; PEDRYCZ; GOMIDE, 1998), são aqueles que procuram encontrar, no conjunto de dados, grupos cujos elementos apresentam alguma similaridade. Exemplos de classificadores não supervisionados são o *k-means*, o *isodata* (DUDA; HART; STORK, 2001) e o *fuzzy c-means* (PEDRYCZ; GOMIDE, 1998). Os classificadores supervisionados, por outro lado, são aqueles que se baseiam numa primeira etapa (treinamento) de onde, a partir da escolha de amostras das classes de interesse, extraem-se informações para classificar, numa segunda etapa, os *pixels* da imagem (CENTENO, 2004). Os dois classificadores supervisionados mais populares são o classificador da distância mínima e o classificador da máxima verossimilhança (*MaxVer*) (BARATA; PINA, 2002), mas o último apresenta desempenho superior e, por isso, tem sido utilizado como referência para a avaliação de desempenho de novos classificadores (BARATA; PINA, 2002; ROEMER; KACPRZYNSKI; ORSAGH, 2001).

O *MaxVer* é, portanto, um método de classificação estatístico geral que, tradicionalmente, tem sido utilizado para classificar *pixels*, atribuindo cada um destes à classe mais provável (ALI-AKBAR; SHARIFI; MULDER, 2000).

Os classificadores supervisionados têm desempenho superior aos não supervisionados quando as classes são similares espectralmente (LANDGREBE, 2002) e, por isso, trabalhar-se-á, nesta tese, exclusivamente, com classificadores supervisionados, uma vez que se pretende distinguir culturas agrícolas, tais como, milho, soja e alfafa, por exemplo.

A seguir são descritas, com maior profundidade, as etapas de treinamento e de classificação, típicas em um classificador supervisionado e, na sequência, será abordado o classificador da máxima verossimilhança.

### **2.3.1 Treinamento**

Na fase de treinamento, extraem-se amostras das classes de interesse nas imagens de cada banda espectral. Um aspecto importante do treinamento é que a escolha das amostras é fundamental para o processo de classificação. Dois critérios podem ser estabelecidos para a escolha das amostras: tamanho e qualidade. Do ponto de vista do tamanho, as amostras não devem ser nem muito pequenas (LEE; LANDGREBE, 1993) e nem muito grandes; se forem muito pequenas, não serão representativas da classe e se forem muito grandes, aumenta-se o risco de se ter mistura de classes (CENTENO, 2004; ZHUKOV et al., 1999). Uma forma de aumentar o tamanho das amostras e, ao mesmo tempo, evitar a mistura de classes é tomar várias amostras de pequeno tamanho. O critério da qualidade pede que as amostras sejam puras, ou seja, que contenham apenas uma classe de cobertura vegetal (HSIEH; LEE; CHEN, 2001).

Em Landgrebe e Jimenez (1998) foi apresentado um estudo sobre o número de amostras necessárias para a classificação supervisionada onde se concluiu que, quanto maior a dimensão do problema (neste caso, o número de bandas espectrais consideradas), maior deve ser o número de amostras de treinamento. Segundo os autores, é razoável esperar que o cálculo com altas dimensões contenha mais informação para detectar classes com maior precisão, mas esta vantagem não se verificará se o número de dados de treinamento não aumentar proporcionalmente com a dimensão do problema.

### **2.3.2 Classificação**

Na fase de classificação, utilizando-se as informações obtidas na etapa de treinamento, atribui-se a cada *pixel* um rótulo que o identifique como membro de uma das classes, em função de seus valores do contador digital (CENTENO, 2004).

### **2.3.3 Validação de um classificador**

Um classificador digital ideal pode ser visto como sendo aquele que estima corretamente todos os *pixels* de teste. Desta forma, para avaliar a qualidade de um classificador qualquer, deve-se verificar quantos pontos de teste ele é capaz de estimar e com que acerto o faz.

O ideal seria que nenhum ponto avaliado ficasse sem rótulo, mas é preferível não classificar a fazê-lo com erro, desde que existam informações adicionais para reclassificar tal ponto posteriormente.

De forma semelhante, Fahsi et al. (2000) definiram que a qualidade da classificação pode ser verificada através da comparação dos rótulos atribuídos aos dados de teste com a sua verdade de campo correspondente.

Para avaliar a qualidade de um classificador, existem duas métricas bastante utilizadas: a matriz de confusão e a precisão, que passam a ser descritas a seguir.

- Matriz de confusão

A matriz de confusão representa as diferenças entre a verdade de campo e o resultado do processo de classificação. As linhas da matriz correspondem às classes corretas e as colunas correspondem às classes estimadas pelo classificador (CENTENO, 2004). A tabela 2.1 mostra um exemplo de matriz de confusão considerando 3 classes, milho, soja e alfafa, que foram classificadas (estimadas) como  $\widehat{\text{milho}}$ ,  $\widehat{\text{soja}}$  e  $\widehat{\text{alfafa}}$ , respectivamente.

Tabela 2.1: Exemplo de matriz de confusão.

	Milho	Soja	Alfafa
Milho	4965	343	70
Soja	140	4300	243
Alfafa	51	225	247

Na primeira linha da matriz, pode-se ver que 4965 *pixels* eram milho e foram classificados como tal; 343 eram milho e foram classificados como  $\widehat{\text{soja}}$ ; e 70 eram milho e foram classificados como  $\widehat{\text{alfafa}}$ . Assim num classificador ideal a matriz de confusão teria nulos todos os elementos fora da diagonal principal.

- Precisão

A precisão é a relação entre o número de *pixels* corretamente classificados e o número de *pixels* avaliados e pode ser calculada a partir da matriz de confusão, somando-se os *pixels* da diagonal principal e dividindo-se pela soma de todos os *pixels* da matriz, conforme a equação (2.1),

$$P = \frac{\sum_{i=1}^M M(i, i)}{N} \quad (2.1)$$

onde  $P$  é a precisão,  $M$  é o número de classes consideradas,  $M(i, i)$  é o número de *pixels* corretamente classificados na linha  $i$  e na coluna  $i$ , e  $N$  é o número total de *pixels*.

A precisão fornece uma indicação da qualidade do classificador, mostrando o percentual de pontos corretamente classificados mas não leva em conta os erros de omissão ou falsos negativos, onde *pixels* que pertenciam a uma determinada classe  $c$  são classificados como sendo de outra, e os erros de inclusão ou falsos positivos, onde *pixels* que eram de outra classe e são classificados com sendo de  $c$ .

Convém mencionar que, em sensoriamento remoto, é usual o índice *Kappa* que possui como vantagem, sobre a precisão, o fato de levar em conta os erros de inclusão e os erros de omissão (RICHARDS; JIA, 1999). Nesta tese foi considerada apenas a precisão devido à sua simplicidade de cálculo.

### 2.3.4 Classificador da Máxima Verossimilhança

O classificador da máxima verossimilhança (*MaxVer*) utiliza amostras de treinamento para estimar seus parâmetros, sendo que a melhor estimativa é aquela que maximiza a probabilidade de se obter as mesmas amostras que estão sendo observadas (DUDA; HART; STORK, 2001).

Suponha a existência de  $L$  classes temáticas identificadas pelos rótulos  $C_i$ ,  $i = 1 \dots L$ . O rótulo  $C_i$  pode ser visto, do ponto de vista estatístico, como “o evento de um objeto pertencer à classe  $i$ ” (KETTIG; LANDGREBE, 1976).

Quando se trabalha com imagens provenientes de múltiplos sensores espectrais, considera-se um espaço  $P$ -dimensional, onde  $P$  é o número de fontes consideradas (bandas espectrais). Neste contexto, a variável de entrada  $\mathbf{X}$ , a ser classificada, passa a ser o vetor coluna  $\mathbf{X} = [x_1, x_2, \dots, x_P]^t$  onde  $x_k$ ,  $k = 1, 2, \dots, P$  é uma determinada característica extraída da banda  $k$  como, por exemplo, o nível de cinza de um *pixel* sob análise. Cada banda admite um conjunto de valores possíveis  $\Omega_k$ , tal que  $x_k \in \Omega_k$ .

Seja  $p(\mathbf{X} \setminus C_i)$  a função densidade de probabilidade condicional para a  $i$ -ésima classe, ou seja, a função densidade de probabilidade para  $\mathbf{X}$  dado que a classe observada é  $C_i$  (DUDA; HART; STORK, 2001). Supondo que seja possível assumir que  $p(\mathbf{X} \setminus C_i)$  tem densidade Gaussiana multivariada com média  $\mu_i$  e matriz de covariâncias  $\Sigma_i$

$$p(\mathbf{X} \setminus C_i) \approx N(\mu_i, \Sigma_i) \equiv \frac{1}{\sqrt{(2\pi)^P |\Sigma_i|}} \cdot e^{-\frac{1}{2}(\mathbf{X}-\mu_i)' \Sigma_i^{-1} (\mathbf{X}-\mu_i)} \quad (2.2)$$

onde  $\Sigma_i$  é a matriz de covariâncias,  $|\Sigma_i|$  é o determinante desta matriz,  $\mathbf{X}$  é o vetor de entrada,  $P$  é a dimensão do vetor de entrada e  $\mu_i$  é o vetor das médias.

Embora os valores populacionais  $\mu_i$  e  $\Sigma_i$  não sejam conhecidos, estes podem ser estimados a partir das amostras de treinamento como foi demonstrado por Duda, Hart e Stork (2001)

$$\hat{\mu}_i = \frac{1}{n} \sum_{k=1}^n \mathbf{X}_k \quad (2.3)$$

e

$$\hat{\Sigma}_i = \frac{1}{n} \sum_{k=1}^n (\mathbf{X}_k - \hat{\mu}_i)(\mathbf{X}_k - \hat{\mu}_i)^t \quad (2.4)$$

onde  $\hat{\mu}_i$  é a estimativa de  $\mu_i$ ,  $\hat{\Sigma}_i$  é a de  $\Sigma_i$ ,  $n$  é o número de amostras de treinamento considerado para a classe  $C_i$  e  $\mathbf{X}_k$  é o vetor de elementos da amostra  $k$ .

A equação (2.3) diz que a média populacional pode ser estimada a partir da média aritmética das amostras de treinamento, ou seja, a média amostral, e a equação (2.4) diz que a matriz de covariâncias estimada é a média aritmética de  $n$  matrizes  $(\mathbf{X}_k - \mu_i)(\mathbf{X}_k - \mu_i)^t$ ,  $k = 1 \dots n$ .

Finalmente, a teoria da decisão Bayesiana permite escolher a qual classe o vetor de entrada será atribuído (ZENG; LIU, 2007; DUDA; HART; STORK, 2001; KETTIG; LANDGREBE, 1976)

$$\text{SE } p(\mathbf{X} \in C_l) = \max_j (p(\mathbf{X} \in C_j)) \text{ ENTÃO } \mathbf{X} \in C_l. \quad (2.5)$$

## 2.4 Sistemas *fuzzy*

### 2.4.1 Sistema de inferência *fuzzy*

O conhecimento capturado por um sistema especialista manifesta-se, muitas vezes, de forma imprecisa ou incerta e é desejável que tal sistema seja capaz de realizar inferências não triviais a partir destes dados imprecisos. Por isso, a teoria dos conjuntos *fuzzy* tem sido empregada como ferramenta para o projeto de sistemas especialistas que lidam com a incerteza (ABDULGHAFOR; CHANDRA; ABIDI, 1993). Sabe-se que muita pesquisa tem sido feita visando utilizar o raciocínio aproximado e os conjuntos *fuzzy* no processo de inferência, dando origem aos “Sistemas de inferência *fuzzy*” (*FIS: fuzzy inference systems*) (KLIR; FOLGER, 1988), cujas variáveis de entrada e saída normalmente são valores não-*fuzzy* (números reais ou inteiros). A figura 2.2 ilustra um sistema de inferência *fuzzy* típico.

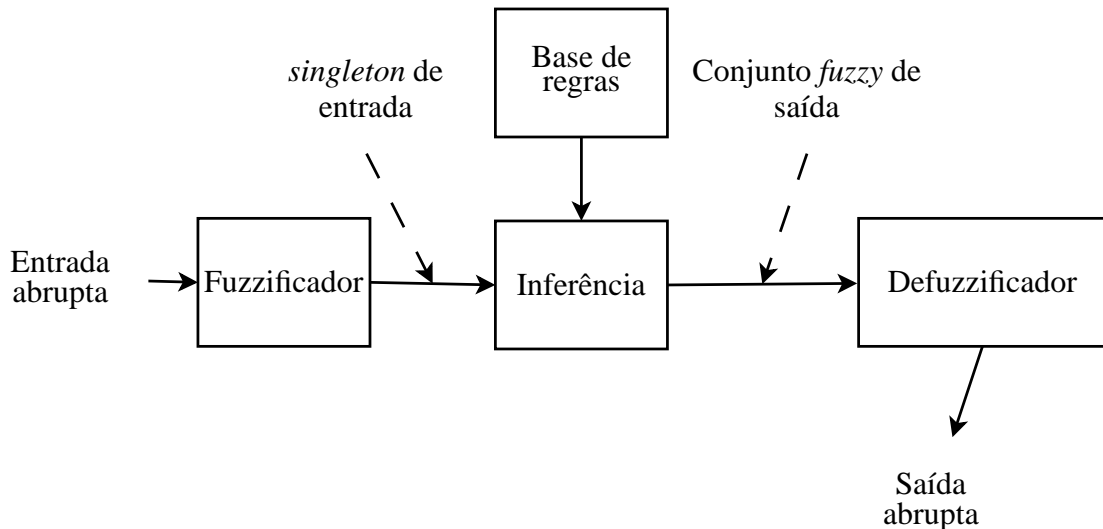


Figura 2.2: Estrutura típica de um sistema de inferência *fuzzy*.

A teoria dos conjuntos *fuzzy* é, portanto, uma ferramenta matemática poderosa para modelar a habilidade humana de se chegar a conclusões quando a informação disponível é imprecisa, incompleta e não totalmente confiável (SOLAIMAN; PIERCE; ULABY, 1999).

Na teoria convencional dos conjuntos, a pertinência de um determinado elemento a um conjunto é 1 ou 0, ou seja, ou o elemento pertence ao conjunto ou não pertence. Na teoria dos conjuntos *fuzzy*, trabalha-se com graus intermediários de pertinência (números reais compreendidos entre 0,0 e 1,0), permitindo que um elemento pertença parcialmente a um conjunto. Logo, um conjunto *fuzzy* pode ser definido como um conjunto de pares ordenados (SOLAIMAN; PIERCE; ULABY, 1999) na forma:

$$A = \{(x, \mu_A(x)), x \in \Omega\} \quad (2.6)$$

onde  $A$  é o conjunto *fuzzy*,  $x$  é um elemento deste conjunto,  $\mu_A(x)$  é denominada função de pertinência e determina com que grau o elemento  $x$  pertence ao conjunto  $A$  ( $\mu_A \in [0, 1]$ ) e  $\Omega$  é o conjunto universo.

Uma notação bastante usada em conjuntos *fuzzy* discretos é dada por  $A = \sum \mu_A(x)/x$  onde o somatório é usado apenas para identificar a união dos pares (pertinência, elemento). Esta notação será usada para descrever as fatias verticais discretas apresentadas neste capítulo.

Existe um tipo particular de conjunto *fuzzy*, bastante utilizado em aplicações práticas, denominado conjunto unitário ou *singleton*, cuja função de pertinência é definida como (DELGADO, 2002):

$$\mu_A(x) = \begin{cases} h & \text{se } x = m \\ 0 & \text{caso contrário} \end{cases} \quad (2.7)$$

onde  $m$  é um valor particular de  $x$  e  $h$  é a altura do *singleton* ( $0 \leq h \leq 1$ ). A figura 2.3 mostra o *singleton* descrito.

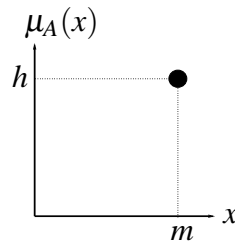


Figura 2.3: Conjunto unitário ou *singleton*.

Um elemento importante que integra um sistema de inferência *fuzzy* é sua base de regras e esta, por sua vez, é formada pelas regras *fuzzy*. A *defuzzificação* permite transformar a saída *fuzzy* de um FIS numa saída escalar. Estes elementos constituintes de um FIS são detalhados a seguir.

#### 2.4.1.1 Regras *fuzzy*

As regras *fuzzy* são muito apropriadas quando o conhecimento do domínio resulta de associações empíricas e experiências do operador humano, ou quando se deseja uma representação lingüística do conhecimento adquirido (PEDRYCZ; GOMIDE, 1998).

Em geral, as regras *fuzzy* assumem a forma “se <antecedente> então <consequente>”, na qual os antecedentes e os consequentes são conjuntos *fuzzy*. As regras *fuzzy* têm, portanto, a seguinte forma:

$$R : \text{SE } x_1 \text{ é } F_1 \text{ op } x_2 \text{ é } F_2 \cdots \text{op } x_p \text{ é } F_p \text{ ENTÃO } Y \text{ é } G \quad (2.8)$$

onde  $F_k$ ,  $k = 1 \cdots P$ , e  $G$  representam termos lingüísticos definidos por conjuntos *fuzzy* nos universos  $\mathbf{X}$  e  $\mathbf{Y}$ , respectivamente. Os termos  $F_k$  são os antecedentes da regra,  $G$  é o consequente e *op* é o operador de agregação dos antecedentes, podendo ser conjuntivo (E) ou disjuntivo (OU).

Alguns sistemas *fuzzy* aceitam regras com condições irrelevantes (*don't care conditions*) nos antecedentes e/ou nos consequentes (ISHIBUCHI; YAMAMOTO, 2004). Para estas regras, denominadas regras genéricas, nem todas as variáveis disponíveis são relevantes para a obtenção da saída inferida por aquela regra:

$$R : \text{SE } x_1 \text{ é } \# \text{ op } x_2 \text{ é } \# \cdots \text{op } x_p \text{ é } F_p \text{ ENTÃO } Y \text{ é } G \quad (2.9)$$



onde os símbolos # indicam a irrelevância das variáveis correspondentes. Deste modo, proposições envolvendo um grande número de variáveis podem ser substituídas por proposições mais simples, sem a presença das variáveis irrelevantes.

### 2.4.1.2 Base de regras

A base de regras (BR) de um sistema de inferência *fuzzy* é o conjunto de regras que determina como as variáveis de entrada se relacionam com a variável de saída. A base de regras pode ser de dois tipos (DELGADO, 2002):

- base de regras específica (BRE): utiliza somente regras específicas nas quais todas as variáveis são relevantes;
- base de regras genérica (BRG): utiliza também as regras genéricas nas quais existem variáveis irrelevantes.

### 2.4.1.3 Inferência escalonada

O método conhecido como inferência escalonada combina cada proposição atômica do fato (variável de entrada) com a proposição associada no antecedente da regra (PEDRYCZ; GOMIDE, 1998; DELGADO, 2002).

A figura 2.4 apresenta um exemplo da inferência escalonada para um sistema *fuzzy* com duas regras e dois antecedentes, onde as entradas são conjuntos *fuzzy*. Convém salientar que no caso mais usual, as entradas do sistema de inferência são *singletons* (figura 2.2) mas no exemplo considerado, as entradas são conjuntos *fuzzy*, pois este modelo particular será utilizado como base para a proposição da inferência escalonada *fuzzy* geral do tipo-2, no capítulo 3. O esquema geral para a inferência escalonada é dado pelos seguintes passos:

**Passo 1** *Matching* das entradas e dos antecedentes: para cada regra, computa-se o nível de similaridade (*matching*) entre cada proposição atômica do antecedente da regra e a proposição atômica correspondente no fato (ou proposição de entrada). Na figura 2.4, por exemplo,  $\mu_{m_1}^1$  é o *matching* da entrada  $A_1$  com o antecedente  $A_1^1$  (primeiro antecedente da primeira regra).

**Passo 2** Agregação dos Antecedentes: para cada regra, computa-se o grau de ativação através da operação de:

- conjunção para regras com proposições atômicas ligadas por E

- disjunção para regras com proposições atômicas ligadas por OU

Na figura 2.4,  $\mu_1$  é agregação de  $\mu_{m_1}^1$  e  $\mu_{m_2}^1$  através da *norma-t*  $t_2$  (operação E).

**Passo 3** Derivação da conclusão individual: para cada regra, computa-se o valor inferido, baseado no resultado da agregação dos antecedentes e na semântica da regra escolhida (e.g. *norma-t* (PEDRYCZ; GOMIDE, 1998)) A figura 2.4 mostra que a conclusão para regra 1 é  $B^{1*}$  e para a regra 2 é  $B^{2*}$ .

**Passo 4** Derivação do resultado final: computa-se o valor inferido do conjunto completo de regras através da agregação dos valores inferidos no passo 3. Pode-se ver na figura 2.4, que o resultado final para o exemplo com dois antecedentes e duas regras é  $B(y)$ , que foi obtido a partir da agregação de  $B^{1*}$  e  $B^{2*}$  pela *norma-s* = *max*.

O esquema de inferência escalonada simplifica muito a computação em sistemas de inferência *fuzzy* e, por isso, a maioria das aplicações práticas utiliza sistemas *fuzzy* com este mecanismo.

Nesta tese, pretende-se partir destes conceitos básicos para a formalização do esquema de inferência escalonado do tipo-2 (capítulo 3), o qual será a base do classificador digital do tipo-2 que está sendo proposto no capítulo 4.

#### 2.4.1.4 Defuzzificação

A *defuzzificação* é o processo que transforma a saída de um sistema *fuzzy*, obtida pela agregação das regras ativadas no processo de inferência, em valores *crisp* (valores abruptos ou não-*fuzzy*). Existem várias técnicas de *defuzzificação* para conjuntos *fuzzy* convencionais, tais como, a média dos máximos (MOM) e o centro de área (COA) (LEE, 1990).

Classificadores digitais (seção 2.3) podem ser construídos utilizando-se os sistemas de inferência *fuzzy* (seção 2.4.1). Geralmente, tais classificadores utilizam conjuntos *fuzzy* convencionais ou do tipo-1. Quando se utiliza um sistema de inferência como classificador, a saída desejada normalmente é um rótulo, correspondendo a uma das classes pré-definidas. Assim, pode não ser interessante utilizar as técnicas convencionais de *defuzzificação* para a obtenção de tais classes, uma vez que, sendo médias ponderadas, estas técnicas tendem a gerar classes diferentes das desejadas.

Em função disto, num sistema classificador com saída não-*fuzzy*, a *defuzzificação* passa a ser uma tomada de decisão, onde se atribui o *pixel* a apenas uma das classes previamente estabelecidas (SCHIEWE, 2003; SOLAIMAN; PIERCE; ULABY, 1999).

Suponha um caso bastante particular, que será importante quando se estiver projetando o classificador do tipo-2 (capítulo 4), onde existam  $L$  regras, uma para cada classe. Neste caso, um possível

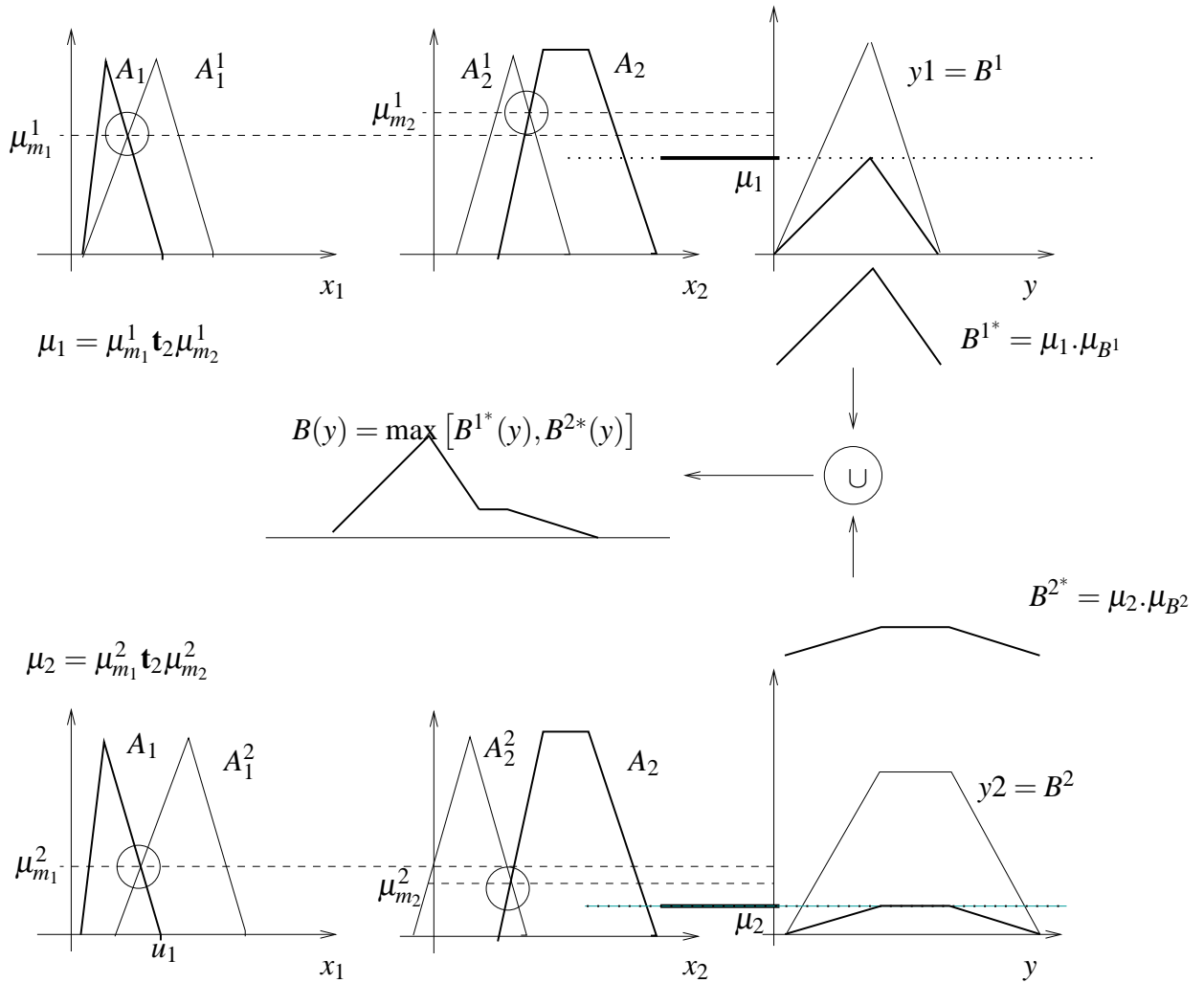


Figura 2.4: Exemplo da inferência escalonada para um sistema *fuzzy* com duas regras e dois antecedentes.

método de decisão considera que a classe escolhida  $l$  será aquela cujo valor de pertinência da relação de implicação *fuzzy* (regra *fuzzy*) aplicada no ponto  $(\mathbf{X}, Y)$  seja a máxima dentre todas as  $j$  regras, ou seja,

$$\text{SE } \mu_{F_1 \times \dots \times F_p \rightarrow G^l}(\mathbf{X}, Y) = \max_j (\mu_{F_1 \times \dots \times F_p \rightarrow G^j}(\mathbf{X}, Y)) \text{ ENTÃO } \mathbf{X} \text{ é } C_l \quad (2.10)$$

onde  $C_l$  é o rótulo da classe escolhida e  $\mathbf{X}$  é o vetor de elementos de entrada (*pixel* que se deseja classificar).

## 2.4.2 Sistema de inferência *fuzzy* do tipo-2

Ao contrário de um sistema de inferência tradicional, como aquele apresentado na seção 2.4.1, que utiliza conjuntos *fuzzy* convencionais, os sistemas de inferência do tipo-2 (TAN; FOO; CHUA, 2007) utilizam conjuntos do tipo-2 (KARNIK; MENDEL; LIANG, 1999). A partir deste ponto,

os conjuntos convencionais serão referenciados como tipo-1 para que se possa diferenciá-los dos tipo-2, mais complexos.

A estrutura típica de um sistema do tipo-2 (*T2 FIS: type-2 fuzzy inference system*) pode ser vista na figura 2.5. Notam-se os seguintes blocos funcionais: o “Fuzzificador”, responsável por converter a entrada abrupta (*crisp*) em um conjunto *fuzzy*; a “Inferência”, responsável por processar o conjunto de entrada, obtendo o tipo-2 de saída; a “Base de regras”, que diz como a entrada se relaciona com a saída; o “Redutor de tipo”, que transforma um conjunto tipo-2 em um tipo-1 e o “Defuzzificador”, que obtém uma saída abrupta a partir do conjunto tipo-1.

Convém mencionar que, de forma semelhante à mostrada na figura 2.2, o usual é que o “Fuzzificador” produza um *singleton* em sua saída. Na figura 2.5, porém, está sendo considerado um conjunto *fuzzy* à saída do “Fuzzificador” pois este modelo servirá de base para o desenvolvimento de um classificador do tipo-2 que trabalhe com entrada *fuzzy*.

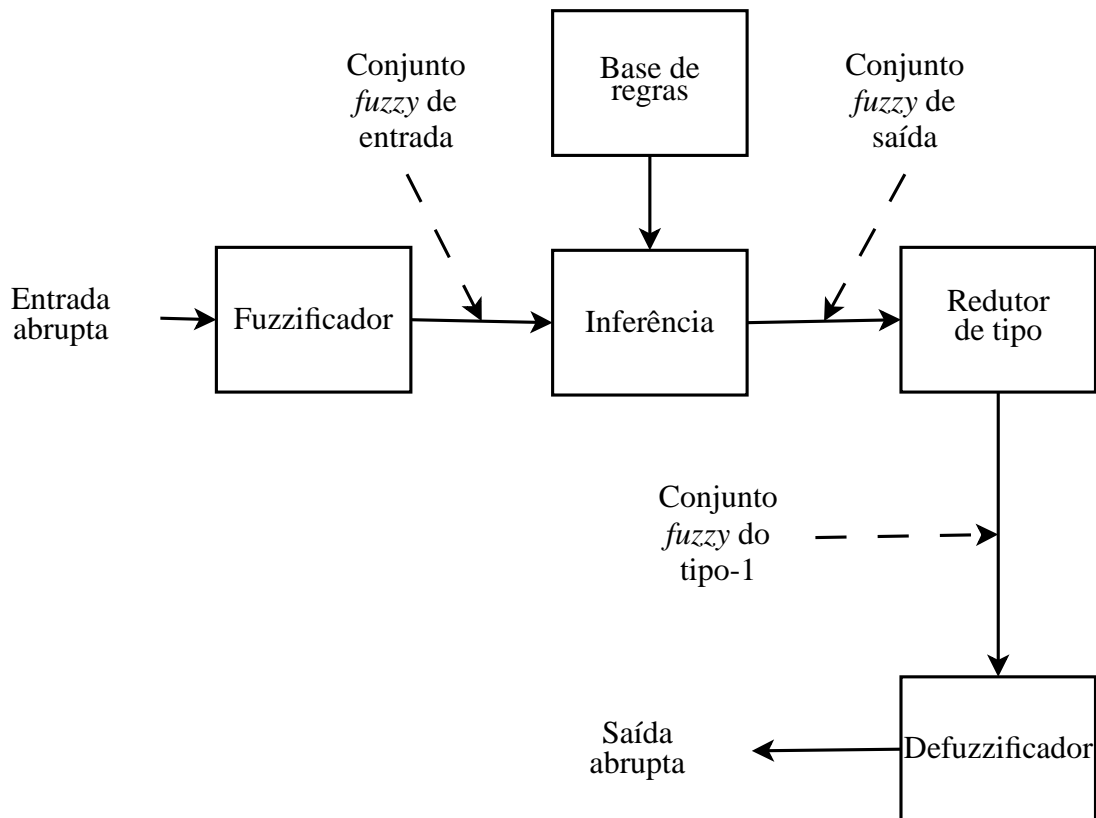


Figura 2.5: Estrutura típica de um sistema de inferência *fuzzy* do tipo-2.

Admita que este *T2 FIS* tenha  $p$  variáveis de entrada  $x_1 \in X_1, x_2 \in X_2, \dots, x_p \in X_p$ , e uma variável de saída  $y \in Y$ . Suponha que sua base de regras seja composta de  $L$  regras, cuja  $l$ -ésima regra seja dada pela equação (2.11)

$$R^l : \text{SE } x_1 \text{ é } \tilde{F}_1^l \text{ E } x_2 \text{ é } \tilde{F}_2^l \text{ E } \dots \text{ E } x_p \text{ é } \tilde{F}_p^l, \text{ ENTÃO } y \text{ é } \tilde{G}^l, \quad (2.11)$$

onde  $\tilde{F}_1^l, \tilde{F}_2^l \dots \tilde{F}_p^l$  são os antecedentes desta regra e  $\tilde{G}^l$  é o consequente. O *til* colocado sobre os nomes dos conjuntos é a simbologia convencionada para indicar que estes são do tipo-2.

A seguir são apresentadas as seguintes definições que auxiliam a compreensão da inferência do tipo-2: conjunto do tipo-2, fatia vertical e função de pertinência principal.

### 2.4.3 Conjunto *fuzzy* do tipo-2

Os conjuntos do tipo-2 foram apresentados pela primeira vez por Zadeh (1975) como uma evolução da sua teoria dos conjuntos *fuzzy*. Posteriormente, tais conjuntos foram estudados com mais profundidade, e sua base teórica foi adequadamente estabelecida (KARNIK; MENDEL, 2001). A principal definição aceita atualmente é a de que os conjuntos do tipo-2 são aqueles cujos graus de pertinência são, eles próprios, conjuntos do tipo-1. Formalmente, um conjunto do tipo-2, denotado por  $\tilde{A}$ , pode ser definido no universo  $X$  como

$$\tilde{A} = \int_{x \in X} \int_{u \in J_x} \mu_{\tilde{A}}(x, u) / (x, u), \quad (2.12)$$

onde  $J_x \subseteq [0, 1]$  é o conjunto dos graus de pertinência primária de  $x \in X$ , com  $u \in J_x, \forall x \in X$ , e  $\mu_{\tilde{A}}(x, u)$  é a função de pertinência do tipo-2 (MENDEL; JOHN, 2002; CHEN et al., 2006). Como se está considerando o caso mais geral, onde o conjunto é contínuo, utiliza-se o símbolo  $\int \int$  para denotar a união de todos os  $x$  e  $u$  admissíveis. Se o conjunto fosse discreto, substituir-se-ia  $\int \int$  por  $\Sigma \Sigma$  (MENDEL; JOHN, 2002).

