

UNIVERSIDADE FEDERAL DE SÃO CARLOS
CENTRO DE CIÊNCIAS EXATAS E TECNOLOGIA
PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIA DA COMPUTAÇÃO

**COMBINAÇÃO DE MÚLTIPLOS CLASSIFICADORES PARA
IDENTIFICAÇÃO DE MATERIAIS EM IMAGENS RUIDOSAS**

MOACIR PEREIRA PONTI JUNIOR

Dissertação apresentada ao Programa de Pós-Graduação em Ciência da Computação do Centro de Ciências Exatas e Tecnologia da Universidade Federal de São Carlos, como parte dos requisitos para a obtenção do título de Mestre em Ciência da Computação.

Campo de pesquisa: Reconhecimento de Padrões.

67 p.

SÃO CARLOS
2004

**Ficha catalográfica elaborada pelo DePT da
Biblioteca Comunitária da UFSCar**

P811cm

Ponti Junior, Moacir Pereira.

Combinação de múltiplos classificadores para
identificação de materiais em imagens ruidosas / Moacir
Pereira Ponti Junior. -- São Carlos : UFSCar, 2004.
67 p.

Dissertação (Mestrado) -- Universidade Federal de São
Carlos, 2004.

1. Processamento de imagens. 2. Reconhecimento de
padrões. 3. Combinação de classificadores. 4. Tomografia
computadorizada. I. Título.

CDD: 006.4 (20^a)

Agradecimentos

A Deus pela vida e pela criação do mundo de forma tão magnífica e perfeita;

Ao Nelson, meu orientador, por guiar-me entre os caminhos da pesquisa com grande conhecimento e uma memória infalível;

A Embrapa, especialmente ao Professor Dr. Paulo Estevão Cruvinel pela aquisição e orientações acerca das imagens utilizadas no trabalho;

Ao Murillo pelas dúvidas solucionadas, revisões, conversas sobre o projeto, métodos e resultados;

A todos os professores do PPG-CC;

A Paula pelo enorme apoio antes e durante este curso;

A meus pais e aos queridos familiares pelo apoio incessante;

À família da Casa de Jesus;

Aos colegas do laboratório GAPIS;

A todos os alunos da turma 2002, principalmente aos amigos: Anderson, Pablo, Eduardo, Ricardo, Bel, Lucas, Matheus, Lúcio, Isabel, Taciana, etc. pela amizade e companheirismo, sem os quais seria muito mais difícil terminar este curso;

Aos colegas das outras turmas, conquistados durante o mestrado;

Aos amigos Emiliano, Eugênio, Baltazar e todos outros que de alguma forma ajudaram e me deram forças durante todo o curso e na execução deste trabalho.

Sumário

1. INTRODUÇÃO	1
1.1 ORGANIZAÇÃO DO TEXTO	3
2. TOMOGRAFIA COMPUTADORIZADA	4
2.1 PRINCÍPIOS BÁSICOS.....	4
2.1.1 COEFICIENTE DE ATENUAÇÃO LINEAR	5
2.2 RECONSTRUÇÃO DE IMAGENS.....	6
2.2.1 APLICAÇÃO DIRETA DA TRANSFORMADA DE FOURIER	6
2.2.2 RETROPROJEÇÃO FILTRADA	8
2.3 IMAGENS MULTIESPECTRAIS DE TOMOGRAFIA COMPUTADORIZADA ..	8
3. CLASSIFICAÇÃO	10
3.1 RECONHECIMENTO DE PADRÕES	10
3.2 CLASSIFICADORES	11
3.2.1 CLASSIFICADOR BAYESIANO LINEAR	12
3.2.2 CLASSIFICADOR LOGÍSTICO.....	14
3.2.3 CLASSIFICADOR PARZEN	16
3.2.4 CLASSIFICADOR k-VIZINHOS MAIS PRÓXIMOS	18
3.3 ESTIMAÇÃO DE PARÂMETROS PARA AS TÉCNICAS DE CLASSIFICAÇÃO PARZEN E k-VIZINHOS MAIS PRÓXIMOS.....	19
3.4 AVALIAÇÃO DO DESEMPENHO DOS CLASSIFICADORES.....	20
3.4.1 VALIDAÇÃO CRUZADA	20
3.4.2 COEFICIENTE KAPPA	21
4. COMBINAÇÃO DE MÚLTIPLOS CLASSIFICADORES	23
4.1 CARACTERÍSTICAS.....	23
4.2 FUNDAMENTOS TEÓRICOS.....	25
4.2.1 REGRA DO PRODUTO	28
4.2.2 REGRA DA SOMA	29
4.2.3 DERIVAÇÕES DAS REGRAS DE PRODUTO E SOMA.....	30
4.2.4 REGRA DO MÁXIMO	30
4.2.5 REGRA DO MÍNIMO	30
4.2.6 REGRA DA MEDIANA	31
4.2.7 REGRA DA VOTAÇÃO POR MAIORIA.....	31
4.2.8 VOTAÇÃO DOS COMBINADORES.....	31
5. MATERIAL E MÉTODOS	33
5.1 AQUISIÇÃO DAS IMAGENS	33
5.1.1 TOMÓGRAFO	33
5.1.2 CORPO DE PROVA (PHANTOM)	34
5.1.3 IMAGENS OBTIDAS E CARACTERÍSTICAS.....	35
5.2 CLASSIFICAÇÃO DAS IMAGENS	38

5.2.1	FERRAMENTAS UTILIZADAS	38
5.2.2	CLASSIFICAÇÃO DAS IMAGENS.....	39
6.	RESULTADOS	41
6.1	AVALIAÇÃO DOS CLASSIFICADORES	42
6.1.1	COEFICIENTE KAPPA DOS CLASSIFICADORES E COMBINADORES	42
6.1.2	ESTIMAÇÃO DE ERRO POR VALIDAÇÃO CRUZADA DOS CLASSIFICADORES E COMBINADORES	44
6.2	IMAGENS CLASSIFICADAS	46
6.2.1	IMAGEM 40 keV	46
6.2.2	IMAGEM 60 keV	47
6.2.3	IMAGEM 85 keV	47
6.2.4	IMAGEM 662 keV	48
6.2.5	IMAGEM 40-60 keV	48
6.2.6	IMAGEM 40-85 keV	49
6.2.7	IMAGEM 40-662 keV	49
6.2.8	IMAGEM 60-85 keV	50
6.2.9	IMAGEM 60-662 keV	50
6.2.10	IMAGEM 85-662 keV	51
6.2.11	IMAGEM 40-60-85 keV	51
6.2.12	IMAGEM 40-60-662 keV	52
6.2.13	IMAGEM 40-85-662 keV	52
6.2.14	IMAGEM 60-85-662 keV	53
6.2.15	IMAGEM 40-60-85-662 keV	53
7.	DISCUSSÃO	54
8.	CONCLUSÕES.....	62
8.1	TRABALHOS FUTUROS	64
	REFERÊNCIAS BIBLIOGRÁFICAS	65

Índice de Figuras

FIGURA 2.1 – Modelo Usado na Reconstrução da Imagem $f(x,y)$	7
FIGURA 3.2 – Esquema de Estimação de Erro Utilizando <i>Leave-One-Out Cross-Validation</i>	21
FIGURA 4.1 – Diagrama de um Sistema de Combinação de Classificadores.....	26
FIGURA 5.1 – Minitomógrafo	34
FIGURA 5.2 – Diagrama da Construção do Corpo de Prova (<i>Phantom</i>) Imageado.....	35
FIGURA 5.3 – Curvas do Coeficiente de Atenuação Linear em Função da Energia para a Água, Alumínio, Cálcio e Fósforo.	36
FIGURA 5.4 – Imagens Individuais Adquiridas Com Energias de 40, 60, 85 e 662 KEV.	37
FIGURA 6.1 – Comparação dos Classificadores e Combinadores pela Média do Kappa.....	43
FIGURA 6.2 – Comparação dos Classificadores e Combinadores pela Média do Erro	45
FIGURA 6.3 – Imagens Temáticas para a Classificação da Banda de 40 KEV	46
FIGURA 6.4 – Imagens Temáticas para a Classificação da Banda de 60 KEV	47
FIGURA 6.5 – Imagens Temáticas para a Classificação da Banda de 85 KEV	47
FIGURA 6.6 – Imagens Temáticas para a Classificação da Banda de 662 KEV	48
FIGURA 6.7 – Imagens Temáticas para a Classificação das Bandas de 40 e 60 KEV	48
FIGURA 6.8 – Imagens Temáticas para a Classificação das Bandas de 40 e 85 KEV	49
FIGURA 6.9 – Imagens Temáticas para a Classificação das Bandas de 40 e 662 KEV	49
FIGURA 6.10 – Imagens Temáticas para a Classificação das Bandas de 60 e 85 KEV	50
FIGURA 6.11 – Imagens Temáticas para a Classificação das Bandas de 60 e 662 KEV	50
FIGURA 6.12 – Imagens Temáticas para a Classificação das Bandas de 85 e 662 KEV	51
FIGURA 6.13 – Imagens Temáticas para a Classificação das Bandas de 40, 60 e 85 KEV ...	51
FIGURA 6.14 – Imagens Temáticas para a Classificação das Bandas de 40, 60 e 662 KEV .	52
FIGURA 6.15 – Imagens Temáticas para a Classificação das Bandas de 40, 85 e 662 KEV .	52
FIGURA 6.16 – Imagens Temáticas para a Classificação das Bandas de 60, 85 e 662 KEV .	53
FIGURA 6.17 - Imagens Temáticas para a Classificação das Bandas 40, 60, 85 e 662 KEV .	53
FIGURA 7.1 – Divisão do Espaço de Atributos por Diferentes Classificadores: (a) Bayesiano Linear, (b) Logístico, (c) K-Vizinhos, (d) Parzen	56
FIGURA 7.2 – Divisão do Espaço de Atributos pelos Combinadores: (a) Produto (maior performance), (b) Mínimo (menor performance).....	57
FIGURA 7.3 – Divisão do Espaço de Atributos da Imagem de 60 e 662 KEV Pelos Combinadores: (a) Soma, (b) Máximo	59

Índice de Tabelas

TABELA 3.1 – Interpretação para Resultados do Coeficientes Kappa.	22
TABELA 5.1 – Coeficientes de Atenuação Linear em Função das Energias 40, 60, 85 E 662 KEV Para a Água, Alumínio, Cálcio e Fósforo.	36
TABELA 5.2 – Cores Associadas às Classes nas Imagens Temáticas	39
TABELA 6.1 – Coeficientes Kappa dos Classificadores para Cada Imagem.....	42
TABELA 6.2 – Coeficientes Kappa dos Combinadores para Cada Imagem.....	43
TABELA 6.3 – Estimação do Erro dos Classificadores para Cada Imagem	44
TABELA 6.4 – Estimação do Erro dos Combinadores para Cada Imagem.....	45

Resumo

A identificação de materiais em imagens vem sendo explorada por diversas áreas para aplicações muito interessantes. Neste trabalho utilizamos imagens multiespectrais ruidosas obtidas por tomografia computadorizada adquiridas com múltiplas energias no propósito de estudos de ciência do solo. Através das imagens foi possível desenvolver uma sistema de reconhecimento de padrões de forma a identificar os materiais do corpo imageado. Para esta tarefa foram utilizadas diversas técnicas de classificação. Os classificadores individuais: Parzen, k-vizinhos mais próximos, logístico e Bayesiano linear foram combinados de modo a estudar o comportamento de técnicas de combinação de classificadores. Foram utilizados os combinadores de regra fixa: votação por maioria, máximo, mínimo, mediana., soma e produto. Foi também proposto e utilizado um segundo estágio de combinação, utilizando a votação dos combinadores. O desempenho dos classificadores foi analisado através da estimação de erro pelo método da validação cruzada *leave-one-out* e pelo coeficiente Kappa. Foram demonstradas as vantagens da utilização de múltiplas energias nos problemas de identificação de imagens e estudados os comportamentos de cada método de combinação. Os resultados apontaram que a combinação de classificadores possibilita melhor capacidade de generalização e resultados mais estáveis que os classificadores individuais, aproveitando as informações fornecidas por todos os classificadores individuais, inclusive os mais fracos, sendo recomendada na classificação de dados escassos ou difíceis em que apresentam ambigüidade ou altos índices de ruído.

Abstract

Material identification in images has been explored in multiple areas and very interesting applications are arising in this field. This work uses noisy multispectral images from a computerized tomograph scanner acquired with multiple energies for soil sciences applications and develops a recognition system to identify materials on the scanned body. Techniques of statistical classification were used. The individual classifiers: Parzen, k-nearest neighbors, logistic and linear Bayesian were combined in order to study the behavior of classifier combination techniques. For this task, we used the fixed rules combiners: majority voting, maximum, minimum, median, sum and product. Also, a second stage of combination was considered and used, the majority voting of combiners. The performance of the classifiers was analyzed through the leave-one-out cross-validation error estimation method and the Kappa coefficient. The advantages of the use of multiple energies in the problems of identification of images and the behavior of each combination method are also demonstrated. The results pointed out that the combination of classifiers gives better capacity of generalization and more stable results than the individual classifiers, using information supplied for all individual classifiers, including the weakest one, being recommended in classification of scarce, difficult discrimination data, on the presence of ambiguity or high noise levels.

1. Introdução

Aprendizado é o processo pelo qual é adquirido o conhecimento sobre o meio, enquanto a memória é o processo pelo qual este conhecimento é codificado, armazenado e depois resgatado (KANDEL; SCHWARTZ, 1987). O aprendizado é parte determinante do desenvolvimento humano, começa na vida intra-uterina e estende-se até a morte. É resultado da integração de ações pré-programadas da espécie, dos estímulos recebidos do ambiente e impostos pelas tarefas e das reações do organismo e suas conseqüências. Ao longo desse processo, todas as informações recebidas pelo indivíduo são interpretadas e associadas a um significado. Ocorre, a partir daí, a fixação do conhecimento pela memória e a capacidade de associá-lo a diferentes estímulos. (SHUMWAY-COOK; WOOLLACOTT, 2003)

Reconhecimento de Padrões é o estudo de como as máquinas podem observar o ambiente e aprender a distinguir padrões de interesse entre uma série de informações. Em quase 50 anos de pesquisas mantém sua característica multidisciplinar que permite explorar aplicações em diversas áreas (JAIN et al., 2000). Desde então muitas técnicas surgiram, mas o desafio até hoje permanece em generalizar as técnicas, de modo a superar as dificuldades e disponibilizar aos usuários ferramentas sólidas para trabalhar.

O desenvolvimento de um sistema de reconhecimento de padrões pode ser realizado em diversas abordagens, cujas mais conhecidas são: estatística, estrutural, sintática e redes neurais e envolve essencialmente três aspectos: aquisição de dados e pré-processamento, representação dos dados e tomada de decisão. Na abordagem estatística, cada padrão é representado em termos de n atributos ou medidas e é visto como um ponto em um espaço n -dimensional. Para classificar, divide-se este espaço de atributos utilizando um método de classificação, gerando limites de decisão de forma a separar os dados. Quando há um problema de classificação muito difícil, com poucos dados ou altas taxas de ruído, o uso de apenas um classificador torna-se insuficiente. Para aproveitar as vantagens de diversos esquemas de classificação, foi proposto combinar múltiplos classificadores para melhorar a performance. As primeiras formulações teóricas de modelos de combinação surgiram por volta de 1990. No início os classificadores eram utilizados seqüencialmente, de forma que

cada um oferecesse informação complementar ao outro (NADAL et. al., 1990). As primeiras aplicações voltaram-se para reconhecimento de escrita manual, uma tarefa difícil, para a qual diversos métodos são propostos (XU et. al., 1992). Desde então, muitas formas de combinar classificadores foram desenvolvidas, e uma das mais intuitivas é através de medidas estatísticas em que, por exemplo, a partir dos resultados das classificações individuais (probabilidades *a posteriori*), atribui-se o padrão à classe mais votada, ou à probabilidade *a posteriori* mais próxima da média (KITTLER et. al., 1998).

É possível pensar em muitas aplicações que utilizem técnicas de classificação e muitas são as informações possíveis de se classificar, entre elas imagens digitais. O uso de imagens para classificação é usado em medicina (diagnósticos clínicos em raios-X, ultrassonografia, etc.), biotecnologia (sequenciamento genético em chips de DNA), biologia (reconhecimento de microorganismos e células), sensoriamento remoto (classificação de terrenos, detecção de fenômenos meteorológicos, etc.). As imagens objeto de classificação podem ser obtidas por diferentes sensores, desde câmeras comuns até aparelhos de tomografia computadorizada.

A Tomografia Computadorizada (TC) é uma técnica muito interessante, pois utiliza um método não-destrutivo para obter informações sobre a estrutura de um corpo. A tomografia computadorizada iniciou-se em 1972 com as pesquisas de Godfrey Hounsfield, que, em 1979, foi laureado com o prêmio Nobel de medicina por suas pesquisas (Allan Cormack dividiu o prêmio com Hounsfield por ter descoberto alguns algoritmos independentemente). A evolução dos computadores auxiliou na evolução da tomografia, pois, quanto mais rápidos e mais poderosos se tornavam, os aparelhos de TC se tornam mais velozes e capazes de realizar exames jamais imaginados anteriormente como a angiotomografia, colono-grafia e reconstruções em 3D. Apesar da TC ter sido extensamente utilizada na medicina e áreas afins, há uma imensa gama de outras aplicações que são exploradas atualmente, como em Ciências de Solos (CRUVINEL et. al., 1990). A utilização de imagens de TC para identificação de materiais também vem sendo recentemente explorada (ROUX; UDUPA, 2003).

De acordo com o cenário apresentado, em uma aplicação de Ciências de Solo, imagens multiespectrais ruidosas de TC foram adquiridas e utilizadas para o estudo de técnicas de classificação, em especial as de combinação de múltiplos classificadores, no intuito de observar o comportamento e o potencial destas técnicas com as imagens sob a presença de ruído, apresentando formalmente os modelos utilizados e a relevância de seu uso para este tipo de problema.

1.1 ORGANIZAÇÃO DO TEXTO

Esta dissertação está organizada da seguinte maneira:

- Capítulo 1 trata da introdução ao objetivo do trabalho;
- Capítulos 2, 3 e 4 explanam uma revisão da literatura, apresentando a teoria dos métodos utilizados neste trabalho;
- Capítulo 5 descreve os materiais e métodos utilizados, detalhando a forma de aquisição das imagens e realização dos experimentos;
- Capítulo 6 apresenta os resultados obtidos;
- Capítulo 7 discute os resultados apresentados;
- Capítulo 8 realiza a conclusão do trabalho através de todos os resultados obtidos, comparações com a literatura e oferece sugestões para trabalhos futuros.

2. Tomografia Computadorizada

Este trabalho utilizou imagens obtidas através de tomografia computadorizada (TC) de transmissão. Iremos relatar os principais aspectos desta tecnologia e as características que a tornam interessante de ser explorada na tarefa de classificação de imagens. Detalhes teóricos podem ser encontrados em (KAK; SLANEY, 1988), (MACOVSKI, 1983) e (PARKER, 1990).

2.1 PRINCÍPIOS BÁSICOS

Tomografia computadorizada é um método não-destrutivo para obter informações sobre a estrutura de um corpo através do imageamento de uma secção deste objeto. O impacto desta técnica na medicina foi revolucionário, mas, apesar de difundido em larga escala para aplicações médicas, há inúmeras outras aplicações que utilizam os métodos da tomografia computadorizada. A base da tomografia é reconstruir a imagem de um corpo através de suas projeções, que representam as informações derivadas da transmissão ou reflexão de energias em um objeto, incidindo em um determinado ângulo, gerando uma imagem bidimensional da seção transversal de um corpo.

A solução para o problema de reconstruir uma função pelas suas projeções foi publicada por Radon, em 1917. Porém, foi apenas em 1973 que Hounsfield concretizou o imageamento por tomografia computadorizada (HOUNSFIELD, 1973), que viria a revolucionar o diagnóstico médico e a investigação da estrutura de objetos.

Os tomógrafos podem ser classificados em gerações, sendo que até o presente momento temos cinco gerações. O desenvolvimento das arquiteturas referentes a estas gerações deve-se, principalmente, à busca de menor tempo de aquisição das projeções para a tomografia. Isto é estudado de forma a viabilizar as aplicações em diversas áreas do conhecimento humano como, por exemplo, em aplicações médicas, onde o paciente não pode ficar exposto por um tempo prolongado às radiações incidentes. Os tempos de aquisição são significativamente menores nas gerações mais recentes com relação às primeiras, no entanto, o tamanho da imagem, entre outros parâmetros, exerce influência sobre o tempo de aquisição.