Diferentemente dos conjuntos do tipo-1, onde a pertinência é um número real no intervalo  $[0, 1]$ ; nos tipo-2, a pertinência é um conjunto do tipo-1. Também se diz que os do tipo-2 possuem duas pertinências, a primária e a secundária, e, portanto, a natureza de tais conjuntos é tridimensional, ou seja, suas funções de pertinência são definidas no  $\mathfrak{R}^3$  (PAREEK; KAR, 2006; JOHN; MENDEL; CARTER, 2006) dificultando, sobremaneira, o esboço, o manuseio e o entendimento destes (MENDEL; JOHN, 2002). Em suma, um conjunto *fuzzy* do tipo-2 possui um grau de liberdade a mais do que o do tipo-1 (a terceira dimensão) e, portanto, tem o potencial para carregar mais informação do que o conjunto *fuzzy* tradicional (KARNIK; MENDEL; LIANG, 1999).

Atualmente há duas linhas principais de pesquisa envolvendo conjuntos do tipo-2: os conjuntos intervalares (MENDEL, 2000) e os conjuntos gerais (COUPLAND; JOHN, 2007b; LUCAS; CENTENO; DELGADO, 2007). Os gerais são aqueles que seguem estritamente a definição apresentada acima, não se impondo nenhuma restrição na construção de tais conjuntos. Os intervalares são simplificações dos gerais, onde se admite que as funções de pertinência secundária são constantes e

iguais a um. Tais simplificações reduzem significativamente o custo computacional das operações envolvendo conjuntos do tipo-2 e permitiram, ao longo dos anos, o surgimento de uma série de aplicações envolvendo conjuntos do tipo-2 (LIANG; MENDEL, 2000a, 2000b; KARNIK; MENDEL; LIANG, 1998).

#### 2.4.4 Fatia vertical

De acordo com John, Mendel e Carter (2006), um conjunto do tipo-2 pode ser seccionado por planos, conhecidos como fatias verticais, que passam por todo  $x \in X$ . Se o conjunto for contínuo, o número de planos a serem considerados depende do grau de discretização desejado, já que seria inviável computacionalmente trabalhar com infinitas fatias.

Quando se toma um valor particular  $x' \in X$ , a equação (2.12) pode ser reescrita como

$$\begin{aligned} \int_{x \in X} \int_{u \in J_{x'}} \mu_{\tilde{A}}(x = x', u) / (x = x', u) \\ = \int_{u \in J_{x'}} \mu_{\tilde{A}}(x', u) / (x', u). \end{aligned} \quad (2.13)$$

o que produz

$$\int_{u \in J_{x'}} \mu_{\tilde{A}}(x', u) / (x', u) = \int_{u \in J_{x'}} f_{x'}(u) / u / x', \quad (2.14)$$

onde

$$\mu_{\tilde{A}}(x = x', u) \equiv \mu_{\tilde{A}}(x') = \int_{u \in J_{x'}} f_{x'}(u) / u \quad (2.15)$$

é a fatia vertical de  $\mu_{\tilde{A}}(x, u)$  em  $x = x'$  (JOHN; MENDEL; CARTER, 2006; TAHAYORI; TET-TAMANZI; ANTONI, 2006),  $J_{x'} \subseteq [0, 1]$  é o conjunto dos graus de pertinência primária de  $x'$  e  $f_{x'}(u)$ ,  $0 \leq f_{x'}(u) \leq 1$ , é uma função  $f$  do grau de pertinência primária  $u$ , que identifica os graus de pertinência secundária de  $x'$  em  $\tilde{A}$  (MENDEL; JOHN, 2002).

Um conjunto do tipo-2 pode ser reconstruído a partir de suas fatias verticais por meio da união destas, conforme mostra a equação (2.16) (MENDEL; JOHN, 2002)

$$\tilde{A} = \int_{x \in X} \mu_{\tilde{A}}(x) / x = \int_{x \in X} \int_{u \in J_x} f_x(u) / u / x. \quad (2.16)$$

o que implica que, para cada  $x$  em particular ( $x'$ ), podem ser associados muitos graus de pertinência primária  $u_1, u_2, u_3, \dots, u_{n_A}$ , onde  $n_A$  é a cardinalidade do suporte de  $\mu_{\tilde{A}}(x')$ ; contudo, a cada grau primário, pode-se associar apenas um grau secundário,  $a_1 = f_{x'}(u_1), a_2 = f_{x'}(u_2), a_3 = f_{x'}(u_3) \dots a_{n_A} = f_{x'}(u_{n_A})$ .

## 2.4.5 Função de pertinência principal

Aqueles graus de pertinência primária, cujos graus secundários são iguais a um, recebem o nome especial de “graus principais” e, no caso particular onde, para cada  $x_i$ , houver apenas um grau principal, tem-se a “função de pertinência principal” (*PMF: Principal Membership Function*) (LIANG; MENDEL, 2001; KARNIK; MENDEL, 2001), que é um conjunto do tipo-1 dado pela equação (2.17).

$$\mu_{\tilde{A}_{princ}} = \mu_A = \begin{cases} u & \forall x \mid f_x(u) = 1 \\ 0 & \text{caso contrário} \end{cases} \quad (2.17)$$

Considere o conjunto discreto  $\tilde{A}_D =$

$$\begin{aligned} & (1/0 + 0,8/0,25 + 0,6/0,5 + 0,4/0,75 + 0,2/1)/1 + \\ & \quad (1/0,5)/1,5 + \\ & (0,2/0 + 0,4/0,25 + 0,6/0,5 + 0,8/0,75 + 1/1)/2 + \\ & \quad (1/0,5)/2,5 + \\ & (1/0 + 0,7/0,25 + 0,5/0,5 + 0,3/0,75 + 0,1/1)/3 \end{aligned}$$

que está definido no universo  $X = \{1, 1,5, 2, 2,5, 3\}$ , conforme apresentado na figura 2.6.

Na figura 2.7, o mesmo conjunto foi desenhado de maneira ligeiramente diferente, rotacionando-se o eixo de coordenadas, de forma que o desenho fique com o plano  $x \times u$  no papel e que o terceiro eixo ( $f(u)$ ) seja representado saindo do papel. Esta forma diferente de se desenhar o conjunto do tipo-2 foi utilizada para que fique semelhante ao GEFOU, que será apresentado oportunamente (seção 3.2).

No gráfico da figura 2.7 a intensidade dos níveis de cinza representa o grau de pertinência secundária (níveis mais escuros representam pertinência secundária maior). Pode-se ver que  $\tilde{A}_D$  é composto de cinco fatias verticais, definidas no plano  $u \times f(u)$ ,  $\mu_{\tilde{A}}(x_i) = \sum f_{x_i}(u)/u$ :

$$\begin{aligned} \mu_{\tilde{A}}(1) &= 1/0 + 0,8/0,25 + 0,6/0,5 + 0,4/0,75 + 0,2/1,0, \\ \mu_{\tilde{A}}(1,5) &= 1,0/0,5, \end{aligned}$$

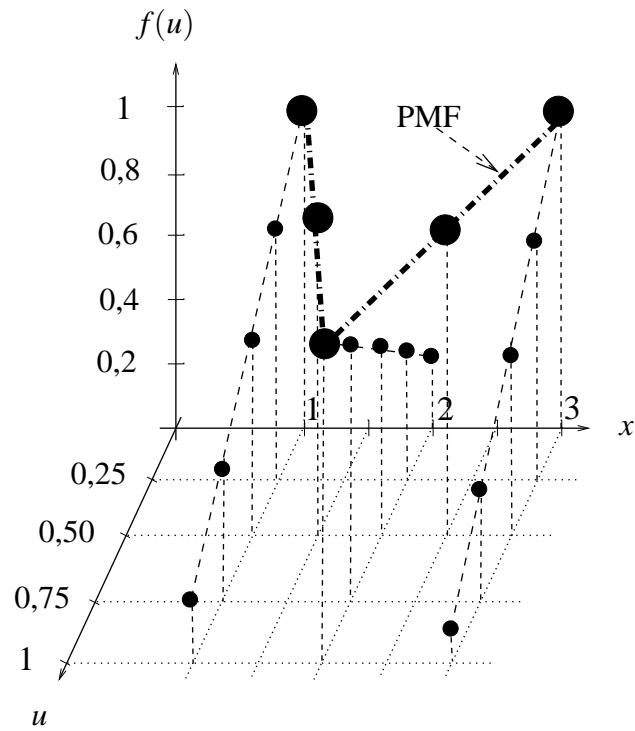


Figura 2.6: Um conjunto *fuzzy* discreto do tipo-2.

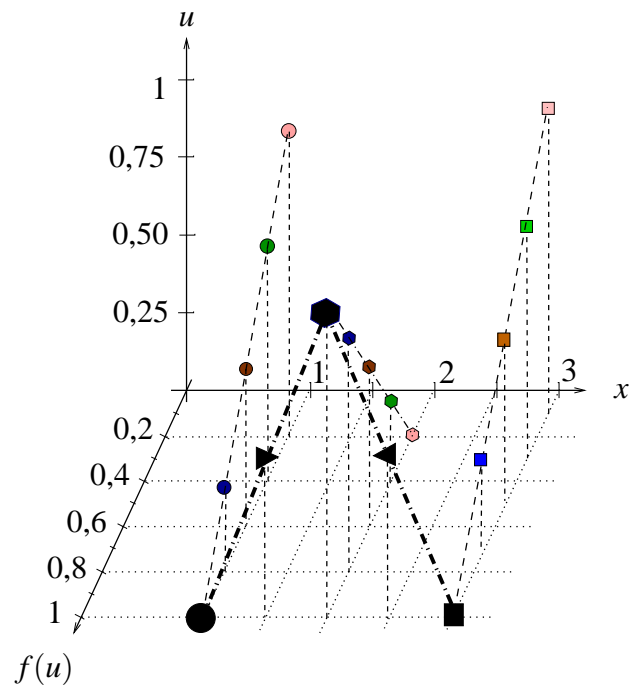


Figura 2.7: Um conjunto *fuzzy* discreto do tipo-2 rotacionado.



$$\begin{aligned}\mu_{\tilde{A}}(2) &= 0,2/0,0 + 0,4/0,25 + 0,6/0,5 + 0,8/0,75 + 1,0/1,0, \\ \mu_{\tilde{A}}(2,5) &= 1,0/0,5 \text{ e} \\ \mu_{\tilde{A}}(3) &= 1,0/0 + 0,7/0,25 + 0,5/0,5 + 0,3/0,75 + 0,1/1,0.\end{aligned}$$

Neste caso, a PMF de  $\tilde{A}_D$  é o conjunto do tipo-1, definido no plano  $x \times u$ ,  $A = \sum u/x_i = 0/1 + 0,5/1,5 + 1/2 + 0,5/2,5 + 0/3$ , correspondendo aos pontos mais escuros onde em  $f(u) = 1$ .

## 2.4.6 Operações sobre conjuntos do tipo-2

Considere dois conjuntos *fuzzy* do tipo-2,  $\tilde{A}$  e  $\tilde{B}$ , ambos definidos no universo  $X$ . Operações sobre estes conjuntos são facilmente calculadas utilizando-se suas fatias verticais  $\mu_{\tilde{A}}(x)$  e  $\mu_{\tilde{B}}(x)$  respectivamente, que são muito práticas do ponto de vista da implementação (MENDEL; JOHN, 2002; INNOCENT; JOHN, 2002).

Seja  $\mu_{\tilde{A}}(x)$  e  $\mu_{\tilde{B}}(x)$  representadas, para cada  $x$ , como  $\mu_{\tilde{A}}(x) = \int_u f_x(u)/u$  e  $\mu_{\tilde{B}}(x) = \int_w g_x(w)/w$ , onde  $u, w \in J_x \subseteq [0, 1]$  são os graus de pertinência primários de  $\tilde{A}$  e  $\tilde{B}$ , e  $f_x(u)$  e  $g_x(w)$  são os graus de pertinência secundários de  $\tilde{A}$  e  $\tilde{B}$ , respectivamente. De acordo com Karnik e Mendel (2001), é possível definir algumas operações binárias sobre estes conjuntos do tipo-2, como a união e a interseção, apresentadas a seguir.

### 2.4.6.1 União de dois conjuntos *fuzzy* do tipo-2

A união de dois conjuntos *fuzzy* do tipo-2  $\tilde{A}$  e  $\tilde{B}$  é um novo conjunto do tipo-2  $\tilde{A} \cup \tilde{B}$  que pode ser representado como

$$\tilde{A} \cup \tilde{B} = \int_{x \in X} \mu_{\tilde{A} \cup \tilde{B}}(x)/x, \quad (2.18)$$

onde  $\mu_{\tilde{A} \cup \tilde{B}}(x)$  é uma fatia vertical de  $\tilde{A} \cup \tilde{B}$  definida em um valor particular  $x \in X$ . Como proposto por Karnik e Mendel (2001), a fatia vertical  $\mu_{\tilde{A} \cup \tilde{B}}(x)$  pode ser obtida pela união de todas as fatias verticais,  $\mu_{\tilde{A}}(x)$  e  $\mu_{\tilde{B}}(x)$ , definidas no mesmo  $x \in X$ , como mostrado pela equação (2.19):

$$\begin{aligned}\mu_{\tilde{A} \cup \tilde{B}}(x) &= \mu_{\tilde{A}}(x) \sqcup \mu_{\tilde{B}}(x) \\ &= \int_u \int_w f_x(u) \star g_x(w) / u \vee w,\end{aligned} \quad (2.19)$$

onde  $\sqcup$  é o operador *join*,  $\star$  é uma *norma-t* e  $\vee$  é a *norma-s max* (KLEMENT; MESIAR; PAP, 2000). Nesta tese adota-se o *min* ( $\wedge$ ) como *norma-t*.

Combinando as equações 2.18 e 2.19 pode-se obter a equação (2.20), que é a expressão final para se calcular a união de dois conjuntos *fuzzy* do tipo-2,  $\tilde{A}$  e  $\tilde{B}$ , a partir de suas fatias verticais  $\mu_{\tilde{A}}$  e  $\mu_{\tilde{B}}$  (LUCAS; CENTENO; DELGADO, 2007).

$$\tilde{A} \cup \tilde{B} = \int_{x \in X} \int_u \int_w f_x(u) \wedge g_x(w) / u \vee w / x. \quad (2.20)$$

Aspectos computacionais relacionados ao cálculo da união envolvem a discretização dos conjuntos *fuzzy* contínuos do tipo-2. Logo, para cada  $x_i$ ,  $i = 1 \cdots N$ , combina-se cada  $u_k$ ,  $k = 1 \cdots n_A$ , com cada  $w_j$ ,  $j = 1 \cdots n_B$ , através do operador *max*, onde  $N$  é a cardinalidade de  $X$ ,  $n_A$  e  $n_B$  são a cardinalidade do suporte de  $\mu_{\tilde{A}}(x_i)$  e  $\mu_{\tilde{B}}(x_i)$ , respectivamente. Adicionalmente, cada  $f_x(u_k)$  é combinado com cada  $g_x(w_j)$  através do operador *min*.

Como discutido por Karnik e Mendel (2001), se mais de um cálculo de  $u_k \vee w_j$  produzir o mesmo ponto  $(u_k \vee w_j)^*$ , utiliza-se aquele que tiver o maior grau de pertinência secundária  $[f_x(u_k) \wedge g_x(w_j)]_{\max} / (u_k \vee w_j)^*$ .

#### 2.4.6.2 Interseção de dois conjuntos *fuzzy* do tipo-2

De maneira semelhante ao que foi feito com a união, pode-se substituir o operador *join*  $\sqcup$  na equação (2.19) pelo operador *meet*  $\sqcap$  (KARNIK; MENDEL, 2001), obtendo-se a equação (2.21) que permite calcular a interseção de dois conjuntos *fuzzy* do tipo-2

$$\tilde{A} \cap \tilde{B} = \int_{x \in X} \int_u \int_w f_x(u) \star g_x(w) / u \star w / x. \quad (2.21)$$

ou seja

$$\tilde{A} \cap \tilde{B} = \int_{x \in X} \int_u \int_w f_x(u) \wedge g_x(w) / u \wedge w / x. \quad (2.22)$$

já que a *norma-t* utilizada é o *min*.

Da mesma forma que na união, se mais de um cálculo de  $u_k \wedge w_j$  resultar no mesmo ponto  $(u_k \wedge w_j)^*$ , toma-se aquele que tiver o maior grau de pertinência.

### 2.4.6.3 Redução de tipo

Na seção 2.4.1.4 foi apresentada a *defuzzificação* de um conjunto *fuzzy* do tipo-1 que permite obter um valor não-*fuzzy* a partir do conjunto tipo-1. Pode-se dizer que a *defuzzificação* reduz o tipo do conjunto, transformando o tipo-1 em um número real (tipo-0).

Em conjuntos do tipo-2 o mesmo também se aplica, ou seja, é possível reduzir tais conjuntos de tipo-2 para tipo-1 e de tipo-1 para tipo-0, se for necessário. Vale comentar que toda redução, embora muitas vezes seja uma operação necessária (e.g.: *defuzzificação*), representa perda de informação.

Karnik, Mendel e Liang (1999) apresentaram a técnica tradicional para reduzir um conjunto *fuzzy* do tipo-2 que foi denominada “Redução de Tipo Baseada no Centróide” (CTR: *Centroid Type Reduction*).

A CTR é obtida através do cálculo do centróide de um conjunto do tipo-2, o que leva a um conjunto do tipo-1 chamado “conjunto do tipo reduzido”, i.e., o centróide de um conjunto do tipo-2 é um do tipo-1.

Seja  $\tilde{B}$  um conjunto *fuzzy* do tipo-2 e  $V_{x_j} = \{\theta_i/u_i\}, i = 1 \cdots d$ , uma fatia vertical de  $\tilde{B}$ , em  $x = x_j \in X$ , onde  $d$  é o número total de elementos de cada  $V_{x_j}$ ,  $j = 1 \cdots N$ . Considere toda possível combinação  $U_k = \{u_1^k, \dots, u_N^k\}$ ,  $k = 1, \dots, d^N$ , onde se toma exatamente um elemento  $u_j^k$  de cada  $V_{x_j}$ . Tal combinação pode ser vista como sendo a função de pertinência de um conjunto do tipo-1  $B'_k$ , denominado conjunto embutido, que tem o mesmo domínio de  $\tilde{B}$ , e o centróide dado por  $C_{B'_k} = \frac{\sum_{j=1}^N x_j \cdot u_j^k}{\sum_{j=1}^N u_j^k}$ .

Portanto, o conjunto do tipo reduzido, que representa o centróide de  $\tilde{B}$ , é um conjunto do tipo-1  $C_{\tilde{B}} = \{\phi_k/C_{B'_k}\}$ , com  $k = 1 \cdots d^N$ , cujo suporte é dado pelos centróides  $C_{B'_k}$  de todos os conjuntos embutidos  $B'_k$ . Considerando a combinação  $\Theta_k = \{\theta_1^k, \dots, \theta_N^k\}$ , onde se toma exatamente um elemento  $\theta_j^k$  de cada  $V_{x_j}$ , os graus de pertinência de  $C_{\tilde{B}}$  são calculados como  $\phi_k = \{\theta_1^k \mathbf{t} \theta_2^k \mathbf{t} \cdots \mathbf{t} \theta_N^k\}$ , onde  $\mathbf{t}$  é uma *norma-t*. Se mais de uma combinação resultar no mesmo ponto  $(C_{B'_k})^*$ , utiliza-se aquele que tiver maior grau de pertinência. Isto pode ser verificado através dos valores em negrito na equação 2.23.

Para um exemplo do cálculo da CTR considere o conjunto do tipo-2 discreto

$$\tilde{B}_D = (0,8/0,25 + 0,6/0,5)/1 + (0,8/0,5 + 0,6/0,75)/2 + (0,2/0 + 0,4/0,2)/3$$

que tem três fatias verticais definidas em  $x_1 = 1$ ,  $x_2 = 2$  e  $x_3 = 3$ . Pode-se construir oito conjuntos embutidos do tipo-1:  $B'_1 = 0,25/1 + 0,5/2 + 0/3$ ,  $B'_2 = 0,25/1 + 0,5/2 + 0,2/3$ ,  $B'_3 = 0,25/1 + 0,75/2 + 0/3$ ,  $B'_4 = 0,25/1 + 0,75/2 + 0,2/3$ ,  $B'_5 = 0,5/1 + 0,5/2 + 0/3$ ,  $B'_6 = 0,5/1 +$

$0,5/2 + 0,2/3$ ,  $B'_7 = 0,5/1 + 0,75/2 + 0/3$  e  $B'_8 = 0,5/1 + 0,75/2 + 0,2/3$ , cujos centróides são, respectivamente,

$$\begin{aligned}
 C_{B'_1} &= \frac{1 \times 0,25 + 2 \times 0,50 + 3 \times 0,0}{0,25 + 0,50 + 0,0} = 1,6667, \\
 C_{B'_2} &= \frac{1 \times 0,25 + 2 \times 0,50 + 3 \times 0,2}{0,25 + 0,50 + 0,2} = 1,9474, \\
 C_{B'_3} &= \frac{1 \times 0,25 + 2 \times 0,75 + 3 \times 0,0}{0,25 + 0,75 + 0,0} = 1,7500, \\
 C_{B'_4} &= \frac{1 \times 0,25 + 2 \times 0,75 + 3 \times 0,2}{0,25 + 0,75 + 0,2} = 1,9583, \\
 C_{B'_5} &= \frac{1 \times 0,50 + 2 \times 0,50 + 3 \times 0,0}{0,50 + 0,50 + 0,0} = 1,5000, \\
 C_{B'_6} &= \frac{1 \times 0,50 + 2 \times 0,50 + 3 \times 0,2}{0,50 + 0,50 + 0,2} = 1,7500, \\
 C_{B'_7} &= \frac{1 \times 0,50 + 2 \times 0,75 + 3 \times 0,0}{0,50 + 0,75 + 0,0} = 1,6000, \\
 C_{B'_8} &= \frac{1 \times 0,50 + 2 \times 0,75 + 3 \times 0,2}{0,50 + 0,75 + 0,2} = 1,7931.
 \end{aligned}$$

O conjunto do tipo reduzido é, portanto,

$$\begin{aligned}
 C_{\tilde{B}} &= 0,8 \wedge 0,8 \wedge 0,2/1,6667 + 0,8 \wedge 0,8 \wedge 0,4/1,9474 + \\
 &0,8 \wedge 0,6 \wedge 0,2/1,7500 + 0,8 \wedge 0,6 \wedge 0,4/1,9583 + \\
 &0,6 \wedge 0,8 \wedge 0,2/1,5000 + 0,6 \wedge 0,8 \wedge 0,4/1,7500 + \\
 &0,6 \wedge 0,6 \wedge 0,2/1,6000 + 0,6 \wedge 0,6 \wedge 0,4/1,7931
 \end{aligned}$$

$$\begin{aligned}
 C_{\tilde{B}} &= 0,2/1,6667 + 0,4/1,9474 + 0,2/1,7500 + 0,4/1,9583 + \\
 &0,2/1,5000 + 0,4/1,7500 + 0,2/1,6000 + 0,4/1,7931
 \end{aligned} \tag{2.23}$$

$$\begin{aligned}
 C_{\tilde{B}} &= 0,2/1,6667 + 0,4/1,9474 + 0,4/1,9583 + 0,2/1,5000 + \\
 &0,4/1,7500 + 0,2/1,6000 + 0,4/1,7931.
 \end{aligned}$$

## 2.5 Trabalhos relacionados

Em Solaiman, Pierce e Ulaby (1999) foi proposto um classificador de cobertura vegetal que utiliza a inferência *fuzzy* do tipo-1 e que gera, além dos mapas temáticos, mapas de confiança, que permitem avaliar a qualidade da classificação efetuada.

Em Ishibuchi, Nakashima e Murata (1999) foi feito um estudo sobre o desempenho de classificadores *fuzzy* cujas regras foram obtidas utilizando-se algoritmos genéticos (AG). Neste trabalho,

cada regra da base de regras foi vista como sendo um indivíduo do AG e, portanto, tinha um valor de *fitness* associado. O autor comenta que a implementação do sistema era simples porque as funções de pertinência eram fixas mas, apesar disto, o desempenho obtido foi comparável ao de outros métodos de classificação como técnicas de aprendizado de máquina e redes neurais.

Ishibuchi e Yamamoto (2004) trabalharam com classificadores do tipo-1 mas utilizaram regras um pouco diferentes das apresentadas na equação 2.4.1.1, acrescentando peso a estas (número real no intervalo  $[0, 1]$ ). Tais regras foram obtidas a partir dos dados de treinamento criando um conjunto de regras e, a partir deste, foram selecionadas algumas para compor a base de regras. Diversas técnicas foram testadas para fazer a seleção das regras mas a que produziu melhores resultados utilizou algoritmos genéticos. Uma característica interessante das regras é que estas possuíam antecedentes irrelevantes (*don't care*). A classificação foi realizada de maneira semelhante à mostrada na equação 2.10, técnica referida pelo autor como *single winner-based method* (SWM) (método baseado num único vencedor). Foram utilizados quatro conjuntos de dados provenientes de uma base de dados conhecida (UCI ML) para testar o classificador.

Ishibuchi e Yamamoto (2005) comentam que, em geral, as regras utilizadas pelos classificadores *fuzzy* são de três tipos: 1) regras com um único conseqüente e sem peso, 2) com um único conseqüente e com peso e 3) com múltiplos conseqüentes. No artigo em questão foram utilizadas as regras do segundo tipo (um único conseqüente com peso) e fez-se uma comparação entre o método de um único vencedor (SWM) com o “método do voto ponderado” (WVM: *weighted vote method*). No voto ponderado, somam-se as contribuições de cada regra cujo conseqüente seja o mesmo, admitindo-se que exista mais de uma regra com o mesmo conseqüente (rótulo representando uma determinada classe). Vence a classe cuja soma seja a maior dentre todas as somatórias de todas as classes.

Também é possível construir classificadores que utilizam conjuntos do tipo-2 (seção 2.4.2). Ocorre que os classificadores do tipo-2 costumam utilizar conjuntos intervalares devido ao seu custo computacional menor do que o apresentado pelo conjuntos gerais do tipo-2. Liang e Mendel (2001), por exemplo, construíram um classificador do tipo-2 para tráfego de vídeo MPEG que possui uma base de regras semelhante à utilizada por classificadores do tipo-1, mas os antecedentes são conjuntos do tipo-2 ao invés de conjuntos *fuzzy* convencionais. Os conjuntos T2 são intervalares com função de pertinência primária Gaussiana com incerteza na média e no desvio-padrão. Os autores concluíram que os classificadores T2 eram substancialmente melhores do que os seus equivalentes do tipo-1 em termos de robustez e taxa de erros de classificação.

Coupland e John (2007a) utilizaram geometria computacional para propor uma nova representação dos conjuntos *fuzzy* do tipo-1 e do tipo-2. Também propuseram novos algoritmos para várias operações sobre estes conjuntos, incluindo a defuzzificação dos conjuntos do tipo-1 e dos conjuntos

intervalares do tipo-2.

Como exemplo de classificador do tipo-2 geral pode-se citar Mitchell (2005) que introduziu uma medida de similaridade entre dois conjuntos *fuzzy* do tipo-2 e que é calculada a partir da média ponderada das similaridades entre as funções de pertinência embutidas de cada conjunto (tais funções são obtidas através da amostragem dos conjuntos do tipo-2 em pontos específicos). Com esta nova medida mostrou-se que conjuntos do tipo-2 fornecem uma linguagem natural para formular problemas de classificação e, a partir disso, construiu-se um classificador que compara, através da nova medida, um padrão desconhecido com um conjunto de protótipos, conhecidos, obtidos através de amostras de treinamento. A comparação que produzir a maior similaridade define a classe do padrão desconhecido.

Zeng e Liu (2007) fizeram um estudo do estado da arte do reconhecimento de padrões utilizando conjuntos *fuzzy* do tipo-2. Neste estudo os autores citam, entre outros, os trabalhos de Liang e Mendel (2001) e Mitchell (2005) e afirmam que a incerteza pode estar presente na medição do sinal (e.g.: ruído) ou no nível semântico, devido a diferenças nos significados dos termos linguísticos, que pode variar de especialista para a especialista. Neste contexto os conjuntos T2 são bastante úteis pois permitem modelar as duas formas de incerteza. Abordagens híbridas, que empregam os sistemas *fuzzy* e a estatística, também são mencionadas.

## 2.6 Conclusão

Neste capítulo, foram vistos vários termos relacionados com o sensoriamento remoto, tais como, reflectância, assinatura espectral, valores do contador digital e resolução espacial. Também foram estudados os classificadores digitais supervisionados e as métricas que permitem fazer a validação deles. Foram apresentados os sistemas de inferência *fuzzy* do tipo-1 e do tipo-2 e definições importantes para o entendimento dos últimos: conjuntos do tipo-2, fatias verticais e função de pertinência principal. Finalmente, foi comentada a possibilidade de se construir classificadores *fuzzy* utilizando conjuntos do tipo-2 (T2). Vale destacar que os conjuntos do tipo-2 têm natureza tridimensional, enquanto que os do tipo-1 são bidimensionais, ou seja, os primeiros possuem um grau de liberdade a mais do que os últimos e podem, portanto, carregar mais informação. Da análise dos trabalhos relacionados citados anteriormente, conclui-se que a dimensão adicional permite que os sistemas do tipo-2 obtenham precisões maiores do que as obtidas pelos sistemas que utilizam exclusivamente conjuntos *fuzzy* convencionais (T1). Além disso, os conjuntos T2 admitem incerteza nos termos linguísticos, ou seja, pode-se trabalhar com especialistas com opiniões diferentes (o T2 permite a caracterização desta incerteza) enquanto que os T1, por trabalharem com pertinências no intervalo  $[0, 1]$ , não admitem situações de incerteza como estas.

# 3 *Proposta de Formalização do Sistema de Inferência Escalonada Fuzzy Geral do Tipo-2*

## 3.1 Introdução

Quando se propõe a trabalhar com conjuntos gerais do tipo-2, um autor precisa lidar com uma série de questões em aberto, tais como: a simplificação de algumas operações existentes (e.g. *defuzzificação*), a proposição de operações (e.g. supremo de conjunto geral e extensão cilíndrica de fatia vertical) e a definição de um mecanismo de inferência adequado.

A principal contribuição desta tese é a formalização de um mecanismo de inferência para sistemas *fuzzy* baseados em conjuntos do tipo-2 denominado “sistema de inferência escalonada *fuzzy* geral do tipo-2”. Para dar suporte a este mecanismo, foram propostas: uma ferramenta de visualização para conjuntos do tipo-2 gerais (a “mancha” de incerteza geral), a translação de uma fatia vertical, o supremo de um conjunto *fuzzy* geral do tipo-2, a extensão cilíndrica de uma fatia vertical e o *singleton* do tipo-2. Além disso, para que a inferência geral fosse possível era necessário diminuir seu custo computacional e, para isso, propôs-se também uma técnica de *defuzzificação* de conjuntos gerais baseada nos centróides das fatias verticais.

## 3.2 “Mancha” de Incerteza Geral

Conforme já destacado, é difícil entender e visualizar um conjunto do tipo-2 devido à sua natureza tridimensional. Então, para tornar mais fácil o entendimento deste tipo de conjunto, esta tese propõe que se projete o tipo-2 no plano, substituindo a terceira dimensão por tons de cinza, que fornecem uma indicação visual sobre os graus de pertinência secundária (LUCAS; CENTENO; DELGADO, 2007).

O desenho foi denominado “Mancha” de Incerteza Geral ou GEFOU (*General Footprint of Uncertainty*) e pode ser considerado uma extensão do FOU (*Footprint Of Uncertainty*), bastante

conhecido e utilizado por autores que trabalham com conjuntos *fuzzy* do tipo-2 (MENDEL; JOHN, 2002; KARNIK; MENDEL; LIANG, 1999; KARNIK; MENDEL, 2001). No FOU trabalha-se apenas com duas cores, preto e branco, o que não permite identificar os graus de pertinência secundária mas, apenas, os limites do conjunto. O FOU é bastante útil em conjuntos intervalares onde a pertinência secundária é sempre unitária. No caso dos conjuntos gerais, o FOU pode ser visto como uma projeção do GEFOU sobre o plano  $x \times u$ .

A figura 3.1 mostra o GEFOU de um conjunto contínuo do tipo-2 que guarda semelhanças com o conjunto discreto  $\tilde{A}_D$  (e.g.: ambos têm PMF's triangulares) apresentado na figura 2.7<sup>1</sup>. Regiões mais escuras representam valores de pertinências secundárias mais altos, áreas brancas representam graus secundários nulos e a fina linha preta representa a PMF (função de pertinência principal). Pode-se verificar que o conjunto tipo-2 mostrado tem uma PMF triangular e graus secundários que são máximos sobre a PMF, mas diminuem linearmente (ao longo das retas verticais paralelas ao eixo  $u$ ) conforme se afasta da PMF.

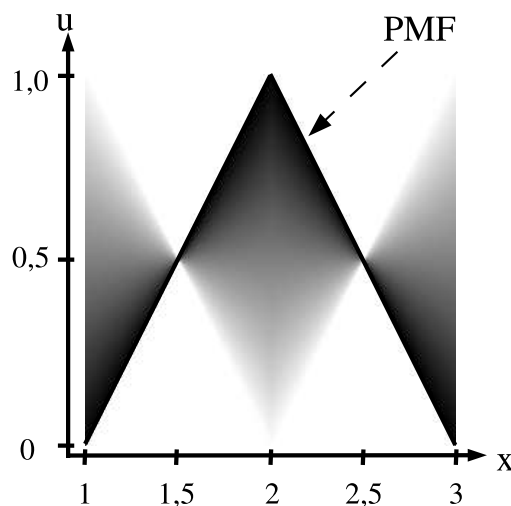


Figura 3.1: GEFOU de um conjunto *fuzzy* geral do tipo-2 com PMF triangular.

A equação (2.20) mostrou como calcular a união de dois conjuntos *fuzzy* gerais do tipo-2  $\tilde{A}$  e  $\tilde{B}$ . Suponha os seguintes conjuntos gerais  $\tilde{A}$  com  $\mu_{\tilde{A}_{princ}} = e^{-\frac{(x-a)^2}{2\sigma^2}}$  e  $\tilde{B}$  com  $\mu_{\tilde{B}_{princ}} = e^{-\frac{(x-b)^2}{2\sigma^2}}$ . A “mancha” de incerteza geral (GEFOU) permite visualizar os dois conjuntos e a união  $\tilde{A} \cup \tilde{B}$ , conforme se vê na figura 3.2 (a PMF Gaussiana pode ser identificada pela linha fina no centro de cada “mancha” de incerteza).

A interseção de dois conjuntos *fuzzy* gerais do tipo-2 foi definida na equação (2.22). Utilizando-se os mesmos conjuntos gerais  $\tilde{A}$  e  $\tilde{B}$  apresentados no exemplo da união, pode-se obter a “mancha”

<sup>1</sup>Vale destacar que o GEFOU, apesar de representar conjuntos contínuos do tipo-2, é uma imagem e, portanto, é desenhado com níveis discretos. No exemplo da figura, foram utilizados 256 elementos  $x_i \in X$ , 256 graus de pertinência primária  $u_k$  e 256 graus secundários  $f_x(u_k)$ .



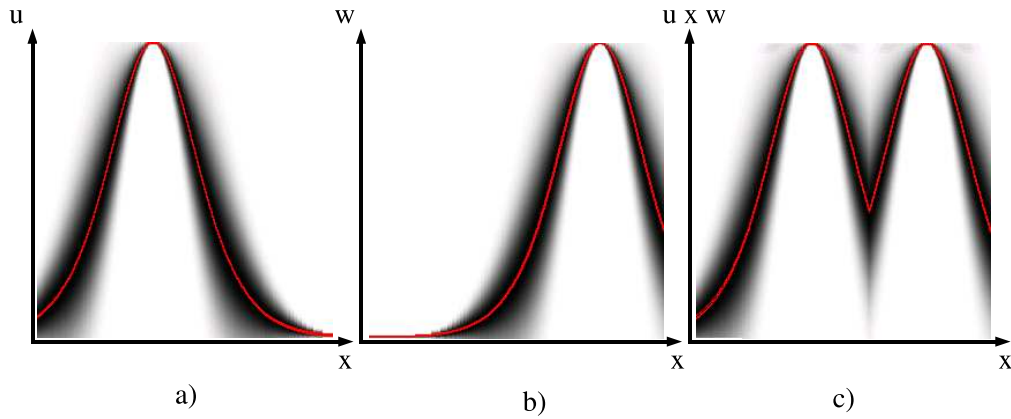


Figura 3.2: União de dois conjuntos *fuzzy* gerais do tipo-2. GEFOU de: a)  $\tilde{A}$ ; b)  $\tilde{B}$ ; c)  $\tilde{A} \cup \tilde{B}$ .

de incerteza geral para a interseção. As figuras 3.3(a) e 3.3(b) ilustram o GEFOU de  $\tilde{A}$  e  $\tilde{B}$ , respectivamente e a figura 3.3(c) mostra o GEFOU da interseção  $\tilde{A} \cap \tilde{B}$ .

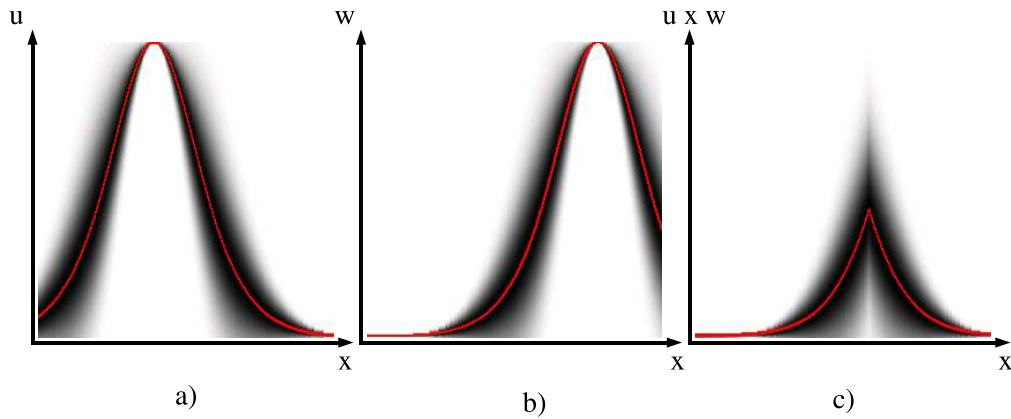


Figura 3.3: Interseção de dois conjuntos *fuzzy* gerais do tipo-2. GEFOU de: a)  $\tilde{A}$ ; b)  $\tilde{B}$ ; c)  $\tilde{A} \cap \tilde{B}$ .

### 3.3 Singleton do tipo-2

Nesta tese define-se um *singleton* do tipo-2  $\tilde{B}$  como (LUCAS; CENTENO; DELGADO, 2008b))

$$\tilde{B} = \int_{x \in X} \int_{u \in J_x} \mu_{\tilde{B}}(x, u) / (x = x', u = 1) = 1/1/x', \quad (3.1)$$

onde  $1/1/x'$  significa que o grau de pertinência primário é 1 e que o grau secundário também é 1 em  $x = x'$ .

Um *singleton* do tipo-2 definido em  $x = x'$  é portanto um ponto no espaço cujas coordenadas são  $x = x'$ ,  $u = 1$  e  $\mu(x, u) = 1$ . A figura 3.4(a) mostra o *singleton* do tipo-2 discutido até agora (desenhado em três dimensões) e a figura 3.4(b) mostra o GEFOU correspondente deste *singleton*.

É fácil verificar que o GEFOU simplificou o desenho do *singleton* ao trocar seu terceiro eixo,  $\mu(x, u) = 1$ , por um ponto preto, o que significa que o grau secundário é 1, de acordo com a definição do GEFOU (ver seção 3.2).

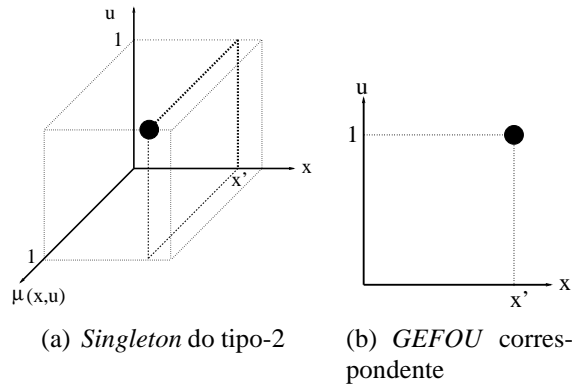


Figura 3.4: Um *singleton* do tipo-2 definido em  $x = x'$ .

### 3.4 Translação de fatia vertical

Uma fatia vertical  $\mu_{\tilde{A}}(x_1)$ , definida em  $x_1 \in X$ , pode ser transladada para uma nova posição  $x_2 \in X$ , simplesmente movendo-se a fatia para a nova posição, conforme mostra a equação (3.2) (LUCAS; CENTENO; DELGADO, 2007).

$$\text{transl}_{x_2 \in X} [\mu_{\tilde{A}}(x_1)] = \left[ \int_{u \in J_{x_1}} f_{x_1}(u)/u \right] / x_2 = \int_{u \in J_{x_1}} f_{x_1}(u)/u / x_2 \quad (3.2)$$

### 3.5 Supremo de um conjunto *fuzzy* geral do tipo-2

O supremo de um conjunto  $\tilde{A}$ , do tipo-2, denotado por  $\text{sup} [\tilde{A}]$ , é uma fatia vertical definida em  $x = x'$  e pode ser calculado aplicando-se a operação *join* (seção 2.4.6.1) unária  $\sqcup_{x \in X}$  sobre as fatias verticais que compõem o conjunto  $(\mu_{\tilde{A}}(x_1) \cdots \mu_{\tilde{A}}(x_N))$ , conforme a equação (3.3) (LUCAS; CENTENO; DELGADO, 2007).

$$\begin{aligned} \text{sup} [\tilde{A}] &= [\sqcup_{x \in X} \mu_{\tilde{A}}(x)] / x' = \\ &= \{ \cdots \{ \text{transl}_{x=x'} [\mu_{\tilde{A}}(x_1)] \sqcup \cdots \sqcup \text{transl}_{x=x'} [\mu_{\tilde{A}}(x_N)] \} \}. \end{aligned} \quad (3.3)$$

ou seja

$$\sup [\tilde{A}] = \int_{u \in J_{x_1}} f_{x_1}(u)/u/x' \sqcup \dots \sqcup \int_{u \in J_{x_N}} f_{x_N}(u)/u/x'. \quad (3.4)$$

### 3.6 Extensão cilíndrica de uma fatia vertical

Uma fatia vertical  $\mu_{\tilde{A}}(x_1)$  definida, por exemplo, em  $x_1 = 0$ , pode ser estendida cilíndricamente ao longo do eixo  $x$  através da cópia de  $\mu_{\tilde{A}}(x_1)$  ao longo deste eixo, como definido pela equação (3.5).

$$\text{cyl}_{x \in X} [\mu_{\tilde{A}}(x_1)] = \int_{x \in X} \left[ \int_{u \in J_{x_1}} f_{x_1}(u)/u \right] / x \quad (3.5)$$

A figura 3.5(a) mostra a “mancha” de incerteza de um conjunto do tipo-2  $\tilde{A}$  e a figura 3.5(b) mostra a “mancha” de incerteza do supremo deste conjunto. Finalmente, em 3.5(c), pode-se ver a “mancha” do supremo estendido cilíndricamente ao longo do eixo  $x$ .

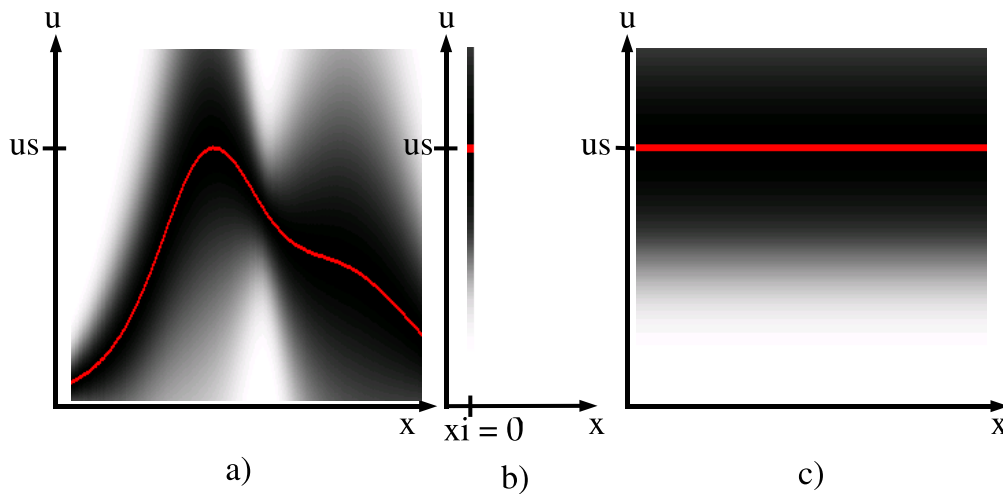


Figura 3.5: GEFOU de: a) Conjunto do tipo-2  $\tilde{A}$ ; b) Supremo  $\sup [\tilde{A}]$ ; c) Extensão cilíndrica  $\text{cyl}_{x \in X} [\mu_{\tilde{A}}(x_1 = 0)]$ .

### 3.7 Redução de tipo baseada no centróide das fatias verticais

Muitas vezes é necessário *defuzzificar* um conjunto *fuzzy*, seja este do tipo-1 ou do tipo-2. Neste último caso, é bastante conhecida a técnica proposta por Mendel, baseada no cálculo dos centróides dos conjuntos embutidos (CTR) (KARNIK; MENDEL; LIANG, 1999), descrita em detalhes na

seção 2.4.6.3. Ocorre que esta técnica, apesar de fornecer resultados exatos, consome muito tempo e é, por isso, de pouco uso em sistemas práticos (LUCAS; CENTENO; DELGADO, 2007).

Nos últimos tempos, vêm sendo propostas simplificações no processo de *defuzzificação* (COUPLAND; JOHN, 2006; MENDEL, 2000) que produzem resultados aproximados mas, em contrapartida, têm como vantagem a rapidez do cálculo. Nesta tese também é proposta uma técnica, denominada de “Redução de tipo baseada no centróide das fatias verticais” (*VSCTR: Vertical Slice Centroid Type Reduction*), onde se calculam apenas os centróides das fatias verticais ( $C_j$ ) de um conjunto geral do tipo-2 ( $\tilde{B}$ ). Estes centróides são, então, reinterpretados como sendo as pertinências de um conjunto do tipo-1 composto pelos elementos  $x \in X$ . Portanto, o conjunto do tipo-1  $C_{\tilde{B}}$ , proveniente da redução de tipo do conjunto do tipo-2  $\tilde{B}$ , pode ser obtido por

$$C_{\tilde{B}} = \int_{x \in X} C_j / x = \int_{x \in X} \frac{\int_{u \in J_x} u \cdot f_x(u)}{\int_{u \in J_x} f_x(u)} / x. \quad (3.6)$$

Para que fique clara a diferença entre a VSCTR, que está sendo proposta nesta tese, e a CTR, proposta por Karnik, Mendel e Liang (1999), considere um conjunto *fuzzy* do tipo-2 que possua três fatias verticais definidas em  $x_1$ ,  $x_2$ , e  $x_3$ . Admita que para  $x_1$  se tenham duas pertinências primárias  $u_1$  e  $u_2$ , para  $x_2$ , uma pertinência primária  $u_3$  e, finalmente, para  $x_3$ , novamente duas pertinências primárias  $u_1$  e  $u_2$ , conforme mostrado na figura 3.6 (o eixo de pertinência secundária  $\mu$  não está representado na figura).

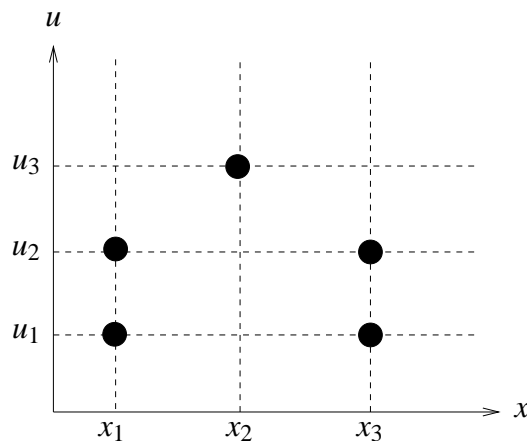


Figura 3.6: Conjunto *fuzzy* do tipo-2.

Para se obter a CTR, deve-se calcular os centróides dos conjuntos embutidos, e estes últimos são construídos tomando-se um único ponto em cada fatia vertical (seção 2.4.6.3). Na figura 3.7 nota-se que, para o exemplo considerado, foi possível construir quatro conjuntos embutidos (em linhas grossas tracejadas):  $e_1$ , que passa pelas coordenadas  $(x_1, u_1)$ ,  $(x_2, u_3)$ ,  $(x_3, u_1)$ ;  $e_2$ , que passa por  $(x_1, u_1)$ ,  $(x_2, u_3)$ ,  $(x_3, u_2)$ ;  $e_3$ , por  $(x_1, u_2)$ ,  $(x_2, u_3)$ ,  $(x_3, u_1)$  e  $e_4$ , por  $(x_1, u_2)$ ,  $(x_2, u_3)$ ,

$(x_3, u_2)$ . O centróide de  $e_1$  é  $c_1$ , o de  $e_2$  é  $c_2$ , de  $e_3$  é  $c_3$  e de  $e_4$  é  $c_4$ . O conjunto reduzido será formado por estes centróides e suas pertinências secundárias correspondentes, ou seja, este conjunto é representado no plano  $x \times \mu$  e, portanto, não pôde ser representado na figura, que mostra apenas o plano  $x \times u$ .

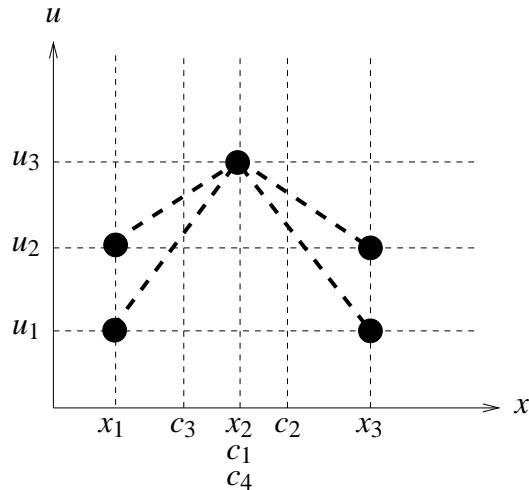


Figura 3.7: Exemplo de obtenção da CTR.

Na VSCTR calculam-se os centróides das fatias verticais, como mostrado na figura 3.8, onde  $c_1$  é o centróide da primeira fatia,  $c_2$  é o centróide da segunda e  $c_3$  é o da terceira. O conjunto reduzido para a VSCTR é definido no plano  $x \times u$  e pôde ser representado na figura (linha grossa cheia).

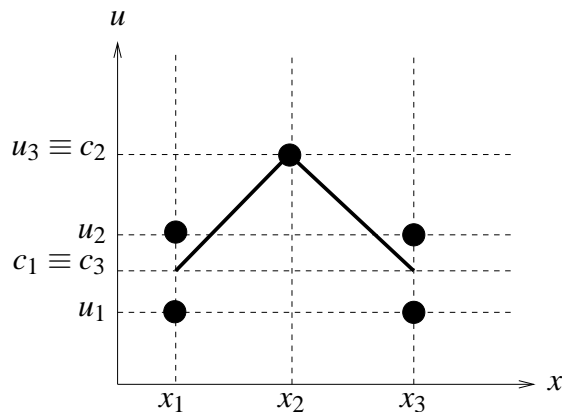


Figura 3.8: Exemplo de obtenção da VSCTR.

É interessante notar que os conjuntos reduzidos obtidos pela CTR e pela VSCTR são diferentes e definidos, inclusive, em planos diferentes, como visto. Mas a defuzzificação de tais conjuntos, visando obter um ponto  $x_i$ , pode produzir números semelhantes, dependendo da geometria de tais conjuntos.

Como exemplo numérico de cálculo da VSCTR, pode-se considerar o mesmo conjunto discreto  $\tilde{B}_D = (0,8/0,25 + 0,6/0,5)/1 + (0,8/0,5 + 0,6/0,75)/2 + (0,2/0 + 0,4/0,2)/3$  considerado na seção 2.4.6.3. Como este conjunto tem três fatias verticais, são calculados três centróides, um para cada fatia,  $C_1 = \frac{0,25 \times 0,8 + 0,5 \times 0,6}{0,8 + 0,6} = 0,3571$ ,  $C_2 = \frac{0,5 \times 0,8 + 0,75 \times 0,6}{0,8 + 0,6} = 0,6071$  e  $C_3 = \frac{0 \times 0,2 + 0,2 \times 0,4}{0,2 + 0,4} =$

0,1333. O conjunto do tipo reduzido, obtido a partir dos centróides das fatias verticais, é  $C_{\tilde{B}} = 0,3571/1 + 0,6071/2 + 0,1333/3$ . É importante comentar que o resultado obtido pela VSCTR geralmente é diferente do obtido pela CTR, produzindo conjuntos de tipo reduzido diferentes. Mas, apesar disso, os valores não-*fuzzy* obtidos pela *defuzzificação* dos dois conjuntos reduzidos são similares. Admitindo a *defuzzificação* pelo centróide (KARNIK; MENDEL; LIANG, 1999), o valor produzido pela VSCTR é  $y_c = \frac{1 \times 0,36 + 2 \times 0,61 + 3 \times 0,13}{0,36 + 0,61 + 0,13} = 1,7961$  que é bem similar ao produzido pela CTR, que é  $y_c = (1,6667 \times 0,2 + 1,9474 \times 0,4 + 1,9583 \times 0,4 + 1,5000 \times 0,2 + 1,7500 \times 0,4 + 1,6000 \times 0,2 + 1,7931 \times 0,4) / (0,2 + 0,4 + 0,4 + 0,2 + 0,4 + 0,2 + 0,4) = 1,7877$ .

Outro exemplo de cálculo da VSCTR, agora para um conjunto contínuo, está apresentado na figura 3.9(a), que mostra um conjunto do tipo-2  $\tilde{B}$  que foi reduzido em um conjunto do tipo-1  $C_{\tilde{B}}$  (figura 3.9(b)) utilizando-se os centróides das fatias verticais. Neste exemplo contínuo, a vantagem, em termos computacionais, da VSCTR em relação à CTR fica evidente, pois na VSCTR a quantidade de centróides a serem calculados é de  $m = N + 1$  vezes, onde  $N$  é o número de fatias verticais do conjunto *fuzzy* que se quer reduzir e na CTR, seriam necessários  $m = (\prod_{j=1}^N M_j) + 1$  cálculos de centróides, onde  $M_j$ ,  $j = 1 \dots N$ , é a cardinalidade do suporte de cada fatia vertical. No exemplo da figura 3.9 ( $N = 256$ ), a VSCTR necessitaria calcular apenas  $m = 256 + 1 = 257$  centróides mas a CTR precisaria calcular  $m = 256^{256} + 1!$  Para mais detalhes sobre a redução de tipo proposta, pode-se consultar (LUCAS; CENTENO; DELGADO, 2007).

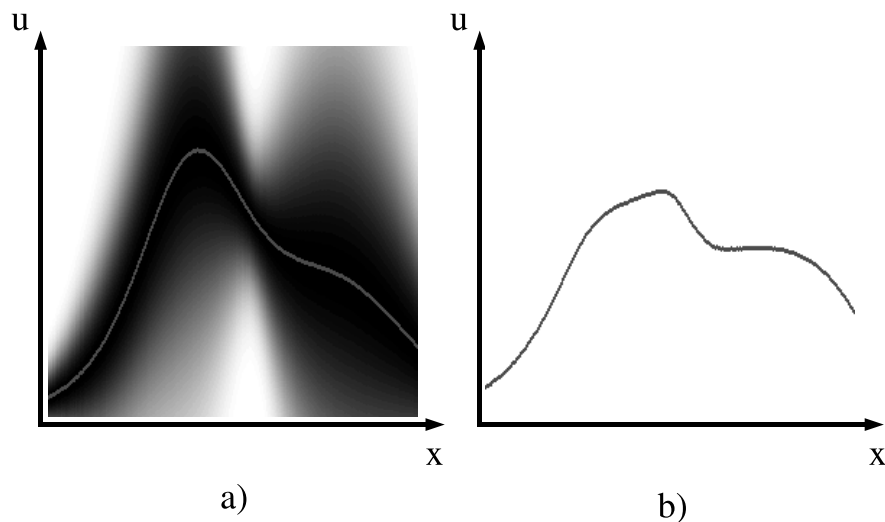


Figura 3.9: Redução de tipo de um conjunto do tipo-2. a) GEFOU de um conjunto do tipo-2  $\tilde{B}$ . b) Conjunto reduzido, do tipo-1,  $C_{\tilde{B}}$ .

### 3.8 Um sistema de inferência *fuzzy* geral do tipo-2

Embora os sistemas de inferência do tipo-2 já venham sendo estudados há algum tempo (MENDEL; LIU, 2006; TAN; FOO; CHUA, 2007), geralmente tais sistemas têm operado sobre conjuntos

intervalares, que são mais fáceis de se trabalhar e possuem custo computacional menor do que os conjuntos gerais (LIANG; MENDEL, 2000a, 2000b; KARNIK; MENDEL; LIANG, 1998).

Para estabelecer um mecanismo de inferência capaz de operar sobre conjuntos gerais do tipo-2, nesta tese buscou-se inspiração na inferência escalonada, usualmente utilizada em sistemas do tipo-1 (seção 2.4.1.3), criando-se um mecanismo que foi denominado “inferência escalonada *fuzzy* geral do tipo-2”.

A inferência geral do tipo-2 (LUCAS; CENTENO; DELGADO, 2008a, 2008b), de forma semelhante à do tipo-1, determina, portanto, que a saída inferida final deve ser obtida pelos seguintes passos: *matching* das entradas e dos antecedentes, agregação dos antecedentes, saída inferida, agregação das regras e *defuzzificação*, quando for necessária. Cada um dos passos é descrito em detalhes a seguir.

Considerando o fato  $x_1 \in F_1$ ,  $x_2 \in F_2, \dots$ ,  $x_p \in F_p$  e um sistema de inferência *fuzzy* com  $p$  variáveis de entrada  $(x_1, x_2, \dots, x_p)$ , uma variável de saída  $y$  e uma base de regras formada por  $L$  regras do tipo

$$R^l : \text{SE } x_1 \in \tilde{F}_1^l \text{ op } x_2 \in \tilde{F}_2^l \text{ op } \dots \text{ op } x_p \in \tilde{F}_p^l, \text{ ENTÃO } y \in \tilde{G}^l. \quad (3.7)$$

onde *op* é o operador de agregação dos antecedentes.

### 1. *Matching* das entradas e dos antecedentes

O *matching*  $\tilde{M}_i^l$  da entrada  $\tilde{F}_i$  com o antecedente da regra  $\tilde{F}_i^l$  pode ser calculado pela composição *sup-t* estendida (KARNIK; MENDEL; LIANG, 1999),  $\tilde{M}_i^l = \sup [\tilde{F}_i \cap \tilde{F}_i^l]$ , que, por sua vez, é composta de duas operações: a interseção da entrada com o antecedente e o supremo do conjunto resultante. Sabe-se, também, que a interseção  $\tilde{I}$ , de  $\tilde{F}_i$  e  $\tilde{F}_i^l$ , pode ser calculada pela operação *meet* (seção 2.4.6.2) como

$$\tilde{I} = \tilde{F}_i \cap \tilde{F}_i^l = \int_{x \in X} \int_u \int_w f_x(u) \wedge g_x(w) / u \wedge w / x. \quad (3.8)$$

onde  $u, w \in J_x \subseteq [0, 1]$  são os graus de pertinência primária de  $\tilde{F}_i$  e  $\tilde{F}_i^l$ , e  $f_x(u)$  e  $g_x(w)$  são os graus de pertinência secundária de  $\tilde{F}_i$  e  $\tilde{F}_i^l$ , respectivamente, calculados em cada  $x$ . O supremo de  $\tilde{I}$ , denotado por  $\sup [\tilde{I}]$ , é uma fatia vertical, definida em  $x = x'$ , e pode ser calculada utilizando-se a operação *join* unária  $\sqcup_{x \in X}$ , agrupando-se as fatias de  $\tilde{I}$  duas a duas (LUCAS; CENTENO; DELGADO, 2007)

$$\begin{aligned}
\sup [\tilde{I}] &= [\sqcup_{x \in X} \mu_{\tilde{I}}(x)] / x' = \\
&= \{ \cdots \{ \text{transl}_{x=x'} [\mu_{\tilde{I}}(x_1)] \sqcup \text{transl}_{x=x'} [\mu_{\tilde{I}}(x_2)] \} \\
&\quad \cdots \sqcup \text{transl}_{x=x'} [\mu_{\tilde{I}}(x_N)] \} \}
\end{aligned} \tag{3.9}$$

onde  $\mu_{\tilde{I}}(x_i)$  são as fatias verticais de  $\tilde{I}$ ,  $i = 1 \cdots N$ , e  $\text{transl}_{x=x'} [\mu_{\tilde{I}}(x_i)]$  é a translação da fatia vertical definida em  $x = x_i$  para uma nova posição  $x = x'$ . Como todas as fatias foram transladadas para a mesma posição  $x = x'$ , pode-se operá-las duas a duas.

A figura 3.10 mostra um exemplo de obtenção do *matching*  $\tilde{M}_i^l$  (em c) a partir da interseção  $\tilde{I}$  (em b) calculada a partir da entrada  $\tilde{F}_i$  com o antecedente  $\tilde{F}_i^l$  (em a).

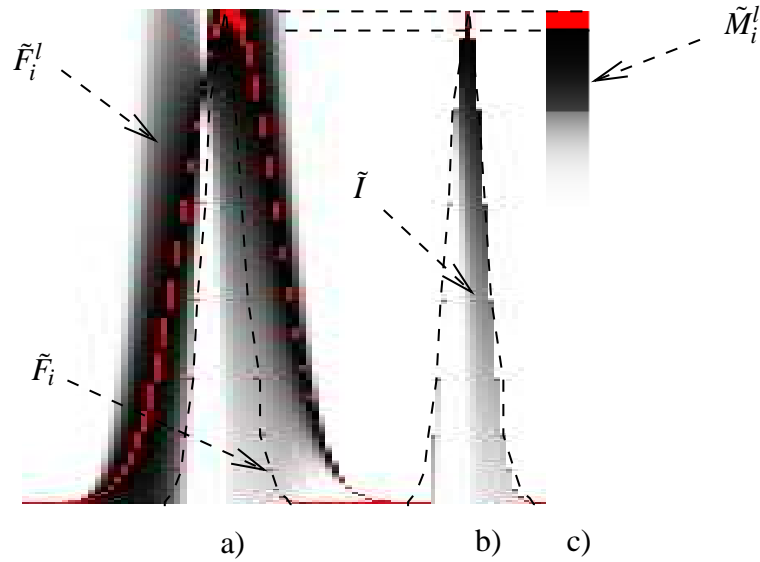


Figura 3.10: Exemplo de obtenção do *matching*  $\tilde{M}_i^l$  (em c) a partir da interseção  $\tilde{I}$  (em b) calculada a partir da entrada  $\tilde{F}_i$  com o antecedente  $\tilde{F}_i^l$  (em a).

## 2. Agregação dos antecedentes

A agregação dos antecedentes  $FS^l$ , também conhecida como grau de ativação da regra, mede a contribuição da  $l$ -ésima regra ao processo de inferência total (PEDRYCZ; GOMIDE, 1998). Considerando  $op = E$ , a agregação pode ser calculada pela interseção de todos os *matchings* desta regra, ou seja,  $FS^l = (\tilde{M}_1^l \cap \tilde{M}_2^l) \cap \cdots \cap \tilde{M}_p^l$ , que é uma interseção de fatias verticais definidas em  $x = x'$ :

$$\begin{aligned}
FS^l &= \left[ \int_u \int_v p_{x'}(u) \wedge q_{x'}(v) / u \wedge v / x' \right] \cap \cdots \cap \int_w r_{x'}(w) / w / x' \\
FS^l &= \int_{u \times v \times w} (p_{x'}(u) \wedge q_{x'}(v)) \wedge \cdots \wedge r_{x'}(w) / (u \wedge v) \wedge \cdots \wedge w / x'
\end{aligned} \tag{3.10}$$



onde  $u, v, w \in J_x \subseteq [0, 1]$  são os graus de pertinência primários de  $\tilde{M}_1^l$ ,  $\tilde{M}_2^l$  e  $\tilde{M}_p^l$ , respectivamente, e  $p_{x'}(u)$ ,  $q_{x'}(v)$  e  $h_{x'}(w)$  são os graus de pertinência secundários correspondentes, em  $x = x'$ .

### 3. Saída inferida de cada regra

Assumindo-se que a semântica da regra (PEDRYCZ; GOMIDE, 1998) seja dada pela *norma-t* = *min*, pode-se obter a saída  $\tilde{B}^l$ , inferida pela  $l$ -ésima regra, através: 1) da extensão cilíndrica do grau de ativação ao longo do eixo  $y$ ; e 2) do cálculo da interseção entre o conjunto tipo-2 resultante e o consequente da regra,  $\tilde{G}^l$ . Desta forma, a saída inferida  $\tilde{B}^l$  pode ser obtida pela equação (3.11).

$$\tilde{B}^l = \text{cyl}_{y \in Y} [FS^l] \cap \tilde{G}^l. \quad (3.11)$$

A extensão cilíndrica da fatia vertical  $FS^l$  ao longo do eixo  $y$  pode ser obtida pela cópia desta fatia ao longo deste eixo, como definido pela equação (3.12) (LUCAS; CENTENO; DELGADO, 2007).

$$\text{cyl}_{y \in Y} [FS^l] = \int_{y \in Y} \left[ \int_{u \in J_{y'}} f_{y'}(u)/u \right] / y \quad (3.12)$$

A figura 3.11 apresenta um exemplo de saída inferida  $\tilde{B}^l$  cujo consequente  $\tilde{G}^l$  tem uma PMF (função de pertinência principal) Gaussiana e o grau de ativação  $FS^l$  é um conjunto tipo-2 com a PMF constante dada por  $\mu_{A_{princ}} = 0.45 \quad \forall x \in X$ .

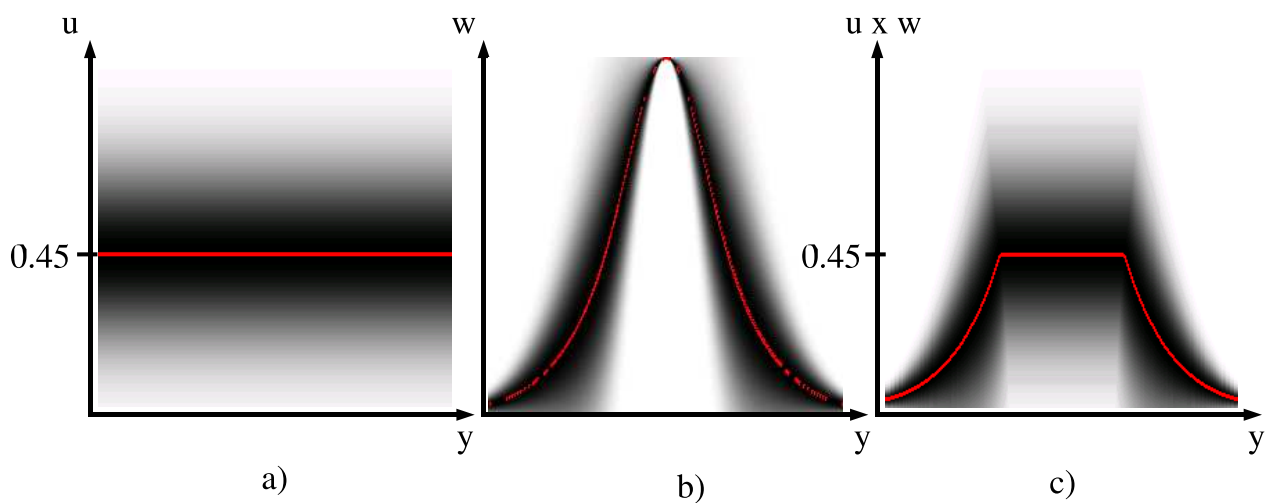


Figura 3.11: A  $l$ -ésima regra de um sistema de inferência do tipo-2. O GEFOU de: a) Extensão cilíndrica do grau de ativação ao longo do eixo  $y$ ; b) Consequente do tipo-2; c) Saída obtida da  $l$ -ésima regra.

### 4. Agregação das regras

A saída *fuzzy*  $\tilde{B}$ , inferida a partir de toda a base de regras, é calculada pela agregação das saídas  $\tilde{B}^l$ , obtidas por cada regra individualmente. A saída final é, portanto, calculada pela agregação das  $L$  saídas:  $\tilde{B} = \tilde{B}^1 \cup \tilde{B}^2 \cup \dots \cup \tilde{B}^L$ , sendo um conjunto do tipo-2.