Os algoritmos matemáticos mais comuns para reconstrução tomográfica são baseados em dados da projeção. No caso de tomografia por transmissão utilizada neste trabalho, estas projeções representam a atenuação dos raios-X ou raios- γ que passam através de um objeto. Os dados da projeção, por natureza, são um resultado da interação entre a radiação e a substância da qual o objeto em questão é composto.

2.1.1 COEFICIENTE DE ATENUAÇÃO LINEAR

Este conceito está bem relacionado com a TC, e seu uso foi observado como sendo muito conveniente em aplicações de engenharia (CRUVINEL et. al., 1990) bem como de reconhecimento de padrões (HOMEM et. al., 2000). O coeficiente de atenuação linear pode ser entendido como o grau de dificuldade que um feixe de fótons, de uma radiação com determinada energia, encontra para atravessar um corpo. Esta dificuldade irá depender de vários parâmetros, como a energia empregada, a composição e a densidade dos materiais encontrados no corpo em estudo.

Durante a passagem pelo corpo, o feixe é atenuado devido aos fótons que estão sendo absorvidos pelos átomos do material, ou sendo dispersos, afastados de seu sentido original do curso. Os mecanismos responsáveis por estas duas ocorrências são os efeitos fotoelétrico e Compton como relatado em (KAK; SLANEY, 1988). A interação com efeito fotoelétrico é predominante para baixas energias e elementos de elevado número atômico. Este efeito consiste na interação entre os fótons e os elétrons da camada mais externa do material. O elétron absorve a energia de um fóton, usando-a em parte para superar a força que o mantém preso ao átomo, e a energia restante aparece como energia cinética. O efeito Compton, diferentemente do fotoelétrico, ocorre normalmente com elétrons livres, ou fracamente ligados. A consequência desta interação é o espalhamento do fóton para um sentido diferente do original, com alguma perda de energia, absorvida pelo elétron, que também é espalhado. É predominante para radiações de energias intermediárias. Os efeitos Compton e fotoelétrico são dependentes da energia. Isto significa que a probabilidade de um fóton se perder de seu feixe original, devido à absorção ou espalhamento, é uma função da energia deste fóton. Porém, a absorção fotoelétrica é mais dependente da energia do que o espalhamento do efeito Compton.

2.2 RECONSTRUÇÃO DE IMAGENS

A reconstrução é o último passo e resulta na principal finalidade da tomografia que é a geração da imagem que esperamos observar. Há diversos métodos e cada técnica de imageamento se baseia em diferentes teorias. Aqui descreveremos brevemente os métodos de reconstrução chamados genericamente de métodos da transformada, válidos para os fenômenos onde as projeções podem ser descritas pela integral de linha, que representa a integral de algum parâmetro do objeto ao longo de uma linha. Como descreve KAK e SLANEY (1988), um exemplo de processo que gera esta situação é a atenuação de raios-X. Neste caso o objeto é modelado como uma distribuição bidimensional (ou tridimensional) da constante de atenuação do raio-X e uma integral de linha representa a atenuação total sofrida pelo feixe de raio-X enquanto viaja em linha reta através do objeto. Tais métodos são soluções analíticas para a obtenção da inversa da transformada de Radon (PARKER, 1990). Toda a formulação teórica envolve funções contínuas, mas é realizada discretização para a implementação digital.

Seja $f(x, y)$ uma função representando uma seção transversal de um corpo em estudo, o problema consiste na recuperação da informação contida em $f(x, y)$ através de projeções desta função em diferentes ângulos.

2.2.1 APLICAÇÃO DIRETA DA TRANSFORMADA DE FOURIER

O método apresentado é conhecido como Aplicação Direta da Transformada de Fourier. Para isto, utilizaremos a teoria descrita por KAK e SLANEY (1988) e aplicada na reconstrução de imagens em CRUVINEL et. al. (1990). A teoria, apesar de ser apresentada formalmente através de funções contínuas, é discreta na aplicação real, sendo possível utilizar a Transformada Rápida de Fourier (conhecida como FFT) ao invés da Transformada de Fourier.

Para iniciar, utiliza-se a transformada de Radon da função $f(x, y)$:

$$g(t, \theta) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x, y) \delta(x \cos \theta + y \sin \theta - t) dx dy \quad (2.1)$$

onde $\delta(\cdot)$ é a função delta de Dirac, como ilustrado na Figura 2.1, onde, para um θ fixo, a Equação 2.1 é a integral de linha sobre a reta L a qual está a uma distância t da origem. Esta integral é denominada raio soma ou raio projeção. O conjunto completo de raios soma para cada ângulo θ recebe a denominação de projeção.

Tomando-se a transformada de Fourier (TF) unidimensional de uma projeção, nota-se que é igual a uma fatia da TF bidimensional do objeto original. Esta conclusão permite, através dos dados da projeção, estimar o objeto realizando uma transformada de Fourier inversa (TFI) bidimensional.

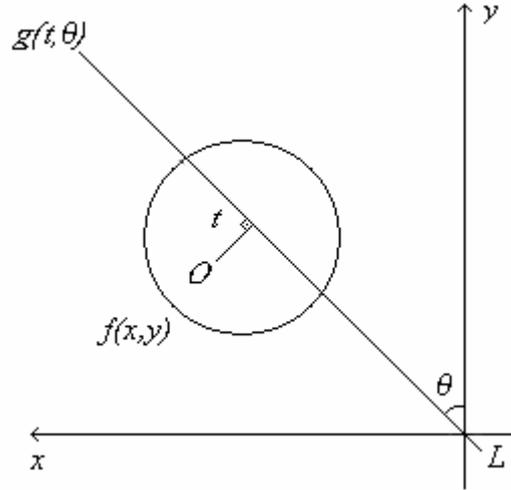


FIGURA 2.1 – Modelo usado na reconstrução da imagem $f(x,y)$

Sendo $\mathfrak{T}_1(x)$ a Transformada de Fourier de x , a transformada unidimensional de $g(t, \theta)$ na primeira variável é dada pela equação (2.4):

$$G_1(\rho, \theta) = \mathfrak{T}_1\{g(t, \theta)\} = \int_{-\infty}^{+\infty} g(t, \theta) e^{-j2\pi\rho t} dt \quad (2.2)$$

$$G_1 = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} f(x, y) \left[\int_{-\infty}^{+\infty} \delta(x \cos \theta + y \sin \theta - t) e^{-j2\pi\rho t} dt \right] dx dy \quad (2.3)$$

$$G_1 = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} f(x, y) e^{-j2\pi\rho(x \cos \theta + y \sin \theta)} dx dy \quad (2.4)$$

Pela definição da TF:

$$G_1(\rho, \theta) = F(u, v) \quad (2.5)$$

onde $u = \rho \cos \theta$ e $v = \rho \sin \theta$

A imagem reconstruída pode ser obtida por:

$$f(x, y) = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} F(u, v) e^{j2\pi(ux+vy)} du dv \quad (2.6)$$

2.2.2 RETROPROJEÇÃO FILTRADA

Este método consiste em filtrar as projeções de cada ângulo através da convolução no domínio do espaço e realizar em seguida a retroprojeção que é equivalente a encontrar as reconstruções elementares correspondentes a cada projeção filtrada mencionada acima.

Substituindo-se a equação (2.5) em (2.6) e realizando mudança de variáveis, temos:

$$f(x, y) = \int_0^\pi \int_{-\infty}^{+\infty} F(\rho \cos \theta, \rho \sin \theta) e^{j2\pi \rho t} |\rho| d\rho d\theta \quad (2.7)$$

onde $t = x \cos \theta + y \sin \theta$

Pode ser ainda mostrado por:

$$f(x, y) = \int_0^\pi \left[\int_{-\infty}^{+\infty} G_1(\rho, \theta) \cdot |\rho| \cdot e^{j2\pi \rho t} d\rho \right] d\theta \quad (2.8)$$

E finalmente por:

$$f(x, y) = \int_0^\pi \left[g(t, \theta) * \mathfrak{T}_1^{-1} \{ |\rho| \} \right] d\theta \quad (2.9)$$

observando que a TF da função rampa não existe, por ser um problema não integrável, e deste modo usa-se na prática outros filtros como o Ram-Lak.

A Equação 2.9 representa o método da Convolução-Retroprojeção que filtra os dados de cada ângulo das projeções no domínio do espaço e realiza a retroprojeção

A Equação 2.8 pode ser usada diretamente e é conhecida como Retroprojeção Filtrada, método usado para a reconstrução das imagens deste trabalho.

2.3 IMAGENS MULTIESPECTRAIS DE TOMOGRAFIA COMPUTADORIZADA

Para que tenhamos uma imagem digital, é necessário adquiri-la por um sistema de imageamento que mede, através de seus sensores, a luz incidente. A resposta de um sensor depende não só da quantidade de luz incidente, mas também da frequência da mesma. Um sensor é então caracterizado por sua curva de resposta espectral, a qual dá a intensidade da resposta do sensor para cada frequência. Portanto, uma imagem multiespectral representa uma

coleção de imagens de uma mesma cena, em um mesmo instante, obtida por sensores com respostas espectrais diferentes. Cada imagem que compõe uma imagem multiespectral é inserida em um “canal” ou “banda”. As imagens que são obtidas por apenas um sensor são ditas imagens monocromáticas, ou simplesmente imagens.

Uma imagem pode ser definida como uma função bidimensional na forma $f(x, y)$. O valor (ou amplitude) de f em coordenadas espaciais (x, y) é um número escalar positivo cujo sentido físico é determinado pela origem da imagem. Quando uma imagem é gerada por um processo físico, seus valores são proporcionais à energia irradiada de uma fonte física – neste caso raios-X e raios- γ . As imagens são formadas, portanto, por transmissão de energia através do objeto onde teríamos uma função coeficiente de atenuação linear no ponto (x, y) .

Considere um objeto, uma fonte emissora de radiação e um detector. Seja I_0 o número de fótons emitidos pela fonte e que incidem sobre o objeto. Ainda, seja I o número de fótons que atravessam o objeto em estudo atingindo o detector em um intervalo de tempo t . A relação entre I e I_0 é dada por:

$$\frac{I}{I_0} = e^{-\int_{\text{raio}} \mu(x, y) ds} \quad (2.10)$$

onde $\mu(x, y)$ é a função coeficiente de atenuação linear no ponto (x, y) .

3. *Classificação*

Classificação de imagens é o tema principal desta pesquisa, que visa estudar diversas abordagens explorando-as em uma aplicação de análise de materiais em imagens de tomografia computadorizada, cujas características foram expostas no capítulo 2. Este capítulo introduz e descreve a teoria dos métodos utilizados nos experimentos, escolhidas por possuírem bases teóricas diferentes e, por consequência, gerarem resultados bem diferentes uns dos outros, de forma que forneçam informações complementares uma às outras e possam ser exploradas pelos combinadores descritos no capítulo 4.

3.1 RECONHECIMENTO DE PADRÕES

“Reconhecimento de padrões – o ato de, tendo informações desconhecidas, tomar uma decisão baseada na categoria (ou classe) de um padrão – tem sido crucial para nossa sobrevivência, e através dos dez últimos milhões de anos, desenvolvemos sistemas neurais e cognitivos altamente sofisticados para estas tarefas.” (DUDA et. al., 2000)

A idéia de capacitar uma máquina para classificar padrões está sendo desenvolvida desde meados do século XX e muitas aplicações surgiram para incentivar as pesquisas: reconhecimento de fala, reconhecimento de caracteres, identificação de impressões digitais, análise de terrenos em imagens de satélite, entre outras. Diversas áreas tratam destes problemas e algumas delas são Reconhecimento de Padrões, Aprendizado de Máquina e Visão Computacional.

O objetivo principal de reconhecimento de padrões é classificar objetos de interesse em uma categoria ou classe dentre um número finito de categorias ou classes. Os objetos de interesse são chamados, genericamente, de padrões.

Para realizar a classificação é necessário encontrar medidas e características inerentes a cada classe que possam ser usadas para diferenciar um objeto do outro. A estas características chamamos atributos. Para cada aplicação diversas características podem ser utilizadas, a depender do número de informação que temos. Por exemplo, no caso de uma aplicação de reconhecimento facial, poderíamos utilizar como atributos: posição dos olhos,

boca e nariz; formato da face; distância entre os olhos; etc. O importante é utilizar atributos que possam diferenciar o melhor possível os diferentes objetos em estudo.

Após a escolha dos atributos, é preciso escolher o modelo para a classificação, que pode ser dividida em dois tipos: supervisionada e não supervisionada. Neste trabalho utilizamos apenas classificação supervisionada, que consiste em utilizar amostras conhecidas de cada classe para encontrar um classificador de forma que ele possa ser utilizado posteriormente para rotular objetos desconhecidos. O processo de geração do classificador através de amostras conhecidas é chamado treinamento e o conjunto destas amostras é chamado conjunto de treinamento. A classificação então pode ser realizada, e pode ser entendida, de maneira geral, pela partição do espaço de atributos em um número finito de regiões de tal forma que objetos de uma mesma classe recaiam, pelo menos em tese, sempre dentro de uma mesma região.

Tendo os dados rotulados é preciso então testar a eficiência da classificação, o que pode ser feito através de testes em amostras cujas classes são conhecidas. Este conjunto de amostras é chamado de conjunto de testes. Uma forma de analisar a performance do reconhecimento é a matriz de confusão (ou matriz de erros), técnica utilizada geralmente em aplicações de sensoriamento remoto que representa as amostras pertencentes em cada classe e as classes para as quais foram atribuídas (CONGALTON; GREEN, 1999).

Este é o resumo de uma tarefa de reconhecimento de padrões, mas há ainda diversas outras etapas que podem ou não fazer parte do processo como: pré-processamento, extração de atributos, seleção de atributos, entre outros. Na literatura há livros excelentes que tratam de diversos detalhes e apresentam bases teóricas para o reconhecimento de padrões, entre os quais podemos destacar: (DUDA et. al., 2000), (FUKUNAGA, 1990), (DEVJVER; KITTLER, 1982) e (DUDA; HART, 1973). Outra fonte interessante é o artigo recente de JAIN et. al. (2000), que faz revisão da área de reconhecimento estatístico de padrões, revelando as recentes pesquisas e desafios da área.

3.2 CLASSIFICADORES

Como apresentado anteriormente, é preciso definir os modelos de classificação a serem utilizados. Cada modelo é baseado em uma teoria diferente e neste trabalho utilizamos apenas classificadores baseados em teorias estatísticas. Serão apresentados dois classificadores

baseados em funções discriminantes: Bayesiano linear e logístico, e dois classificadores não-paramétricos baseados em estimação de densidade: Parzen e k vizinhos mais próximos.

3.2.1 CLASSIFICADOR BAYESIANO LINEAR

A representação deste classificador é dada pela forma de funções discriminantes, ele baseia-se na Teoria de Decisão Bayesiana sob densidade normal, e assume matrizes de covariância iguais para todas as classes. Os detalhes teóricos estão bem fundamentados na literatura (DUDA; HART, 1973).

Entre as mais conhecidas maneiras de se representar um classificador de padrões é em termos de um conjunto de funções discriminantes $g_i(\mathbf{x}), i = 1, \dots, c$, para c classes. O processo de classificação é realizado pela geração de funções discriminantes para cada vetor de atributos \mathbf{x} observado. Este vetor é atribuído à classe ω_i que forneça a máxima função discriminante.

Utiliza-se a probabilidade *a posteriori* de, dado \mathbf{x} , a classe correta ser ω_i , ou seja $P(\omega_i | \mathbf{x})$. Desta forma, estabelecendo-se $g_i(\mathbf{x}) = P(\omega_i | \mathbf{x})$, a máxima função discriminante corresponderá à máxima probabilidade *a posteriori*. A escolha da função discriminante não é única. Em tratando-se de obter-se o mínimo erro, qualquer das escolhas abaixo geram resultados idênticos, por terem o denominador comum (DUDA et. al., 2000):

$$g_i(\mathbf{x}) = P(\omega_i | \mathbf{x}) = \frac{p(\mathbf{x} | \omega_i)P(\omega_i)}{\sum_{j=1}^c p(\mathbf{x} | \omega_j)P(\omega_j)} \quad (3.1)$$

$$g_i(\mathbf{x}) = p(\mathbf{x} | \omega_i)P(\omega_i) \quad (3.2)$$

$$g_i(\mathbf{x}) = \ln p(\mathbf{x} | \omega_i) + \ln P(\omega_i) \quad (3.3)$$

onde \ln corresponde ao logaritmo natural.

Assim como funções discriminantes podem ser reescritas em uma variedade de formas, as *regras de decisão* também podem. O objetivo das regras de decisão é dividir o espaço de atributos em c *regiões de decisão*, $\mathfrak{R}_1, \dots, \mathfrak{R}_c$. Se $g_i(\mathbf{x}) > g_j(\mathbf{x})$ para todo $i \neq j$,

então \mathbf{x} está em \mathcal{R}_i , e a decisão será atribuir \mathbf{x} à classe ω_i . As regiões são separadas por *limites de decisão*.

A estrutura de um classificador bayesiano é determinada pelas densidades condicionais e probabilidades *a priori*. É conveniente utilizar-se da função de densidade multivariada normal (ou Gaussiana). Isto porque o tratamento matemático é simples, pois pode ser determinada completamente através da média e da variância, possuindo propriedades de interesse para problemas de classificação.

A densidade normal multivariada em d dimensões é dada por:

$$p(\mathbf{x}) = \frac{1}{(2\pi)^{d/2} |\boldsymbol{\Sigma}|^{1/2}} \exp \left[-\frac{1}{2} (\mathbf{x} - \boldsymbol{\mu})' \boldsymbol{\Sigma}^{-1} (\mathbf{x} - \boldsymbol{\mu}) \right] \quad (3.4)$$

onde \mathbf{x} é um vetor de d componentes, $\boldsymbol{\mu}$ é o vetor média de d componentes, $\boldsymbol{\Sigma}$ é a matriz de covariância (de tamanho $d \times d$), onde os componentes da diagonal principal σ_{ii} são as variâncias de x_i e os elementos fora da diagonal σ_{ij} são as covariâncias de x_i e x_j . Os elementos $|\boldsymbol{\Sigma}|$ e $\boldsymbol{\Sigma}^{-1}$ são, respectivamente, determinante e inversa da matriz de covariância. O termo $(\mathbf{x} - \boldsymbol{\mu})' \boldsymbol{\Sigma}^{-1} (\mathbf{x} - \boldsymbol{\mu})$ é chamado de distância quadrática de Mahalanobis entre \mathbf{x} e $\boldsymbol{\mu}$.

Para simplificar, podemos abreviar a equação 3.4 para $p(\mathbf{x}) \approx N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, caracterizando a densidade normal. Formalmente temos:

$$\boldsymbol{\mu} \equiv E[\mathbf{x}] = \int \mathbf{x} p(\mathbf{x}) d\mathbf{x} \quad (3.5)$$

e

$$\boldsymbol{\Sigma} \equiv E[(\mathbf{x} - \boldsymbol{\mu})(\mathbf{x} - \boldsymbol{\mu})'] \quad (3.6)$$

onde os valores esperados do vetor e da matriz podem ser obtidos calculando-se os valores esperados de seus componentes, ou seja, $\mu_i = E[x_i]$ e $\sigma_{ij} = E[(\mathbf{x} - \boldsymbol{\mu})(\mathbf{x} - \boldsymbol{\mu})']$

Como observado, podemos ter uma classificação que oferece o mínimo erro utilizando funções discriminantes na forma da equação 3.3. A expressão da equação 3.3 pode ser avaliada considerando que as densidades são normais multivariadas, ou seja, $p(\mathbf{x} | \omega_i) \approx N(\boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i)$. Neste caso, da equação 3.4, teremos:

$$g_i(\mathbf{x}) = -\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu})' \boldsymbol{\Sigma}_i^{-1}(\mathbf{x} - \boldsymbol{\mu}) - \frac{d}{2} \ln 2\pi - \frac{1}{2} \ln |\boldsymbol{\Sigma}_i| + \ln P(\omega_i) \quad (3.7)$$

Se assumirmos que as classes possuem matrizes de covariância iguais, mas arbitrárias, teremos uma situação geométrica na qual as amostras cairão em *clusters* hiperelipsoidais de igual tamanho e formato, e o *cluster* para a i -ésima classe estará centrado sobre o vetor média $\boldsymbol{\mu}_i$. Sendo os termos $|\boldsymbol{\Sigma}|$ e $(d/2) \ln 2\pi$ independentes de i na equação 3.4, e assumindo probabilidade *a priori* igual para todas as classes, os termos podem ser ignorados, resultando em:

$$g_i(\mathbf{x}) = -\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu})' \boldsymbol{\Sigma}^{-1}(\mathbf{x} - \boldsymbol{\mu}) \quad (3.8)$$

Isto significa que a decisão ótima pode ser obtida através da distância de Mahalanobis entre \mathbf{x} e os c vetores média, atribuindo \mathbf{x} à categoria mais próxima em distância. Este classificador resulta em uma discriminante linear. Assume-se probabilidades *a priori* igual para todas as classes quando não há qualquer informação *a priori*, pois qualquer conhecimento prévio pode alterar fortemente o resultado, gerando melhores performances.