Para uma exemplo com duas regras, com dois antecedentes cada, considere a figura 3.12, onde  $x_1$  é  $\tilde{F}_1$  E  $x_2$  é  $\tilde{F}_2$  é o fato (entrada do sistema de inferência),  $\tilde{M}_1^1 = \sup [\tilde{F}_1 \cap \tilde{F}_1^1]$  é o *matching* entre o primeiro antecedente da primeira regra ( $\tilde{F}_1^1$ ) e a primeira entrada ( $\tilde{F}_1$ ),  $\tilde{M}_2^1 = \sup [\tilde{F}_2 \cap \tilde{F}_2^1]$  é o *matching* entre o segundo antecedente da primeira regra e a segunda entrada ( $\tilde{F}_2$ ),  $FS^1$  é o grau de ativação da primeira regra,  $\tilde{G}^1$  é o consequente da primeira regra,  $\tilde{B}^1$  é a saída inferida pela primeira regra,  $\tilde{M}_1^2 = \sup [\tilde{F}_1 \cap \tilde{F}_1^2]$  ( $\tilde{F}_1^2$  é o primeiro antecedente da segunda regra),  $\tilde{M}_2^2 = \sup [\tilde{F}_2 \cap \tilde{F}_2^2]$  ( $\tilde{F}_2^2$  é o segundo antecedente da segunda regra),  $FS^2$  é o grau de ativação da segunda regra,  $\tilde{G}^2$  é o consequente da segunda regra,  $\tilde{B}^2$  é a saída inferida pela segunda regra, e, finalmente,  $\tilde{B}$  é a saída inferida a partir de toda a base de regras (saída do sistema de inferência).

### 3.9 Conclusão

Neste capítulo foi proposta a formalização do “sistema de inferência escalonada *fuzzy* geral do tipo-2” e, para atender à essa proposição, diversas etapas foram cumpridas. Para facilitar a visualização e a compreensão de conjuntos gerais do tipo-2, foi apresentada uma ferramenta de visualização, denominada de “mancha” de incerteza geral. Para dar suporte ao mecanismo de inferência escalonada para sistemas com conjuntos *fuzzy* do tipo-2 geral, foram definidos, e formalizados, termos importantes como a translação de fatia vertical, o supremo de um conjunto *fuzzy* geral do tipo-2, a extensão cilíndrica de uma fatia vertical e do *singleton* do tipo-2. Finalmente, para reduzir o custo computacional da inferência geral, foi apresentada uma técnica de *defuzzificação* de conjuntos *fuzzy* gerais do tipo-2 denominada VSCTR (redução de tipo baseada nos centróides das fatias verticais).

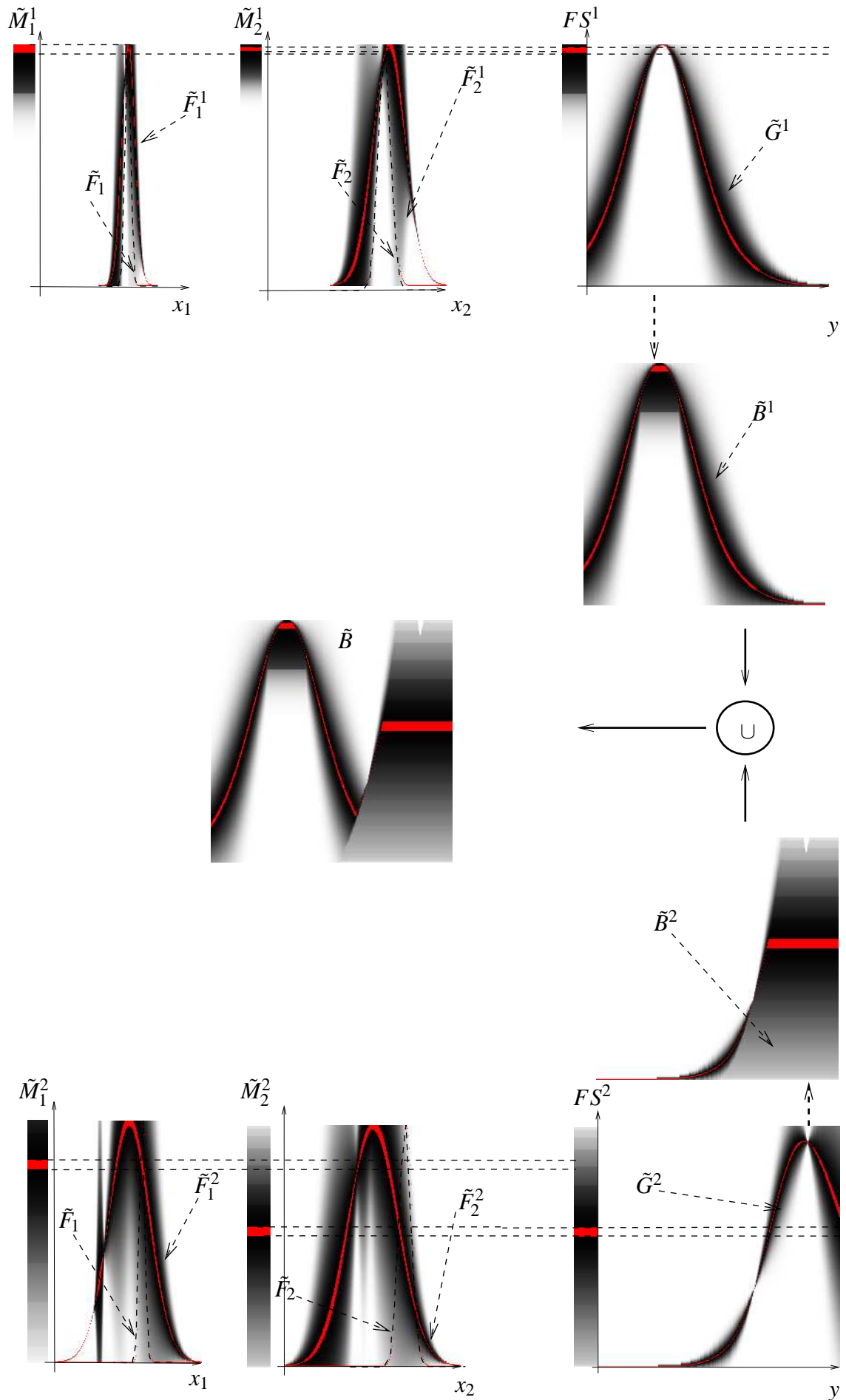


Figura 3.12: Exemplo da inferência escalonada *fuzzy* geral do tipo-2 para um sistema *fuzzy* com duas regras, com dois antecedentes cada.

## 4 Metodologia de Desenvolvimento do Classificador Fuzzy do Tipo-2 Geral

### 4.1 Introdução

Neste capítulo é apresentada a metodologia para obtenção de classificadores *fuzzy* gerais do tipo-2 propostos para a classificação de cobertura vegetal. Os classificadores serão obtidos a partir da adaptação do sistema de inferência formalizado no capítulo anterior.

### 4.2 O classificador *fuzzy* do tipo-2 geral

Neste capítulo propõe-se adaptar o sistema de inferência geral do tipo-2 estabelecido no capítulo anterior para que trabalhe como classificador de cobertura vegetal (LUCAS; CENTENO; DELGADO, 2008b). No entanto vale salientar que o método proposto é geral e poderia ser aplicado a outros problemas de classificação. A técnica sugerida para construir o classificador do tipo-2 é composta de duas etapas: a fase de projeto e a de avaliação. No projeto, algumas amostras de treinamento de diferentes bandas espectrais são selecionadas para construir os conjuntos do tipo-2 que formarão os antecedentes das regras, enquanto que os consequentes são formados por *singletons* do tipo-2 utilizados como rótulos para cada uma das classes temáticas.

Na fase de avaliação, o mecanismo de inferência atua, produzindo a classificação. A avaliação considera quatro blocos principais: “Fuzzificador”, “Mecanismo de inferência”, “Redutor de tipo” e “Tomador de decisão”. A figura 4.1 ilustra a concepção geral do classificador do tipo-2 que está sendo proposto nesta tese.

#### 4.2.1 Fase de projeto

##### 4.2.1.1 Obtenção dos antecedentes das regras

No método proposto, quando se constrói a base de regras, cada antecedente  $\tilde{F}_k^l$  é obtido a partir da banda  $k$ ,  $k = 1 \cdots P$ , de tal forma que o número de antecedentes é igual ao número de bandas

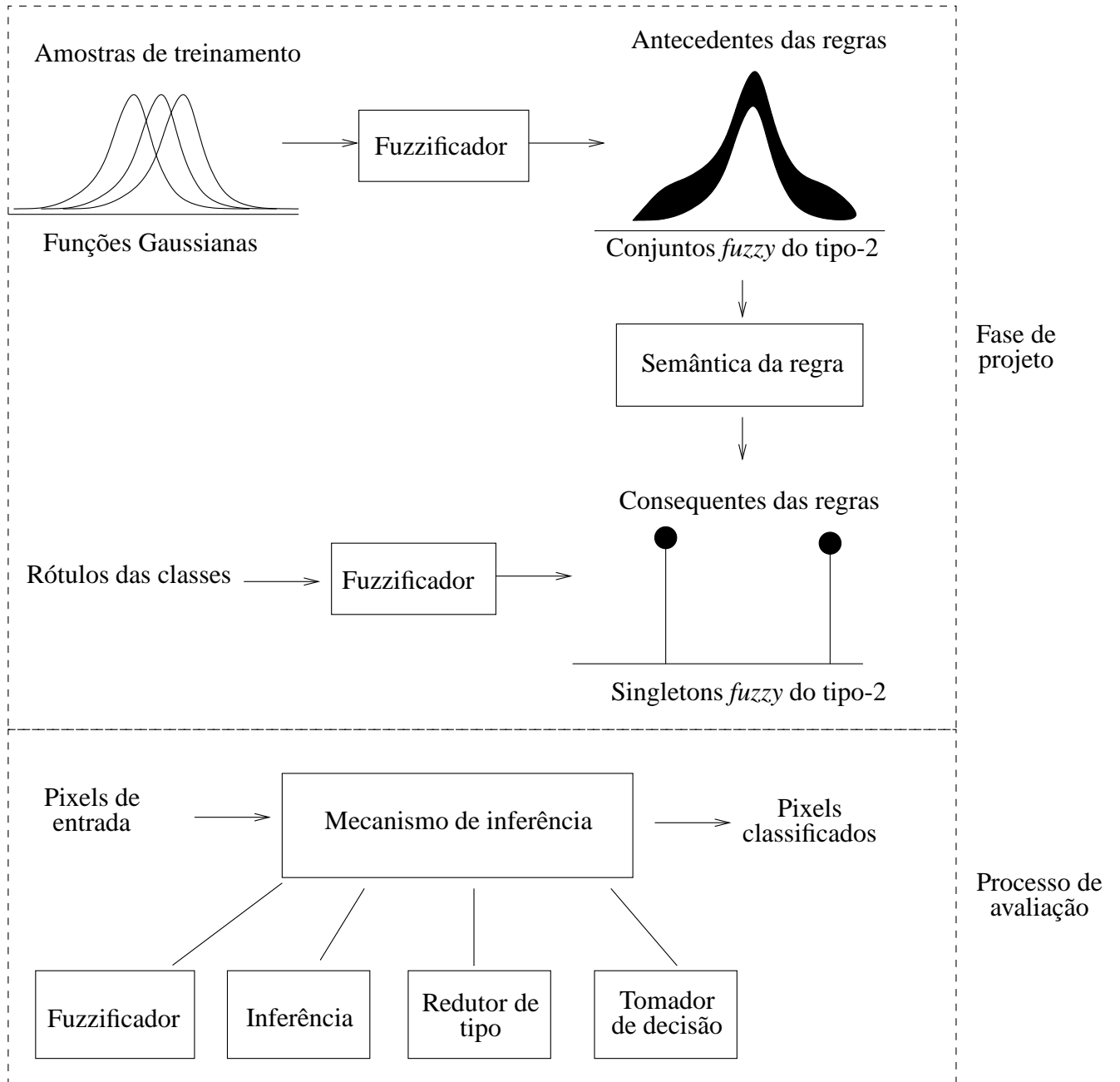


Figura 4.1: Aspectos gerais do classificador do tipo-2 proposto.

espectrais consideradas (LUCAS; CENTENO; DELGADO, 2008b).

No contexto de interesse para esta tese, que é o da cobertura vegetal, a fase de projeto considera as bandas espectrais provenientes de um equipamento de sensoriamento remoto e seleciona algumas amostras de treinamento para construir os conjuntos do tipo-2 que melhor representam cada classe em cada banda espectral. A figura 4.2 mostra um exemplo da extração de algumas amostras de treinamento de uma determinada banda em um conjunto de dados de cobertura vegetal. No exemplo da figura 4.2, as amostras rotuladas com **S** representam soja, as rotuladas com **M** representam milho e as com **A** representam alfafa.

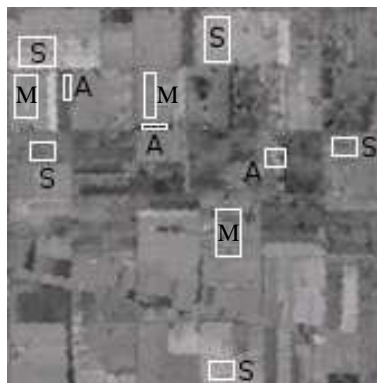


Figura 4.2: Amostras de treinamento obtidas de uma banda espectral.

Suponha que se deseje construir o conjunto  $\tilde{F}$  descrito anteriormente o qual representa o antecedente de uma das regras apresentadas na equação (3.7) e admita que, por construção, tal conjunto possua uma PMF (função de pertinência principal) dada pela função Gaussiana

$$\mu_{\tilde{F}_{princ}} = \mu_F = \begin{cases} u_{princ}(x) = e^{-\frac{(x-a_{princ})^2}{2 \cdot \sigma_{princ}^2}} & \forall x \mid f_x(u_{princ}(x)) = 1 \\ 0 & \text{caso contrário} \end{cases} \quad (4.1)$$

onde  $a_{princ}$  é a média dos valores dos *pixels*  $x_i$  de todas as amostras de uma determinada classe (cada retângulo na figura 4.2) tomadas em conjunto, como se fossem uma grande amostra, e  $\sigma_{princ}$  é o desvio-padrão destes dados em torno de  $a_{princ}$ .

Ocorre que, um antecedente do tipo-2 possui pertinência secundária e, para projetar um classificador que utilize tal antecedente, deve-se definir qual informação será disposta nestes graus secundários. Para isso, sugere-se utilizar as diferenças que existem entre as amostras de uma mesma classe, oriundas, por exemplo, das regiões de sombra existentes em partes das imagens ou dos diferentes níveis de crescimento da cultura agrícola sob estudo (SCHOWENGERDT, 1983). Desta forma, para tentar capturar tais diferenças, propõe-se que as amostras de treinamento da classe (os retângulos na figura 4.2) sejam tomadas individualmente e sejam construídas funções Gaussianas que as representem. Desta forma, a Gaussiana

$$u_j(x) = e^{-\frac{(x-a_j)^2}{2\sigma_j^2}} \quad (4.2)$$

onde  $j$ ,  $j = 1 \dots J$ , é uma das  $J$  amostras de treinamento que estão sendo consideradas para a classe em questão ( $J = 5$  para a soja na figura 4.2),  $a_j$  é a média dos *pixels* da amostra  $j$ , e  $\sigma_j$  é o desvio-padrão destes dados em torno de  $a_j$  (LUCAS; CENTENO; DELGADO, 2008b).

Sugere-se, então, construir as fatias verticais de  $\tilde{F}$ ,  $\mu_{\tilde{F}}(x_i)$ , levando-se em conta a dispersão  $\sigma(x_i)$ , que existe entre as funções Gaussianas  $u_j(x)$  no ponto  $x_i$ , ou seja,

$$\mu_{\tilde{F}}(x_i) = e^{-\frac{(u(x_i) - PMF(x_i))^2}{2\sigma^2(x_i)}} \quad (4.3)$$

onde  $PMF(x_i)$  é o valor da PMF em  $x = x_i$  e  $u(x_i)$ , são os graus de pertinência primários da fatia vertical  $\mu_{\tilde{F}}(x_i)$  também em  $x = x_i$ . A dispersão total nos dados de treinamento,  $\sigma(x_i)$ , pode ser calculada por

$$\sigma(x_i) = \sqrt{\frac{\sum_{j=1}^J \left( u_j(x_i) - \frac{\sum_{j=1}^J u_j(x_i)}{J} \right)^2}{J}}. \quad (4.4)$$

A figura 4.3 mostra a construção de um antecedente do tipo-2 a partir das cinco amostras de soja mostradas na figura 4.2. Em 4.3(a), pode-se ver a obtenção da PMF; em (b), as funções Gaussianas, uma para cada amostra; em (c), uma comparação da PMF, em linha tracejada, com as Gaussianas, em linhas cheias e, em (d), o conjunto tipo-2 resultante. Nota-se que, quando se afasta da região central do conjunto do tipo-2, ou seja, quando se aproxima dos pontos  $x=0$  ou  $x=255$ , a pertinência secundária tende a ser definida exclusivamente pela PMF e isso ocorre porque a dispersão entre as funções Gaussianas  $u_j(x)$  é nula nestes pontos.

Pode-se dizer que os antecedentes das regras, produzidos conforme a técnica sugerida, caracterizam a classe temática que estão representando, ou seja, é como se aqueles “emitissem sua opinião” sobre qual é a aparência da classe na banda considerada: a PMF tem uma opinião e cada função Gaussiana (cada amostra tomada individualmente) tem outra, às vezes semelhante à da PMF e, outras vezes, bem diferente. A caracterização feita pela PMF tem mais peso pois foi feita por todas as amostras tomadas em conjunto e, por isso, recebeu pertinência secundária máxima. Já as caracterizações feitas pelas funções Gaussianas tem menos peso e contribuem, através da sua dispersão, para compor os demais graus secundários. Desta forma, a pertinência secundária do antecedente, em cada fatia vertical, será máxima sobre a PMF e irá decrescendo não linearmente (Gaussiana) ao longo das retas verticais paralelas ao eixo  $u$ . A PMF, construída desta forma, captura o comportamento médio da classe temática, na banda espectral considerada. Caso se procedesse da mesma

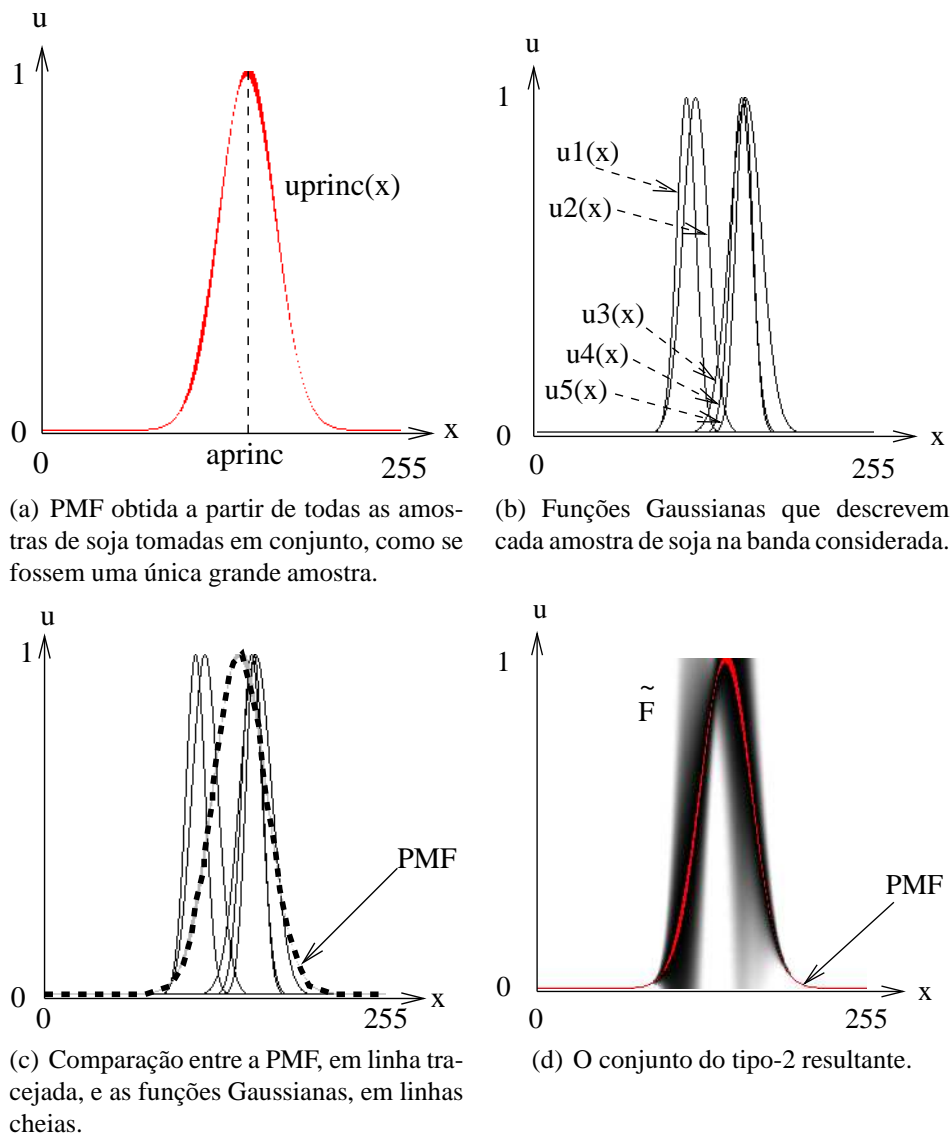


Figura 4.3: Construção de um conjunto *fuzzy* do tipo-2 a partir das amostras de treinamento.



forma para as demais classes e bandas, poder-se-ia construir um classificador do tipo-1 como o discutido na seção 2.5, com os antecedentes *fuzzy* sendo definidos exclusivamente pelas PMFs obtidas. Tal classificador será utilizado nos testes do capítulo 4. A diferença para o tipo-2 proposto é que este inclui o grau secundário como forma de caracterizar a dispersão entre as amostras.

#### 4.2.1.2 Definição dos consequentes

É comum definirem-se os consequentes das regras de um classificador *fuzzy* como *singletons*, cada um representando uma classe temática (BÁRDOSSY; SAMANIEGO, 2002; LIANG; MENDEL, 2001). A mesma idéia é utilizada nesta tese, de forma que, para três classes — soja, milho e alfafa — poder-se-ia definir os consequentes em  $y = 0$ , para a soja (rótulo  $C_s$  na figura 4.4(a)); em  $y = 127$ , para o milho (rótulo  $C_m$  na figura 4.4(b)) e  $y = 255$ , para a alfafa (rótulo  $C_a$  na figura 4.4(c)) (LUCAS; CENTENO; DELGADO, 2008b).

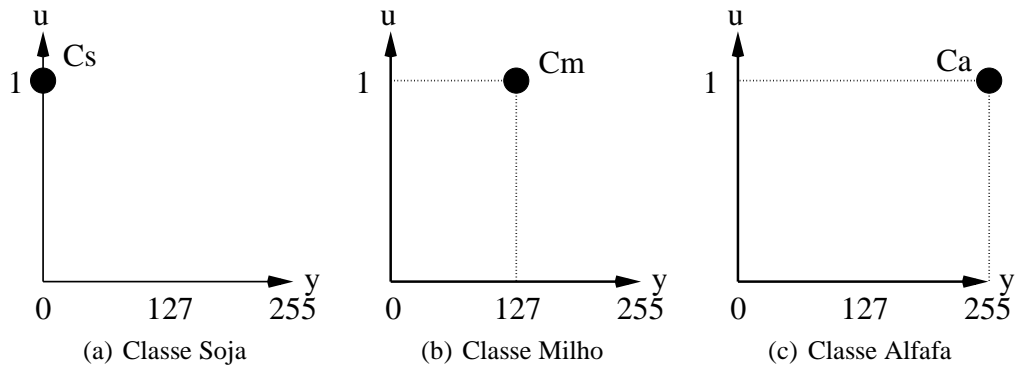


Figura 4.4: Consequentes,  $C_s$ ,  $C_m$  e  $C_a$ , para uma base de regras com três classes, soja, milho e alfafa, respectivamente.

Portanto, independentemente do número total de bandas consideradas, cada classe será representada por seu *singleton* correspondente, com a localização deste (o valor particular de  $y$ ) sendo irrelevante, já que, o que realmente interessa é que exista apenas um único  $y$  para cada classe que esteja sendo discriminada.

#### 4.2.1.3 Obtenção da base de regras

O classificador do tipo-2 é um sistema de inferência adaptado, e, portanto, grande parte do que foi falado no capítulo 3 ainda se aplica, ou seja, o classificador tem  $p$  variáveis de entrada  $x_1 \in X_1$ ,  $x_2 \in X_2, \dots, x_p \in X_p$ , e uma variável de saída  $y \in Y$  e a base de regra é composta de  $L$  regras, cuja  $l$ -ésima regra é dada por

$$R^l : \text{SE } x_1 \text{ é } \tilde{F}_1^l \text{ op } x_2 \text{ é } \tilde{F}_2^l \text{ op } \dots \text{ op } x_p \text{ é } \tilde{F}_p^l, \text{ ENTÃO } y \text{ é } \tilde{G}^l,$$

onde  $\tilde{F}_1^l, \tilde{F}_2^l \dots \tilde{F}_p^l$  são os antecedentes desta regra,  $\tilde{G}^l$  é o consequente e  $op$  é o operador de agregação dos antecedentes (E, OU, etc). Os antecedentes são conjuntos do tipo-2 mas, diferentemente do que acontecia com um sistema de inferência tradicional, o consequente não é mais um conjunto *fuzzy* do tipo-2, mas sim um *singleton* do tipo-2 (seção 3.3).

Uma das grandes vantagens do método proposto é o reduzido número de regras uma vez que o número total de regras ( $L$ ) que compõem a base de regras é igual ao número de classes, i.e., há uma regra para cada classe, e isto é independente do número de bandas espectrais. Portanto, quando se trabalha com 3 classes, haverá três regras ( $L = \text{número de classes} = 3$ ).

Um exemplo de base de regras, obtida quando se considera apenas uma banda e a etapa de projeto descrita, é:

$$\begin{aligned} R^1 &: \text{ SE } x \text{ é } \tilde{S} \text{ ENTÃO a classe é } C_s \\ R^2 &: \text{ SE } x \text{ é } \tilde{C} \text{ ENTÃO a classe é } C_m \\ R^3 &: \text{ SE } x \text{ é } \tilde{A} \text{ ENTÃO a classe é } C_a \end{aligned}$$

onde  $\tilde{S}$  poderia ser o conjunto do tipo-2 mostrado na figura 4.3(d) e  $C_s$ ,  $C_m$  e  $C_a$  são rótulos que identificam as classes soja, milho e alfafa, respectivamente (como mostrado na figura 4.4).

Na primeira regra, por exemplo,  $\tilde{S}$  é uma fonte de evidência que diz que se o *pixel*  $x$  é soja então ele será classificado como soja. As demais regras, para milho e para alfafa, têm interpretações semelhantes. Obviamente, haverá ocasiões nas quais mais de uma regra será ativada porque o *pixel* sob classificação assemelha-se a mais de uma classe temática.

Quando se trabalha com duas bandas espectrais, as regras serão da forma:

$$\begin{aligned} R^1 &: \text{ SE } x_1 \text{ é } \tilde{S}_1 \text{ op } x_2 \text{ é } \tilde{S}_2 \text{ ENTÃO a classe é } C_s \\ R^2 &: \text{ SE } x_1 \text{ é } \tilde{M}_1 \text{ op } x_2 \text{ é } \tilde{M}_2 \text{ ENTÃO a classe é } C_c \\ R^3 &: \text{ SE } x_1 \text{ é } \tilde{A}_1 \text{ op } x_2 \text{ é } \tilde{A}_2 \text{ ENTÃO a classe é } C_a \end{aligned} \quad (4.5)$$

onde  $\tilde{S}_k$  são os antecedentes do tipo-2 construídos a partir das amostras de treinamento de soja extraídos da  $k$ -ésima banda. De maneira semelhante,  $\tilde{M}_k$  e  $\tilde{A}_k$  são obtidos a partir das amostras de milho e alfafa, respectivamente, presentes na banda  $k$ . Agora a regra 1 diz que se houver evidências ( $\tilde{S}_1$  e  $\tilde{S}_2$ ) de que a classe seja soja, então o *pixel*  $x = [x_1 \ x_2]$  será classificado como soja (assumindo  $op = E$ ). O mesmo se aplica à regra 2 e à regra 3. Deve-se notar que, provavelmente, a classificação que utiliza duas bandas espectrais é mais confiável do que aquela que utiliza apenas uma banda, porque há duas evidências que dão suporte à decisão, ou seja, espera-se que quanto mais evidências se tenha para tomar uma decisão, mais confiança se tenha nela.

Em etapas preliminares do desenvolvimento, foram obtidas regras secundárias através da partição *fuzzy* em grade (ISHIBUCHI; NAKASHIMA; MURATA, 1999) onde se cruzam os antecedentes das regras principais (equação 4.5), como em

$$R^1 : \text{ SE } x_1 \text{ é } \tilde{S}_1 \text{ op } x_2 \text{ é } \tilde{M}_2 \text{ ENTÃO a classe é } C_s \text{ OU } C_c,$$

mas tais regras foram abandonadas por apresentaram diversas desvantagens, tais como: o número de regras explodia para muitos antecedentes (muitas bandas espectrais), introduzia-se complexidade desnecessária no sistema e, mais importante, não se notavam ganhos no desempenho.

#### 4.2.2 Processo de avaliação

Após a definição dos parâmetros do classificador do tipo-2, conhecida como fase de projeto, o sistema está pronto para inferir saídas a partir das entradas apresentadas, isto é, está pronto para classificar as entradas. A classificação é, geralmente, feita utilizando-se amostras de teste, que devem ser diferentes das amostras utilizadas para treinar o sistema. A figura 4.5 mostra exemplos de amostras de teste extraídas da mesma banda considerada no exemplo de treinamento que, como se pode ver, são diferentes das amostras de treinamento apresentadas na figura 4.2.

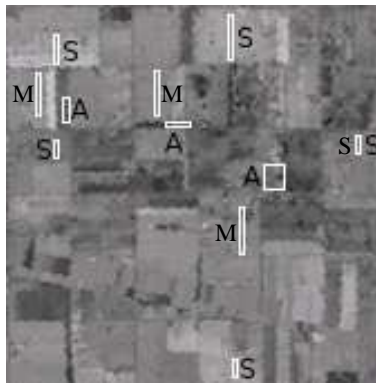


Figura 4.5: Amostras de teste para a banda espectral usada no treinamento.

A fase de classificação, também conhecida como processo de avaliação, foi adaptada de um sistema de inferência típico do tipo-2 (figura 2.5), produzindo o mecanismo de inferência do classificador tipo-2 apresentado na figura 4.6.

Pode-se ver na figura 4.6 que o classificador proposto é composto de quatro blocos principais — “Fuzzificador”, “Mecanismo de Inferência”, “Redutor de tipo” e “Tomador de decisão” — que serão descritos a seguir.

- *Fuzzificador*

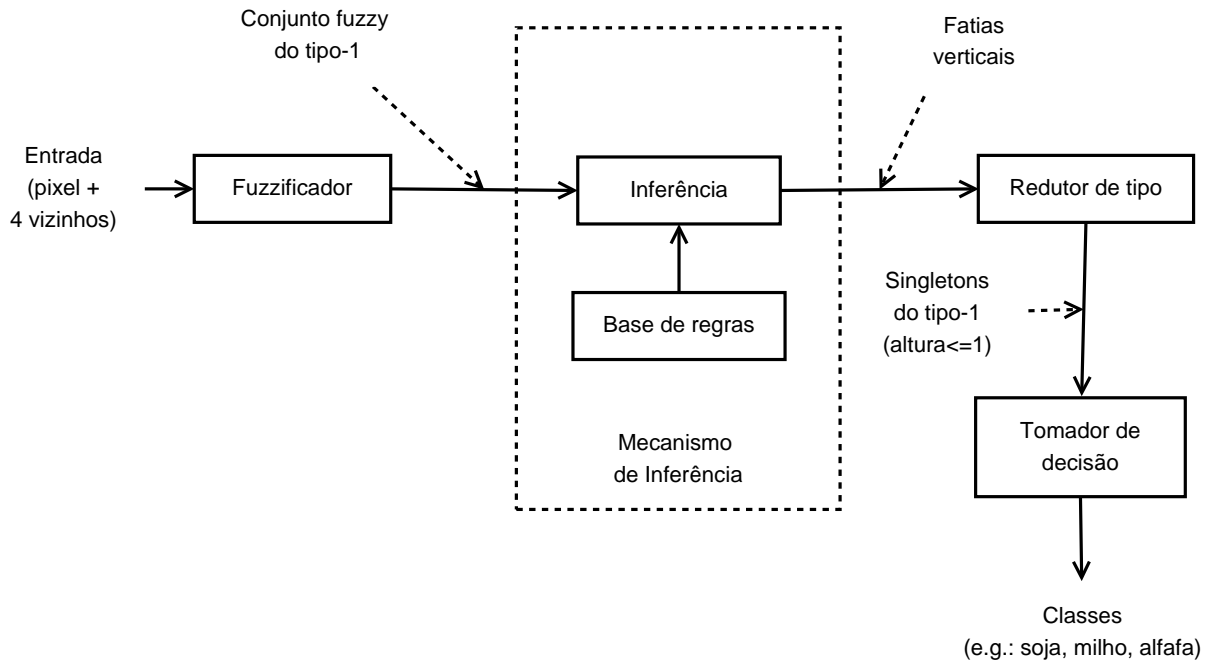


Figura 4.6: Mecanismo de inferência do classificador *fuzzy* do tipo-2 proposto.

O classificador proposto utiliza a informação espectral de uma cena, i.e., o conjunto de imagens da mesma cena, cada uma representando uma banda espectral diferente. Cada *pixel* desta “imagem” multiespectral tem suas coordenadas espaciais, como de costume, mas, também, a coordenada espectral  $\lambda$  (comprimento de onda), que é quantizada em algumas bandas espectrais discretas. Para uma imagem composta de  $P$  bandas, há  $P$  níveis de cinza associados com cada *pixel*, um para cada banda (SCHOWENGERDT, 1983). Portanto, a entrada do sistema é um *pixel* representado por um vetor de  $P$  elementos  $[x_1, x_2, \dots, x_P]$ , cada um obtido de uma banda espectral. Este *pixel*  $P$ -dimensional, e seus quatro vizinhos, são transformados, pelo *fuzzificador*, em  $P$  conjuntos do tipo-1, um para cada  $x_i$ ,  $i = 1 \dots P$ . Cada conjunto do tipo-1 é criado tomando-se a média e o desvio-padrão do *pixel* de entrada e de seus quatro vizinhos e construindo-se uma função de pertinência Gaussiana através da equação

$$\mu_i(x) = e^{-\frac{(x-a_i)^2}{2\sigma_i^2}} \quad (4.6)$$

onde  $a_i$  é a média entre os níveis de cinza do *pixel* de entrada e seus vizinhos para a banda  $i$ ,  $\sigma_i$  é o desvio-padrão destes dados e  $x$  é o valor do nível de cinza de cada um dos *pixels* considerados.

Como exemplo, considere um *pixel* 7-dimensional (proveniente de 7 bandas espectrais) cujos níveis de cinza sejam dados pelo vetor  $[120, 101, 111, 139, 98, 120, 123]$ . A tabela 4.1 mostra, na coluna da esquerda, a banda de onde o *pixel* foi extraído; na coluna central o nível de cinza do *pixel* e, na coluna da direita, os 4-vizinhos correspondentes.

A partir da tabela 4.1 e da equação 4.6, constroem-se os conjuntos *fuzzy* do tipo-1, conforme se vê na figura 4.7.

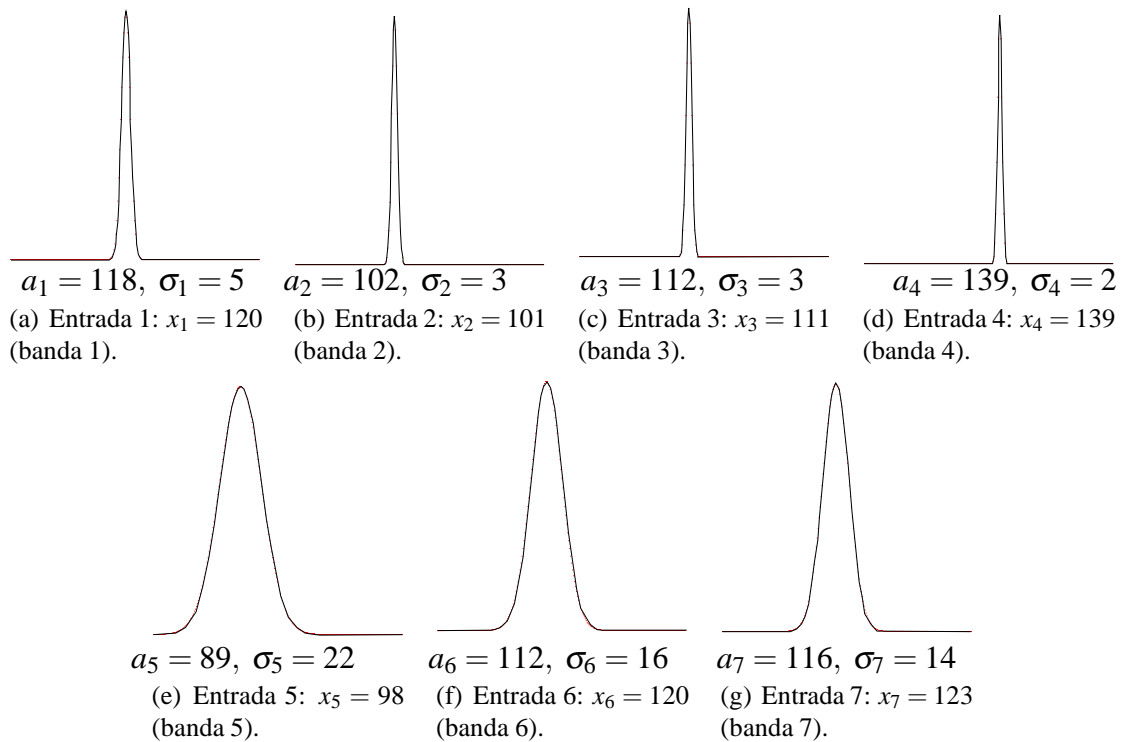


Figura 4.7: Exemplo da construção da entrada 7-dimensional do tipo-1.

Entradas cuja Gaussiana seja mais estreita (menor desvio-padrão) são mais homogêneas como, por exemplo a entrada 4, onde o nível de cinza do *pixel* e de seus 4-vizinhos é semelhante. Por outro lado, entradas cuja Gaussiana seja mais larga, como a entrada 5, são menos homogêneas.

- Mecanismo de inferência

Na seção 3.8 foi sugerida a “inferência escalonada *fuzzy* geral do tipo-2” (LUCAS; CENTENO; DELGADO, 2007, 2008a, 2008b) inspirada na inferência escalonada utilizada em sistemas de inferência *fuzzy* convencionais (seção 2.4.1.3). Agora, propõe-se modificar a inferência do tipo-2 para que seja utilizada em classificadores, obtendo-se as seguintes etapas:

Tabela 4.1: *Pixel* de entrada e seus 4-vizinhos.

Banda	Pixel	Seus 4-vizinhos
1	120	112, 124, 114, 122
2	101	104, 100, 107, 99
3	111	112, 110, 118, 111
4	139	139, 138, 142, 135
5	98	71, 116, 57, 102
6	120	97, 132, 90, 123
7	123	102, 132, 98, 126

*matching* das entradas com os antecedentes, agregação dos antecedentes e saída inferida. O mecanismo adaptado será detalhado a seguir.

### 1. *Matching* das entradas e dos antecedentes

O *matching*  $\tilde{M}_i^l$  do fato  $\tilde{F}_i$  e do antecedente  $\tilde{F}_i^l$  pode ser calculado pela composição *sup-t* estendida (KARNIK; MENDEL; LIANG, 1999):  $\tilde{M}_i^l = \text{sup} [\tilde{F}_i \cap \tilde{F}_i^l]$ , que é definida em  $x_1 = x_1'$  (LUCAS; CENTENO; DELGADO, 2007, 2008a).

Como exemplo, admita o antecedente  $\tilde{F}_i^l$  mostrado na figura 4.8(a), à direita, e uma entrada *fuzzy* convencional  $\tilde{F}_i$ , na figura 4.8(a), à esquerda, em linha tracejada. A interseção de  $\tilde{F}_i^l$  e  $\tilde{F}_i$  produz o conjunto do tipo-2  $\tilde{I}$  mostrado na figura 4.8(b), à direita, e o supremo deste conjunto produz o *matching*  $\tilde{M}_i^l$  que é a fatia vertical vista na figura 4.8(b), à esquerda.

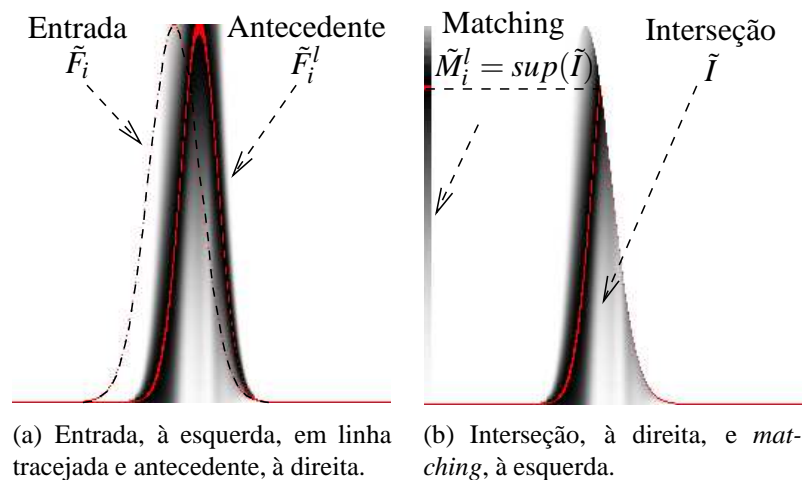


Figura 4.8: Um exemplo de *matching* entre um antecedente do tipo-2 e uma entrada *fuzzy* convencional.

### 2. Agregação dos antecedentes

A agregação dos antecedentes leva à “obtenção do grau de ativação da regra” ( $FS^l$ ) que mede a contribuição da  $l$ -ésima regra ao processo de inferência total (PEDRYCZ; GOMIDE, 1998), podendo ser calculada por  $FS^l = \tilde{M}_1^l \text{ op } \tilde{M}_2^l \text{ op } \dots \text{ op } \tilde{M}_p^l$ .  $FS^l$  é uma fatia vertical porque  $\tilde{M}_i^l$ ,  $i = 1 \dots p$ , são todas fatias verticais. Para exemplificar, considere uma base de regras E e 7 bandas espectrais. A figura 4.9 mostra os sete *matchings*,  $M_1 \dots M_7$ , da regra e o grau de ativação  $FS$  correspondente. Para auxiliar a compreensão, cada rótulo,  $M_1 \dots M_7$  e  $FS$ , foi posicionado à direita da fatia vertical correspondente e na mesma altura dos graus primários principais (aqueles cujos graus secundários são iguais a 1). Pode-se notar que os graus principais de  $FS$  são iguais aos de  $M_6$  e isso ocorre porque a agregação dos antecedentes foi feita pela *norma-t = min* (regras E) que produz, como resultado, os menores graus dentre todos os *matchings* considerados.

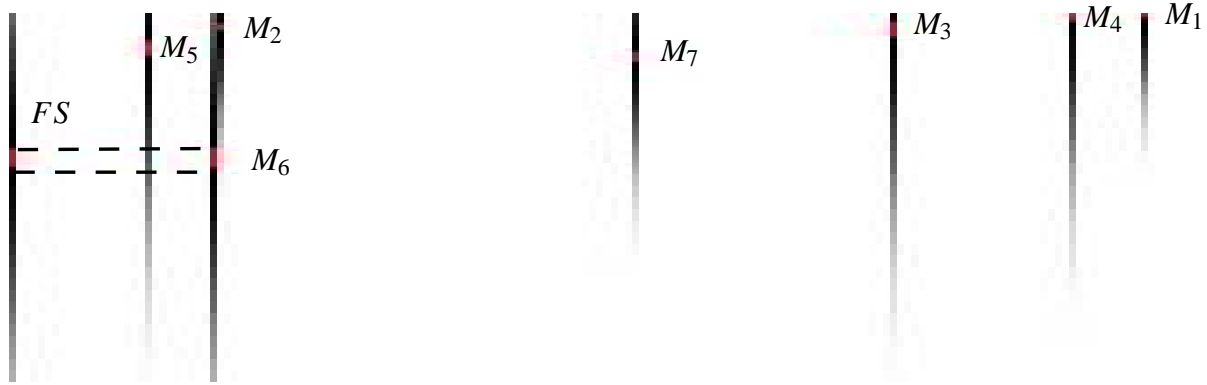


Figura 4.9: Obtenção do grau de ativação, FS, de uma regra com sete antecedentes, onde, M1 a M7 são os *matchings* destes antecedentes com o *pixel* 7-dimensional de entrada.

Convém mencionar que, neste exemplo, quando um dos antecedentes de uma das regras não produz *matching* com a entrada correspondente, o grau de ativação desta regra será nulo, também devido ao uso da *norma-t* na regra E.

Além disso, se, numa base de regras, nenhuma das regras for ativada, a saída do sistema de inferência será nula e o *pixel* não poderá ser classificado. Nesta situação, o classificador deverá atribuir o rótulo “não-classificado” ao *pixel*.

Ainda mais, se as regras forem compostas de apenas um antecedente (uma única banda espectral) e, mesmo assim, não houver ativação de nenhuma regra é porque não há cobertura do *pixel* na banda considerada.

Não se deveria construir uma base de regras com problemas de cobertura como o citado. Portanto, quando se agrega os antecedentes pelo operador E, não se deve utilizar uma banda espectral com problemas de cobertura pois, nestes casos, não adianta agregar novas bandas, pois os *matchings* nulos continuarão existindo, já que o valor nulo prevalece.

### 3. Saída inferida por cada regra

Assumindo-se que a semântica da regra seja dada por uma *norma-t* (i.e. uma relação de conjunção *fuzzy*) (PEDRYCZ; GOMIDE, 1998), pode-se obter a saída  $\tilde{B}^l$ , da regra  $l$ , pelas seguintes operações: 1) extensão cilíndrica do grau de ativação da regra ao longo do eixo  $y$ ; 2) cálculo da interseção do conjunto tipo-2 resultante com o consequente  $\tilde{G}^l$ , logo,

$$\tilde{B}^l = \text{cyl}_{y \in Y} [FS^l] \cap \tilde{G}^l. \quad (4.7)$$

A figura 4.10(b) mostra um exemplo de saída inferida  $\tilde{B}^l$  obtida a partir do consequente  $\tilde{G}^l$ , que é um *singleton* rotulado como  $C_s$  e posicionado em  $y' = 0$  (figura 4.10(a)) e da extensão cilíndrica  $\text{cyl}_{y \in Y} [FS^l]$ , que é um conjunto do tipo-2 com uma função de pertinência principal (PMF) dada por  $\mu_{A_{princ}} = 0.15 \forall y \in Y$  (figura 4.10(a)).

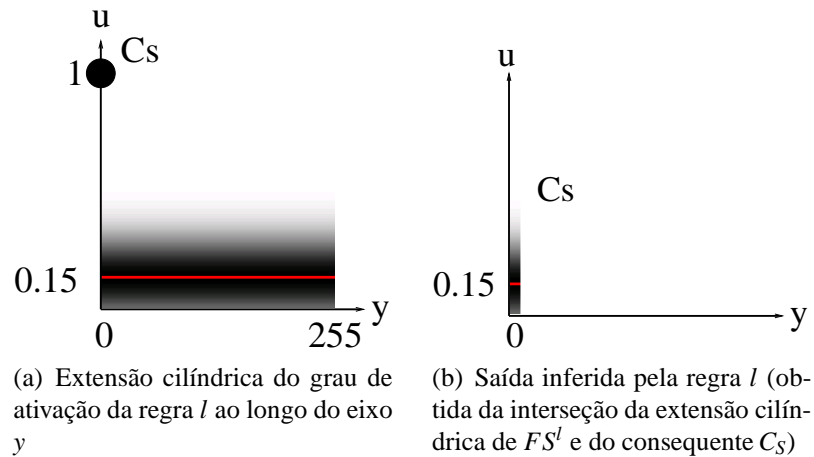


Figura 4.10: Exemplo de saída inferida para a regra  $l$ .

- Redutor de tipo

Como visto, a saída de cada regra é uma fatia vertical em  $y = y'$  e tal saída é reduzida a um *singleton* do tipo-1 (*singleton* convencional) ao ser processada pelo “Redutor de tipo” (figura 4.6).

Na seção 3.7 foi proposta uma técnica para reduzir um conjunto do tipo-2 a um do tipo-1 com o objetivo de *defuzzificar* o tipo-2. Esta técnica, chamada de “Redução de tipo através dos centróides das fatias verticais” (VSCTR) representa uma contribuição importante para diminuir o custo computacional da *defuzzificação* de um conjunto *fuzzy* geral do tipo-2.

Como a saída de cada regra do classificador tipo-2 é uma fatia vertical, resgata-se o VSCTR para reduzir tal saída a um *singleton* do tipo-1, cuja amplitude  $u = h$ ,  $0 \leq h \leq 1$ , é obtida como sendo a abcissa do centróide da fatia em  $y = y'$ .

Como exemplo, considere a fatia vertical em  $y = y' = 0$  apresentada na figura 4.11(a) cujo grau de pertinência principal ocorre em  $u = 0,15$ . Quando se olha esta fatia de lado (quando se olha para o plano  $\mu \times u$  a partir de um ponto qualquer em  $y$ ), obtém-se a figura 4.11(b) e, a partir desta, é possível calcular o centróide  $C$  e obter a abcissa  $h$ . Na figura 4.11(c) pode-se ver o *singleton* do tipo-1 posicionado em  $y = y' = 0$  cuja altura é  $h$ .

- Tomador de decisão

Quando se está classificando um *pixel*, mais de uma regra pode ser ativada e, nestes casos, um tomador de decisão é necessário para escolher entre as saídas destas regras, uma vez que, a classificação pressupõe que o *pixel* seja atribuído a apenas uma das classes pré-definidas. Tal processo de tomada de decisão será discutido a seguir.

Como já explicado, cada regra  $l$  produz uma saída obtida através da interseção entre a extensão cilíndrica do grau de ativação da regra,  $FS^l$ , e o consequente,  $\tilde{G}^l$ , desta regra. No caso do classificador do tipo-2 que está sendo proposto, os consequentes são *singletons* do tipo-2 e,



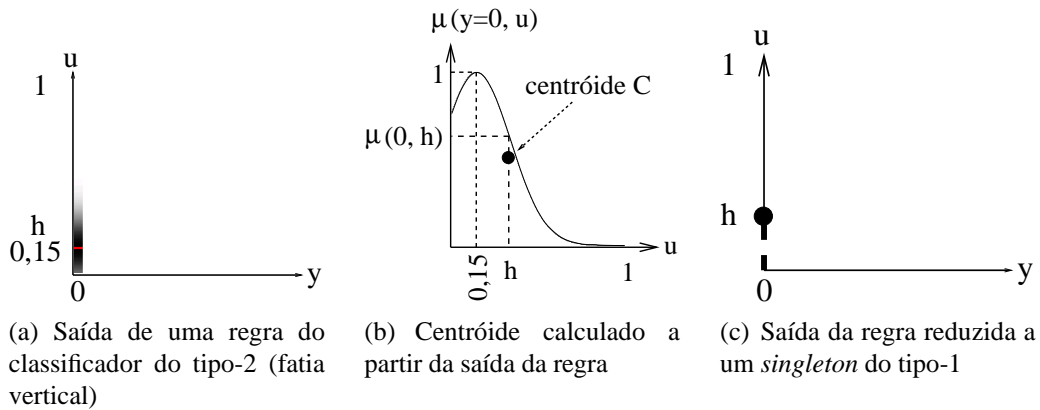


Figura 4.11: Exemplo de *defuzzificação* da saída de uma das regras do classificador geral do tipo-2.

portanto, as saídas são fatias verticais que são reduzidas a *singletons* do tipo-1 com alturas diferentes.

O “Tomador de decisão” mostrado na figura 4.6 é, então, alimentado com  $L$  *singletons* do tipo-1, cada um produzido pelo “Redutor de tipo” agindo sobre a saída de cada uma das  $l$  regras ativadas,  $l = 1 \dots L$ .

No exemplo que vem sendo considerado, com três classes temáticas — soja, milho e alfafa — a base de regras será composta de três regras, uma para cada classe. A primeira regra produz, como saída, um *singleton* do tipo-1, em  $y = 0$ , que representa o quanto o *pixel* sob análise se assemelha à classe soja. A segunda regra produz um *singleton* em  $y = 127$  que representa a semelhança do *pixel* com o milho e a terceira regra, com *singleton* em  $y = 255$ , mostra a semelhança com a alfafa.

Para realizar a classificação, o estágio “Tomador de decisão” escolherá a classe associada com o maior *singleton*: classe milho (rótulo  $C_m$ ) com altura  $h_2$  no exemplo apresentado na figura 4.12. Se houver mais de um *singleton* com a altura máxima, a classificação será impossível e, nestes casos, o classificador atribuirá o rótulo “não-classificado” ao *pixel*. Desta forma, define-se  $NC$  como a quantidade de *pixels* não classificados obtida por um classificador em uma determinada tarefa de classificação de pontos de testes previamente escolhidos.

No decorrer do desenvolvimento do classificador do tipo-2 e, também posteriormente, durante os testes, perceberam-se algumas limitações do modelo proposto inicialmente e, para superar tais limitações, que serão descritas oportunamente, o modelo evoluiu ao longo de quatro versões.

A seguir são descritas as quatro versões do classificador *fuzzy* geral do tipo-2: 1) com base de regras específica e operador E (BRE-E); 2) com base de regras específica e operador OU (BRE-OU); 3) com base de regras específica e operador E seguido de OU (BRE-E/OU); 4) com base de regras genérica e operador E (BRG-E).

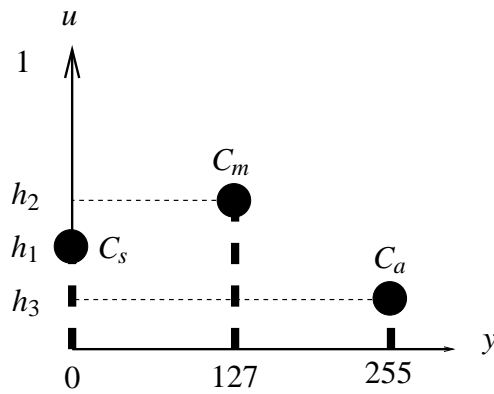


Figura 4.12: Exemplo de tomada de decisão: classe = milho.

### 4.2.3 Base de regras específica com operador de agregação E (BRE-E)

Quando se agrega os antecedentes pelo operador E obtém-se o classificador com base de regras específica com operador E (BRE-E) cuja base de regras, para duas bandas espectrais, pode ser obtida a partir da equação (4.5) substituindo-se o operador de agregação  $op$  pelo E:

$$\begin{aligned}
 R^1 : & \text{ SE } x_1 \text{ é } \tilde{S}_1 \text{ E } x_2 \text{ é } \tilde{S}_2 \text{ ENTÃO a classe é } C_s \\
 R^2 : & \text{ SE } x_1 \text{ é } \tilde{M}_1 \text{ E } x_2 \text{ é } \tilde{M}_2 \text{ ENTÃO a classe é } C_m \\
 R^3 : & \text{ SE } x_1 \text{ é } \tilde{A}_1 \text{ E } x_2 \text{ é } \tilde{A}_2 \text{ ENTÃO a classe é } C_a
 \end{aligned} \tag{4.8}$$

Por exigir que várias fontes de evidência apontem para a mesma classificação, o classificador BRE-E tem, como vantagem, o alto nível de confiança desta classificação. Por outro lado, a principal desvantagem é a tendência de aparecerem *pixels* não classificados pois, basta uma das evidências não apontar para a mesma decisão (nível de ativação nulo), para que a classificação não seja feita. Esta tendência é tanto maior quanto mais fontes participarem da decisão, pois o consenso se torna mais difícil.

A desvantagem apontada é consequência do uso das normas-t dos conjuntos *fuzzy* (agregação dos antecedentes pelo operador E) (KLIR; FOLGER, 1988; PEDRYCZ; GOMIDE, 1998), que são operadores conjuntivos e, portanto, “severos” (BLOCH, 1996), exigindo o consenso.

### 4.2.4 Base de regras específica com operador de agregação OU (BRE-OU)

Como visto, o classificador anterior pode apresentar um grande número de pontos sem classificação devido a agregação ser feita por uma *norma-t* (operador *meet*, seção 2.4.6.2). Então, para reduzir a quantidade destes *pixels* não classificados (NC) do sistema anterior, buscou-se utilizar as *normas-s* (operador *join*, seção 2.4.6.1), que são exemplos de operadores disjuntivos ou

“indulgentes” (BLOCH, 1996) que se baseiam na maior pertinência para agregar os antecedentes, independente da fonte de evidência (banda) correspondente ser confiável ou não.

A base de regras do sistema seria semelhante à da equação (4.8) bastando-se substituir o operador de agregação pelo OU. Então, para duas bandas espectrais, utilizar-se-ia o seguinte conjunto de regras

$$\begin{aligned}
 R^1 : & \quad \text{SE } x_1 \text{ é } \tilde{S}_1 \text{ OU } x_2 \text{ é } \tilde{S}_2 \text{ ENTÃO a classe é } C_s \\
 R^2 : & \quad \text{SE } x_1 \text{ é } \tilde{M}_1 \text{ OU } x_2 \text{ é } \tilde{M}_2 \text{ ENTÃO a classe é } C_m \\
 R^3 : & \quad \text{SE } x_1 \text{ é } \tilde{A}_1 \text{ OU } x_2 \text{ é } \tilde{A}_2 \text{ ENTÃO a classe é } C_a
 \end{aligned}
 \tag{4.9}$$

o mesmo valendo para qualquer número de bandas superior a duas.

No classificador que utilizava a agregação pelo operador E da seção 4.2.3, quando um *pixel* não possuía cobertura em uma das bandas (não ativava nenhuma das regras), a adição de novas bandas não solucionava o problema, pois o *matching* nulo permanecia, conforme discutido em detalhes na seção 4.2.2. Quando se utiliza o operador OU, por outro lado, ao se adicionar uma nova banda que tenha cobertura para o *pixel* anteriormente não classificado (NC) devido a *matching* nulo, garante-se que a base de regras produzirá uma saída diferente de zero e isto ocorre porque o OU escolhe o máximo entre os *matchings* e, portanto, o classificador BRE-OU tem o potencial para reclassificar NCs provenientes de *matching* nulos.

Imagina-se, porém, que o uso das *normas-s* degrade o desempenho do classificador, pois, apesar de deixar poucos pontos sem classificação, esta estaria baseada apenas em uma fonte de evidência (aquela que produziu o maior *matching*), nem sempre confiável.

#### 4.2.5 Base de regras específica com operador de agregação E seguido de OU (BRE-E/OU)

O classificador com operador E, descrito na seção 4.2.3, e o que utiliza a agregação pelo OU, na seção 4.2.4, são bons pontos de partida para o projeto de um classificador *fuzzy* geral do tipo-2 mas, não são candidatos sérios para aplicações práticas, uma vez que, o primeiro pode apresentar um grande número de pontos não classificados (NC) e o segundo pode apresentar degradação do desempenho.

Seria interessante combinar o bom desempenho do classificador com operador E com a baixa quantidade de pontos não classificados do operador OU e, para isso, sugere-se uma solução hierarquizada, chamada de BRE-E/OU, onde seriam feitas duas classificações: a primeira, com maior precisão, utilizando o E, equação (4.8), e a segunda, apenas para os *pixels* não-classificados, utilizando o operador OU, equação (4.9).

#### 4.2.6 Base de regras genérica com operador de agregação E (BRG-E)

O classificador “E seguido de OU” (BRE-E/OU), sugerido na seção anterior, pretendeu conciliar o bom desempenho do E com o baixo número de pontos não classificados do OU, produzindo um sistema útil. Apesar disso, a transição do operador E, excessivamente restritivo, para o operador OU, excessivamente flexível, parece ser muito abrupta.

Tentando-se obter uma alternativa de transição mais suave de um operador de agregação para o outro, buscou-se manter a classificação hierarquizada, mas, ao invés de se utilizar o operador E seguido de OU, propõe-se utilizar o operador E com uma base de regras que vai se tornando mais genérica em cada nível hierárquico, conforme descrito a seguir. Numa primeira etapa, tenta-se a decisão com todos os antecedentes possíveis, ou seja, obtidos de todas as bandas disponíveis (e.g.: 7 bandas) mas, se um determinado *pixel* não puder ser classificado desta forma, porque nenhuma regra foi ativada, procura-se, em cada regra, um antecedente que possua *matching* nulo e elimina-se este antecedente. Após a redução no número de antecedentes, tenta-se a reclassificação, com um antecedente a menos em cada regra (e.g.: 6). Observe-se que os antecedentes que foram eliminados no total (considerando-se todas as regras) não são provenientes, necessariamente, da mesma banda espectral em todas as regras: pode ter sido eliminado o sétimo antecedente da primeira regra, o sexto da segunda e o quinto da terceira, por exemplo. Se a classificação ainda não for possível porque, novamente, nenhuma regra foi ativada, procura-se outra vez, em cada regra, dois antecedentes causadores de *matchings* nulos para serem eliminados e reclassifica-se com dois antecedentes a menos. O processo é repetido até que o *pixel* possa ser reclassificado, mesmo que seja por uma única banda. Se esta última banda sobrevivente também não puder reclassificar (e.g.: porque o seu *matching* é nulo ou porque há duas, ou mais, classes empatadas), aí sim, o *pixel* é rotulado como “não-classificado” (NC). O pseudo-código a seguir mostra o algoritmo que está sendo proposto:

```
para pixel_avaliacao de 1 a total faça
  total_antecedentes <= total de bandas {em cada regra}
  inferência tipo 2 para pixel_avaliacao
  enquanto regra ativada = 0 e total_antecedentes > 1
    total_antecedentes <= total_antecedentes -1;
    inferência tipo 2 para pixel_avaliacao
  fim enquanto
  se regra ativada = 0 ou há conflito nas classes
    erro na classificação
  senão
    ponto é classificado {correta ou incorretamente}
  fim se
```

fim para

O método foi denominado de “classificador com base de regras genérica e operador de agregação E” (BRG-E) e admite a existência de variáveis de entrada irrelevantes conforme discutido na seção 2.4.1.1. Deve-se notar que a base de regras do classificador BRG-E é dinâmica uma vez que ela se altera, *pixel a pixel*, dependendo de quais antecedentes foram eliminados. Este aspecto dinâmico da base de regras compromete, de certa maneira, a interpretabilidade.

É interessante observar que o novo classificador é capaz de reclassificar somente NC provenientes de *matchings* nulos mas não consegue reclassificar NC de confusão, uma vez que os últimos possuem duas, ou mais, regras empatadas com mesmo nível de saída (há dúvida entre duas, ou mais, classes) e, portanto, não há antecedente com nível nulo para ser eliminado.

O classificador tipo-1 a ser utilizado como referência para fins de comparação de desempenho também utiliza a base de regras genérica e operador de agregação E, possuindo, como principal diferença para o classificador tipo-2 descrito, o fato de utilizar conjuntos *fuzzy* convencionais (tipo-1) nos antecedentes e *singletons* do tipo-1 nos consequentes.

### 4.3 Conclusão

Neste capítulo, foram apresentadas e discutidas em detalhes algumas contribuições desta tese: a adaptação da inferência escalonada *fuzzy* geral do tipo-2 ao problema da classificação digital, o “método de classificação hierárquica” com base de regras específica, com operador E seguido de OU, ou base genérica, com operador E, ambos com base de regras composta de poucas regras (igual ao número de classes que se quer discriminar). Para testar o mecanismo de inferência, foram propostas quatro versões de um classificador *fuzzy* geral do tipo-2: BRE-E, BRE-OU, BRE-E/OU e BRG-E.

## 5 *Estudo de casos*

### 5.1 Introdução

Neste capítulo são apresentados os conjuntos de dados que serão utilizados para validar os classificadores propostos no capítulo anterior. Foi realizado o estudo de cinco casos, cada um deles com dados provenientes de regiões geográficas diferentes e, muitas vezes, obtidos com equipamentos de sensoriamento diferentes.

### 5.2 Caso 1

O primeiro caso refere-se a uma região plana agricultável localizada em *Tippecanoe County*, Indiana, Estados Unidos. As imagens relativas à esta região foram obtidas pelo satélite mapeador temático *Landsat* em 17 de julho de 1986 (WU; DOERSCHUK, 1995). Foram utilizadas sete imagens com  $169 \times 169$  pixels, cada uma correspondendo a uma das bandas espectrais do mapeador, com comprimentos de onda variando de  $0,45\mu m$  a  $12,5\mu m$ , conforme mostra a figura 5.1. Para este conjunto de dados, as três classes temáticas consideradas foram soja, milho e alfafa.

A verdade de campo para *Tippecanoe County* também foi fornecida como pode ser visto na figura 5.2, onde áreas em branco são as cultivadas com milho, em cinza-claro são as áreas cultivadas com soja, em cinza-escuro tem-se alfafa e em preto, as regiões de fundo da imagem (LANDGREBE; BIEHL, 2001).

### 5.3 Caso 2

O segundo conjunto de dados é composto de nove bandas espectrais com  $145 \times 145$  pixels cada, conforme se vê na figura 5.3. Este conjunto foi obtido pelo espectrômetro aero-embarcado AVIRIS da NASA em junho de 1992 e a cena imageada é de uma paisagem mista de agricultura e floresta no site de testes de *Indian Pine* a 10km a oeste de West Lafayette, Estados Unidos. Cada uma das

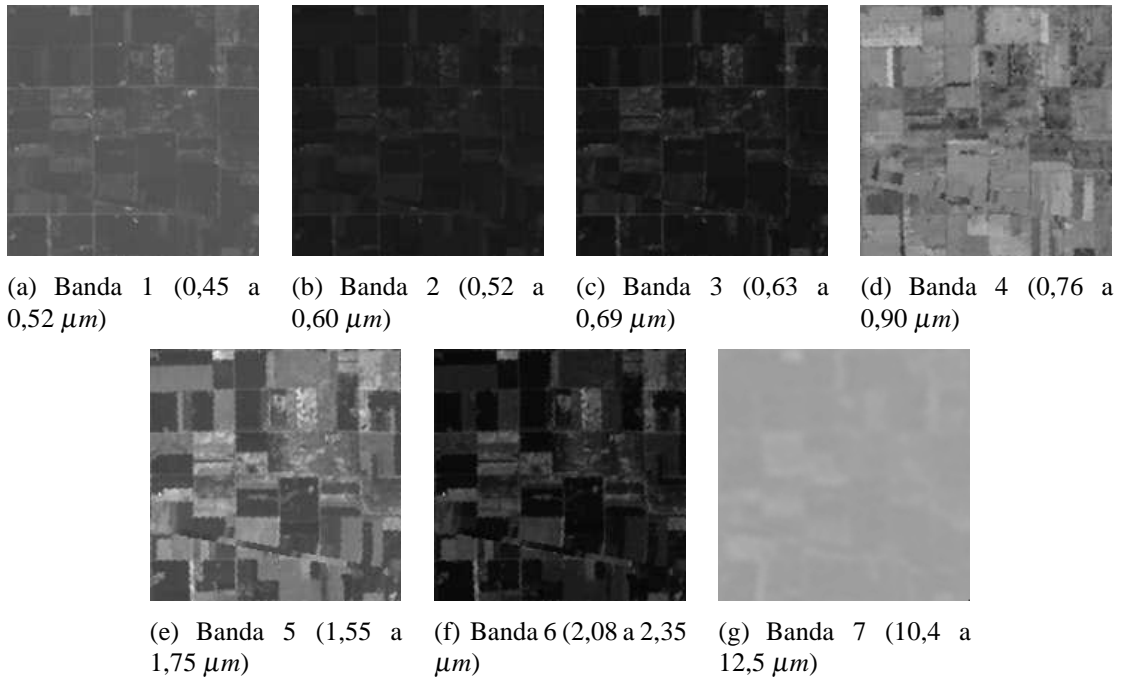


Figura 5.1: O primeiro conjunto de dados com sete bandas espectrais obtidas pelo satélite *Landsat TM* em 17 de julho de 1986.