3.2.2 CLASSIFICADOR LOGÍSTICO

Desenvolvido por ANDERSON (1982), modela as probabilidades *a posteriori* das c classes através de funções lineares em \mathbf{x} , assegurando ao mesmo tempo de que somem 1 e permaneçam no intervalo $[0,1]$. Este intervalo é garantido através da especificação do modelo em termos de $c-1$ log-probabilidades como demonstrado por HASTIE et. al. (2001). A classificação é feita através de um conjunto de valores formados de transformações lineares de variáveis. Estas transformações lineares podem ser determinadas por um procedimento de máxima-verossimilhança. De acordo com WEBB (1999) esta é uma técnica intermediária entre as funções discriminantes lineares e métodos não-lineares, como Parzen windows e k vizinhos mais próximos.

A suposição básica para este classificador é que a razão log-verossimilhança é linear para qualquer par de probabilidades:

$$\log\left(\frac{p(\mathbf{x} | \omega_i)}{p(\mathbf{x} | \omega_c)}\right) = \beta_{i0} + \boldsymbol{\beta}_i' \mathbf{x}, \quad i = 1, \dots, c-1 \quad (3.9)$$

As probabilidades *a posteriori* para este caso são modeladas por funções logísticas:

$$p(\omega_i | \mathbf{x}) = \frac{\exp(\beta'_{i0} + \boldsymbol{\beta}'_i \mathbf{x})}{1 + \sum_{i=1}^{c-1} \exp(\beta'_{i0} + \boldsymbol{\beta}'_i \mathbf{x})}, \quad i = 1, \dots, c-1 \quad (3.10)$$

$$p(\omega_c | \mathbf{x}) = \frac{1}{1 + \sum_{i=1}^{c-1} \exp(\beta'_{i0} + \boldsymbol{\beta}'_i \mathbf{x})} \quad (3.11)$$

onde $\beta'_{i0} = \beta_{i0} + \log(p(\omega_i)/p(\omega_c))$.

A regra de decisão depende unicamente das funções lineares $\beta'_{i0} + \boldsymbol{\beta}'_i \mathbf{x}$, atribuindo \mathbf{x} à classe ω_j se:

$$\max\{\beta'_{i0} + \boldsymbol{\beta}'_i \mathbf{x}\} = \beta'_{j0} + \boldsymbol{\beta}'_j \mathbf{x} > 0 \quad i = 1, \dots, c-1 \quad (3.12)$$

senão atribui \mathbf{x} à classe ω_c .

As transformações lineares são usualmente obtidas por máxima-verossimilhança. A verossimilhança das observações é dada por:

$$L = \prod_{s=1}^c \prod_{r=1}^{n_s} p(\mathbf{x}_{sr} | \omega_s) \quad (3.13)$$

onde \mathbf{x}_{sr} é a observação r da classe s e n_s é o número de observações para a classe s . Maximizar L é equivalente a maximizar:

$$\log(L') = \sum_{i=1}^c \sum_{r=1}^{n_i} \log(p(\omega_i | \mathbf{x}_{ir})) \quad (3.14)$$

o gradiente de $\log(L')$ relativo aos parâmetros é:

$$\frac{\partial \log(L')}{\partial \beta'_{j0}} = n_j - \sum_{\mathbf{x}} p(\omega_j | \mathbf{x}) \quad (3.15)$$

$$\frac{\partial \log(L')}{\partial (\beta'_j)_l} = \sum_{r=1}^{n_j} (x_{jr})_l - \sum_{\mathbf{x}} p(\omega_j | \mathbf{x}) x_l \quad (3.16)$$

Para classes separáveis, a máxima verossimilhança é conseguida em um ponto no infinito no espaço de parâmetros, mas o algoritmo termina quando a separação completa ocorre (WEBB, 1999).

3.2.3 CLASSIFICADOR PARZEN

Parzen-window é uma técnica não-paramétrica conhecida originalmente para estimação de densidade e parte da idéia que, para estimar a densidade em \mathbf{x} , forma-se uma seqüência de regiões $\mathfrak{R}_1, \mathfrak{R}_2, \dots$ que contêm \mathbf{x} - a primeira região a ser usada com uma amostra, a segunda com duas amostras, e assim por diante. Seja V_n o volume de \mathfrak{R}_n , k_n o número das amostras que caem em \mathfrak{R}_n , e $p_n(\mathbf{x})$ a n -ésima estimativa para $p(\mathbf{x})$, teremos:

$$p_n(\mathbf{x}) = \frac{k_n/n}{V_n} \quad (3.17)$$

A abordagem Parzen-window de estimação pode ser introduzida supondo que uma região \mathfrak{R} é um hipercubo d -dimensional cujos lados tem tamanho h e volume dado por $V_n = h_n^d$.

Podemos obter uma expressão analítica para k_n - o número de amostras que caem no hipercubo - definindo a seguinte função janela:

$$\varphi(\mathbf{u}) = \begin{cases} 1 & |u_j| \leq 1/2 \\ 0 & \text{caso contrário} \end{cases}, \quad j = 1, \dots, d \quad (3.18)$$

Esta função define um hipercubo centrado na origem. A função $\varphi((\mathbf{x} - \mathbf{x}_i)/h_n)$ é igual a 1 se o ponto \mathbf{x}_i cai dentro de um hipercubo de volume V_n centrado em \mathbf{x} , e igual a 0 caso contrário. O número de amostras no hipercubo é dado por:

$$k_n = \sum_{i=1}^n \varphi\left(\frac{\mathbf{x} - \mathbf{x}_i}{h_n}\right) \quad (3.19)$$

Ao substituirmos esta na equação 3.17 obteremos a estimação:

$$p_n(\mathbf{x}) = \frac{1}{n} \sum_{i=1}^n \frac{1}{V_n} \varphi\left(\frac{\mathbf{x} - \mathbf{x}_i}{h_n}\right) \quad (3.20)$$

Podemos tratar esta estimação de densidade como consistindo da superposição de N cubos de tamanho h , com cada cubo centrado em uma das amostras. Em geral, se a função satisfaz $\varphi(\mathbf{u}) \geq 0$ e $\int \varphi(\mathbf{u})d(\mathbf{u}) = 1$, e mantiver-se a relação $V_n = h_n^d$, então a estimação em (3.20) também irá satisfazer as condições necessárias.

Examinando o efeito que o tamanho da janela h_n exerce sobre $p_n(\mathbf{x})$, e definindo a função $\delta_n(\mathbf{x})$

$$\delta_n(\mathbf{x}) = \frac{1}{V_n} \varphi\left(\frac{\mathbf{x}}{h_n}\right) \quad (3.21)$$

podemos reescrever $p_n(\mathbf{x})$ como a média:

$$p_n(\mathbf{x}) = \frac{1}{n} \sum_{i=1}^n \delta_n(\mathbf{x} - \mathbf{x}_i) \quad (3.22)$$

sendo $V_n = h_n^d$, é possível ver claramente que h_n afeta a amplitude e o tamanho de $\delta_n(\mathbf{x})$, e sua escolha exerce um efeito crucial na estimação.

O classificador baseado no método de Parzen-window (FUKUNAGA, HAYES, 1989) estima as densidades para cada categoria e classifica um ponto de teste pelo objeto que corresponde à máxima *a posteriori*. As regiões da decisão para este classificador dependem naturalmente da escolha da função janela (também chamada de *kernel*), como visto anteriormente. O kernel Gaussiano é uma escolha apropriada devido à independência estatística do modelo, entre outras características (DUDA et. al., 2000), como visto na seção 3.2.1.

Segundo DUDA et. al. (2000) quando há amostras suficientes, é assegurada a convergência de uma densidade complexa. Entretanto, o número das amostras necessárias pode ser muito grande, muito maior do que seria requerido se fosse conhecida a forma da densidade desconhecida. Há pouca ou nenhuma redução de dados o que conduz a um alto custo computacional para o armazenamento e treinamento. Além disso, a demanda para um

grande número amostras cresce exponencialmente com a dimensionalidade do espaço de atributos¹, restringindo a aplicação prática.

3.2.4 CLASSIFICADOR k-VIZINHOS MAIS PRÓXIMOS

O classificador k-vizinhos mais próximos foi desenvolvido a partir das técnicas de vizinhos mais próximos até ser estabelecido por PATRICK e FISCHER (1970) um classificador generalizado para múltiplas classes. Assim como o Parzen, é uma técnica não paramétrica que se baseia em uma teoria de estimação de densidade, da qual se extrai diretamente um classificador. Segundo HASTIE et. al. (2001), este classificador geralmente obtém sucesso quando a superfície de decisão ideal é irregular. Sua precisão depende ainda da quantidade de amostras.

É possível encontrar um classificador baseado diretamente na técnica dos k vizinhos mais próximos por uma modificação demonstrada em BISHOP (1995). Voltando à equação (3.17) como ponto de partida, vamos agora fixar k_n e variar o volume V_n . Considere uma pequena hipersfera centrada em um ponto \mathbf{x} , de forma que seja possível aumentar o raio desta hipersfera até que ela contenha precisamente k_n amostras. Estas k_n amostras são os k_n vizinhos mais próximos de \mathbf{x} . Se a densidade é alta na região próxima a \mathbf{x} , então as células serão pequenas, se a densidade é baixa então a célula aumenta até entrar em uma região de alta densidade. Neste caso k age como o parâmetro de suavização e assim como o h no caso da densidade estimada por Parzen windows, há um valor ótimo para a escolha de k . (DUDA, et al., 2000)

Suponha que um conjunto de amostras contenha n_j pontos na classe ω_j e n pontos no total, ou seja, $n = \sum_j n_j$. É desenhada uma hipersfera em volta de \mathbf{x} , a qual circunda k pontos (independente da classe a qual pertencem). Se esta esfera de volume V contém k_j pontos da classe ω_j , podemos usar (3.17) para obter aproximações para a densidade condicional na forma:

$$p(\mathbf{x} | \omega_j) = \frac{k_j}{n_j V} \quad (3.23)$$

A densidade incondicional pode ser similarmente estimada por:

¹ Este fenômeno é conhecido como “*curse of dimensionality*” e possui diversos aspectos extensamente estudados na literatura, apresentados com mais detalhes em JAIN et. al. (2000).

$$p(\mathbf{x}) = \frac{k}{nV} \quad (3.24)$$

A probabilidades *a priori* podem ser estimadas por:

$$P(\omega_j) = \frac{n_j}{n} \quad (3.25)$$

Então, para minimizar a probabilidade de erro de classificação de um novo vetor \mathbf{x} , o atribuímos à classe mais votada entre os k_n pontos de menor distância ao ponto \mathbf{x} . Este procedimento é conhecido como a *regra de classificação dos k vizinhos mais próximos*. No caso de $k = 1$ temos o caso particular denominado *regra do vizinho mais próximo*, que simplesmente atribui \mathbf{x} à classe do ponto mais próximo no espaço de atributos.

3.3 ESTIMAÇÃO DE PARÂMETROS PARA AS TÉCNICAS DE CLASSIFICAÇÃO PARZEN E k-VIZINHOS MAIS PRÓXIMOS

Os classificadores não-paramétricos citados anteriormente, Parzen e k vizinhos mais próximos (seções 3.2.3 e 3.2.4), necessitam da escolha de um valor inicial que define substancialmente o desempenho da classificação. No caso do classificador Parzen é necessário escolher o tamanho do kernel (h_n) e no k vizinhos mais próximos faz-se necessário definir a quantidade de vizinhos (k). Para que seja possível obter uma boa performance sem escolher estas medidas por tentativa e erro, FUKUNAGA e HUMMELS (1989) desenvolveram um método que possibilita estimar estes parâmetros através de um procedimento de estimação de erro por validação cruzada para cada amostra.

O trabalho de Fukunaga e Hummels visou obter uma versão da estimação de erro por validação cruzada para técnicas não paramétricas, mais especificamente para Parzen e k-vizinhos mais próximos e afirmava que a estimação dos parâmetros (kernel e quantidade de vizinhos) feita usando os mesmos dados usados para o treinamento do classificador pode gerar estimações de erro com viés alto, muito otimistas. Por isto apresentam uma forma de utilizar validação cruzada *leave-one-out* para estimar o erro com diferentes tamanhos de kernel e k -amostras. Desta idéia, utilizamos métodos de *leave-one-out* para testar as diversas escolhas possíveis de h_n e k , utilizando os parâmetros que obtiveram o menor erro global.

3.4 AVALIAÇÃO DO DESEMPENHO DOS CLASSIFICADORES

A avaliação do desempenho dos classificadores é um passo importante em um sistema de reconhecimento de padrões. Através de métodos de medição do erro obtemos números que indicam quais foram as performances obtidas pelos classificadores que utilizamos.

Entre os diversos métodos de estimar o erro, iremos apresentar a validação cruzada e o coeficiente Kappa.

3.4.1 VALIDAÇÃO CRUZADA

Quando um classificador é utilizado, geralmente, espera-se que a utilização de maior número dados em seu treinamento melhore sua exatidão, e que usando mais dados em seu teste a estimação de sua taxa de erro será também mais precisa. No entanto, há um limite prático para o tamanho do conjunto de dados disponível. O tamanho limitado de dados disponíveis e a necessidade de dividi-lo em conjuntos de teste e treinamento, são fatores que podem limitar a performance das fases de treinamento e teste da classificação de padrões.

Uma técnica utilizada para pequenos conjuntos de dados, que funciona de forma a vencer esta dificuldade, é a validação cruzada (*cross-validation*), formalizada por STONE (1974). Neste procedimento, ao invés de utilizar uma parte dos dados para treinamento e outra para teste, são utilizados todas as n amostras de um conjunto de dados para treinamento. Para estimar o erro são realizados n experimentos. Em cada um deles 1 amostra do conjunto de dados é retirada e o classificador é treinado utilizando as $n-1$ amostras restantes. A amostra que foi retirada é então testada. Este procedimento é repetido deixando de fora uma amostra diferente a cada iteração, até que se tenham n diferentes classificações cada qual usando $n-1$ amostras (Figura 3.2). A taxa de erro da classificação final é dada por:

$$E = \frac{1}{n} \sum_{i=1}^n e(i) \quad (3.26)$$

onde $e(i)$ é uma função do erro em cada iteração, igual a 0 para classificações corretas e 1 para erros.

Desta forma o treinamento aproveitará todo o conjunto de amostras e a estimação da taxa de erro será tão exata quanto possível, pois todas as amostras foram utilizadas para o teste. Apesar das vantagens apresentadas, este procedimento possui alto custo computacional.

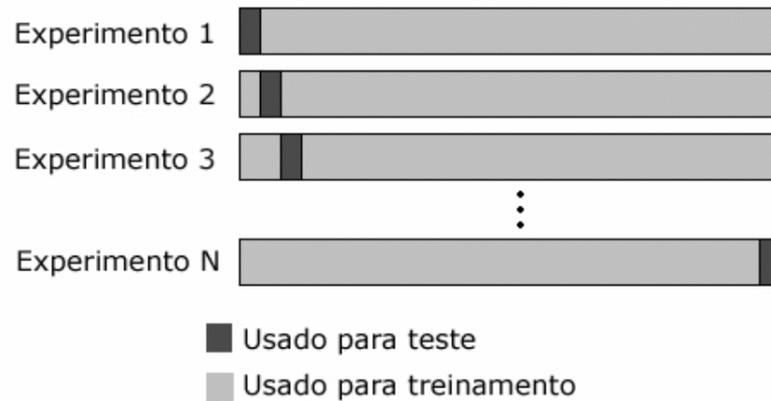


FIGURA 3.2 – Esquema de estimativa de erro utilizando *leave-one-out cross-validation*

3.4.2 COEFICIENTE KAPPA

Através de uma matriz de confusão (ou matriz de erros), onde é possível visualizar a taxa de erro comparando com cada classe, coeficientes de concordância podem ser calculados para avaliar o desempenho dos classificadores. Estes coeficientes são sensíveis não só à perda do desempenho global, mas também às variações de abstenção² (não utilizada neste trabalho) e confusão³ médias, o que torna interessante utilizá-los.

O coeficiente Kappa foi desenvolvido por Cohen, citado por CONGALTON (1991) e é bastante utilizado em aplicações de sensoriamento remoto, apresentando-se muito interessante para medir o desempenho em sistemas de reconhecimento de padrões. Este método determina a concordância esperada *a posteriori*, ou seja, dadas amostras previamente rotuladas pela amostragem, qual a concordância destes rótulos com os gerados pelo classificador. A concordância esperada somente poderá ser determinada após a construção de uma matriz de confusão, e esta por sua vez é obtida após a verificação dos resultados obtidos na classificação digital. Os valores que caem na diagonal principal da matriz de confusão correspondem aos acertos, e fora da diagonal principal é possível analisar a confusão entre as classes. A fórmula para obter o coeficiente é dada por:

$$K = \frac{P_o - P_c}{1 - P_c} \quad (3.27)$$

ou

² Amostras não classificadas ou rejeitadas pelo classificador por diferirem muito do restante das amostras. A utilização de abstenção/rejeição é opcional.

³ Amostras de uma determinada classe que são classificadas como sendo de outra classe.

$$K = \frac{N \sum_{i=1}^M X_{ii} - \sum_{i=1}^M X_{+i} X_{i+}}{N^2 - \sum_{i=1}^M X_{+i} X_{i+}} \quad (3.28)$$

onde

$$P_O = \sum_{i=1}^M X_{ii} \quad (3.29)$$

é a concordância observada e:

$$P_C = \sum_{i=1}^M (X_{i+} X_{+i}) \quad (3.30)$$

é a concordância esperada, sendo que X_{i+} é a soma dos elementos da linha i e X_{+i} a soma dos elementos da coluna i , realizando uma concordância entre os elementos pertencentes a uma classe e a quais classes cada um deles foi atribuído. N é o total de observações e M o total de classes.

O coeficiente Kappa pode apresentar valores entre -1 e +1. A interpretação dos coeficientes de concordância é subjetiva. CONGALTON (1991) apresenta uma sugestão de interpretação (Tabela 3.1) que é utilizada normalmente. No entanto, cada aplicação requer um nível diferente de acerto, o que modifica a interpretação.

TABELA 3.1 – Interpretação para resultados do coeficientes Kappa.

Resultado	Interpretação do desempenho do classificador
$K \leq 0$	Péssimo
$0 < K \leq 0,2$	Mau
$0,2 < K \leq 0,4$	Razoável
$0,4 < K \leq 0,6$	Bom
$0,6 < K \leq 0,8$	Muito Bom
$0,8 < K \leq 1,0$	Excelente

4. *Combinação de Múltiplos Classificadores*

Uma das propriedades mais importantes de um classificador está naturalmente em sua capacidade de resposta ao tentar reconhecer novos padrões, ou seja, sua capacidade de generalização. Muitas vezes são implementados vários e diferentes classificadores, cada um deles possuindo diferentes limiares de decisão e generalizando de modo diferenciado, e são escolhidos então os classificadores que se apresentam mais eficazes (frente a um conjunto de testes), para realizar a tarefa de futuras classificações. Contudo, como foi descrito por vários autores, essa estratégia pode desperdiçar informações ao ignorar classificadores (por menos eficazes que sejam). Para utilizar plenamente a informação de todo um grupo de classificadores, a saída de cada classificador pode ser combinada com as restantes, permitindo uma decisão final que em grande parte dos casos melhora as capacidades de generalização e a estabilidade da classificação.

4.1 CARACTERÍSTICAS

Combinar múltiplos classificadores para resolver um problema de reconhecimento de padrões é um procedimento conveniente em casos particulares. Alguns deles são apresentados por JAIN et. al. (2000):

1. Acesso a diferentes classificadores, cada um deles desenvolvidos em um diferente contexto e utilizando diferentes representações/descrições de um mesmo problema como, por exemplo, a identificação de pessoas pela voz, face e impressão digital,
2. Disponibilidade de mais de um conjunto de treinamento, coletados em tempo ou ambientes diferentes. Estes conjuntos podem usar ainda diferentes atributos.
3. Diferentes classificadores treinados em um mesmo conjunto de dados que podem não apenas ter performances diferentes, mas apresentar diferenças locais, de forma que cada classificador tenha uma região no espaço de atributos para a qual seu desempenho é o melhor,

4. No caso de redes neurais, é possível ter redes com diferentes inicializações. Ao invés de utilizar apenas uma e descartar as outras, a combinação poderia utilizar as vantagens de cada uma delas.

Estes casos sugerem o uso de combinadores devido à disponibilidade de classificadores ou dados. No entanto, para muitas aplicações a escolha de um classificador que possua um bom desempenho é suficiente para resolver o problema. Quando há problemas complexos a serem resolvidos, envolvendo um grande conjunto de classes, conjuntos de atributos com dimensionalidades e características diferentes ou ainda dados ruidosos, a escolha de um único classificador pode se tornar difícil, pois provavelmente limitaria a capacidade de reconhecimento do sistema. Exemplos de aplicações complexas, e para as quais já foram exploradas técnicas de combinação de classificadores são: reconhecimento de caracteres manuscritos (XU et. al., 1992) e reconhecimento de pessoas (KITTLER et. al., 1998). Nestes casos e outros casos complexos, combinar classificadores pode ser interessante:

“Freqüentemente um classificador combinado dá resultados melhores do que classificadores individuais, por combinar – com o uso de alguma técnica – as decisões independentes de cada classificador e, por consequência, as vantagens dos classificadores individuais na solução final, resultando em considerável melhora no acerto geral.” (DUIN; TAX, 2000)

Um grande número de esquemas de combinação foi proposto na literatura, incluindo métodos de *ranking*, *Borda count* e *regressão logística* (HO et. al., 1994), *teoria da incerteza de Dempster-Shafer* (XU et. al., 1992), *bagging* (BREIMAN, 1996) e *regras fixas* (KITTLER et. al., 1998). Cada modelo realiza suas suposições sobre os atributos, os classificadores e faz exigência de determinadas informações.

Segundo XU et. al. (1992), cada classificador oferece um nível diferente de informação como saída, e que será usada pelo combinador. São os níveis:

1. Abstrato: o classificador r apenas indica um rótulo j , que é a classe escolhida para o padrão;
2. Ranking: r faz um ranking com todas as classes, colocando em uma lista na qual a classe do topo é a primeira escolha.
3. Medida (confiança): r atribui uma medida a cada classe, que é o grau de confiança de x pertencer a cada classe.

O nível 3 (medida) é o que oferece informação mais relevante e o nível 1 (abstrato) menos informação sobre a decisão a ser tomada. Iremos tratar problemas referentes ao nível 3 pois todos os classificadores aqui abordados fornecem este nível de informação.

A combinação de classificadores é especialmente útil se os classificadores individuais são teoricamente independentes em suas representações (JAIN et. al., 2000), ou seja, se possuem atributos diferentes, como num caso do reconhecimento de uma pessoa, os atributos: voz, face e impressão digital. Apesar de muitas aplicações possuírem esta característica, esta condição de independência pode ser garantida pelo uso de atributos diferentes, uso de conjuntos de treinamento diferentes, pela diferença entre os classificadores individuais ou pode ser criado artificialmente através de técnicas como *bagging* (BREIMAN, 1996), *boosting* (SCHAPIRE, 1990) e *stacking* (WOLPERT, 1992). Uma das mais usadas é o *stacking* que insere pesos nos vetores de atributos dependendo da localização do objeto, de forma a “empilhar” os diferentes métodos de treinamento e melhorar a performance da classificação quando combinados.

Na Figura 4.1 pode-se verificar o funcionamento de um sistema de múltiplos classificadores aplicado a imagens. Um conjunto de classificadores é aplicado a uma imagem de entrada. Cada um deles produz uma classificação de um conjunto de classes. Um subconjunto de classes é escolhido (entre as melhores classificadas), por limiares de decisão contidos em cada classificador. A união destes subconjuntos passa pelo combinador, que produz a classificação final dos padrões.

4.2 FUNDAMENTOS TEÓRICOS

Apesar de uma área recente, existem diversos trabalhos sendo publicados. Em livro recente, KUNCHEVA (2004) classifica a área de combinação de classificadores como uma bagunça e admite que não há consensos e embasamento apropriado que regule a maioria destas técnicas. Desta forma, neste trabalho foi realizado o estudo dos métodos de combinação de forma a utilizar a mesma base teórica para todas as regras de combinação, como será apresentado.

Fundamentando de maneira geral a teoria dos sistemas de combinação de classificadores por regra fixa, considere um problema onde um objeto y tem de ser atribuído a uma das c classes possíveis. Temos R classificadores, cada qual representando o padrão atual

por um vetor de atributos distintos. Vamos supor que o vetor de atributos usado pelo i ésimo classificador é \mathbf{x}_i . No espaço de atributos, cada classe ω_k , $k = 1, \dots, c$ é modelada pela função de densidade de probabilidade $p(\mathbf{x}_i | \omega_k)$ e a probabilidade *a priori* de sua ocorrência é denotada por $P(\omega_k)$.

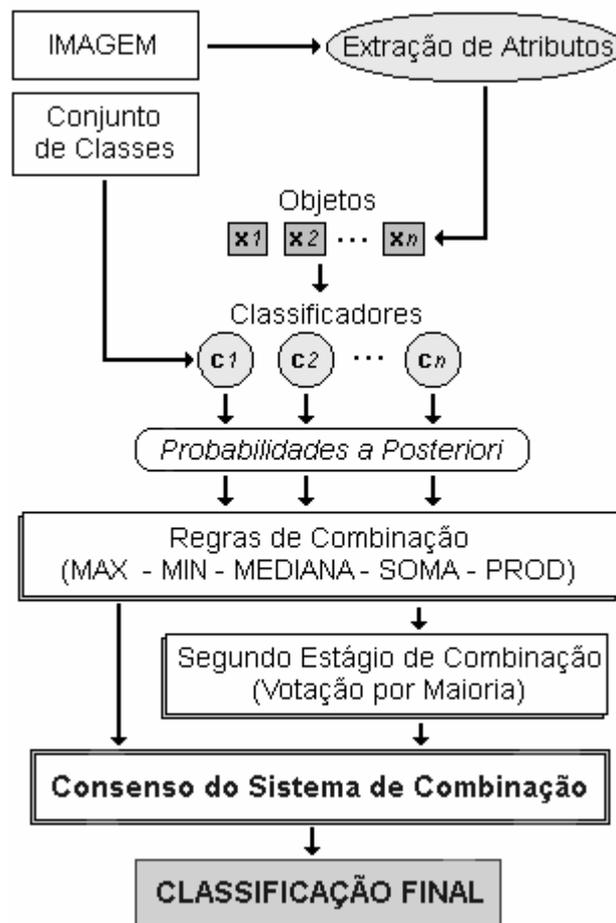


FIGURA 4.1 – Diagrama de um sistema de combinação de classificadores

De acordo com a teoria Bayesiana, dados os vetores de características para cada representação dos R classificadores: \mathbf{x}_i , $i = 1, \dots, R$ o padrão \mathbf{y} tem de ser atribuído à classe ω_j que provê o máximo da probabilidade *a posteriori*. Ou seja:

$$\text{atribuir } \mathbf{y} \rightarrow \omega_j \text{ se} \\ P(\omega_j | \mathbf{x}_1, \dots, \mathbf{x}_R) = \max_k P(\omega_k | \mathbf{x}_1, \dots, \mathbf{x}_R) \quad (4.1)$$

A regra de decisão Bayesiana (4.1) indica que, a fim de utilizar corretamente toda a informação disponível para alcançar uma decisão, é essencial computar as probabilidades das

várias hipóteses considerando todas as medidas simultaneamente. Esta exigência é naturalmente correta para o problema de classificação, mas pode ser impraticável. A computação das probabilidades *a posteriori* depende do conhecimento de medidas estatísticas de ordem elevada descrita em termos de funções densidade de probabilidade conjuntas $p(\mathbf{x}_1, \dots, \mathbf{x}_R | \omega_k)$, as quais seriam difíceis de inferir. De forma a simplificar a regra acima, iremos expressá-la em termos das decisões executadas pelos classificadores individuais, cada um explorando somente a informação dada pelo seu vetor \mathbf{x}_i . Desta forma buscaremos, através de regras de combinação simples, uma regra de decisão de fácil tratamento computacional.

Descrevendo a probabilidade *a posteriori* $P(\omega_k | \mathbf{x}_1, \dots, \mathbf{x}_R)$ usando o teorema de Bayes temos:

$$P(\omega_k | \mathbf{x}_1, \dots, \mathbf{x}_R) = \frac{p(\mathbf{x}_1, \dots, \mathbf{x}_R | \omega_k)P(\omega_k)}{\sum_{j=1}^c p(\mathbf{x}_1, \dots, \mathbf{x}_R | \omega_j)P(\omega_j)} \quad (4.2)$$

onde o denominador representa $p(\mathbf{x}_1, \dots, \mathbf{x}_R)$, a densidade de probabilidade conjunta incondicional.

Considere as probabilidades *a posteriori* determinadas por cada classificador i como $p_{ik}(\mathbf{x})$, $i = 1, \dots, R$, limitadas entre 0 e 1 e calculadas para cada objeto \mathbf{x} para as classes k , $k = 1, \dots, c$. Estes números são calculados de forma que:

$$\sum_k^c p_{ik}(\mathbf{x}) = 1, \quad i = 1, \dots, R \quad (4.3)$$

A confiança $q_k(\mathbf{x})$ para a classe k é dada por:

$$q'_k(\mathbf{x}) = \text{regra}(p_{ik}(\mathbf{x}))$$

$$q_k(\mathbf{x}) = \frac{q'_k(\mathbf{x})}{\sum_k q'_k(\mathbf{x})} \quad (4.4)$$

Diversas regras podem ser usadas na Equação 4.4, e elas serão abordadas nos próximos itens. A classificação final é dada pela seguinte equação:

$$\omega(\mathbf{x}) = \arg \max_k (q_k(\mathbf{x})) \quad (4.5)$$

As teorias de combinação de classificadores fazem muitas suposições acerca dos dados a serem classificados e dos erros produzidos pelos classificadores individuais. Isto se torna inevitável para que as equações sejam aproximadas e tenham embasamento teórico correto. Nos próximos itens, examinaremos com mais detalhe cada uma das regras a partir das regras do produto e da soma, que constituem os esquemas básicos de combinação dos quais podem ser derivados diversos outros. As implicações de suas suposições serão discutidas no Capítulo 7.

4.2.1 REGRA DO PRODUTO

Para fundamentar teoricamente esta regra, assumimos que as representações dos dados usadas por cada classificador possuem erros estatisticamente independentes, ou seja, cometem erros diferentes na classificação (o uso de classificadores com características bem diferentes ou representações diferentes para os vetores de atributos pode garantir isto em muitos casos). Podemos definir a regra do produto a partir de:

$$p(\mathbf{x}_1, \dots, \mathbf{x}_R | \omega_k) = \prod_{i=1}^R p(\mathbf{x}_i | \omega_k) \quad (4.6)$$

onde $p(\mathbf{x}_i | \omega_k)$ é o modelo para a i -ésima representação. Substituindo (4.6) em (4.2):

$$P(\omega_k | \mathbf{x}_1, \dots, \mathbf{x}_R) = \frac{P(\omega_k) \prod_{i=1}^R p(\mathbf{x}_i | \omega_k)}{\sum_j^c P(\omega_j) \prod_{i=1}^R p(\mathbf{x}_i | \omega_j)} \quad (4.7)$$

e aplicando (4.7) em (4.1), em termos das probabilidades *a posteriori* obtidas pelos classificadores individuais, teremos:

$$\begin{aligned} & \text{atribuir } \mathbf{y} \rightarrow \omega_j \text{ se} \\ P(\omega_j) \prod_{i=1}^R P(\omega_j | \mathbf{x}_i) &= \max_{k=1}^c P(\omega_k) \prod_{i=1}^R P(\omega_k | \mathbf{x}_i) \end{aligned} \quad (4.8)$$

A regra acima quantifica a probabilidade de uma hipótese combinando as probabilidades *a posteriori* geradas por cada classificador por meio de uma regra do produto. Segundo KITTNER et. al (1998) e DUIN e TAX (2000), esta regra é severa pois depende da independência dos atributos de cada classificador a seus respectivos erros, improvável em aplicações práticas mas que pode ser aproximada em grande parte dos casos.

4.2.2 REGRA DA SOMA

Estabelecida a partir da regra do Produto (4.8), supõe que as probabilidades *a posteriori* computadas por cada classificador não irão se desviar muito das probabilidades *a priori*. KITTLER et. al (1998) considera que esta suposição é rígida mas pode ser satisfeita quando a informação disponível é altamente ambígua, por exemplo, devido a um alto nível de ruído. Com isto, podemos expressar as probabilidades *a posteriori* na forma:

$$P(\omega_k | \mathbf{x}_i) = P(\omega_k)(1 + \delta_{ki}) \quad (4.9)$$

onde δ_{ki} satisfaz $\delta_{ki} \ll 1$. Substituindo (4.9) nas probabilidades *a posteriori* em (4.8):

$$P(\omega_k) \prod_{i=1}^R P(\omega_k | \mathbf{x}_i) = P(\omega_k) \prod_{i=1}^R (1 + \delta_{ki}) \quad (4.10)$$

Se expandirmos o produto e negligenciarmos os termos de segunda ou mais alta ordens, podemos aproximar o lado direito da Equação (4.10) por:

$$P(\omega_k) \prod_{i=1}^R (1 + \delta_{ki}) = P(\omega_k) + P(\omega_k) \sum_{i=1}^R \delta_{ki} \quad (4.11)$$

Finalmente, substituindo (4.11) e (4.9) na regra do produto (4.8), obtemos a regra da soma:

$$\text{atribuir } \mathbf{y} \rightarrow \omega_j \text{ se} \\ P(\omega_j) + \sum_{i=1}^R P(\omega_j | \mathbf{x}_i) = \max_{k=1}^c \left[P(\omega_k) + \sum_{i=1}^R P(\omega_k | \mathbf{x}_i) \right] \quad (4.12)$$

KITTLER et. al. (1998) avisa que a suposição que dá origem a Equação (4.9) não será realista na maioria das aplicações. Quando o vetor de observações fornecer informação discriminatória suficiente e as probabilidades *a posteriori* crescerem, a aproximação do produto para a regra da soma irá introduzir erros grosseiros de aproximação. No entanto o autor demonstra que esta aproximação será interessante, mostrando que a regra da soma é menos sensível ao erro dos classificadores do que a regra do produto.

4.2.3 DERIVAÇÕES DAS REGRAS DE PRODUTO E SOMA

Através da relação:

$$\prod_{i=1}^R P(\omega_k | \mathbf{x}_i) \leq \min_{i=1}^R P(\omega_k | \mathbf{x}_i) \leq \frac{1}{R} \sum_{i=1}^R P(\omega_k | \mathbf{x}_i) \leq \max_{i=1}^R P(\omega_k | \mathbf{x}_i) \quad (4.13)$$

estabelecida por KITTLER et. al. (1998), as regras de produto e soma podem ser aproximadas para seus limites máximos e mínimos. Pode-se também estabelecer uma maneira de obter valores binários a partir dos classificadores individuais:

$$\Delta_{ki} = \begin{cases} 1 & \text{se } P(\omega_k | \mathbf{x}_i) = \max_{j=1}^c P(\omega_j | \mathbf{x}_i) \\ 0 & \text{caso contrário} \end{cases} \quad (4.14)$$

resultando em informações no nível abstrato de informação (XU et. al., 1992), como visto na seção 3.3, permitindo utilizar estratégias outras que combinar probabilidades *a posteriori*. Nas próximas seções vamos descrever as regras derivadas destas relações.

4.2.4 REGRA DO MÁXIMO

A partir da Equação (4.12), aproximando a soma pelo máximo das probabilidades *a posteriori* e assumindo iguais probabilidades *a priori*, obtemos:

$$\begin{aligned} & \text{atribuir } \mathbf{y} \rightarrow \omega_j \text{ se} \\ \max_{i=1}^R P(\omega_j | \mathbf{x}_i) &= \max_{k=1}^c \max_{i=1}^R P(\omega_k | \mathbf{x}_i) \end{aligned} \quad (4.15)$$

4.2.5 REGRA DO MÍNIMO

A partir da Equação (4.8), limitando o produto das probabilidades *a posteriori* pelo seu máximo e assumindo iguais probabilidades *a priori*, obtemos:

$$\begin{aligned} & \text{atribuir } \mathbf{y} \rightarrow \omega_j \text{ se} \\ \min_{i=1}^R P(\omega_j | \mathbf{x}_i) &= \max_{k=1}^c \min_{i=1}^R P(\omega_k | \mathbf{x}_i) \end{aligned} \quad (4.16)$$

4.2.6 REGRA DA MEDIANA

A regra da soma, supondo conhecimento *a priori* igual para as classes, pode ser calculada pela média das probabilidades *a posteriori* de cada classe sobre todas as saídas dos classificadores:

$$\begin{aligned} &\text{atribuir } \mathbf{y} \rightarrow \omega_j \text{ se} \\ &\frac{1}{R} \sum_{i=1}^R P(\omega_j | \mathbf{x}_i) = \max_{k=1}^c \frac{1}{R} P(\omega_k | \mathbf{x}_i) \end{aligned} \quad (4.17)$$

Então a regra atribui um padrão à classe cuja média das probabilidades *a posteriori* é máxima. Se qualquer das saídas dos classificadores para alguma classe tiver um valor muito diferente dos demais irá afetar a média e poderá levar a uma decisão incorreta. Baseando a decisão da combinação na mediana das probabilidades *a posteriori* obtemos uma estimativa mais robusta e obtemos a regra:

$$\begin{aligned} &\text{atribuir } \mathbf{y} \rightarrow \omega_j \text{ se} \\ &\underset{i=1}{\overset{R}{\text{med}}} P(\omega_j | \mathbf{x}_i) = \max_{k=1}^c \underset{i=1}{\overset{R}{\text{med}}} P(\omega_k | \mathbf{x}_i) \end{aligned} \quad (4.18)$$

4.2.7 REGRA DA VOTAÇÃO POR MAIORIA

Partindo da Equação (4.12), supondo iguais probabilidades *a priori* e obtendo informações dos classificadores a partir de (4.14), podemos definir:

$$\begin{aligned} &\text{atribuir } \mathbf{y} \rightarrow \omega_j \text{ se} \\ &\sum_{i=1}^R \Delta_{ji} = \max_{k=1}^c \sum_{i=1}^R \Delta_{ki} \end{aligned} \quad (4.19)$$

Note que para cada classe a soma no lado direito da equação acima simplesmente conta os votos recebidos para esta hipótese pelos classificadores individuais. A classe que recebe o maior número de votos é então selecionada por um consenso ou decisão da maioria.

4.2.8 VOTAÇÃO DOS COMBINADORES

A Votação dos Combinadores representa um segundo estágio de combinação em que, tendo disponíveis informações de diversos combinadores, verifica-se qual a classe mais votada entre os combinadores aplicando-se a Equação (4.19) na saída dos classificadores, ou seja, tomando-se a moda das classes escolhidas.

Com este procedimento, espera-se obter um método que forneça resultados mais constantes, realizando sempre performances boas ou ótimas. Permitindo ainda observar se os combinadores em geral acertaram. Esta técnica é otimista no sentido que supõe que a maioria dos combinadores irá realizar uma classificação correta dos padrões.

5. *Material e Métodos*

O presente trabalho propõe um método para a identificação de materiais em imagens multiespectrais ruidosas de tomografia de transmissão utilizando múltiplas energias. A proposta consistiu em, tendo adquirido imagens tomográficas de um corpo em estudo através de várias energias e, a partir da construção de imagens multiespectrais com as bandas de energias correspondentes, utilizar a informação pertinente a cada banda do espectro para explorar técnicas de classificação e combinação de classificadores. Através dos experimentos, foram feitas comparações e estudos teóricos sobre as técnicas utilizadas.

5.1 AQUISIÇÃO DAS IMAGENS

5.1.1 TOMÓGRAFO

O minitomógrafo utilizado para a aquisição das imagens é um equipamento de primeira geração em que a fonte de raio-X ou raio- γ (proveniente de materiais com decaimento radioativo como Césio e Amerício), e o detector (de cristal) são fixos, enquanto são empregados os movimentos de rotação e translação ao corpo em estudo. O detector é acoplado a um pré-amplificador e a uma fotomultiplicadora. Um microcomputador é associado ao sistema e a saída é dada em forma de uma matriz retangular contendo as projeções tomográficas – cada linha da matriz corresponde a uma projeção e o número de colunas corresponde ao número de pontos por projeção. O software de reconstrução gera esta matriz, cujos valores são inteiros, representando os coeficientes de atenuação linear multiplicados por um fator 1000. Este minitomógrafo é o resultado de um trabalho desenvolvido para a Embrapa, cujos detalhes podem ser vistos em CRUVINEL et. al. (1990). Veja foto na Figura 5.1.