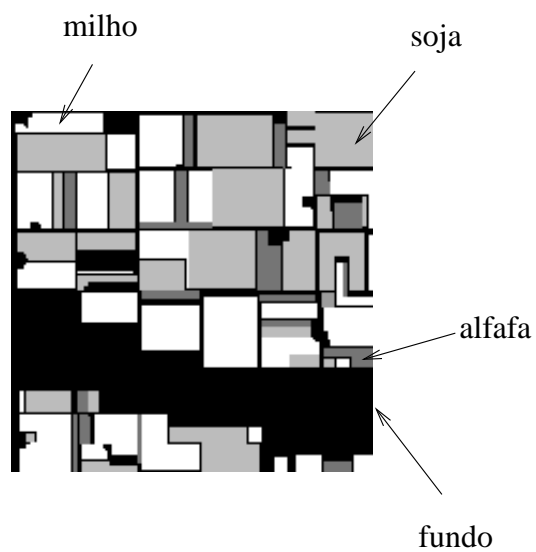


Figura 5.2: Verdade de campo para *Tippecanoe County*.

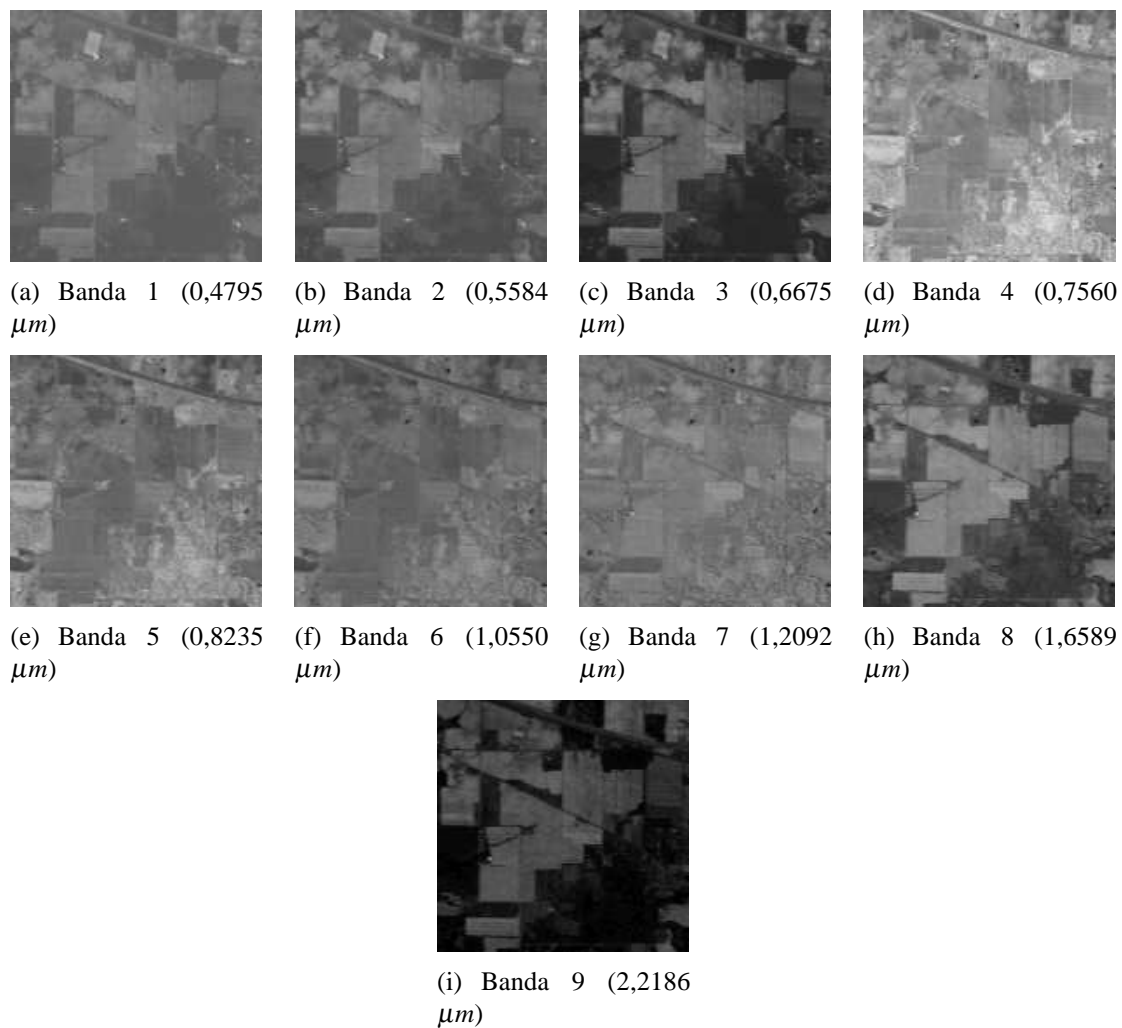


Figura 5.3: O segundo conjunto de dados com nove bandas espectrais obtido pelo espectrômetro AVIRIS em junho de 1992.



bandas possui aproximadamente 10nm de largura e é centrada no comprimento de onda mostrado na figura 5.3 (LANDGREBE, 1998).

A verdade de campo para o segundo conjunto de testes pode ser vista na figura 5.4, onde áreas em branco são as cultivadas com milho, em cinza-claro são as áreas cultivadas com soja, em cinza-escuro tem-se floresta e em preto, as regiões de fundo da imagem (LANDGREBE, 1998).

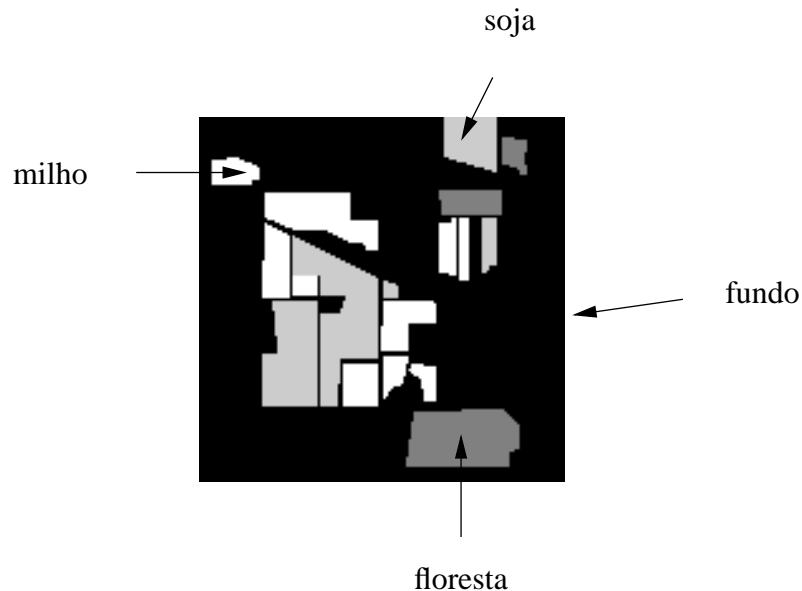


Figura 5.4: Verdade de campo para o segundo conjunto de testes.

## 5.4 Caso 3

O terceiro conjunto de dados é composto por doze bandas espectrais com resoluções de  $220 \times 949$  pixels e foi obtido a partir de um scanner aero-embarcado que sobrevoou o extremo sul de *Tippecanoe County*, Indiana, Estados Unidos, em junho de 1966. Apesar destes dados serem antigos, eles permanecem contemporâneos (LANDGREBE, 1997). As doze bandas espectrais que compõem o terceiro conjunto podem ser vistas na figura 5.5.

A figura 5.6 apresenta a verdade de campo para o terceiro conjunto de testes onde se pode ver que áreas em branco são as cultivadas com milho, em cinza-claro são as áreas cultivadas com soja, em cinza-escuro são as regiões com aveia e em preto, as regiões de fundo da imagem (LANDGREBE, 1997).

## 5.5 Caso 4

O quarto conjunto de dados corresponde à região de Beverly, Massachussets, Estados Unidos e foi obtido pelo satélite *Landsat TM*, sendo composto de 5 imagens de  $512 \times 512$  pixels correspon-

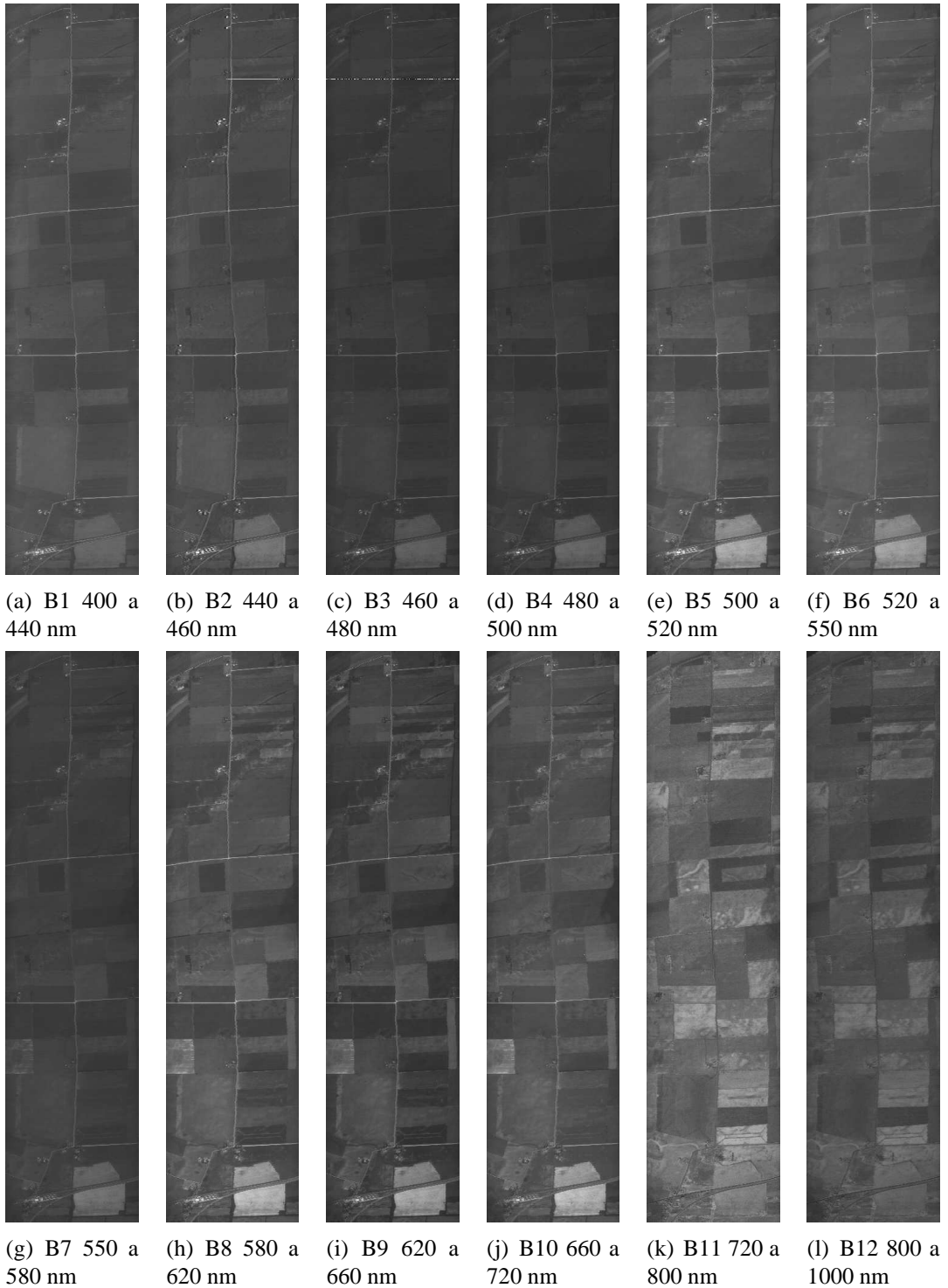


Figura 5.5: O terceiro conjunto de dados com doze bandas espectrais obtido por um *scanner* aerombarcado em junho de 1966.

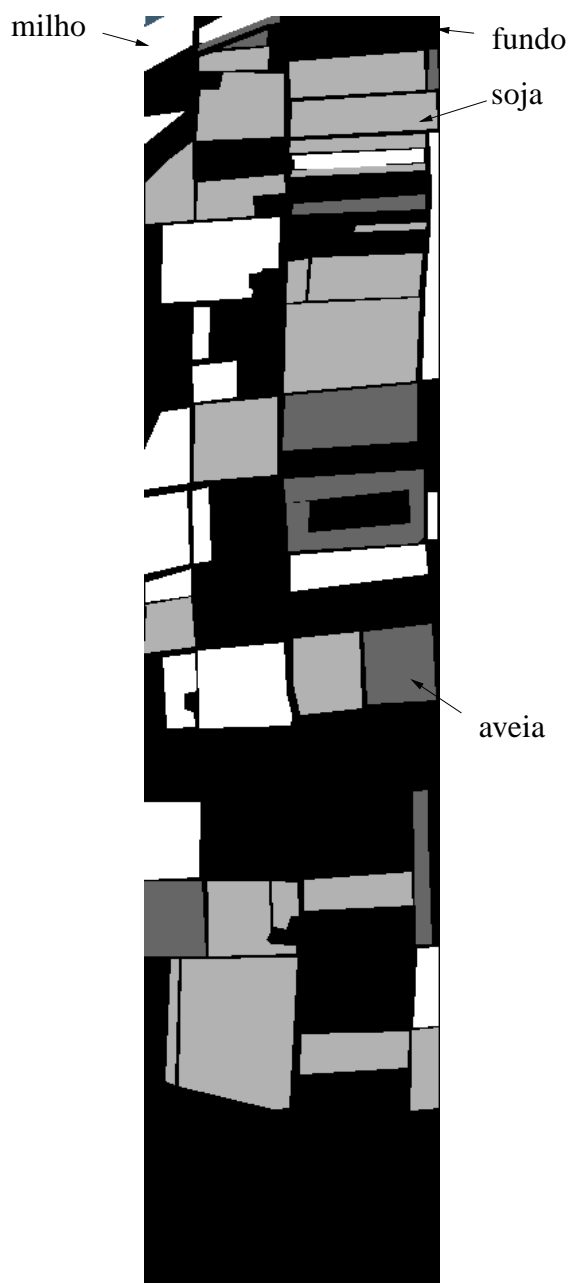


Figura 5.6: Verdade de campo para o terceiro conjunto de testes.

dendo às bandas espectrais azul, verde e vermelho, na faixa visível do espectro eletromagnético, e mais duas bandas no infravermelho do espectro (figura 5.7) (CONGALTON, 2008).

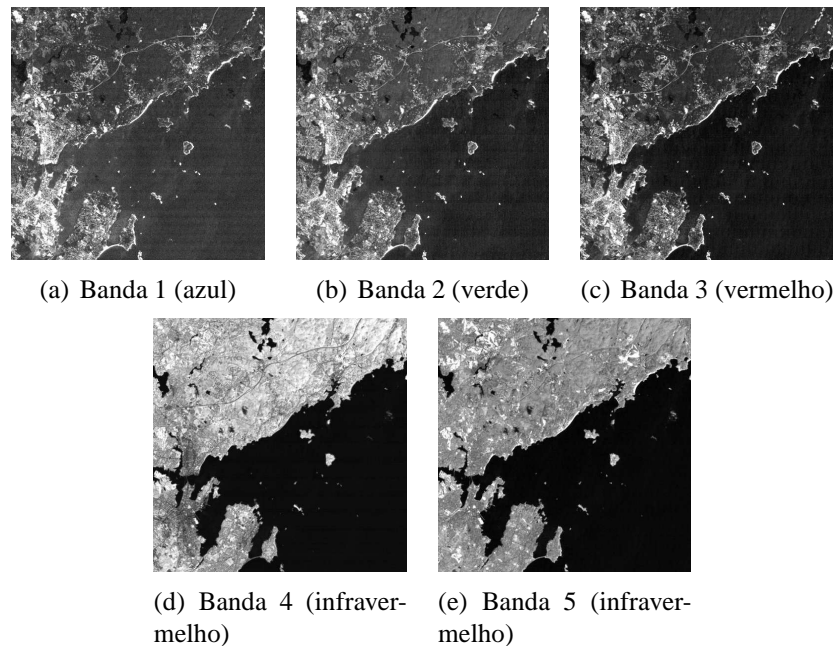


Figura 5.7: O quarto conjunto de dados com cinco bandas espectrais obtidas pelo satélite *Landsat TM*.

A verdade de campo para o quarto conjunto de testes pode ser vista na figura 5.8, onde áreas em branco correspondem às regiões rochosas, em cinza-claro são as regiões cobertas com vegetação rasteira, em cinza-escuro tem-se floresta e em preto, as regiões de fundo da imagem (CONGALTON, 2008).

## 5.6 Caso 5

O quinto conjunto de dados corresponde ao parque Barigui em Curitiba, Paraná, Brasil. O parque Barigüi fica localizado próximo à região central da cidade de Curitiba e faz parte da política municipal de preservação de fundos de vale, cujo objetivo é evitar o assoreamento e a poluição dos rios, bem como dar proteção à mata ciliar e impedir a ocupação irregular das suas margens. O Barigui, cujo nome tem origem indígena e significa “rio do fruto espinhoso”, em alusão às pinhas das araucárias nativas, foi criado em 1972. O parque é constituído por três bosques com floresta primária nativa, florestas secundárias, o rio Barigui, um lago com 230.000 metros quadrados, áreas gramadas, estradas e vias de pedestres asfaltadas e áreas construídas (salão de atos, pavilhão de exposições, parque de diversões, quadras de esporte, etc) (PARQUE..., 2004). As imagens do parque Barigui foram obtidas pelo satélite Ikonos e possuem resolução de  $734 \times 734$  pixels e 1m de resolução espacial, correspondendo às bandas espectrais azul, verde e vermelho, na faixa visível do espectro eletromagnético, conforme pode ser visto na figura 5.9.

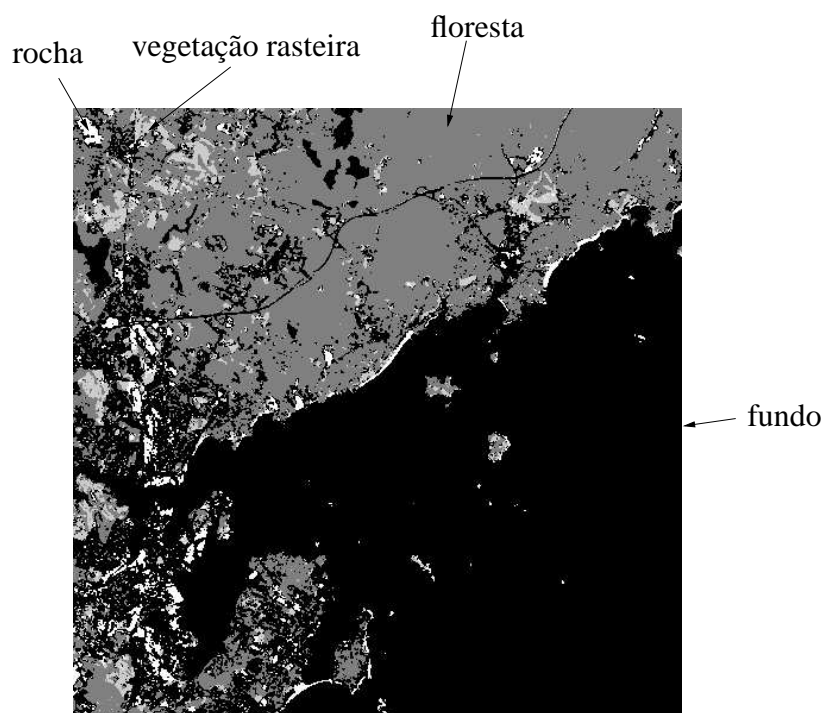


Figura 5.8: Verdade de campo para o quarto conjunto de testes.

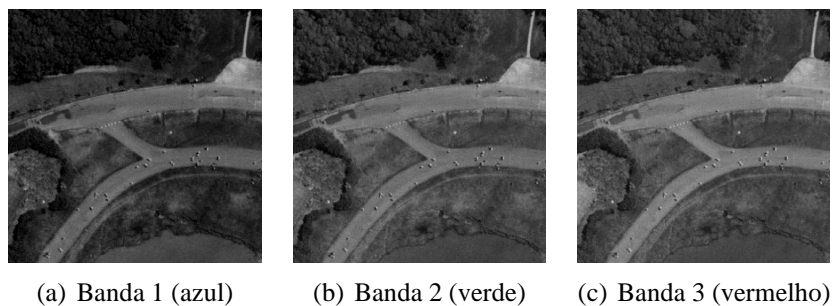


Figura 5.9: O quinto conjunto de dados com três bandas espectrais obtidas pelo satélite Ikonos com resolução espacial de 1m.

A verdade de campo para o quinto conjunto de testes pode ser vista na figura 5.10, onde áreas em branco correspondem aos trechos asfaltados, em cinza-claro são as regiões gramadas, em cinza-escuro tem-se floresta e em preto, as regiões de fundo da imagem (PARQUE... , 2004).

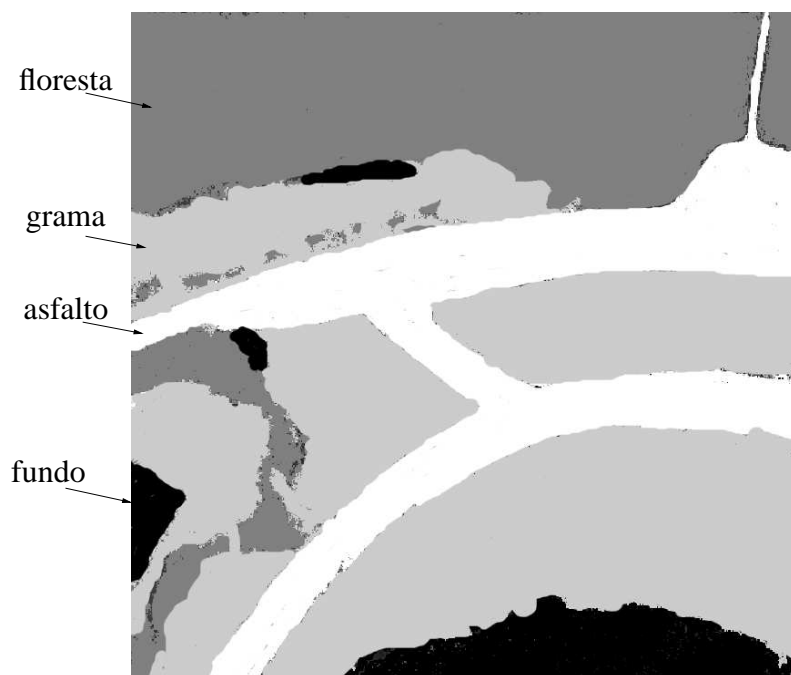


Figura 5.10: Verdade de campo para o quinto conjunto de testes.

## 5.7 Conclusão

Neste capítulo, foram apresentados os cinco conjuntos de dados, provenientes de diferentes equipamentos de sensoriamento remoto, que serão utilizados para testar e validar os classificadores desenvolvidos.

## 6 *Resultados e discussão*

### 6.1 Introdução

Neste capítulo, serão testados os quatro classificadores propostos no capítulo anterior (BRE-E, BRE-OU, BRE-E/OU e BRG-E) utilizando-se quatro testes, um para cada classificador. Estes testes serão realizados utilizando-se as amostras de treinamento e de testes mostradas no capítulo anterior. Os quatro testes têm por objetivo fornecer um indicativo do comportamento dos quatro classificadores propostos mas não são suficientes para que se possa generalizar as conclusões sobre o desempenho dos classificadores.

Então, para que se obtenham conclusões gerais, os classificadores BRE-E/OU e BRG-E serão testados em cinco problemas diferentes cada um com um conjunto de dados específico, contendo culturas agrícolas, tais como soja, milho, alfafa e aveia e, também, com regiões de floresta. Em cada problema será feita a validação cruzada, alternando-se os conjuntos de treinamento e de teste, de forma que serão utilizadas seis amostras para treinamento e uma para testes (cada amostra será composta de 81 *pixels*). Serão mostrados os resultados obtidos por cada um dos classificadores projetados e será feita a comparação destes com os obtidos pelos classificadores de referência: o *MaxVer*<sup>1</sup> e o classificador *fuzzy* tradicional. O objetivo da sequência de testes apresentada é ordenar os classificadores de acordo com o seu desempenho, dizendo qual é o melhor e qual é o pior, considerando todos os problemas abordados. Ocorre que tal ordenação pode não ser estatisticamente significativa, então para avaliar os valores obtidos serão utilizados os testes de Friedman e de Nemenyi (DEMSAR, 2006). No final do capítulo, os resultados são discutidos.

### 6.2 Descrição dos experimentos

Nesta seção são descritos quatro experimentos que permitem avaliar o comportamento dos classificadores BRE-E, BRE-OU, BRE-E/OU e BRG-E descritos no capítulo anterior e, a partir da seção 6.3.1, os resultados obtidos com estes experimentos são apresentados e discutidos.

---

<sup>1</sup>Utilizou-se o *software MultiSpec* que apresenta uma implementação do classificador *MaxVer* (LANDGREBE; BI-EHL, 2001)

Os quatro experimentos, denominados de “Experimento 1”, “2”, “3” e “4”, fornecem indicativos sobre o comportamento dos classificadores BRE-E, BRE-OU, BRE-E/OU e BRG-E, respectivamente, mas, como utilizam amostras de teste e de treinamento particulares, não se pode generalizar as conclusões que são obtidas por tais experimentos.

Como se sabe, a validação de um classificador supervisionado costuma ser feita através da obtenção da matriz de confusão e da estimativa global de qualidade para pontos de teste, previamente selecionados. Sendo que, antes da classificação dos pontos de teste, o sistema deve ter sido treinado através de pontos de treinamento (seção 2.3.3).

Ocorre que a escolha das amostras de treinamento é um processo subjetivo e, portanto, varia de pessoa para pessoa, além do que, algumas amostras de treinamento poderiam favorecer um classificador em detrimento dos outros. Para que a comparação seja justa, será feita uma validação cruzada, ou seja, definem-se algumas amostras, denominadas de amostras de avaliação, para cada classe e faz-se um rodízio onde se escolhe, entre as amostras de avaliação, uma parte para testes e as demais para treinamento (ISHIBUCHI; YAMAMOTO, 2005). Cada vez que se faz o rodízio, executa-se a classificação dos *pixels* de teste (aqueles dentro da amostra de teste) e anota-se o desempenho do classificador. Quando todas as amostras de avaliação tiverem a chance de serem amostras de teste, encerra-se o processo e faz-se a média dos resultados obtidos. Desta forma, obtém-se o desempenho médio e o número médio de pontos não classificados para cada um dos classificadores testados.

Também já se comentou que o classificador BRE-E, atuando sozinho, e o BRE-OU, exclusivamente, não eram candidatos sérios para aplicações práticas. Então, foram propostos dois sistemas práticos: BRE-E/OU e BRG-E. Para validar os dois classificadores propostos, serão utilizados os cinco conjuntos de teste (seção 5) e o desempenho daqueles será comparado com o de dois classificadores usuais, o *fuzzy* convencional e o *MaxVer*.

A validação sugerida será feita através do “Experimento 5”, onde se escolhe, para cada classe, 7 amostras de avaliação com 81 *pixels* cada e a validação cruzada será feita utilizando-se todas as bandas espectrais disponíveis (7 bandas no primeiro conjunto de dados, 9 no segundo conjunto, 12 no terceiro, 5 no quarto e 3 no quinto).

A figura 6.1 mostra as amostras escolhidas para o primeiro conjunto de dados: as rotuladas com  $S0 \dots S6$  são amostras de soja, as rotuladas com  $C0 \dots C6$  são para o milho e as com  $A0 \dots A6$ , para a alfafa. Como se trabalha com 3 classes e 7 amostras por classe, serão necessárias  $7 \times 7 \times 7 = 343$  etapas de validação cruzada para que todas as amostras de avaliação, de todas as classes, tenham a oportunidade de se tornarem amostras de teste.

Na figura 6.2 são vistas as amostras de avaliação escolhidas para o segundo conjunto de dados:  $C0 \dots C6$  para o milho,  $S0 \dots S6$  para a soja e  $W0 \dots W6$  para a madeira. Também são necessárias 343 avaliações neste caso.



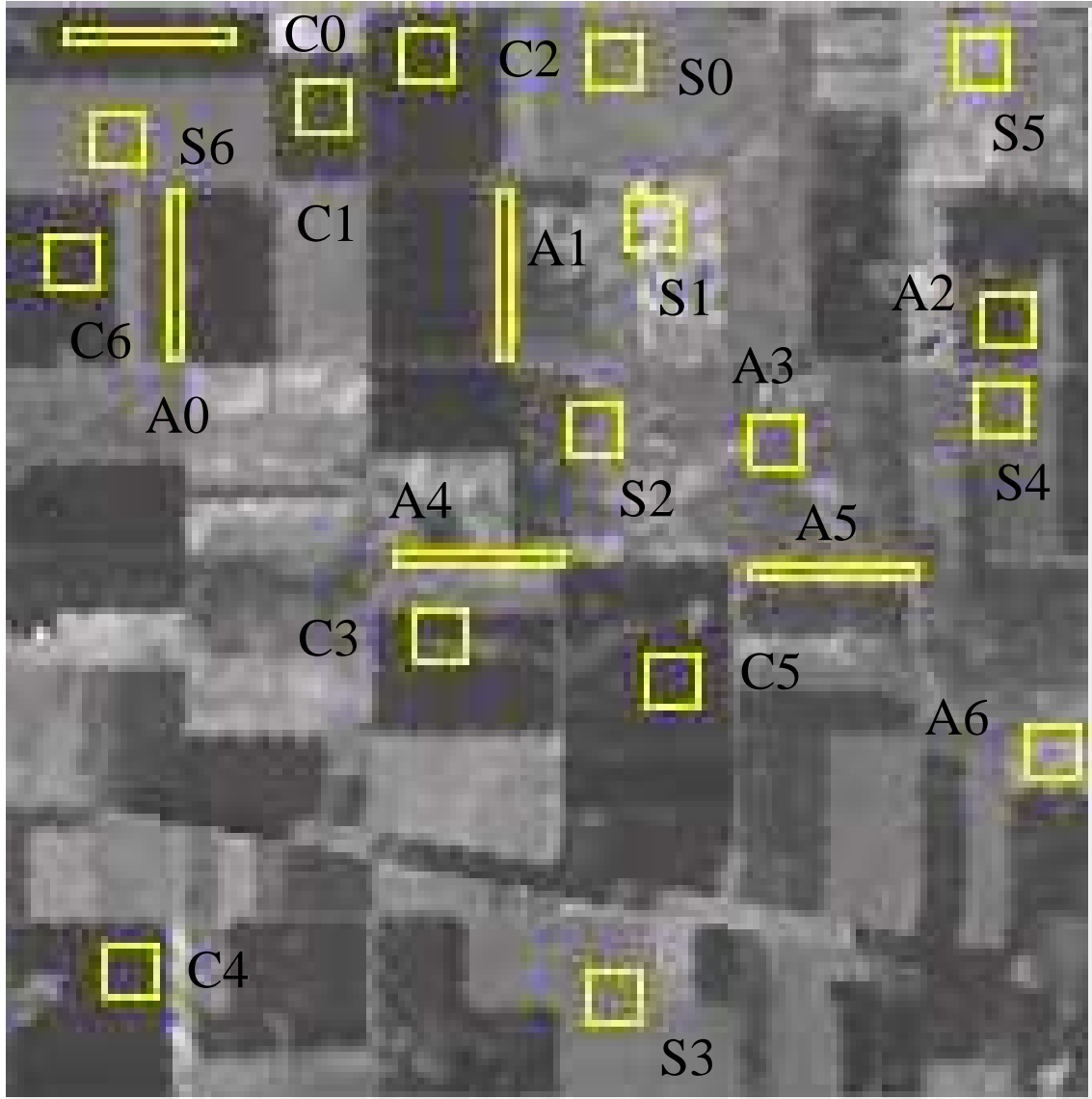


Figura 6.1: Amostras de avaliação para o primeiro conjunto de dados.

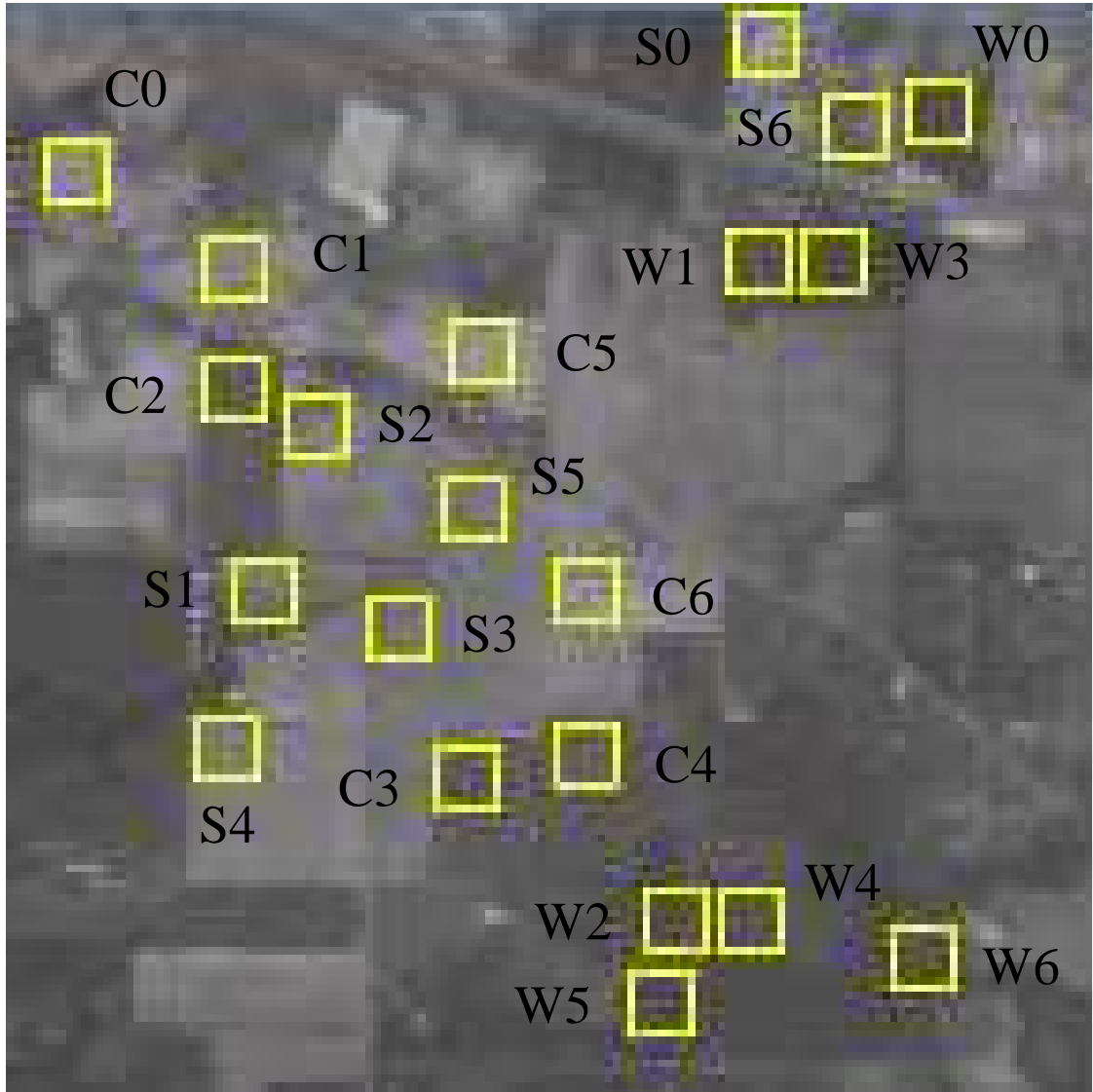


Figura 6.2: Amostras de avaliação para o segundo conjunto de dados.