FIGURA 5.1 – Minitomógrafo

5.1.2 CORPO DE PROVA (PHANTOM)

Corpo de prova ou *phantom* é um objeto confeccionado especialmente para realizar algum experimento. Neste caso, o *phantom* foi construído com uma base de plexiglass – um tipo de polímero – em forma cilíndrica, o qual tem quatro espaços – também em forma cilíndrica – onde foram inseridos quatro materiais: cálcio, fósforo, alumínio e água. O *phantom* tem 60mm de diâmetro, sendo que as amostras de cálcio e água possuem 19mm de diâmetro e as amostras de fósforo e alumínio, 22mm de diâmetro, como ilustrado na Figura 5.2. As tomografias deste corpo de prova foram imageadas utilizando o minitomógrafo de primeira geração citado na seção anterior (5.1.1). Estas imagens constituem os objetos de pesquisa deste trabalho.

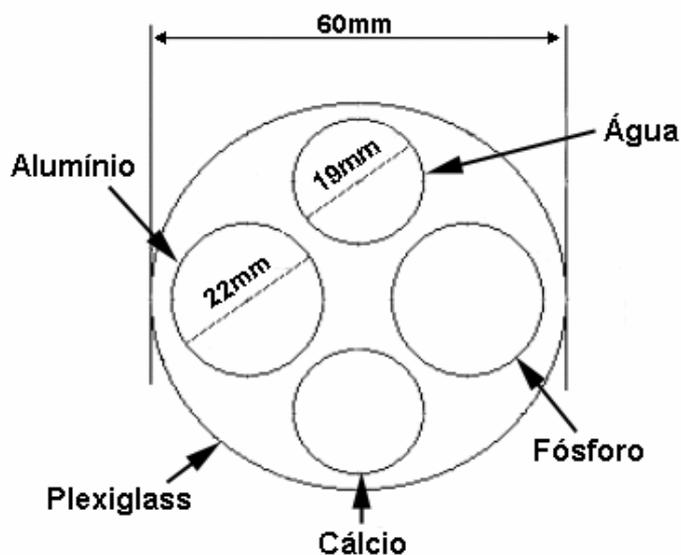


FIGURA 5.2 – Diagrama da construção do corpo de prova (*phantom*) imageado

5.1.3 IMAGENS OBTIDAS E CARACTERÍSTICAS

Foram empregadas para obtenção das projeções, duas fontes de raio- γ (Césio e Amerício) e duas fontes de raio-X disponíveis. As energias relativas às fontes de Césio e Amerício são de 662 keV e 60 keV, respectivamente. A fim de determinar as energias do feixe de raio-X, foi analisado o comportamento dos coeficientes de atenuação linear dos quatro materiais a serem colocados na amostra para se fazer uma comparação dos contrastes entre os mesmos, de forma a obter uma maior capacidade discriminatória através do coeficiente de atenuação. Desta forma, foram escolhidas dentre as possíveis opções, as energias de raio-X de 40 e 85 keV. A Figura 5.3 (HOMEM, 1998) mostra as curvas do coeficiente de atenuação linear em função da energia para os materiais em questão cujos valores podem ser vistos em LIDE (2002), e a Tabela 5.1 mostra os valores teóricos dos coeficientes de atenuação linear de cada material inclusive o plexiglass.

O procedimento de obtenção das imagens foi realizado de forma a obter-se imagens com alto nível de ruído, com translação total de 7,00 cm, passo linear de 0,109 cm, rotação total de 180°, passo angular de 2.81° e tempo de contagem de fótons em 3,00 segundos.

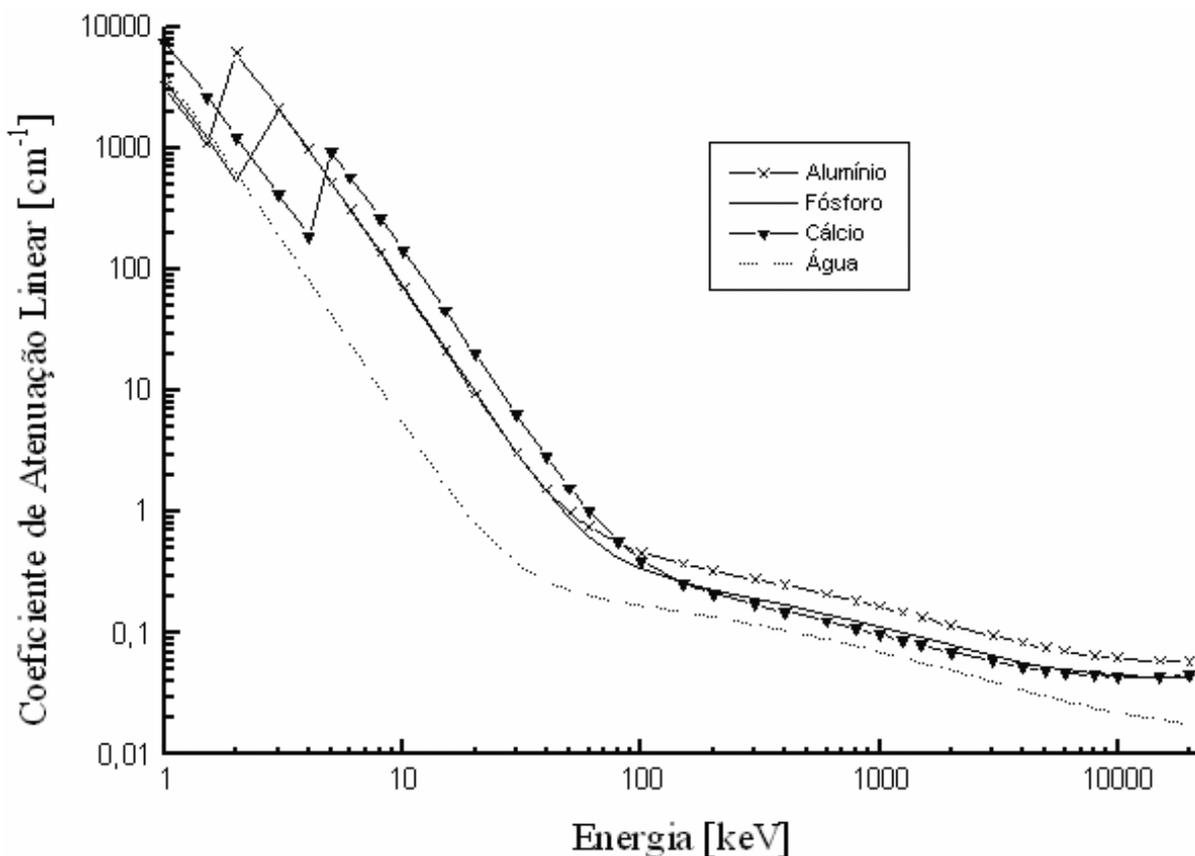


FIGURA 5.3 – Curvas do coeficiente de atenuação linear em função da energia para a água, alumínio, cálcio e fósforo.

TABELA 5.1 – Coeficientes de atenuação linear em função das energias 40, 60, 85 e 662 keV para a água, alumínio, cálcio e fósforo.

Energia (keV)	Coeficientes de Atenuação Linear dos Materiais				
	Água	Alumínio	Cálcio	Fósforo	Plexiglass
40	0.2683	1.5349	2.8365	1.4734	0.2880
60	0.2052	0.7479	1.0013	0.6315	0.2270
85	0.1750	0.6250	0.7750	0.5250	0.1000
662	0.0837	0.2019	0.1205	0.1365	0.0570

Fonte: Embrapa

Uma vez obtidas as projeções, as imagens foram reconstruídas utilizando o método da retro-projeção filtrada. Desta forma, foram montadas quatro imagens de tamanho 65x65 pixels, que foram convertidas para o formato BMP (Bitmap), ideais para a utilização na interface desenvolvida para Matlab de forma que pudesse passar pela classificação. Nesta conversão, os valores dos coeficientes de atenuação contidos nas imagens foram normalizados para um intervalo de 256 valores (intensidades).

Foram obtidas então 15 imagens, sendo 4 imagens monocromáticas – provenientes das quatro energias diferentes (Figura 5.4) – e criadas 11 imagens multiespectrais a partir de diferentes combinações das quatro primeiras. As imagens mostradas são as imagens negativas das imagens originais convertidas para o formato BMP. Trabalhou-se com elas nessa forma para obter melhor visualização, uma vez que a inversão na escala de cinza não altera os resultados decorrentes da classificação, como pode ser facilmente verificado através de uma comparação da confusão média após classificação de uma imagem e de sua correspondente imagem negativa.

Para obter-se coerência entre as bandas no caso do uso de imagens multiespectrais, de forma que as informações relativas a cada espectro se mantenham, as bandas de 40, 60, 85 e 662 keV foram normalizadas, respectivamente, por um fator de 17, 7, 5 e 2. Estes valores foram encontrados através dos valores máximos experimentais dos coeficientes de atenuação linear encontrados em cada energia utilizada para as imagens. Os valores observados para as bandas de 40, 60, 85 e 662 keV foram, respectivamente, 4129, 1740, 1148 e 336.

Os valores máximos experimentais estão abaixo dos valores teóricos e esta diferença é bastante expressiva em alguns casos. Este fato está relacionado principalmente às impurezas presentes nos materiais em estudo. Segundo MACOVSKI (1983), para energias menores, o coeficiente de atenuação possui uma forte dependência com o número atômico do material tornando-se mais evidentes os efeitos oriundos destas impurezas. O ruído também contribui muito para este fenômeno e vários são os processos de formação do ruído no imageamento por tomografia computadorizada. Há o ruído inerente ao próprio processo de aquisição das projeções tomográficas como também é acrescentado algum ruído decorrente do dispositivo eletrônico de aquisição (normalmente ruído Gaussiano). Além disso, artefatos decorrentes da própria digitalização da reconstrução podem ser inseridos nos resultados finais.

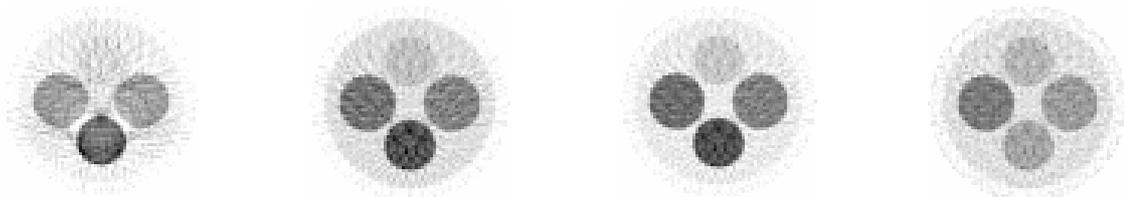


FIGURA 5.4 – Imagens individuais adquiridas com energias de 40, 60, 85 e 662 keV.

5.2 CLASSIFICAÇÃO DAS IMAGENS

5.2.1 FERRAMENTAS UTILIZADAS

Para a utilização das técnicas e implementação da *interface* para a classificação foi utilizado o software Matlab. Esta ferramenta foi escolhida porque reconhecimento de padrões necessita do uso de programas numéricos para estudar dados e avaliar as técnicas utilizadas. Enquanto as técnicas novas estão sendo propostas ainda na literatura é necessária uma plataforma de programação que permita uma execução rápida e flexível. O Matlab oferece todas as características para este tipo de implementação. Além disso, em comparação com outros ambientes estatísticos especializados, oferece uma integração fácil com dados de toda a natureza e possui diversos *toolboxes* especializados e de alta qualidade, facilitando os trabalhos que enfocam teoria de reconhecimento de padrões. O uso de uma ferramenta disponível como Matlab pode ser utilizado e o resultado dos algoritmos compartilhados entre instituições de qualquer área do conhecimento.

Uma das características interessantes do Matlab é a possibilidade de integrar a implementação existente com novas rotinas. Em um projeto desenvolvido pelo grupo de Reconhecimento de Padrões da Universidade de Delft na Holanda, e liderado por Duin, foi criada a *toolbox* Pattern Recognition Tools (PRTools) que atualmente oferece mais de 200 rotinas, cobrindo grande parte da área de reconhecimento de padrões. Conta com grande qualidade de desenvolvimento e constante atualização, visto que é utilizado por pesquisadores de todo o mundo, que fazem contribuições a este projeto. A característica interessante implementada pelo PRTools é tratar os dados de forma bem parecida com as referências teóricas utilizadas. Aproveitando a característica “orientada a objetos” que o Matlab disponibiliza, foram criadas as classes Dataset e Mapping (classes no sentido de orientação a objetos, não confundir com classes de padrões). A primeira armazena os dados a serem classificados bem como outras informações pertinentes sobre estes dados, representados por conjuntos de objetos em forma de matrizes e vetores. A classe Mapping, por sua vez, armazena classificadores treinados, resultados de extração de atributos, projeções não-lineares, etc. Exemplificando a fácil utilização, suponha que tenhamos um Dataset A e um Mapping W. Quando se realiza uma operação $A*W$ obtemos um novo objeto Dataset, agora representando os dados mapeados ou classificados, ou seja o resultado da aplicação da definição do mapeamento de W em A. (DUIN, 2003)

5.2.2 CLASSIFICAÇÃO DAS IMAGENS

Foram realizados experimentos de classificação com as imagens descritas na seção

5.1.3. No processo foram consideradas 6 classes:

- Água,
- Alumínio,
- Fósforo,
- Cálcio,
- Plexiglass,
- Fundo da Imagem.

Assumimos probabilidades *a priori* iguais para todas as classes, pois não havia nenhuma informação sobre as classes. Na construção da imagem temática (imagem classificada), é preciso associar uma cor a cada uma das classes. Para isto foram escolhidas cores para cada classe, relacionadas na Tabela 5.2.

TABELA 5.2 – Cores associadas às classes nas imagens temáticas

Classe	Cor
Água	Azul
Alumínio	Vermelho
Fósforo	Amarelo
Cálcio	Verde
Plexiglass	Cinza Escuro
Fundo	Cinza Claro

Para o treinamento e teste foram adquiridas, para cada classe, 64 amostras em janelas de 8x8 pixels, num total de 384 amostras para cada imagem de 65x65 pixels. Este número de amostras é reduzido e deve-se ao tamanho da imagem gerada pelo tomógrafo. Devido a este fato foi utilizado, além do coeficiente Kappa, o método de estimação de erro por validação cruzada *leave-one-out* de forma a podermos aproveitar todas as amostras para teste e treinamento.

Para cada uma das 15 imagens geradas foi realizado o seguinte processo:

1. Treinamento dos classificadores individuais: bayesiano linear, logístico, parzen e k vizinhos mais próximos (apresentados no Capítulo 3), utilizando a metodologia

descrita acima. Usando cada classificador individual foi classificada a imagem e obtida a estimação do erro da classificação por validação cruzada e o coeficiente Kappa.

2. Os resultados de cada classificador foram combinados usando: votação por maioria, máximo, mínimo, mediana, soma e produto (apresentados no Capítulo 4). Não foi realizado nenhum tratamento nos conjuntos de treinamento para assegurar a independência dos classificadores. Usando os resultados dos combinadores foram classificadas as imagens e obtidas estimções de erros da classificação por validação cruzada, bem como pelo coeficiente Kappa.
3. Com as classes escolhidas por cada combinador, foi aplicada a votação por maioria na saída dos combinadores (descrita na seção 4.2.8), de forma a selecionar a classe mais freqüente entre os combinadores para cada objeto. Para esta técnica também foram classificadas imagens e obtidos o erro da classificação por validação cruzada e o coeficiente Kappa.

6. *Resultados*

Neste capítulo são apresentados os resultados das classificações das imagens tomográficas multiespectrais segundo os classificadores descritos no Capítulo 3 e combinadores descritos no Capítulo 4, de acordo com a metodologia apresentada no Capítulo 5.

Consideramos relevante inserir neste capítulo as estimações do erro, os coeficientes Kappa e as imagens temáticas (classificadas).

Os resultados serão demonstrados em três seções principais. A seção 6.1 apresenta a performance dos classificadores através dos coeficientes Kappa e estimação de erro por validação cruzada. Nesta seção também é demonstrada a média dos erros, e comparação entre os classificadores com o uso destes valores.

A seção 6.2 foi subdividida por imagem e dedica-se à apresentação das imagens temáticas para cada classificador individual e método de combinação de classificadores.

6.1 AVALIAÇÃO DOS CLASSIFICADORES

6.1.1 COEFICIENTE KAPPA DOS CLASSIFICADORES E COMBINADORES

TABELA 6.1 – Coeficientes Kappa dos classificadores para cada imagem

Imagens	k- Vizinhos	Parzen	Logístico	Bayesiano Linear
40	0,6188	0,6344	0,6406	0,6094
60	0,7406	0,7438	0,7125	0,7281
85	0,7750	0,7813	0,7531	0,7688
662	0,7000	0,7094	0,7000	0,7031
40-60	0,7188	0,7688	0,7500	0,7875
40-85	0,7250	0,7844	0,7656	0,8125
40-662	0,6781	0,7875	0,9000	0,9188
60-85	0,7688	0,7813	0,8219	0,8719
60-662	0,8063	0,8656	0,9500	0,9469
85-662	0,8625	0,8875	0,9313	0,9531
40-60-85	0,7469	0,8125	0,9031	0,9375
40-60-662	0,7375	0,8156	0,9438	0,9406
40-85-662	0,7406	0,8563	0,9406	0,9469
60-85-662	0,9675	0,8875	0,9875	0,9844
40-60-85-662	0,7563	0,8563	0,9875	0,9844
<i>Média</i>	0,7562	0,7981	0,8458	0,8596

TABELA 6.2 – Coeficientes Kappa dos combinadores para cada imagem

Imagens	Votação	Máximo	Mínimo	Mediana	Soma	Produto	Votação Comb.
40	0,6344	0,6344	0,6469	0,6094	0,6438	0,6312	0,6438
60	0,7438	0,7281	0,7250	0,7344	0,7375	0,7375	0,7375
85	0,7750	0,7750	0,7594	0,7781	0,7750	0,7781	0,7750
662	0,7031	0,6969	0,7031	0,7125	0,7000	0,7125	0,7000
40-60	0,7813	0,7688	0,7625	0,7906	0,7844	0,7813	0,7969
40-85	0,8125	0,8125	0,8000	0,8219	0,8250	0,8063	0,8250
40-662	0,8781	0,8938	0,8906	0,9063	0,9068	0,9125	0,9129
60-85	0,8469	0,8438	0,8125	0,8750	0,8656	0,8688	0,8750
60-662	0,9188	0,9438	0,9500	0,9625	0,9563	0,9563	0,9594
85-662	0,9219	0,9406	0,9156	0,9500	0,9469	0,9375	0,9438
40-60-85	0,8969	0,9250	0,9000	0,9344	0,9438	0,9500	0,9469
40-60-662	0,9125	0,9375	0,9188	0,9469	0,9500	0,9562	0,9531
40-85-662	0,9281	0,9438	0,9188	0,9219	0,9469	0,9406	0,9500
60-85-662	0,9875	0,9813	0,9875	0,9844	0,9813	0,9844	0,9844
40-60-85-662	0,9469	0,9813	0,9500	0,9688	0,9844	0,9813	0,9875
<i>Média</i>	0,8458	0,8529	0,8442	0,8600	0,8619	0,8635	0,8654

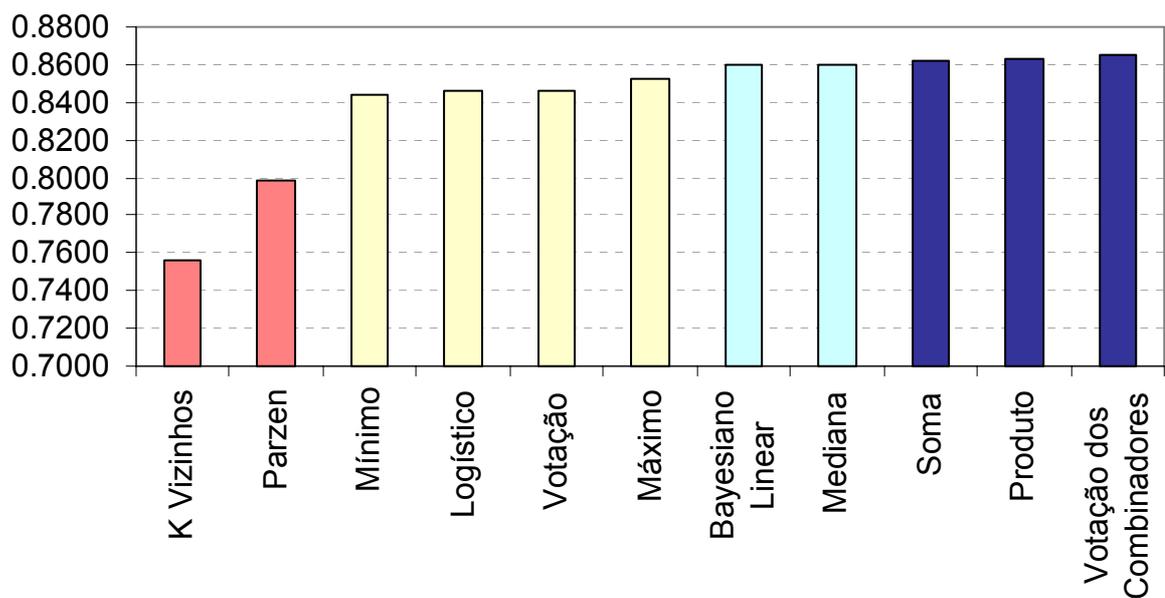


FIGURA 6.1 – Comparação dos classificadores e combinadores pela média do Kappa

6.1.2 ESTIMAÇÃO DE ERRO POR VALIDAÇÃO CRUZADA DOS CLASSIFICADORES E COMBINADORES

TABELA 6.3 – Estimação do erro dos classificadores para cada imagem

Imagens	k- Vizinhos	Parzen	Logístico	Bayesiano Linear
40	0,5495	0,3151	0,4323	0,3385
60	0,4609	0,2240	0,2630	0,2318
85	0,4349	0,1901	0,2240	0,1927
662	0,2995	0,2578	0,2865	0,2474
40-60	0,3542	0,2135	0,2422	0,1875
40-85	0,3151	0,1953	0,2083	0,1563
40-662	0,2344	0,2031	0,0859	0,0781
60-85	0,2943	0,1901	0,1615	0,1120
60-662	0,0391	0,1302	0,0443	0,0469
85-662	0,0625	0,1068	0,0651	0,0391
40-60-85	0,2995	0,1693	0,0911	0,0547
40-60-662	0,2656	0,1849	0,0495	0,0495
40-85-662	0,2422	0,1458	0,0677	0,0443
60-85-662	0,0443	0,1172	0,0208	0,0156
40-60-85-662	0,2344	0,1458	0,0156	0,0130
<i>Média</i>	0,2347	0,1631	0,1189	0,0969

TABELA 6.4 – Estimação do erro dos combinadores para cada imagem

Imagens	Votação	Máximo	Mínimo	Mediana	Soma	Produto	Votação Comb.
40	0,2865	0,2787	0,4069	0,3385	0,2787	0,2396	0,2578
60	0,1927	0,2057	0,1882	0,1680	0,2057	0,1693	0,1797
85	0,1979	0,1823	0,3073	0,1484	0,1693	0,1302	0,1712
662	0,2266	0,2656	0,2246	0,2188	0,2656	0,1908	0,2162
40-60	0,0703	0,0026	0,1484	0,0156	0,0026	0,0026	0,0026
40-85	0,0521	0,0156	0,1537	0,0182	0,0156	0,0104	0,0130
40-662	0,0326	0,0234	0,1589	0,0182	0,0234	0,0183	0,0182
60-85	0,0625	0,0260	0,0189	0,0286	0,0260	0,0156	0,0130
60-662	0,0026	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000
85-662	0,0163	0,0130	0,0078	0,0130	0,0130	0,0078	0,0104
40-60-85	0,0260	0,0026	0,0026	0,0142	0,0026	0,0026	0,0026
40-60-662	0,0052	0,0026	0,1458	0,0078	0,0026	0,0026	0,0026
40-85-662	0,0163	0,0078	0,1484	0,0078	0,0078	0,0078	0,0078
60-85-662	0,0000	0,0000	0,0000	0,0052	0,0000	0,0000	0,0000
40-60-85-662	0,0026	0,0026	0,0026	0,0078	0,0026	0,0026	0,0026
<i>Média</i>	0,0645	0,0536	0,0980	0,0475	0,0526	0,0397	0,0450