As amostras para o terceiro conjunto são apresentadas na figura 6.3:  $C0 \dots C6$  para o milho,  $S0 \dots S6$  para a soja e  $W0 \dots W6$  para a aveia. Novamente, 343 etapas de validação são necessárias.

As amostras para o quarto conjunto são apresentadas na figura 6.4:  $R0 \dots R6$  para rocha,  $F0 \dots F6$  para floresta e  $G0 \dots G6$  para vegetação rasteira. Novamente, 343 avaliações são necessárias.

Finalmente, as amostras para o quinto conjunto são apresentadas na figura 6.5:  $F0 \dots F6$  para floresta,  $A0 \dots A6$  para asfalto e  $G0 \dots G6$  para a grama. Novamente, 343 etapas de avaliação são necessárias.

A seguir são apresentados e discutidos os resultados obtidos em cada um dos cinco experimentos.

## 6.3 Resultados e Discussões

### 6.3.1 Experimento 1

O “Experimento 1”, detalhado a seguir, fornecerá uma indicação de como o aumento do número de bandas influencia a precisão do classificador com base de regras específica e operador E (BRE-E) descrito na seção 4.2.3 e, também, como afeta o número de *pixels* não classificados (NC). Propõe-se treinar o classificador com as amostras da figura 4.2 e avaliá-lo com as amostras da figura 4.5, da seguinte forma: a) com a primeira banda, b) com as duas primeiras bandas, c) com as três primeiras, d) com as quatro, e) com as cinco, f) com as seis e, finalmente, g) com as sete bandas espectrais. O agrupamento de bandas não está levando em conta a capacidade destas para classificar os *pixels* de teste, i.e., as bandas não estão ordenadas por nenhum critério de qualidade (e.g.: distância de Bhattacharyya (LANDGREBE, 1998)) e estão sendo agrupadas simplesmente na ordem em que se apresentam, do menor para o maior comprimento de onda. O motivo de não se ordenar as bandas por critério de qualidade é evitar que tal critério privilegie um classificador em detrimento dos outros.

Como consequência do “Experimento 1”, serão construídos dois gráficos: 1) desempenho em função do número de bandas utilizadas e 2) NC em função do número de bandas. Através da regressão linear dos pontos obtidos nos dois gráficos, traçar-se-á uma reta que permitirá verificar como o desempenho varia em função do número de bandas, ou seja, se direta ou inversamente proporcional e, da mesma forma, como a quantidade de não-classificados varia com o número de bandas.

O “Experimento 1” produziu sete matrizes de confusão, uma para cada conjunto de bandas: a primeira, as duas primeiras, as três, as quatro, as cinco, as seis e, finalmente, para as sete bandas espectrais disponíveis.

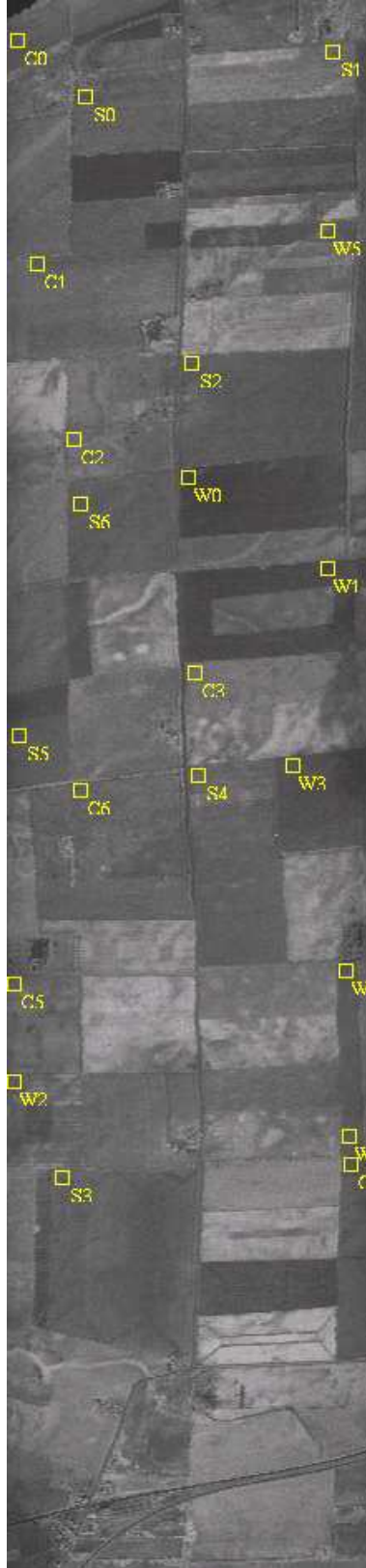


Figura 6.3: Amostras de avaliação para o terceiro conjunto de dados.

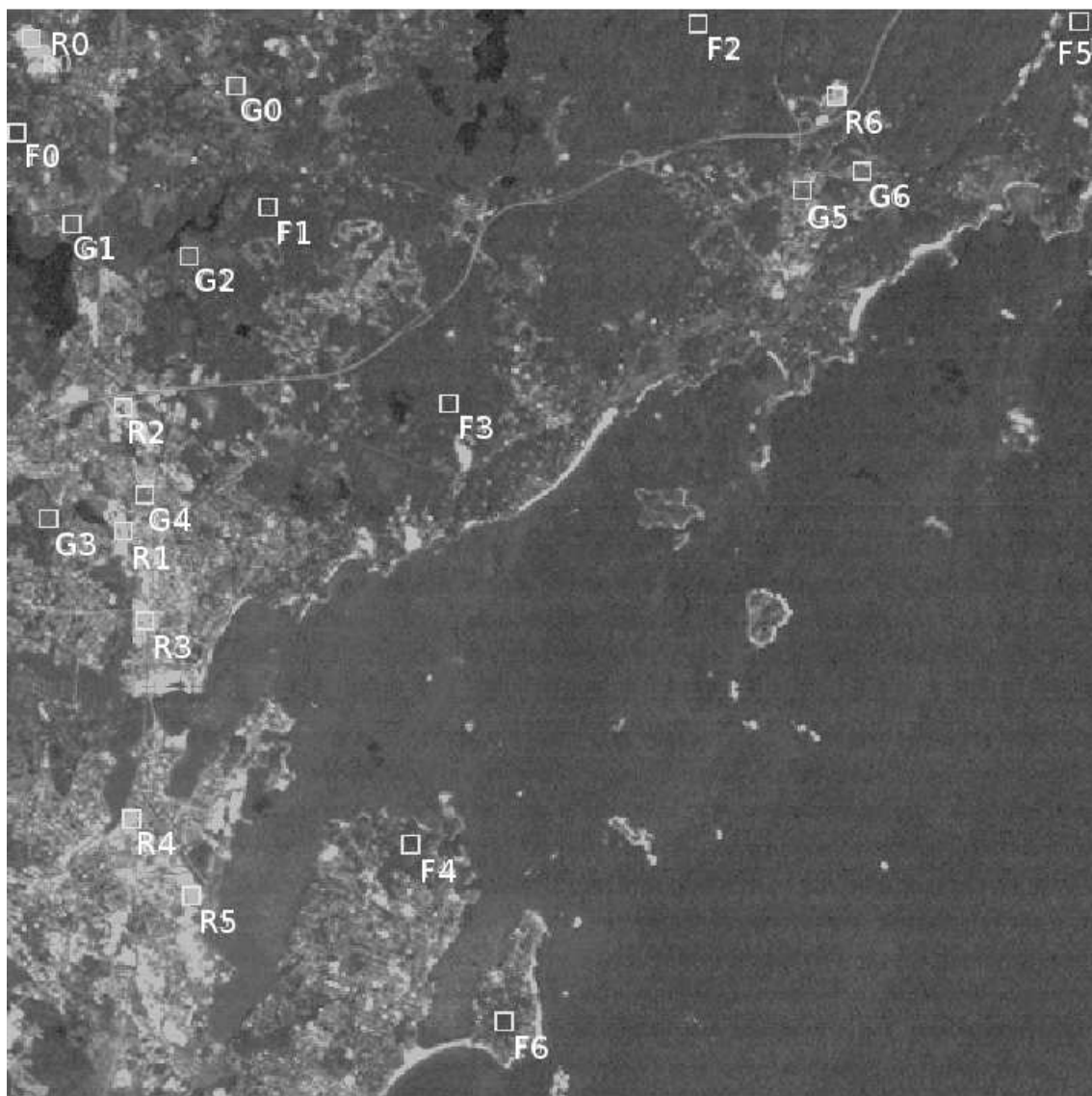


Figura 6.4: Amostras de avaliação para o quarto conjunto de dados.

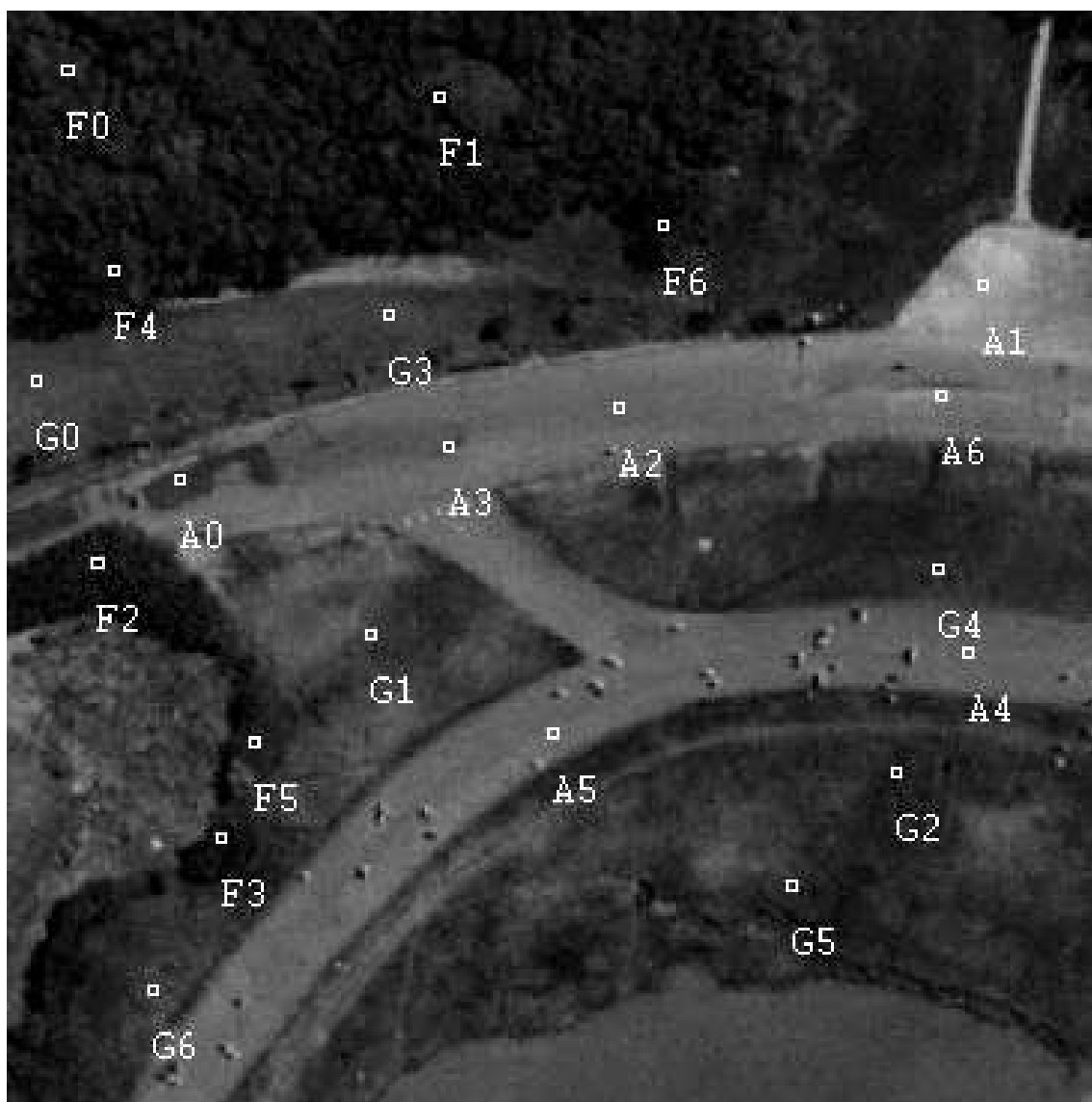


Figura 6.5: Amostras de avaliação para o quinto conjunto de dados.

A tabela 6.1 mostra a matriz de confusão obtida quando se utiliza apenas a primeira banda espectral. Nota-se, na última linha da tabela, que 8 *pixels* de alfafa permaneceram sem classificação. O total de *pixels* sem classificação foi, portanto de 8, que corresponde a 2,1% do total, que é de 388 *pixels*. Calculando-se a precisão, equação 2.1), obtém-se  $P = \frac{48+65+135}{48+50+16+39+65+16+8+3+135} = 0,65$ , o que indica que 65% do *pixels* foram corretamente classificados. Neste cálculo do desempenho, os não-classificados (NC) não foram contabilizados como erros de classificação, já que aqueles poderão ser reclassificados futuramente, utilizando-se outra técnica. Se os NC fossem contados como erro, o desempenho seria de 64%.

Tabela 6.1: Matriz de confusão para o classificador tipo-2 com base de regras específica e operador E para agregação dos antecedentes (BRE-E) utilizando apenas a primeira banda espectral.

	Soja	Milho	Alfafa	Não classificados	P
Soja	48	50	16	0	65%
Milho	39	65	16	0	
Alfafa	8	3	135	8	

A tabela 6.2 mostra a matriz para as duas primeiras bandas, onde se nota que 3 *pixels* de milho e 8 de alfafa permaneceram sem classificação. O total de *pixels* sem classificação foi, portanto, de 11, que corresponde a 2,8% do total e o desempenho global foi de 82%, neste caso.

Tabela 6.2: Matriz de confusão para o classificador tipo-2 BRE-E utilizando duas bandas espectrais.

	Soja	Milho	Alfafa	Não classificados	P
Soja	61	32	21	0	82%
Milho	1	106	10	3	
Alfafa	0	5	141	8	

Na tabela 6.3 pode-se ver a matriz para três bandas onde se nota que 3 *pixels* de soja, 4 de milho e 8 de alfafa permaneceram sem classificação, produzindo um total de 15 *pixels* não classificados (NC) ou 3,9% do total. Para este exemplo, 84% dos pontos de teste foram corretamente classificados.

Tabela 6.3: Matriz de confusão para o classificador tipo-2 BRE-E utilizando três bandas espectrais.

	Soja	Milho	Alfafa	Não classificados	P
Soja	63	31	17	3	84%
Milho	0	105	11	4	
Alfafa	0	0	146	8	

A tabela 6.4, para quatro bandas espectrais, mostra que 2 *pixels* de soja, 4 de milho e 8 de alfafa permaneceram sem classificação (NC foi de 14 ou 3,6%). O desempenho foi de 90%.

A tabela 6.5, para cinco bandas, mostra que 2 *pixels* de milho e 8 de alfafa permaneceram sem classificação (NC de 10 *pixels* ou 2,6%) e que o desempenho foi de 89%.

Tabela 6.4: Matriz de confusão para o classificador tipo-2 BRE-E utilizando quatro bandas espectrais.

	Soja	Milho	Alfafa	Não classificados	P
Soja	86	4	22	2	90%
Milho	0	105	11	4	
Alfafa	0	0	146	8	

Tabela 6.5: Matriz de confusão para o classificador tipo-2 BRE-E utilizando cinco bandas espectrais.

	Soja	Milho	Alfafa	Não classificados	P
Soja	84	0	30	0	89%
Milho	0	108	10	2	
Alfafa	0	0	146	8	

A tabela 6.6, para seis bandas, mostra que 1 *pixel* de milho e 8 de alfafa permaneceram sem classificação (NC de 9 ou 2,3%) e que o desempenho foi de 89%.

Tabela 6.6: Matriz de confusão para o classificador tipo-2 BRE-E utilizando seis bandas espectrais.

	Soja	Milho	Alfafa	Não classificados	P
Soja	84	0	30	0	89%
Milho	0	109	10	1	
Alfafa	0	0	146	8	

Finalmente, a tabela 6.7 mostra a matriz de confusão obtida no experimento 1 quando são utilizadas as sete bandas espectrais disponíveis. Nota-se que 1 *pixel* de soja e 15 de alfafa permaneceram sem classificação. O total de *pixels* sem classificação foi de 16, que corresponde a 4,1% do total (388 *pixels*) e a precisão foi  $P = 0,97$ , ou seja, 97% dos pontos de teste foram corretamente classificados. Na seção 4.2.5, conjecturou-se que, talvez, o classificador tipo-2 com base de regras específica e operador E (BRE-E) não fosse útil em aplicações práticas porque deixaria muitos pontos sem classificação. No exemplo apresentado, o NC foi de 4,1% que, dependendo da aplicação, poderia até ser tolerado.

Com base nas matrizes de confusão obtidas, foram traçados dois gráficos: o primeiro fornece um indicativo da variação do desempenho em função do número de bandas (número de fontes de evidência) que participam da classificação e o segundo mostra a variação da quantidade de *pixels* não-classificados.

A figura 6.6(a) mostra que o desempenho do classificador tende a aumentar quando mais fontes de evidência são utilizadas para se tomar a decisão, o que era esperado já que os antecedentes são agregados pelo operador conjuntivo E, que leva ao aumento do desempenho, uma vez que todas as evidências têm que classificar o *pixel* da mesma forma.

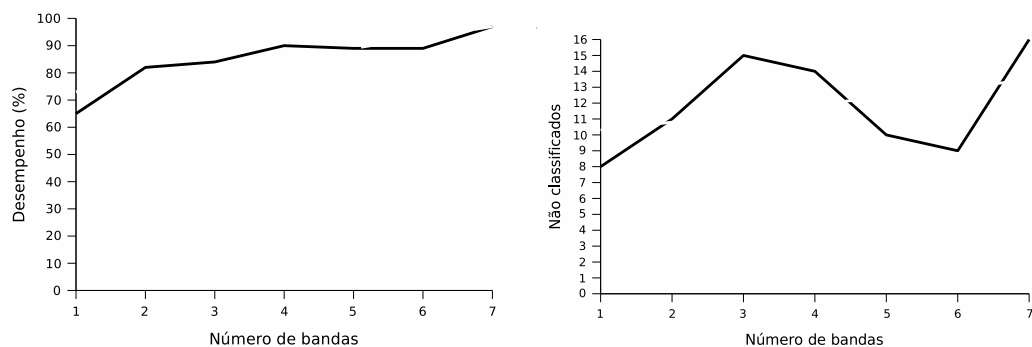
A figura 6.6(b) indica que há uma tendência de crescimento da quantidade de pontos não classificados (NC) com o aumento do número de bandas, como era de se esperar. O aumento do NC



Tabela 6.7: Matriz de confusão para o classificador tipo-2 BRE-E utilizando todas as bandas disponíveis (7 bandas).

	Soja	Milho	Alfafa	Não classificados	P
Soja	103	0	10	1	97%
Milho	0	119	1	0	
Alfafa	0	0	139	15	

é uma desvantagem do operador E, utilizado para agregar os antecedentes. Caso se desejasse trabalhar com NC pequeno, dever-se-ia diminuir o número de bandas participantes, com o ônus de se trabalhar com precisão menor.



(a) Variação do desempenho em função do número de bandas

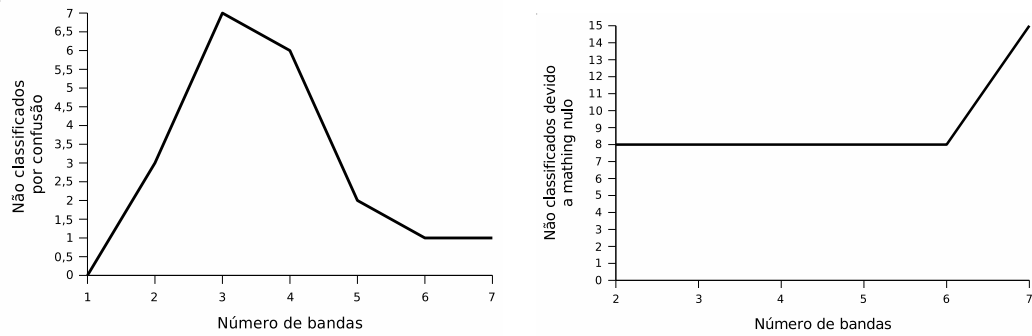
(b) Variação da quantidade de *pixels* não classificados em função do número de bandas

Figura 6.6: Desempenho e pontos não classificados de um classificador BRE-E (“Experimento 1”).

O gráfico que mostra a variação da quantidade de pontos não classificados (NC) com o aumento do número de bandas (figura 6.6(b)), embora apresente tendência de crescimento, possui um comportamento bastante irregular. Para compreender melhor a natureza destas oscilações, os *pixels* sem classificação foram identificados e separados em duas categorias: *pixels* de confusão, que ativam mais de uma regra com a mesma amplitude (seção 2), e *pixels* com *matchings* nulos, que não ativam nenhuma regra (seção 4.2.2). A partir desta identificação, foram construídos dois gráficos: o da figura 6.7(a), para o primeiro tipo, e o da figura 6.7(b), para o segundo.

Na figura 6.7(a) percebe-se que algumas bandas introduzem confusão no classificador (quando se passa de 1 para 2 bandas, ou de 2 para 3, no exemplo em questão) mas, a partir de uma certa quantidade de bandas (três, no exemplo), a dimensão do sistema é tal que a confusão tende a diminuir.

Na figura 6.7(b), pode-se ver que os *pixels* devidos aos *matchings* nulos variam de forma monotônica não-decrescente com o número de bandas utilizadas, como era esperado, uma vez que a agregação de novas bandas, além de não recuperar um *matching* nulo, pode acrescentar outro e isto ocorre porque a agregação dos antecedentes é feita pelo operador E, onde o valor nulo prevalece.



(a) Variação da quantidade de *pixels* de confusão não classificados em função do número de bandas

(b) Variação da quantidade de *pixels* de *matching* nulo não classificados em função do número de bandas

Figura 6.7: Os dois tipos de pontos não classificados apresentados por um classificador BRE-E (“Experimento 1”).

### 6.3.2 Experimento 2

Para que se tenha um indicativo de como o desempenho do classificador com base de regras específica e operador OU (BRE-OU), descrito na seção 4.2.4, é degradado com o uso do operador OU, e também de como a quantidade de não-classificados diminui com o aumento do número de bandas, sugere-se ensaiar o novo classificador através do “Experimento 2”, preparado da mesma forma que o “Experimento 1”, utilizando-se as amostras de treinamento da figura 4.2 e as amostras de teste da figura 4.5 e a mesma agregação de bandas (com a primeira banda, com as duas primeiras, com as três, etc).

O “Experimento 2”, que permite verificar o classificador tipo-2 BRE-OU, foi realizado como sugerido e foram obtidos dois gráficos (figura 6.8): o primeiro fornece um indicativo da variação do desempenho em função do número de bandas (número de fontes de evidência) que participam da classificação e o segundo mostra a variação da quantidade de *pixels* não-classificados.

A figura 6.8(a) mostra que o desempenho do classificador com base de regras específica e operador OU (BRE-OU) aumenta de forma muito lenta com o aumento do número de bandas que participam da decisão, o que ocorre porque a classificação está sendo feita apenas por uma banda (aquela cuja evidência possui o maior *matching*). Também é possível verificar que o desempenho deste classificador é menor do que aquele que utiliza o operador E (na figura 6.6) para qualquer número de bandas consideradas, com exceção de uma banda, onde só há um antecedente e, portanto, não há agregação e, neste caso, os dois sistemas são iguais e possuem o mesmo desempenho. Para sete bandas, por exemplo, o desempenho do E é de 97% e o do OU é de 76%.

A figura 6.8(b) evidencia a vantagem do classificador com operador OU sobre aquele que utiliza o operador E, ou seja, há uma forte diminuição da quantidade de pontos não classificados (NC) com o aumento do número de bandas.

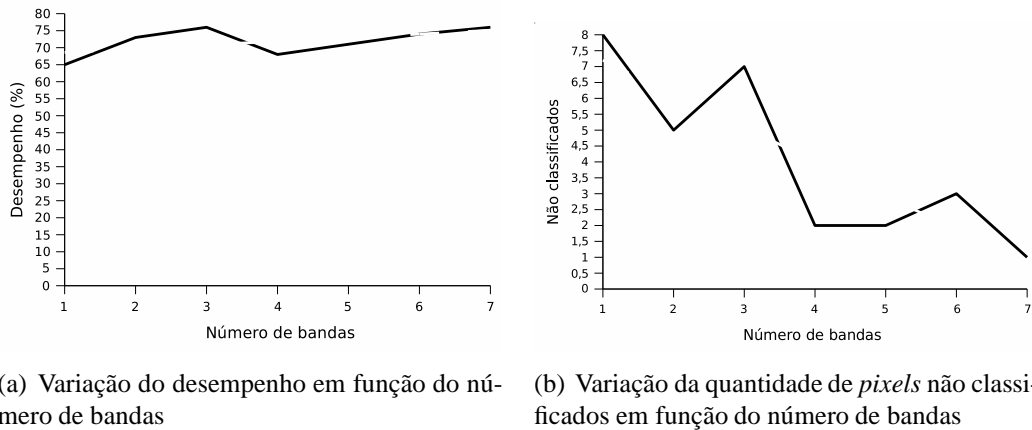


Figura 6.8: Desempenho e pontos não classificados de um classificador do tipo-2 BRE-OU (“Experimento 2”).

### 6.3.3 Experimento 3

Já o “Experimento 3”, conduzido da mesma forma que os “Experimentos 1” e “2”, descritos anteriormente, fornecerá um indicativo do desempenho do classificador com base de regras específica e operador E seguido de OU (BRE-E/OU) apresentado na seção 4.2.5.

O “Experimento 3”, que permite verificar o classificador tipo-2 com base de regras específica com operador E seguido de OU (BRE-E/OU), produziu dois gráficos (figura 6.9): o primeiro, como nos experimentos anteriores, fornece um indicativo da variação do desempenho em função do número de bandas (número de fontes de evidência) que participam da classificação e o segundo mostra a variação da quantidade de *pixels* não-classificados (NC).

A figura 6.9(a) mostra que o desempenho do classificador BRE-E/OU foi igual à obtida pelo operador E atuando sozinho (BRE-E), o que faz sentido porque inicialmente a classificação é feita pelo operador E e, somente a reclassificação dos NC é feita pelo operador OU, ou seja, a maior parte dos pontos é classificada pelo E nos dois sistemas. É claro que se a reclassificação feita pelo OU for muito ruim, com baixo índice de acerto, o desempenho do classificador BRE-E/OU será inferior ao obtido exclusivamente pelo E (esta condição não foi verificada no exemplo apresentado).

A figura 6.9(b) mostra que a reclassificação dos *pixels* pelo operador OU foi bem-sucedida e os NC foram zerados com apenas quatro bandas, o que é melhor do que o resultado apresentado pelo OU atuando sozinho (figura 6.8(b)). Esperava-se incorporar ao classificador E a capacidade do OU em zerar os NC com o aumento do número de evidências (bandas) consideradas e o objetivo foi alcançado. Imagina-se que o fato do BRE-E/OU ter conseguido zerar os *pixels* sem classificação antes do OU, atuando sozinho, tenha sido um caso particular para as amostras de teste e de treinamento utilizadas e não se imagina que este resultado possa ser generalizado.

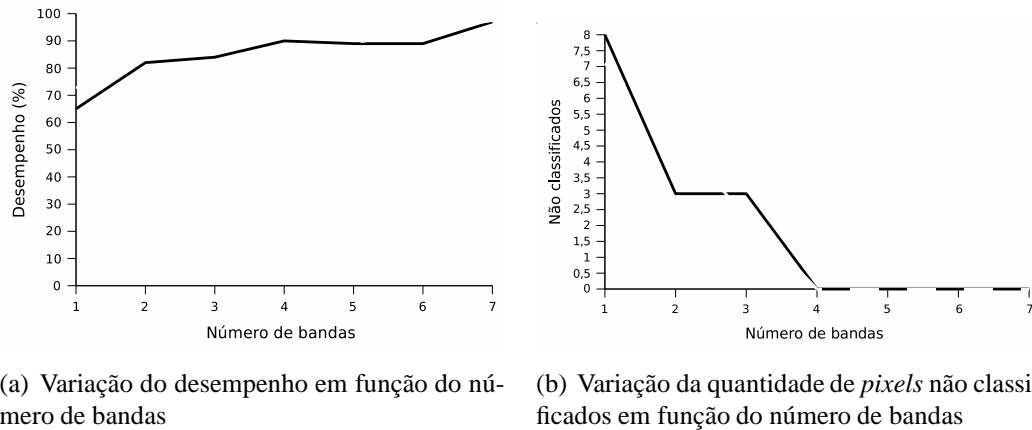


Figura 6.9: Desempenho e pontos não classificados de um classificador do tipo-2 com base de regras específica e operador E seguido de OU, BRE-E/OU (“Experimento 3”).

### 6.3.4 Experimento 4

Finalmente, o “Experimento 4” utilizará as mesmas amostras de treinamento e de testes empregadas no “Experimento 1”, para fornecer um indicativo sobre o desempenho do classificador com base de regras genérica e operador E (BRG-E) descrito na seção 4.2.6.

O “Experimento 4”, utilizado para verificar o classificador tipo-2 com base de regras genérica e operador E (BRG-E), produziu dois gráficos (figura 6.10): o primeiro dá uma indicação sobre a variação do desempenho em função do número de bandas que participam da classificação e o segundo mostra a variação da quantidade de *pixels* não-classificados (NC).

A figura 6.10(a) mostra que o desempenho do classificador BRG-E foi igual à obtida pelo BRE-E/OU que, por sua vez, era igual ao do E atuando sozinho (BRE-E). Os três classificadores apresentaram o mesmo desempenho (seções 6.3.1, 6.3.3 e 6.3.4) porque o estágio inicial de classificação é o mesmo para os três (agregação dos antecedentes utilizando o operador E e todas as bandas disponíveis) e, além disso, este estágio foi capaz de classificar a maior parte dos *pixels* de teste. Possivelmente, para amostras de treinamento e de testes diferentes das utilizadas, a quantidade de pontos classificados pelo estágio inicial diminua e a quantidade reclassificada pelos demais estágios aumente, o que diferenciaria o desempenho dos classificadores.

A figura 6.10(b) mostra que a reclassificação dos *pixels* pelo operador E com número decrescente de bandas (base de regras genérica) foi bem-sucedida e os NC seriam zerados com oito bandas, o que é um pouco pior do que o BRE-E/OU, que zerava o NC com apenas quatro bandas (figura 6.9(b)).

Comparando-se os experimentos “3” e “4”, poder-se-ia concluir que o BRE-E/OU é melhor do que o BRG-E porque, embora o desempenho tenha sido o mesmo, os NC do primeiro classificador foram zerados mais rapidamente. Ocorre, porém, que estes experimentos utilizaram um conjunto

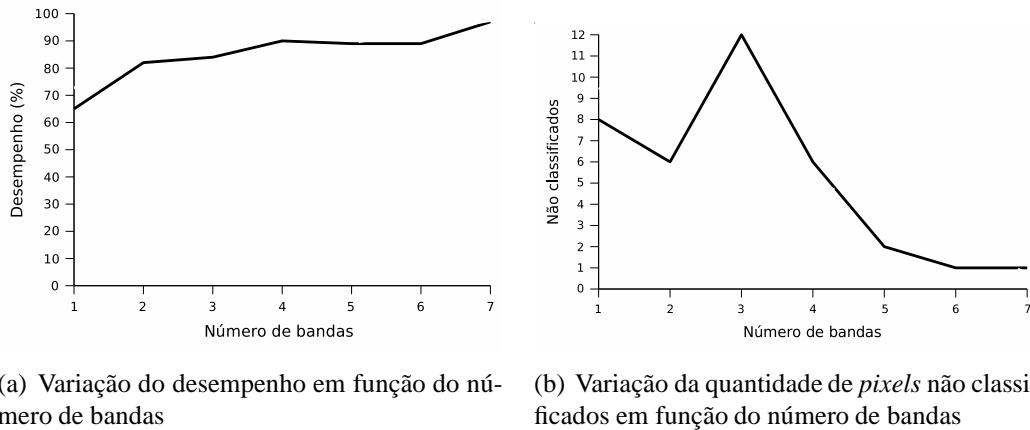


Figura 6.10: Desempenho e pontos não classificados de um classificador do tipo-2 com base de regras genérica e operador E, BRG-E (“Experimento 4”).

de amostras de teste e de treinamento particulares e aquela conclusão não pode ser generalizada, sendo necessário um experimento adicional.

### 6.3.5 Experimento 5

O “Experimento 5”, utilizado para validar os classificadores de comparação foi realizado da forma sugerida na seção 6.2. Foram testados quatro classificadores: o classificador *MaxVer*, um *fuzzy* convencional (T1), o BRE-E/OU e o BRG-E, para os cinco conjuntos de dados (Primeiro, Segundo, Terceiro, Quarto e Quinto) apresentados na seção 5.

O “Experimento 5” produziu a tabela 6.8 que mostra o número médio de *pixels* não classificados apresentado por cada um dos quatro classificadores testados: o *MaxVer*, o *fuzzy* convencional (T1), o BRE-E/OU e o BRG-E.