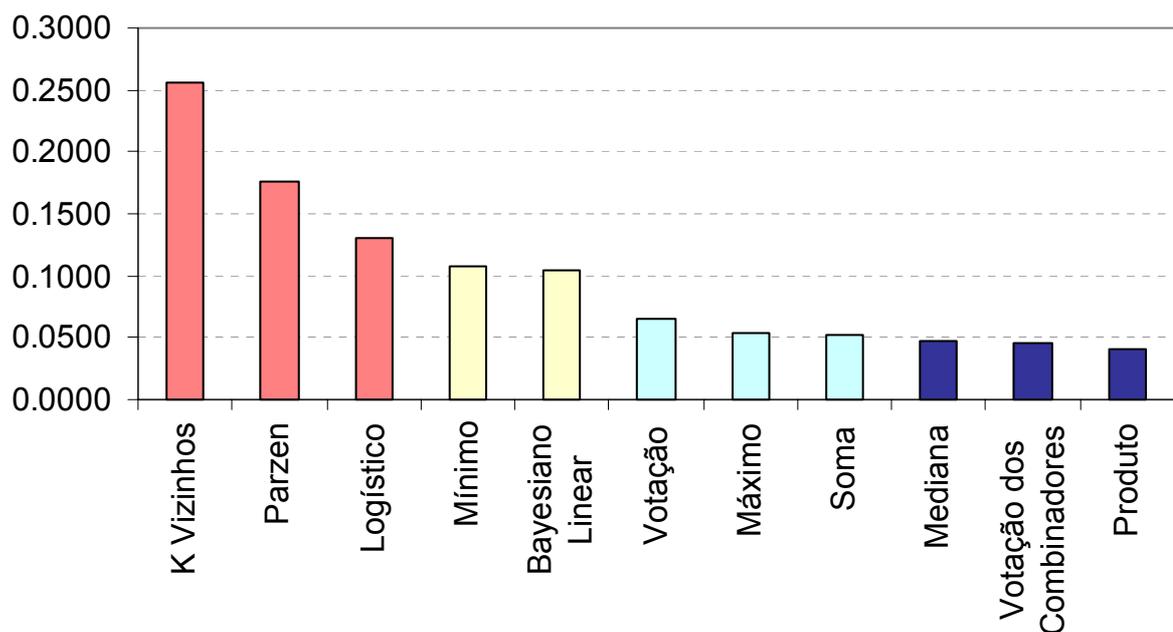


FIGURA 6.2 – Comparação dos classificadores e combinadores pela média do erro

6.2 IMAGENS CLASSIFICADAS

6.2.1 IMAGEM 40 keV

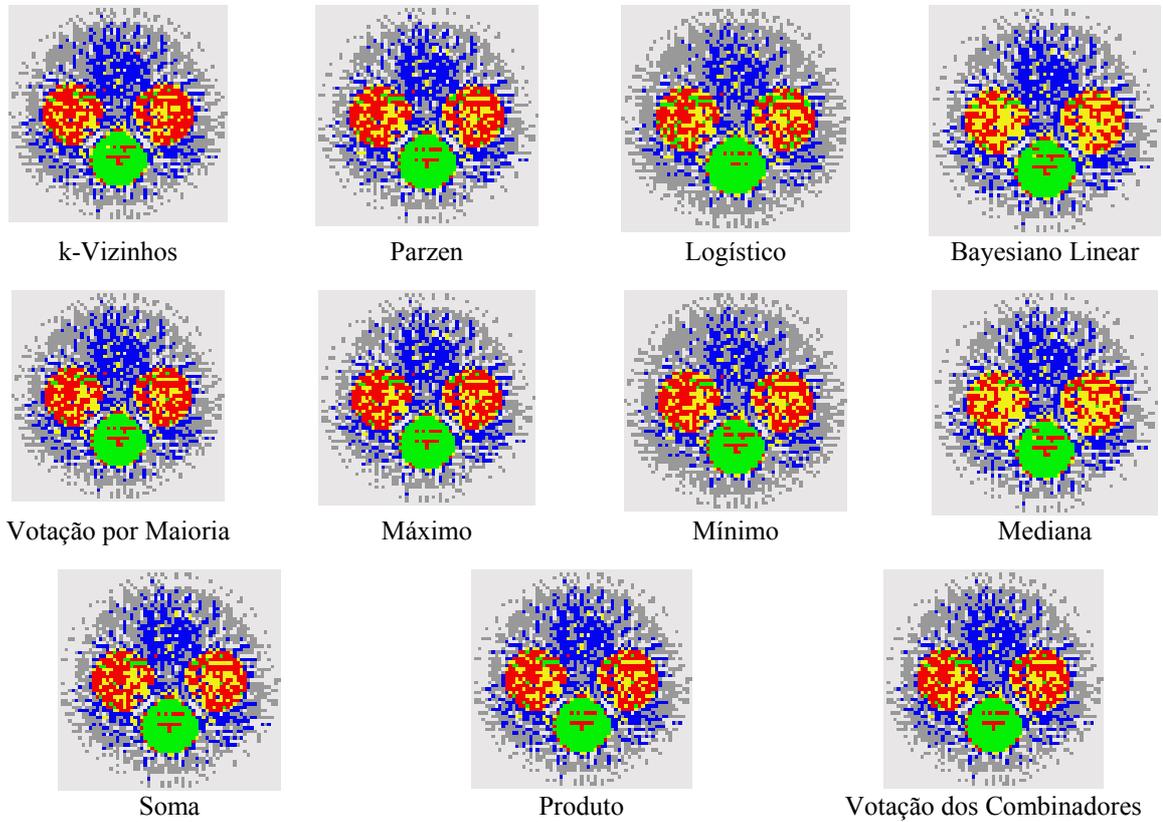


FIGURA 6.3 – Imagens Temáticas para a classificação da banda de 40 keV

6.2.2 IMAGEM 60 keV

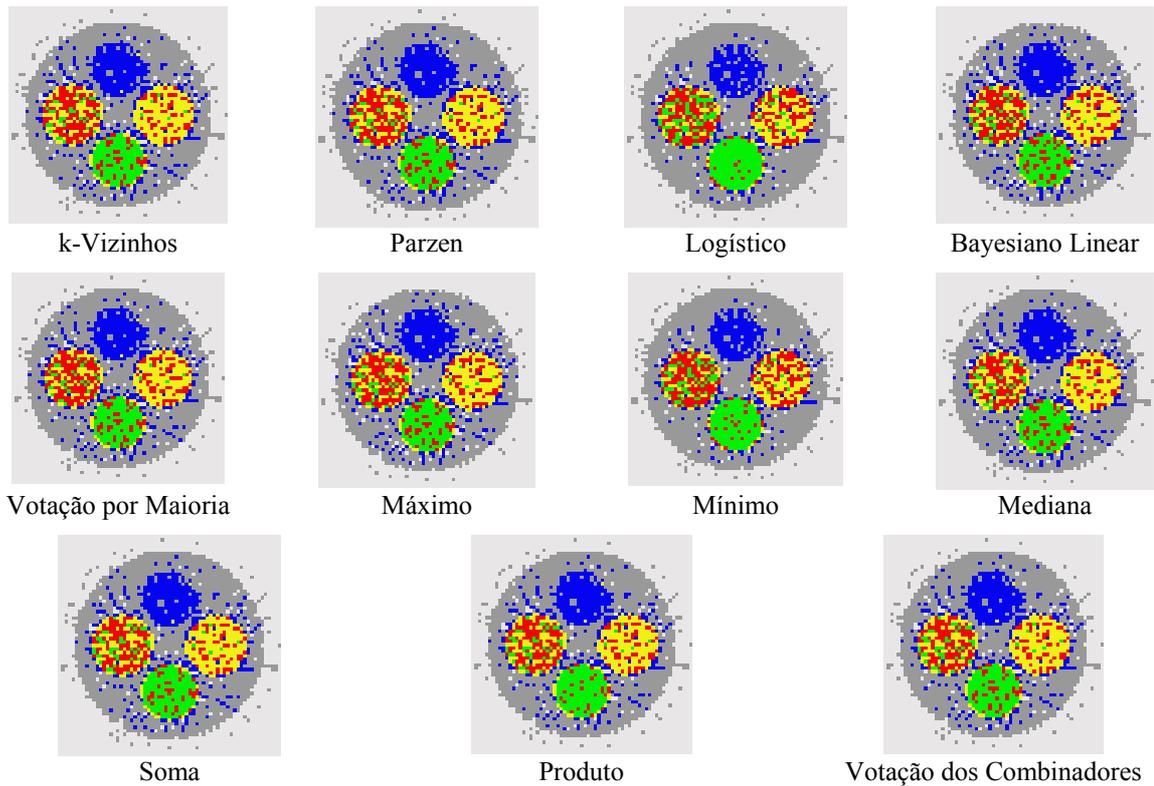


FIGURA 6.4 – Imagens Temáticas para a classificação da banda de 60 keV

6.2.3 IMAGEM 85 keV

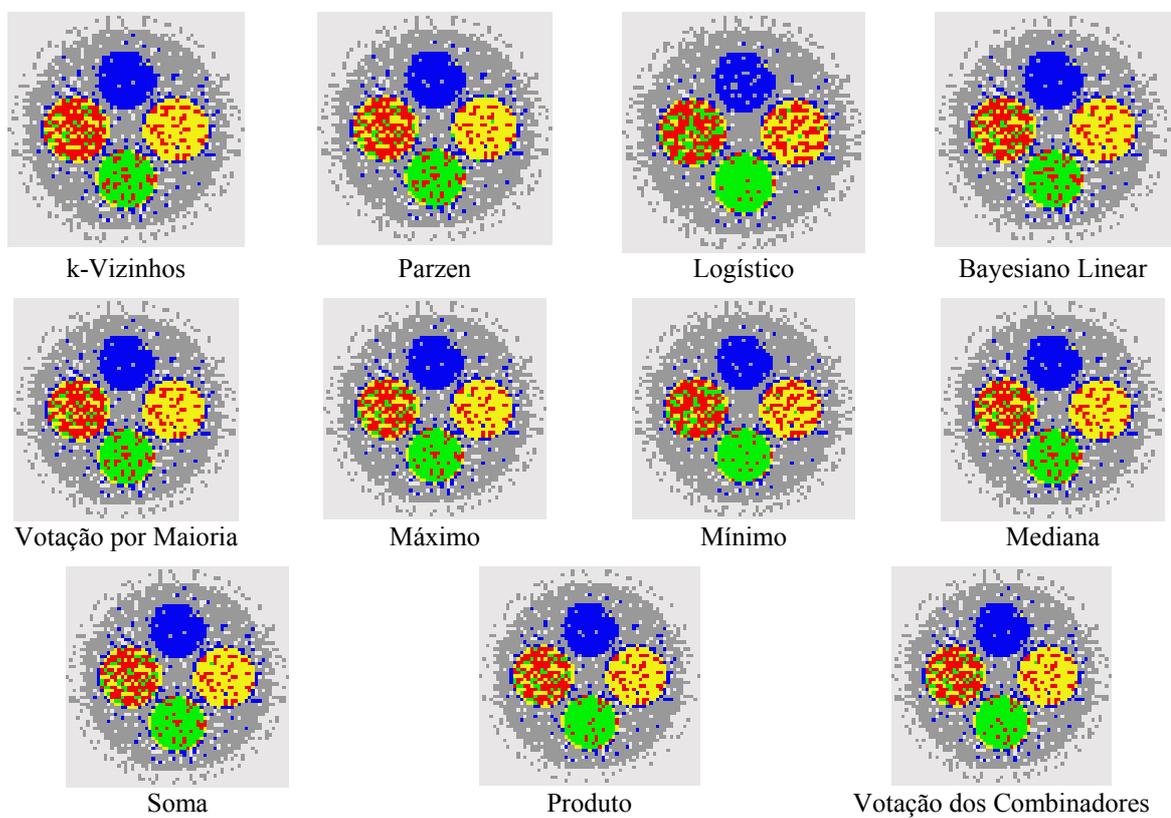


FIGURA 6.5 – Imagens Temáticas para a classificação da banda de 85 keV

6.2.4 IMAGEM 662 keV

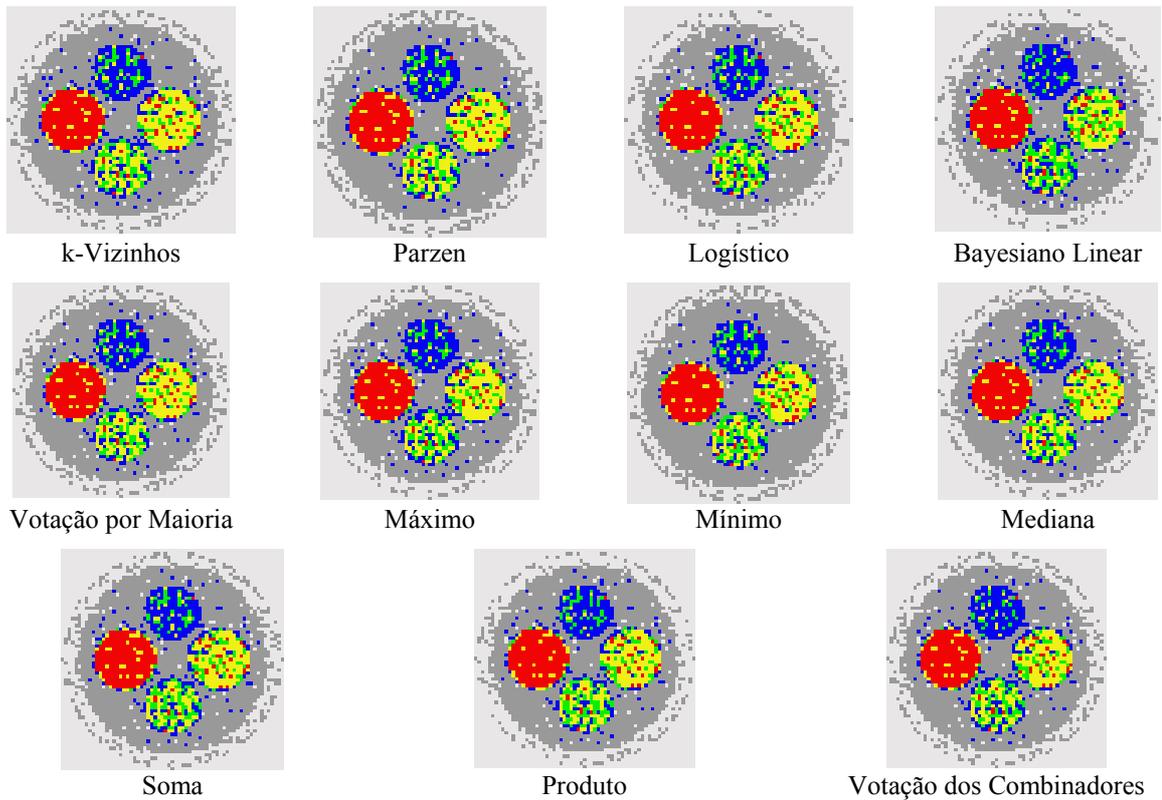


FIGURA 6.6 – Imagens Temáticas para a classificação da banda de 662 keV

6.2.5 IMAGEM 40-60 keV

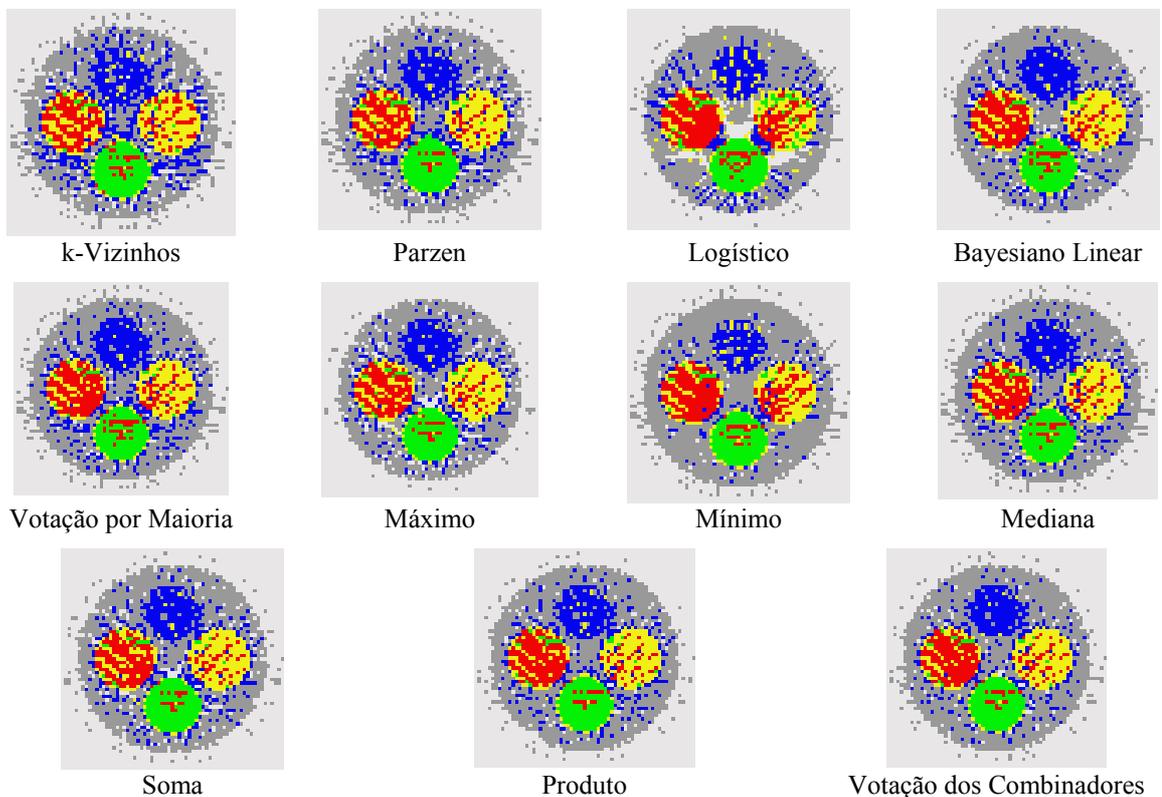


FIGURA 6.7 – Imagens Temáticas para a classificação das bandas de 40 e 60 keV

6.2.6 IMAGEM 40-85 keV

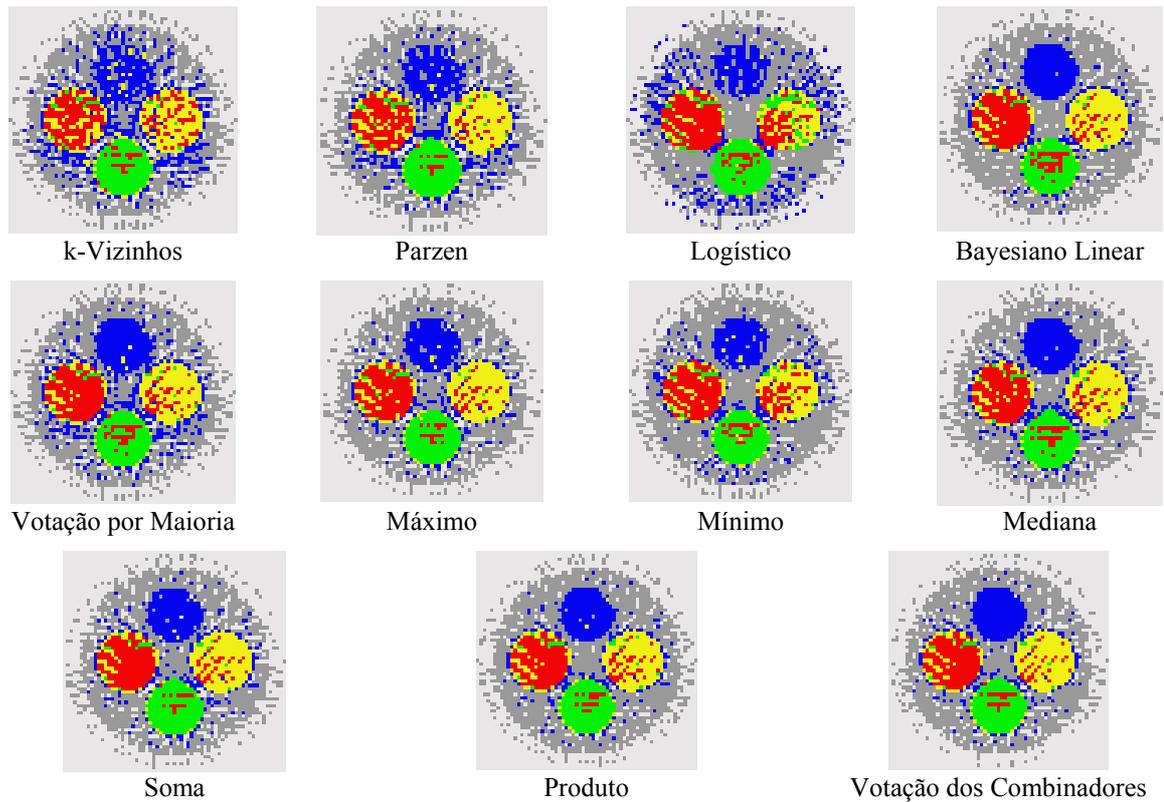


FIGURA 6.8 – Imagens Temáticas para a classificação das bandas de 40 e 85 keV

6.2.7 IMAGEM 40-662 keV

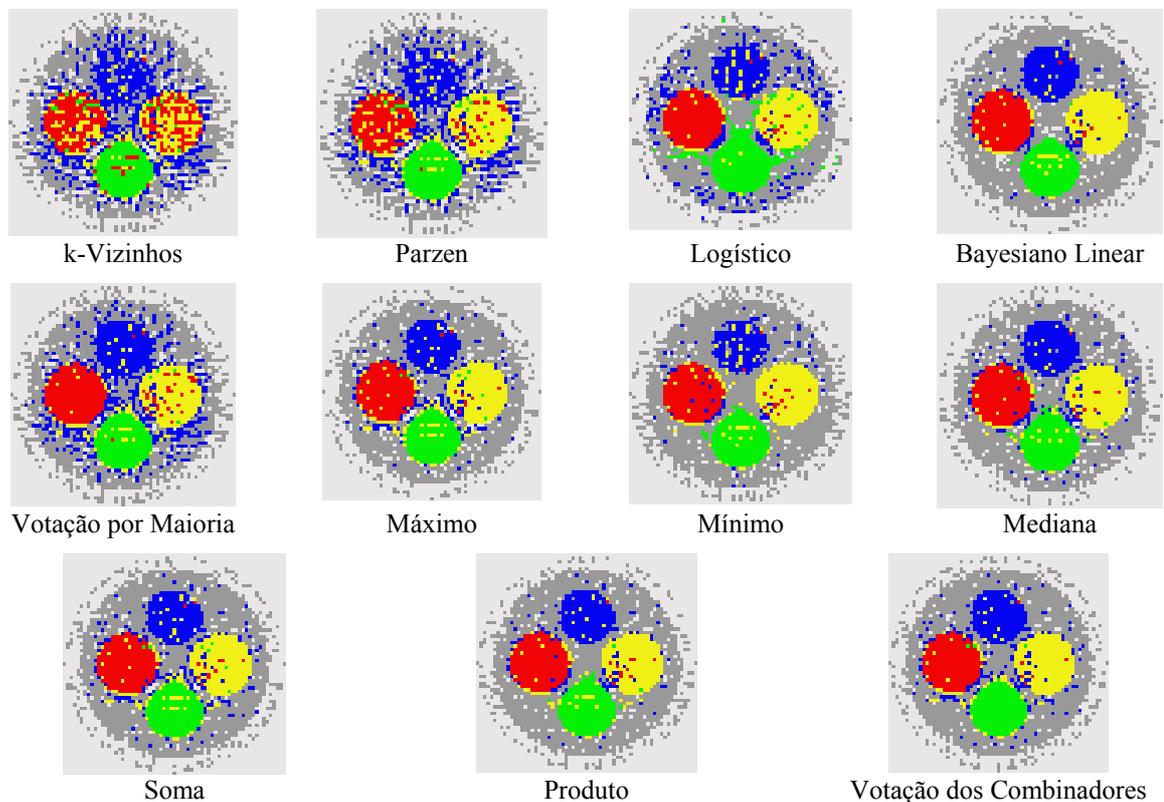


FIGURA 6.9 – Imagens Temáticas para a classificação das bandas de 40 e 662 keV

6.2.8 IMAGEM 60-85 keV

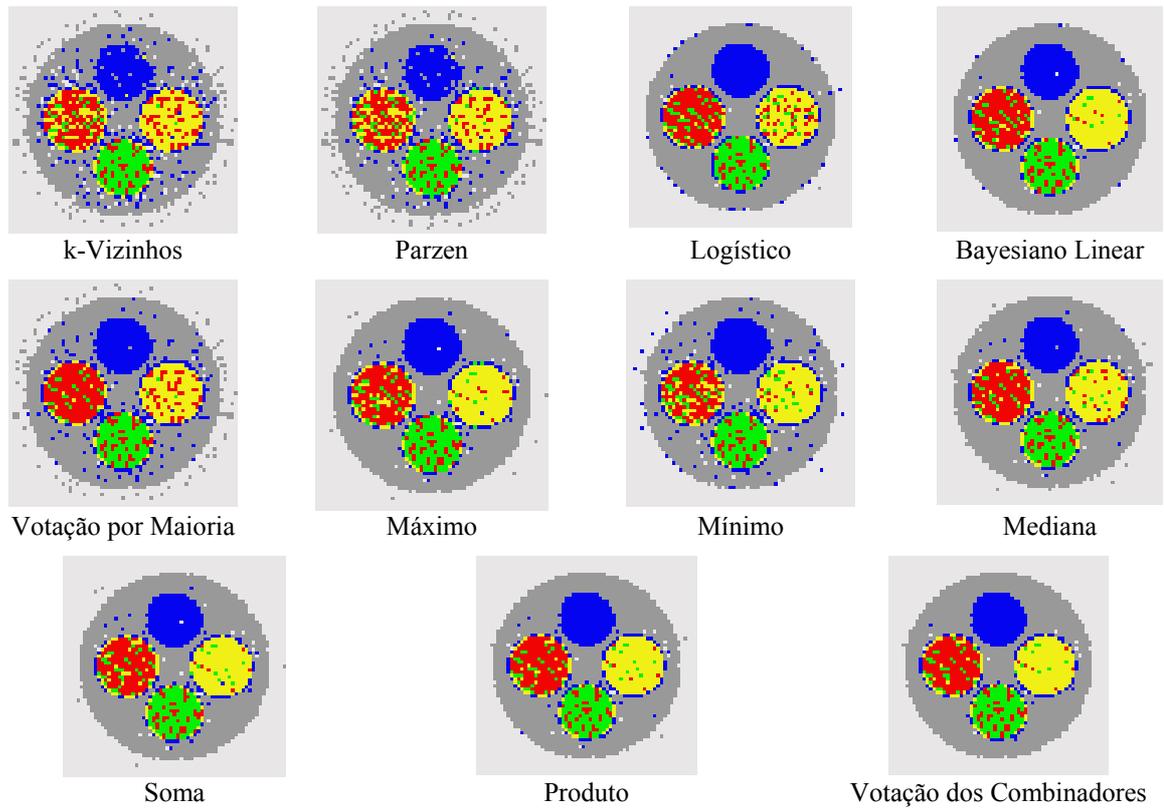


FIGURA 6.10 – Imagens Temáticas para a classificação das bandas de 60 e 85 keV

6.2.9 IMAGEM 60-662 keV

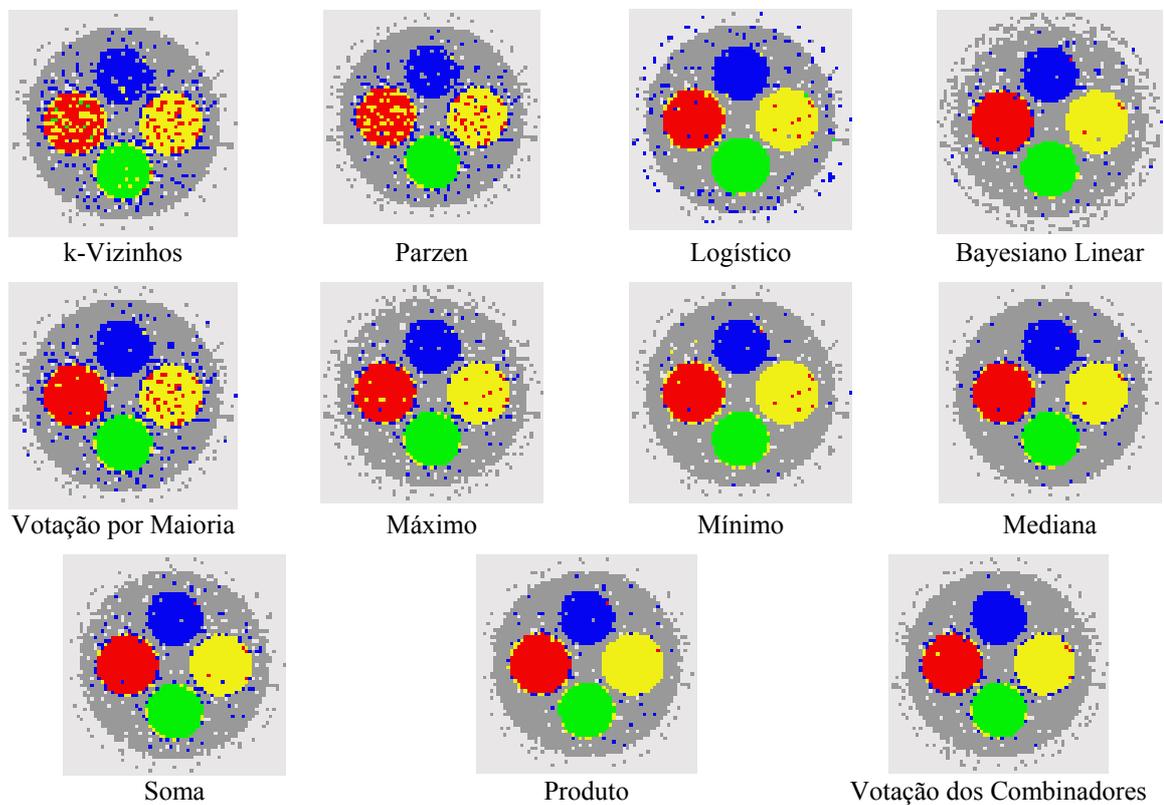


FIGURA 6.11 – Imagens Temáticas para a classificação das bandas de 60 e 662 keV

6.2.10 IMAGEM 85-662 keV

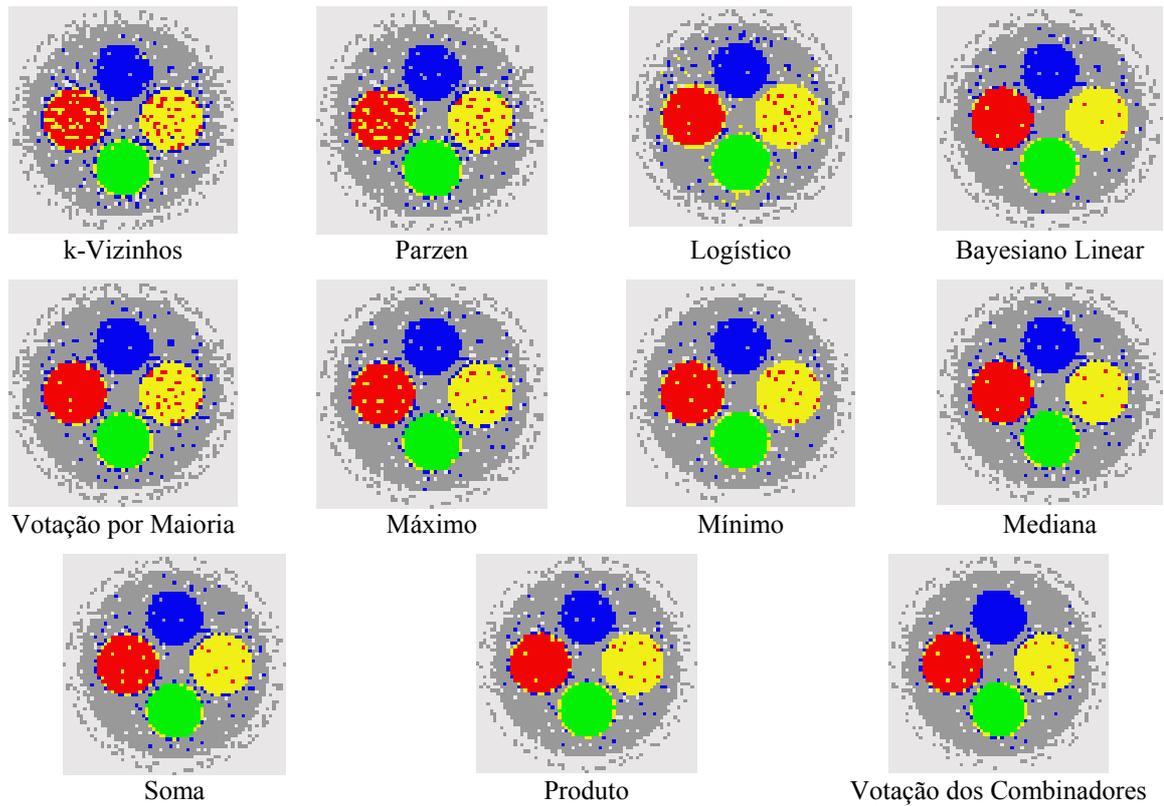


FIGURA 6.12 – Imagens Temáticas para a classificação das bandas de 85 e 662 keV

6.2.11 IMAGEM 40-60-85 keV

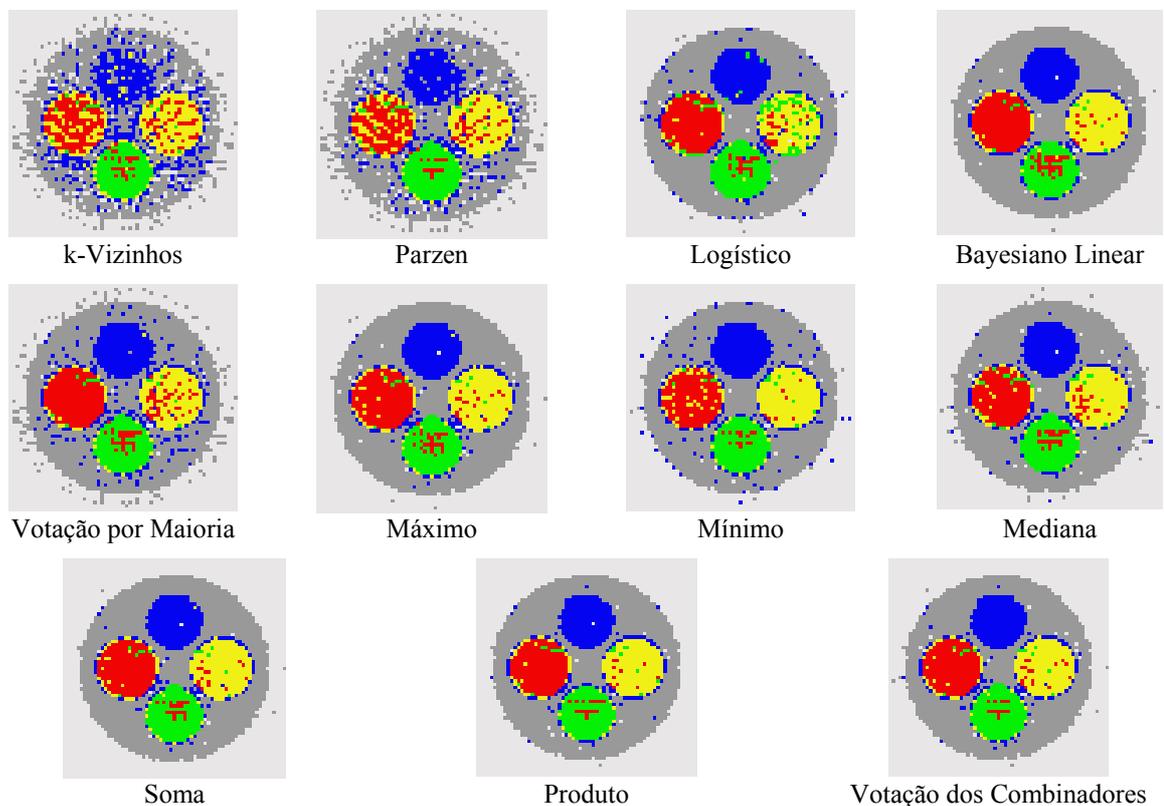


FIGURA 6.13 – Imagens Temáticas para a classificação das bandas de 40, 60 e 85 keV