Tabela 6.8: Comparação entre o número de *pixels* não classificados para os quatro classificadores testados: *MaxVer*, *fuzzy* convencional (T1), BRE-E/OU e BRG-E.

Conjunto de dados	Número de bandas	<i>MaxVer</i>	T1	BRE-E/OU	BRG-E
Primeiro	7	0	1	0	1
Segundo	9	0	1	0	1
Terceiro	12	0	2	0	1
Quarto	5	0	1	0	0
Quinto	3	0	0	0	0

Como já discutido, o classificador tipo-2 BRE-E, embora apresentasse bom desempenho, não era útil em aplicações práticas porque deixava muitos pontos sem classificação (seção 4.2.5). Para resolver este problema, foram propostas duas alternativas do tipo-2: a) com base de regras específica e operador E seguido de OU (BRE-E/OU) e b) com base de regras genérica e operador E (BRG-E). A tabela 6.8 mostra que as alternativas foram eficientes ao reduzir os não-classificados a valores próximos de zero.

A tabela 6.9 apresenta os resultados obtidos no “Experimento 5” onde se nota que o *MaxVer* teve o melhor desempenho para o primeiro conjunto de dados e que o tipo-2 com base de regras genérica (BRG-E) foi o melhor para o segundo, terceiro, quarto e quinto conjuntos.

Tabela 6.9: Comparação entre as precisões (P) dos classificadores: *MaxVer*, *fuzzy* convencional (T1), BRE-E/OU e BRG-E.

Conjunto de dados	Número de bandas	<i>MaxVer</i>	T1	BRE-E/OU	BRG-E
Primeiro	7	0,8643	0,7568	0,7745	0,7763
Segundo	9	0,8455	0,8424	0,8715	0,8740
Terceiro	12	0,8120	0,8398	0,8714	0,8747
Quarto	5	0,9542	0,9373	0,9681	0,9683
Quinto	3	0,8351	0,8285	0,8492	0,8497

### 6.3.5.1 Avaliação estatística

Para que se possa fazer uma análise mais aprofundada da tabela 6.9 utilizou-se o método estatístico proposto por Demsar (2006) (ver também Sheskin (2003)) que permite a comparação de vários classificadores utilizando conjuntos de dados diferentes. O método em questão visa avaliar se os algoritmos que estão sendo comparados são estatisticamente diferentes e, se forem e houver mais de dois algoritmos sendo comparados, quais destes apresentam tal diferença no desempenho.

Demsar (2006) começa a explanação comentando que testes paramétricos, tais como o ANOVA, não podem ser utilizados para comparar classificadores pois estes testes assumem que os dados estejam organizados segundo uma determinada distribuição e tal premissa nem sempre é verdadeira.

Desta forma, a análise começa com o teste de Friedman (DEMSAR, 2006) que é um equivalente não paramétrico do ANOVA. Friedman sugere que se organize os algoritmos segundo um ranking para cada conjunto de dados separadamente, atribuindo-se 1 ao primeiro colocado, 2 ao segundo, e assim sucessivamente. O teste compara os rankings médios dos algoritmos, assumindo, como hipótese nula, a igualdade de todos os classificadores.

Partindo-se da tabela 6.9 e definindo-se os rankings dos classificadores para cada conjunto de dados, pode-se calcular o ranking médio, conforme mostra a tabela 6.10. A primeira linha da tabela, por exemplo, mostra que, para o primeiro conjunto de dados, o *MaxVer* foi o melhor classificador (o que apresenta menor ranking), o tipo-2 com base de regras genérica e operador E (BRG-E) foi o segundo colocado, o tipo-2 com base de regras específica e operador E seguido de OU (BRE-E/OU) foi o terceiro colocado e o tipo-1 foi o pior. A última linha da tabela apresenta o ranking médio dos classificadores e permite ordená-los levando-se em conta todos os conjuntos de dados utilizados. Desta forma, obtém-se o desempenho final dos classificadores: o BRG-E foi o melhor, o BRE-E/OU foi o segundo colocado, o *MaxVer* foi o terceiro e o T1 foi o último.

Tabela 6.10: Comparação entre os rankings médios dos classificadores: *MaxVer*, *fuzzy* convencional (T1), BRE-E/OU e BRG-E.

Conjunto de dados	Número de bandas	<i>MaxVer</i>	T1	BRE-E/OU	BRG-E
Primeiro	7	1	4	3	2
Segundo	9	3	4	2	1
Terceiro	12	4	3	2	1
Quarto	5	3	4	2	1
Quinto	3	3	4	2	1
Ranking médio		2,8	3,8	2,2	1,2

Para verificar se a diferença existente entre os rankings médios dos classificadores é significativa calcula-se a estatística de Friedman através da equação (6.1)

$$\chi_F^2 = \frac{12N}{k(k+1)} \left[ \sum_j R_j^2 - \frac{k(k+1)^2}{4} \right] \quad (6.1)$$

onde  $N$  é o número de conjuntos de dados,  $k$  é a quantidade de classificadores e  $R_j$  é o ranking médio para o  $j$ -ésimo classificador, ou seja,

$$\chi_F^2 = \frac{12 \cdot 5}{4 \cdot (4+1)} \cdot \left[ 2,8^2 + 3,8^2 + 2,2^2 + 1,2^2 - \frac{4 \cdot (4+1)^2}{4} \right] = 10,68$$

Como a estatística de Friedman é considerada muito conservadora, Demsar (2006) sugere o uso da estatística  $F_F$

$$F_F = \frac{(N-1)\chi_F^2}{N(k-1) - \chi_F^2} \quad (6.2)$$

que está distribuída de acordo com a distribuição  $F$  de Snedecor (FONSECA; MARTINS, 1996) com  $k-1$  e  $(k-1) \cdot (N-1)$  graus de liberdade, logo

$$F_F = \frac{(5-1) \cdot 10,68}{5(4-1) - 10,68} = 9,89$$

Comparando-se  $F_F = 9,89$  com o valor tabelado  $F(3, 12) = 3,49$  (FONSECA; MARTINS, 1996), para 3 e 12 graus de liberdade, conclui-se que a hipótese nula deve ser rejeitada ( $F_F > F(3, 12)$ ), o que quer dizer que existe diferença significativa entre os classificadores comparados.

A rejeição da hipótese nula mostra que há diferença entre os classificadores comparados mas ainda resta verificar quais, dentre os classificadores comparados, são diferentes e, para isso, utiliza-se o teste de Nemenyi (DEMSAR, 2006) que utiliza a distância crítica (CD: *Critical Distance*)

$$CD = q_{\alpha} \cdot \frac{k(k+1)}{6N} \quad (6.3)$$

onde  $q_{\alpha}$  são valores críticos tabelados (DEMSAR, 2006). Obtém-se, portanto:

$$CD = 2,560 \cdot \frac{4(4+1)}{6 \cdot 5} = 2,1$$

ao nível de significância de 95%.

A distância crítica é a diferença mínima que os rankings médios de dois classificadores devem possuir para que os últimos sejam considerados estatisticamente diferentes.

De acordo com a sugestão de Demsar (2006) foi construído um gráfico para facilitar a compreensão dos resultados obtidos. A figura 6.11 mostra tal gráfico que apresenta os classificadores posicionados num eixo horizontal de acordo com o valor do seu ranking médio, sendo que o melhor deles é posicionado à direita (BRG-E) e o pior à esquerda (T1). Acima do eixo horizontal é mostrada a distância crítica (CD) e, além disso, os classificadores cujos rankings médios possuam distâncias menores do que a crítica são conectados através de uma linha mais grossa formando grupos. Percebe-se a formação de dois grupos: 1) T1, *MaxVer* e BRE-E/OU e 2) *MaxVer*, BRE-E/OU e BRG-E.

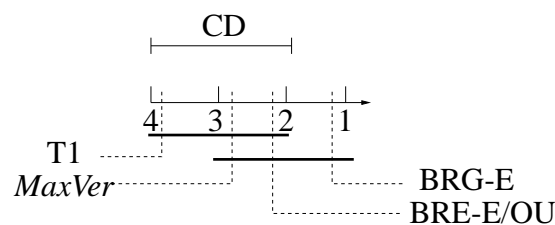


Figura 6.11: Classificadores ordenados pelos rankings médios e agrupados pela distância crítica.

O cálculo da distância crítica revelou que o classificador BRG-E tem desempenho significativamente melhor do que o classificador do tipo-1 já que, como se pode verificar na figura 6.11, não há linha grossa que os conecte. Como mencionado, foi possível identificar a existência de dois grupos de classificadores: o primeiro que inclui o classificador do tipo-1 e o segundo que inclui o BRG-E. Não se pode dizer a qual dos dois grupos o *MaxVer* e o BRE-E/OU pertencem, já que concluir que eles pertencem a ambos os grupos seria uma contradição estatística (DEMSAR, 2006), uma vez que um indivíduo não pode vir de duas populações diferentes.

A afirmação estatística correta seria, portanto, que não foi possível concluir que o BRG-E é superior ao BRE-E/OU e ao *MaxVer*. Na verdade, o gráfico 6.11 apresenta uma incerteza devida ao valor da distância crítica e, se esta incerteza não existisse, as duas linhas grossas vistas no gráfico não teriam sobreposição, ou seja, pode-se afirmar que o gráfico 6.11 admite 3 interpretações:



1. BRG-E, BRE-E/OU e *MaxVer* podem ter o mesmo desempenho e os três serem superiores ao T1

Como mostra a linha grossa que conecta os classificadores BRG-E, BRE-E/OU e *MaxVer* na figura 6.12, estes teriam desempenhos similares e os três seriam superiores ao T1.

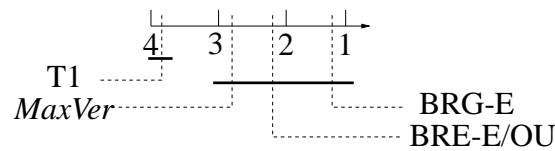


Figura 6.12: Interpretação 1: BRG-E, BRE-E/OU e *MaxVer* teriam o mesmo desempenho e os três seriam superiores ao T1.

2. BRG-E e BRE-E/OU seriam superiores a *MaxVer* e a T1

Na figura 6.13, as duas linhas grossas mostram que BRG-E e BRE-E/OU seriam equivalentes entre si e que *MaxVer* e T1 o seriam entre eles e, além disso, que os primeiros seriam superiores aos segundos.

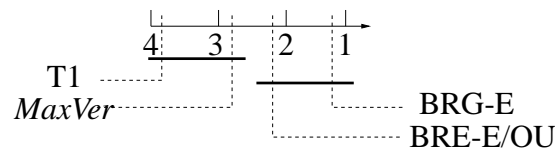


Figura 6.13: Interpretação 2: BRG-E e BRE-E/OU seriam superiores a *MaxVer* e a T1.

3. BRG-E seria superior a BRE-E/OU, *MaxVer* e T1

A linha grossa da figura 6.14 mostra que BRE-E/OU, *MaxVer* e T1 seriam equivalentes e que o BRG-E seria superior aos três primeiros.

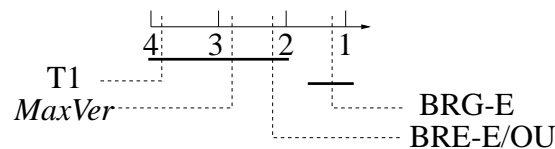


Figura 6.14: Interpretação 3: BRG-E seria superior a BRE-E/OU, *MaxVer* e T1.

Como visto, os experimentos não permitiram concluir que o BRG-E é superior ao BRE-E/OU e ao *MaxVer* e que havia três interpretações possíveis. Foi possível mostrar que o aumento no grau de *fuzzificação* de um classificador *fuzzy* produz um sistema melhor, já que ficou demonstrado que o classificador do tipo-2 BRG-E possui desempenho significativamente melhor do que o do *fuzzy* convencional (T1).

## 6.4 Conclusão

Neste capítulo, foram feitos testes com os quatro classificadores projetados no capítulo anterior: BRE-E, BRE-OU, BRE-E/OU e BRG-E. Tais testes permitiram observar algumas características dos classificadores testados. O BRE-E, por exemplo, permite aumentos no desempenho através da agregação de novas fontes de evidência ao processo de decisão mas, em contrapartida, a quantidade de *pixels* não classificados também aumenta. O BRE-OU, embora tenha apresentado desempenho inferior ao do BRE-E, permite obter poucos pontos não classificados (NC). O BRE-E/OU e o BRG-E uniram as vantagens do BRE-E e do BRE-OU, ou seja, produziram bom desempenho com baixo NC. Novamente, vale salientar que o objetivo destes testes não foi o de obter conclusões gerais sobre os classificadores ensaiados mas sim de verificar certas características destes.

Para generalizar as conclusões, ou seja, para, de fato, dizer qual classificador é melhor e qual é pior, foi feita a validação cruzada dos classificadores BRE-E/OU e BRG-E utilizando-se cinco conjuntos de dados distintos com culturas agrícolas e regiões de floresta. Os classificadores ensaiados foram comparados com os de referência, *MaxVer* e *fuzzy* tipo-1 e, após a realização dos testes de significância estatística, Friedman e Nemenyi, verificou-se que o classificador com base de regras genérica e operador E (BRG-E) foi significativamente melhor do que o tipo-1 e, portanto, justifica-se aumentar o grau de *fuzzificação* de um classificador visando obter desempenhos melhores. Não foi possível afirmar que o BRG-E é melhor do que o *MaxVer*.

## 7 *Conclusões e trabalhos futuros*

Nesta tese, comentou-se sobre a importância dos classificadores de cobertura vegetal para monitorar a degradação das áreas de preservação como florestas, mangues e matas ciliares e para acompanhar o avanço da fronteira agrícola (cana-de-açúcar, soja, trigo, arroz e milho) sobre as áreas de preservação.

Comentou-se, também, que já existem, em uso, bons classificadores estatísticos, como o *MaxVer* e os classificadores *fuzzy* convencionais.

Foi dito que estes classificadores usuais possuíam limitações: o *MaxVer* necessitava de um especialista e o *fuzzy* convencional não modelava adequadamente a incerteza, que pode ser de dois tipos: incerteza de medição (ruído) e incerteza no significado dos termos linguísticos (nível semântico).

Para superar as limitações dos classificadores existentes, começaram a surgir os classificadores *fuzzy* do tipo-2 que, quando comparados com os classificadores *fuzzy* convencionais, representam uma elevação no nível de *fuzzificação* dos últimos, permitindo tratar os dois tipos de incerteza citados.

Logo, foi necessário um estudo aprofundado da teoria dos conjuntos *fuzzy* gerais do tipo-2 que abordou diversos tópicos, tais como, os conjuntos *fuzzy* gerais do tipo-2 e suas especificidades e os sistemas de inferência utilizados até então. Além disso, foi preciso estudar conceitos da área de domínio, como assinatura espectral, resolução espacial de uma imagem de satélite, reflectância e as métricas utilizadas para aferir o desempenho dos classificadores digitais.

Para projetar o classificador pretendido, foi necessária a formalização de um sistema de inferência que trabalhasse com conjuntos gerais do tipo-2 e, durante este processo, contribuições importantes foram dadas, entre elas, o “método de classificação hierárquica” com base de regras específica e genérica, a “mancha” de incerteza geral, a translação de uma fatia vertical, a extensão cilíndrica de um conjunto *fuzzy* geral do tipo-2, o supremo de um destes conjuntos e, finalmente, a *defuzzificação* através dos centróides das fatias verticais (*VSCTR*).

Para validar o mecanismo de inferência, utilizaram-se cinco conjuntos de dados distintos contendo culturas agrícolas e regiões de floresta para testar de forma mais ampla o novo classificador,

nas suas duas principais configurações: com base de regras específica e operador E seguido de OU (BRE-E/OU) e com base de regras genérica e operador E (BRG-E). Os resultados obtidos pelo novo sistema foram comparados com dois classificadores de referência bastante usuais: o *MaxVer* e o classificador *fuzzy* tradicional. Concluiu-se, através do uso dos testes de significância estatística, Friedman e Nemenyi, que o BRG-E foi significativamente melhor do que *fuzzy* do tipo-1 mas não foi possível afirmar que o BRG-E é melhor do que o *MaxVer*. Limitando-se aos conjuntos *fuzzy*, pode-se afirmar que o aumento no grau de *fuzzificação* de um classificador, ou seja, a transição do tipo-1 para o tipo-2, produziu um sistema cujo desempenho é melhor. Além disso, fizeram-se, a partir dos experimentos, algumas observações interessantes:

- o operador E produz sistemas mais precisos ao acrescentar novas fontes de evidência ao processo de decisão
- o classificador hierárquico permite a reclassificação de pontos não classificados
- é possível construir um classificador com poucas regras.

Em suma, foi proposto, desenvolvido e testado um sistema de inferência baseado na teoria dos conjuntos *fuzzy* gerais do tipo-2. Esta inferência, quando aplicada na classificação de cobertura vegetal, mostrou-se eficaz, o que foi corroborado por testes utilizando cinco conjuntos de dados distintos. Convém salientar que a metodologia proposta não se limita unicamente a esta aplicação, podendo ser utilizada em quaisquer problemas de classificação. Além disso, a formalização proposta (principalmente a utilização da “mancha” de incerteza) facilitaria o entendimento do uso de conjuntos do tipo-2 gerais em outras áreas onde os sistemas *fuzzy* vêm sendo aplicados com sucesso, tais como, o controle de plantas industriais (LEE, 1990; WAN; KAMAL, 2006; PAREEK; KAR, 2006), por exemplo.

Os experimentos realizados mostraram que o desempenho do *MaxVer* foi excepcionalmente bom para o primeiro conjunto de dados, sendo bastante superior aos classificadores *fuzzy*, do tipo-1 e do tipo-2. Esta superioridade não se repetiu nos demais quatro conjuntos de dados, onde os classificadores do tipo-2 (BRE-E/OU e BRG-E) foram superiores. Imagina-se que exista uma particularidade no primeiro conjunto de dados que favoreça o classificador estatístico *MaxVer*.

Desta forma pretende-se, em trabalhos futuros, fazer um estudo aprofundado do primeiro conjunto de dados buscando tentar identificar singularidades que favoreçam o *MaxVer*. Sendo possível fazer esta identificação, pretende-se dotar o classificador do tipo-2 de recursos (talvez novas regras ou formas diferentes de se construir os antecedentes) que o tornem competitivo também para estes dados.

Outra questão a ser verificada no futuro diz respeito ao uso dos conjuntos do tipo-2 intervalares. Na sequência dos trabalhos deverá ser investigada a hipótese de que a simplificação dos conjuntos tipo-2 gerais para intervalares traz perda de desempenho.

Nesta tese, a validação cruzada foi feita utilizando-se seis amostras para treinamento e uma para testes (cada amostra possuía 81 *pixels*). Pretende-se, em trabalhos futuros, testar novas proporções para treinamento e testes.

## *Referências Bibliográficas*

ABDULGHAFOR, M.; CHANDRA, T.; ABIDI, M. A. Data fusion through fuzzy logic applied to feature extraction from multi-sensory images. In: *Int'l Conf. Robotics and Automation*. [S.l.]: IEEE, 1993. v. 2, p. 359–366.

ALI-AKBAR; SHARIFI, M. A.; MULDER, N. J. Likelihood-based image segmentation and classification: a framework for the integration of expert knowledge in image classification procedures. *International Journal of Applied Earth Observation and Geoinformation*, v. 2, n. 3, p. 104–119, 2000.

BARATA, T.; PINA, P. Improving classifications rates by modelling the clusters of trainings sets in features space using mathematical morphology operators. In: *16th IEEE Conference on Pattern Recognition*. [S.l.: s.n.], 2002. v. 4, p. 328–331.

BLOCH, I. Information combination operators for data fusion: A comparative review with classification. *IEEE Transactions on Systems, Man, and Cybernetics — Part A: Systems and Humans*, v. 26, n. 1, p. 52–67, 1996.

BÁRDOSSY, A.; SAMANIEGO, L. Fuzzy rule-based classification of remotely sensed imagery. *IEEE Transactions on Geoscience and Remote Sensing*, v. 40, n. 2, p. 362–374, 2002.

CENTENO, J. A. S. *Sensoriamento Remoto e Processamento de Imagens Digitais*. [S.l.]: Editora UFPR, 2004.

CHEN, X. et al. A multi-SVM fusion model using type-2 FLS. In: *IEEE International Conference on Fuzzy Systems*. Vancouver, Canada: [s.n.], 2006. p. 6611–6615.

CONGALTON, R. G. *Exploring a Satellite Image*. University of New Hampshire, Fevereiro 2008. Disponível em: <<http://www.globe.unh.edu>>. Acesso em: 17 jan. 2009.

COUPLAND, S.; JOHN, R. An investigation into alternative methods for the defuzzification of an interval type-2 fuzzy set. In: *IEEE International Conference on Fuzzy Systems*. Vancouver, Canada: [s.n.], 2006. p. 7196–7203.

COUPLAND, S.; JOHN, R. Geometric type-1 and type-2 fuzzy logic systems. *IEEE Transactions on Fuzzy Systems*, v. 15, n. 1, p. 3–15, fev 2007.

COUPLAND, S.; JOHN, R. On the accuracy of type-2 fuzzy sets. In: *IEEE International Conference on Fuzzy Systems*. London, UK: [s.n.], 2007. p. 131–136.

DELGADO, M. R. D. B. da S. *Projeto Automático de Sistemas Nebulosos: Uma Abordagem Co-Evolutiva*. Tese (Doutorado) — Unicamp, 2002.

DEMSAR, J. Statistical comparisons of classifiers over multiple data sets. *Journal of Machine Learning Research*, v. 7, p. 1–30, 2006.

- DUDA, R. O.; HART, P. E.; STORK, D. G. *Pattern Classification*. 2. ed. [S.l.]: Wiley Interscience, 2001.
- FAHSI, A. et al. Incorporation of digital elevation models with landsat-TM data to improve land cover classification accuracy. *Forest Ecology and Management – Elsevier*, v. 128, p. 57–64, 2000.
- FONSECA, J. S. da; MARTINS, G. de A. *Curso de Estatística*. 6. ed. São Paulo: Atlas S.A., 1996.
- HSIEH, P.-F.; LEE, L. C.; CHEN, N.-Y. Effect of spatial resolution on classification errors of pure and mixed pixels in remote sensing. *IEEE Transactions on Geoscience and Remote Sensing*, v. 39, n. 12, p. 2657–2663, 2001.
- INNOCENT, P. R.; JOHN, R. I. Type-2 fuzzy diagnosis. In: *IEEE International Conference on Fuzzy Systems*. [S.l.: s.n.], 2002. v. 2, p. 1326–1330.
- ISHIBUCHI, H.; NAKASHIMA, T.; MURATA, T. Performance evaluation of fuzzy classifier systems for multidimensional pattern classification problems. *IEEE Transactions on Systems, Man, and Cybernetics — Part B: Cybernetics*, v. 29, n. 5, p. 601–618, 1999.
- ISHIBUCHI, H.; YAMAMOTO, T. Comparison of heuristic criteria for fuzzy rule selection in classification problems. *Fuzzy Optimization and Decision Making*, v. 3, n. 2, p. 119–139, 2004.
- ISHIBUCHI, H.; YAMAMOTO, T. Rule weight specification in fuzzy rule-based classification systems. *IEEE Transactions on Fuzzy Systems*, v. 13, n. 4, p. 428–435, 2005.
- ISHIBUCHI, H.; YAMAMOTO, T.; NAKASHIMA, T. An approach to fuzzy default reasoning for function approximation. *Soft Computing*, v. 10, n. 9, p. 850–864, 2006.
- JOHN, R.; MENDEL, J.; CARTER, J. The extended sup-star composition for type-2 fuzzy sets made simple. In: *IEEE International Conference on Fuzzy Systems*. Vancouver, Canada: [s.n.], 2006. p. 7212–7216.
- KARNIK, N. K.; MENDEL, J. M. Operations on type-2 fuzzy sets. *Fuzzy Sets and Systems*, v. 122, p. 327–348, 2001.
- KARNIK, N. K.; MENDEL, J. M.; LIANG, Q. Type-2 fuzzy logic systems. *International Journal on Fuzzy Systems*, v. 7, n. 6, p. 643–658, December 1999.
- KARNIK, N. N.; MENDEL, J. M.; LIANG, Q. Type-2 fuzzy logic systems: type-reduction. In: *IEEE International Conference on Systems, Man, and Cybernetics*. [S.l.: s.n.], 1998. p. 2046–2051.
- KETTIG, R. L.; LANDGREBE, D. A. Classification of multispectral image data by extraction and classification of homogeneous objects. *IEEE Transactions on Geoscience Electronics*, GE-14, n. 1, p. 19–26, 1976.
- KLEMENT, E.; MESIAR, R.; PAP, E. *Triangular Norms*. [S.l.]: Kluwer Academic Publishers, 2000.
- KLIR, G. J.; FOLGER, T. A. *Fuzzy Sets, Uncertainty, and Information*. [S.l.]: Prentice Hall, 1988.
- LANDGREBE, D. Hyperspectral image data analysis as a high dimensional signal processing problem. *IEEE Signal Processing Magazine*, v. 19, n. 1, p. 17–28, 2002.
- LANDGREBE, D.; BIEHL, L. *An Introduction to Multispec*. West Lafayette, 2001. Disponível em: <<http://dynamo.ecn.purdue.edu/~biehl/MultiSpec>>. Acesso em: 01 set. 2008.

- LANDGREBE, D.; JIMENEZ, L. Supervised classification in high dimensional space: Geometrical, statistical and asymptotical properties of multivariate data. *IEEE Transactions on Systems, Man, and Cybernetics. Part C, number 1*, v. 28, p. 39–54, 1998.
- LANDGREBE, D. A. *Multispectral Data Analysis: A Moderate Dimension Example*. West Lafayette, Janeiro 1997. Disponível em: <<http://dynamo.ecn.purdue.edu/~biehl/MultiSpec>>. Acesso em: 01 set. 2008.
- LANDGREBE, D. A. *Multispectral Data Analysis: A Signal Theory Perspective*. West Lafayette, Abril 1998. Disponível em: <<http://dynamo.ecn.purdue.edu/~biehl/MultiSpec>>. Acesso em: 01 set. 2008.
- LEE, C.; LANDGREBE, D. A. Analyzing high dimensional multispectral data. *IEEE Transactions on Geoscience and Remote Sensing*, v. 31, n. 4, p. 792–800, 1993.
- LEE, C. C. Fuzzy logic in control systems: Fuzzy logic controller — part I. *IEEE Transactions on Geoscience and Remote Sensing*, v. 20, n. 2, p. 404–418, 1990.
- LIANG, Q.; MENDEL, J. M. Interval type-2 fuzzy logic systems. In: *IEEE International Conference on Fuzzy Systems*. [S.l.: s.n.], 2000. p. 328–333.
- LIANG, Q.; MENDEL, J. M. Interval type-2 fuzzy logic systems: theory and design. *International Journal on Fuzzy Systems*, v. 8, n. 5, p. 535–550, 2000.
- LIANG, Q.; MENDEL, J. M. MPEG VBR video traffic modeling and classification using fuzzy technique. *IEEE Transactions on Fuzzy Systems*, v. 9, n. 1, p. 183–193, February 2001.
- LUCAS, L.; CENTENO, T.; DELGADO, M. General type-2 fuzzy inference systems: Analysis, design and computational aspects. In: *IEEE International Conference on Fuzzy Systems*. London, UK: [s.n.], 2007. p. 1107–1112.
- LUCAS, L.; CENTENO, T.; DELGADO, M. General type-2 fuzzy classifiers to land cover classification. In: *ACM Symposium on Applied Computing*. Ceará, Brazil: [s.n.], 2008. p. 1743–1747.
- LUCAS, L. A.; CENTENO, T. M.; DELGADO, M. R. Land cover classification based on general type-2 fuzzy classifiers. *International Journal of Fuzzy Systems*, v. 10, n. 3, p. 207–216, set 2008.
- MAULIK, U.; BANDYOPADHYAY, S. Fuzzy partitioning using a real-coded variable-length genetic algorithm for pixel classification. *IEEE Transactions on Geoscience and Remote Sensing*, v. 41, n. 5, p. 1075–1081, May 2003.
- MENDEL, J. M. *Uncertain Rule-Based Fuzzy Logic Systems*. [S.l.]: Prentice Hall PTR, 2000.
- MENDEL, J. M.; JOHN, R. I. B. Type-2 fuzzy sets made simple. *IEEE Transactions on Fuzzy Systems*, v. 10, n. 2, p. 117–127, April 2002.
- MENDEL, J. M.; LIU, F. Super-exponential convergence of the karnik-mendel algorithms used for type-reduction in interval type-2 fuzzy logic systems. In: *IEEE International Conference on Fuzzy Systems*. Vancouver, Canada: [s.n.], 2006. p. 6603–6610.
- MITCHELL, H. B. Pattern recognition using type-II fuzzy sets. *Information Sciences*, v. 170, p. 409–418, 2005.



- PAREEK, U.; KAR, I. N. Estimating compressor discharge pressure of gas turbine power plant using type-2 fuzzy logic systems. In: *IEEE International Conference on Fuzzy Systems*. Vancouver, Canada: [s.n.], 2006. p. 3113–3118.
- PARQUE Barigüi. 2004. Secretaria Municipal do Meio Ambiente. Disponível em: <[www.curitiba.pr.gov.br](http://www.curitiba.pr.gov.br)>. Acesso em: 21 dec. 2004.
- PEDRYCZ, W.; GOMIDE, F. *An Introduction to Fuzzy Sets*. [S.l.]: The MIT Press, 1998.
- RICHARDS, J. A.; JIA, X. *Remote Sensing Digital Image Analysis*. [S.l.]: Springer-Verlag, 1999.
- ROEMER, M. J.; KACPRZYNSKI, G. J.; ORSAGH, R. F. Assessment of data and knowledge fusion strategies for prognostics and health management. In: *Aerospace conference*. [S.l.]: IEEE, 2001. v. 6, p. 2979–2988.
- ROSENFELD, A. From image analysis to computer vision: An annotated bibliography, 1955-1979. *Computer Vision and Image Understanding*, v. 84, p. 298–324, 2001.
- SCHIEWE, J. Integration of multi-sensor data for landscape modeling using a region-based approach. *ISPRS Journal of Photogrammetry & Remote Sensing*, v. 57, p. 371–379, 2003.
- SCHOWENGERDT, R. A. *Techniques for Image Processing and Classification in Remote Sensing*. [S.l.]: Academic Press, Inc., 1983.
- SHESKIN, D. J. *Handbook of Parametric and Nonparametric Statistical Procedures*. 3. ed. [S.l.]: Chapman & Hall/CRC, 2003.
- SOLAIMAN, B.; PIERCE, L. E.; ULABY, F. T. Multisensor data fusion using fuzzy concepts: Application to land-cover classification using ers-1/jers-1 sar composites. *IEEE Transactions on Geoscience and Remote Sensing*, v. 37, n. 3, p. 1316–1326, May 1999.
- TAHAYORI, H.; TETTAMANZI, A. G.; ANTONI, G. D. Approximated type-2 fuzzy set operations. In: *IEEE International Conference on Fuzzy Systems*. Vancouver, Canada: [s.n.], 2006. p. 9042–9049.
- TAN, W. W.; FOO, C. L.; CHUA, T. W. Type-2 fuzzy system for ECG arrhythmic classification. In: *IEEE International Conference on Fuzzy Systems*. London, UK: [s.n.], 2007. p. 859–864.
- USTIN, S. L. et al. Using imaging spectroscopy to study ecosystem processes and properties. *BioScience*, v. 54, n. 6, p. 523–534, 2004.
- WAN, T. W.; KAMAL, D. H. On-line learning rules for type-2 fuzzy controller. In: *IEEE International Conference on Fuzzy Systems*. Vancouver, Canada: [s.n.], 2006. p. 2530–2537.
- WFC, XII World Forestry Congress. Québec City, Canada.: [s.n.], 21 a 28 de Set. 2003. Disponível em: <<http://www.cfm2003.org>>. Acesso em: 02 Set. 2003.
- WU, C. hsin; DOERSCHUK, P. C. Tree aproximations to markov random fields. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, v. 17, n. 4, p. 391–400, 1995.
- ZADEH, L. A. The concept of a linguistic variable and its application to approximate reasoning-I. *Information Sciences*, v. 8, p. 199–249, 1975.
- ZENG, H.; LIU, Z.-Q. Type-2 fuzzy sets for pattern recognition: The state-of-the-art. *Journal of Uncertain Systems*, v. 1, n. 3, p. 163–177, 2007. Disponível em: <[www.jus.org.uk](http://www.jus.org.uk)>. Acesso em: 27 ago. 2008.

ZHUKOV, B. et al. Unmixing-based multisensor multiresolution image fusion. *IEEE Transactions on Geoscience and Remote Sensing*, v. 37, n. 3, p. 1212–1226, 1999.

## RESUMO:

Propõe-se, nesta tese, o desenvolvimento de uma nova ferramenta baseada em conjuntos *fuzzy* gerais do tipo-2 para aplicação em processos de classificação digital de dados. O problema de classificação a ser considerado está relacionado à identificação de regiões de floresta em imagens de satélite com o objetivo de auxiliar em tarefas de monitoramento florestal. O classificador digital desenvolvido utiliza um mecanismo de inferência denominado de “inferência escalonada *fuzzy* geral do tipo-2” para classificar os *pixels* das imagens de satélite de acordo com sua cobertura vegetal. Tal classificador é inovador pois, além de utilizar conjuntos *fuzzy* tipo-2 gerais, pode utilizar tanto uma base de regras específica quanto uma base genérica (ambas de forma hierárquica) para reclassificar pontos que, do contrário, permaneceriam sem classificação. Isto permite a obtenção de uma base de regras compacta (composta de poucas regras). A justificativa para o uso de sistemas de inferência do tipo-2 é que estes, apesar do custo computacional maior, apresentam desempenho superior aos sistemas do tipo-1 equivalentes. Os testes realizados mostram que, de fato, o sistema proposto é melhor do que o classificador *fuzzy* convencional usualmente empregado em aplicações semelhantes e possui desempenho comparável ao classificador estatístico da máxima verossimilhança, sendo uma alternativa viável ao último.

## PALAVRAS-CHAVES

Conjuntos *fuzzy* gerais do tipo-2; sistemas de inferência *fuzzy*; classificador de cobertura vegetal.

## ÁREAS/SUB-ÁREAS DO CONHECIMENTO

1.03.00.00 — 7 Ciência da Computação

1.07.02.06 — 7 Sensoriamento Remoto

2009

Nº: 41