6.2.12 IMAGEM 40-60-662 keV

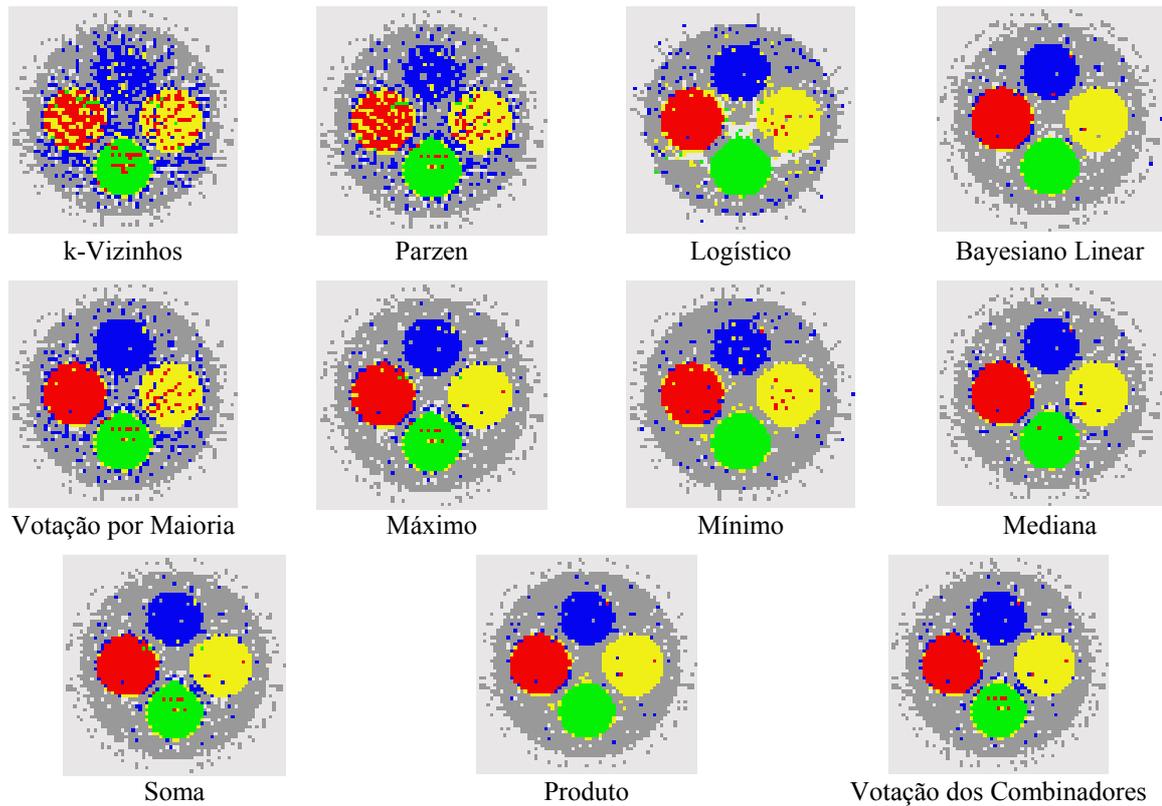


FIGURA 6.14 – Imagens Temáticas para a classificação das bandas de 40, 60 e 662 keV

6.2.13 IMAGEM 40-85-662 keV

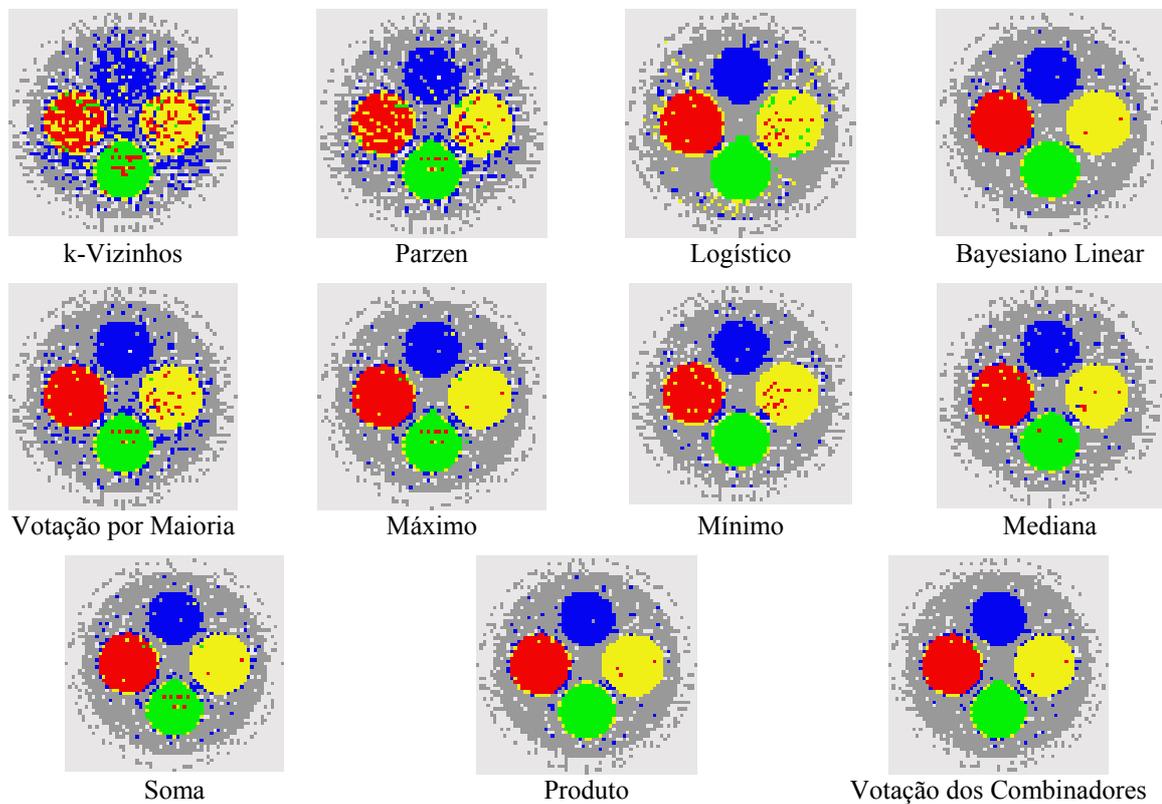


FIGURA 6.15 – Imagens Temáticas para a classificação das bandas de 40, 85 e 662 keV

6.2.14 IMAGEM 60-85-662 keV

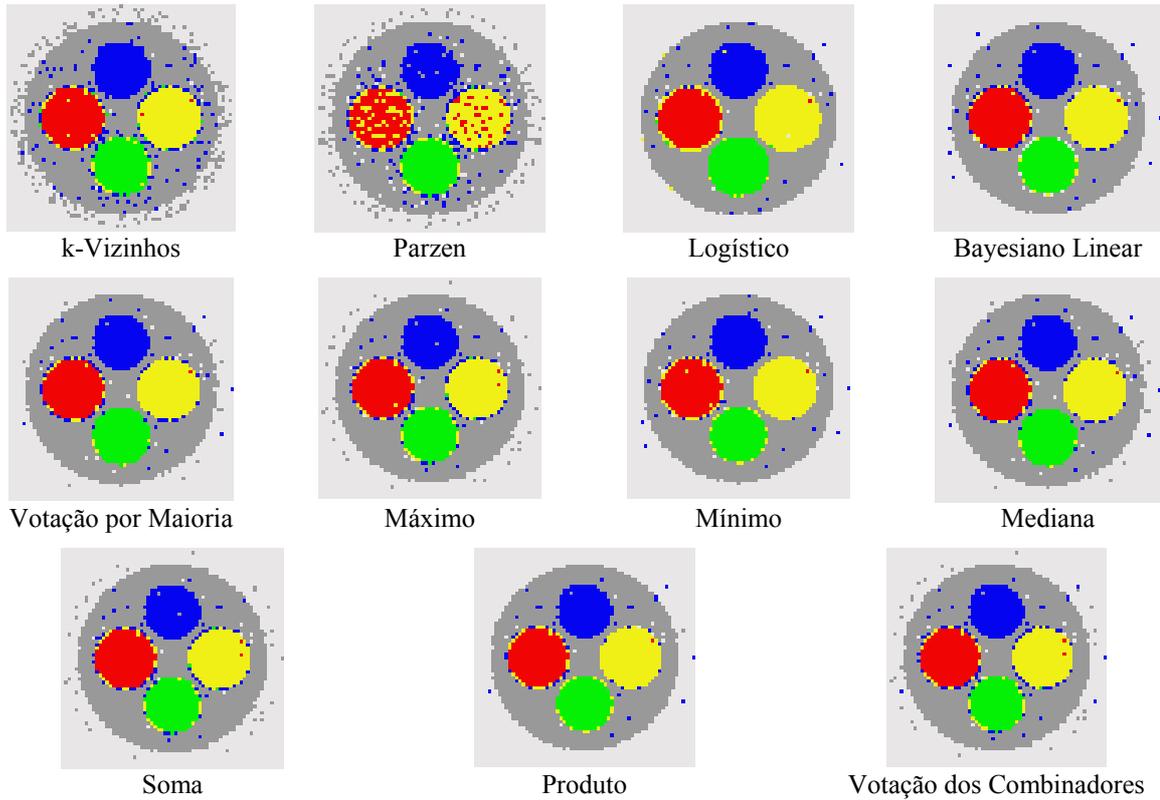


FIGURA 6.16 – Imagens Temáticas para a classificação das bandas de 60, 85 e 662 keV

6.2.15 IMAGEM 40-60-85-662 keV

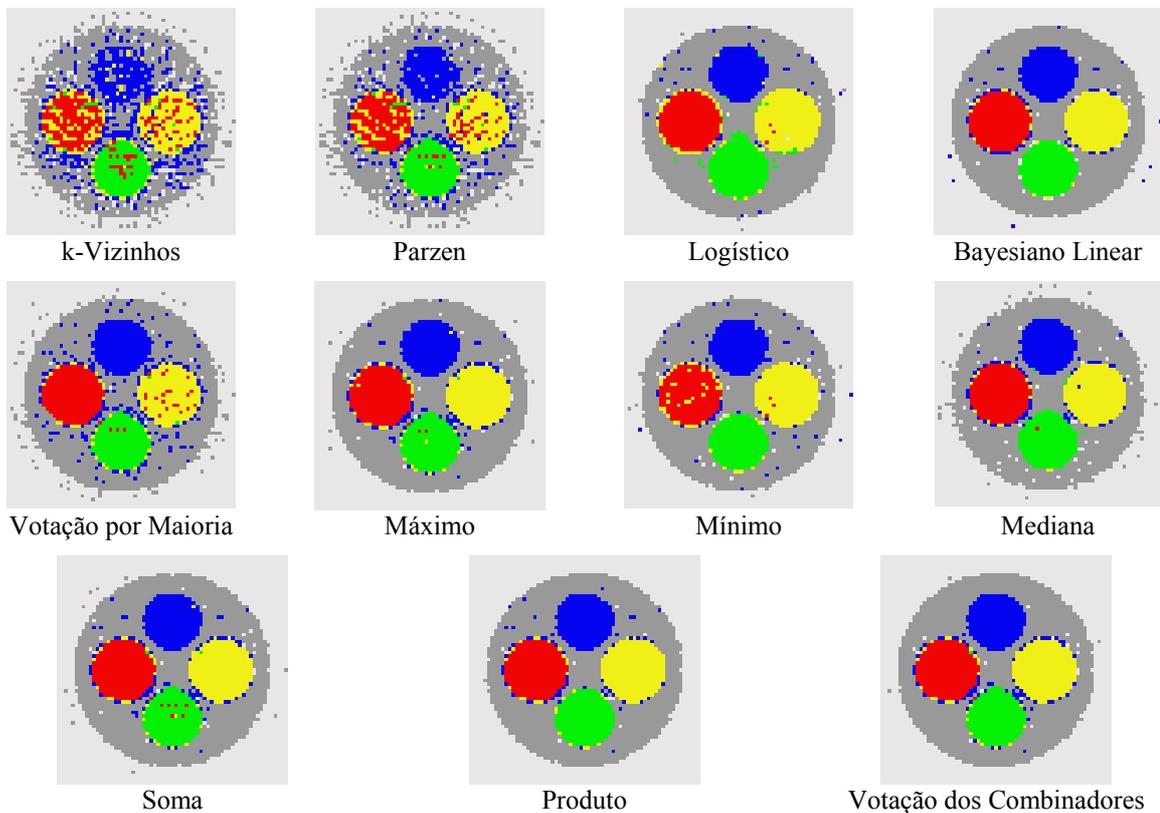


FIGURA 6.17 - Imagens Temáticas para a classificação das bandas 40, 60, 85 e 662 keV

7. *Discussão*

Através dos resultados obtidos e apresentados no Capítulo 6, diversas análises e observações foram feitas de modo a tirar conclusões sobre o desempenho e o comportamento dos classificadores e combinadores atuando sobre as imagens multiespectrais ruidosas de tomografia computadorizada. Iniciamos pela comparação do desempenho com as energias empregadas e o coeficiente de atenuação linear para diversas combinações de bandas. Em seguida analisamos os métodos de avaliação dos classificadores, finalizando com discussões acerca da performance e comportamento dos classificadores e combinadores.

Os resultados da classificação em geral podem ser comparados com os dados da Tabela 5.1, que apresenta os coeficientes de atenuação linear para os materiais que compõem as classes determinadas em cada classificação, e da Figura 5.2 contendo a curva do coeficiente de atenuação linear para os materiais estudados em função da energia empregada. Verifica-se que os coeficientes de atenuação da água e plexiglass, bem como do alumínio e fósforo apresentam-se próximos, principalmente para a energia de 40 keV. Há também proximidade de valores para as energias 60 e 85 keV, mas a diferença é maior. Outra comparação interessante é que, no caso da energia 663 keV, os valores próximos são do cálcio em relação ao fósforo.

As imagens temáticas para a classificação da banda adquirida com a energia de 40 keV (Figura 6.3) mostram claramente a confusão entre as classes água e plexiglass e fósforo e alumínio, o que mostra a baixa discriminação entre estes materiais, já prevista pela análise de seus coeficientes de atenuação linear. No caso da classificação das bandas de 60 e 85 keV (Figuras 6.4 e 6.5), também há confusão entre as classes citadas, porém em menor proporção. Isto foi observado anteriormente por *HOMEM et. al. (2000)*, que utilizou coeficientes de atenuação linear como atributos para classificação. Para a banda de 662 keV (Figura 6.6) verificou-se também a confusão prevista entre cálcio e fósforo, ainda adicionada de uma confusão entre outras classes, muito provavelmente causada pelo fato desta energia gerar valores mais baixos e ser mais afetada pela presença do ruído.

Nas energias individuais, a proximidade entre os coeficientes de atenuação e a presença de ruído nas imagens torna baixo o desempenho da classificação. Mesmo com uma

pequena melhora proporcionada pelos métodos de combinação, como se pode verificar nas tabelas e nas imagens temáticas, a classificação é insatisfatória, sobretudo para a banda de 40 keV (Figura 6.3). Isto era esperado pois as imagens foram adquiridas com baixo tempo de exposição de forma a adicionar ruído e permitir verificar melhor a capacidade de cada técnica de classificação. Nas combinações de 2 ou mais bandas, é possível verificar melhor a diferença entre cada classificador e a eficiência dos combinadores. Mesmo quando utilizada a banda de 40 keV junto com outras, há um aumento no desempenho dos classificadores. Isto porque atributos que obtêm desempenho ruim quando usados individualmente podem, quando combinados, determinar bons resultados, como observado por FUKUNAGA (1990).

Analisando os dois métodos de avaliação de classificação, observamos que o coeficiente Kappa torna mais confusa a análise dos dados, apresentando resultados mais próximos para os diferentes classificadores, enquanto a estimação de erro por validação cruzada (utilizando *leave-one-out*), é mais sensível às diferentes performances. Em um estudo aprofundado sobre a validação cruzada, STONE (1974) verificou que esta técnica gera naturalmente maior variância do que outros índices como o coeficiente Kappa. No entanto, a estimação de erro por *leave-one-out* é aproximadamente livre de viés (*bias*), $E(\hat{\theta}) = \theta$, ou seja, possuindo teoricamente uma estimação $\hat{\theta}$ mais próxima da real θ . A validação cruzada é ainda preferível em casos de poucas amostras (JAIN et. al., 2000). Além disso, mostra em sua estimação não apenas o erro do classificador segundo um conjunto de teste, mas treina e testa usando todas as possibilidades de um conjunto de amostras disponível, medindo também a capacidade de generalização e a estabilidade do classificador (HASTIE et. al., 2001). Devido a estes fatos, tomaremos o método de estimação de erro por validação cruzada como mais preciso e adequado para medir a performance da classificação nos experimentos realizados neste trabalho.

Os classificadores individuais apresentaram um comportamento interessante. Para imagens monoespectrais, o melhor desempenho em geral foi do classificador Parzen. Para imagens multiespectrais, os classificadores logístico e Bayesiano linear superaram os demais em quase todos os casos. Isto demonstra as diferentes capacidades de cada classificador para tratar com diferentes atributos. A Figura 7.1 mostra, a título de exemplo, a divisão do espaço de atributos feita por cada classificador individual para a imagem composta das bandas de 40 e 85 keV, demonstrando a semelhança entre os classificadores logístico e Bayesiano linear

que geram superfícies de decisão mais suaves, enquanto os classificadores Parzen e k-vizinhos mais próximos, que geram superfícies mais irregulares.

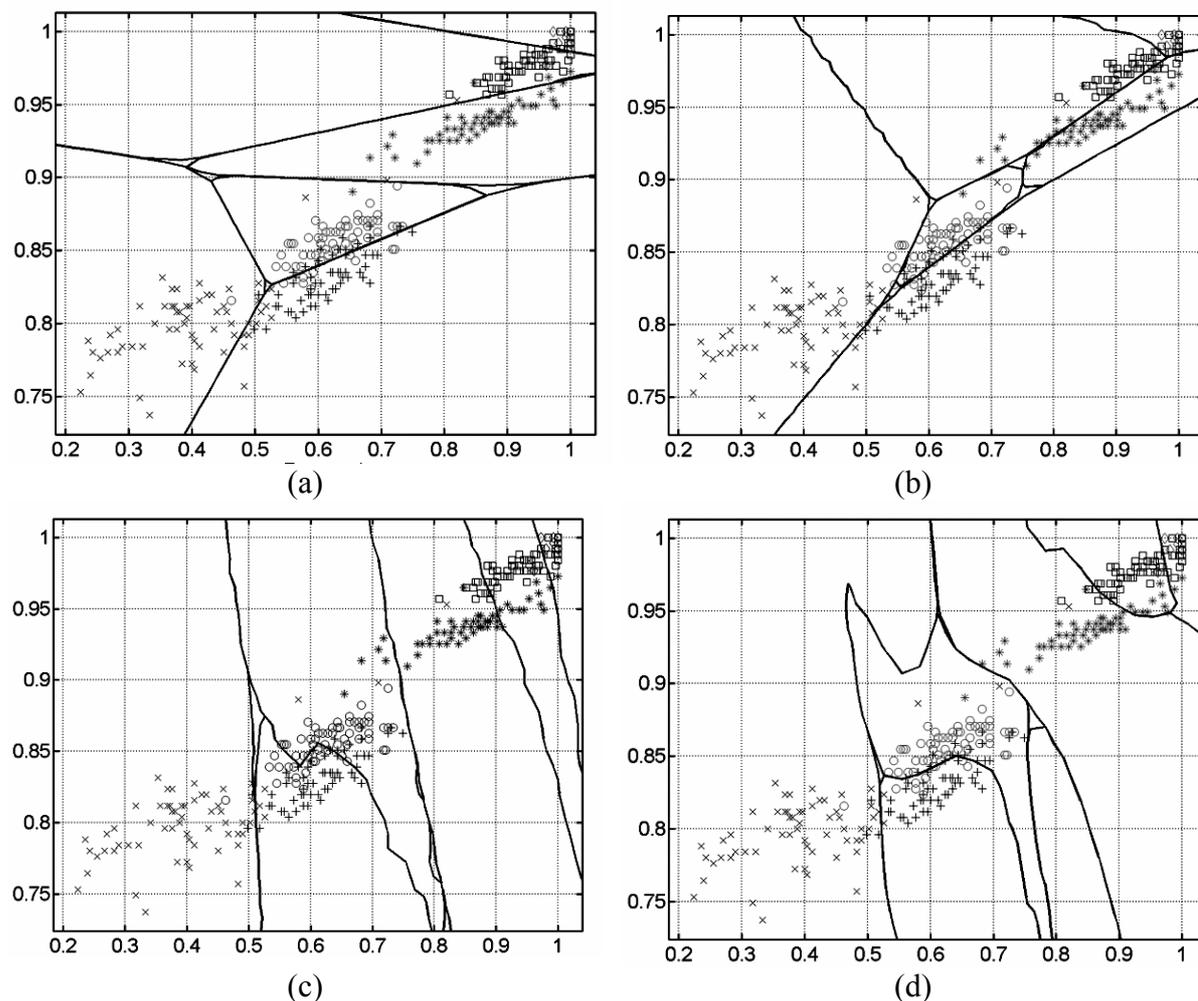


FIGURA 7.1 – Divisão do espaço de atributos por diferentes classificadores: (a) Bayesiano linear, (b) logístico, (c) k-vizinhos, (d) Parzen

O fato do desempenho dos classificadores individuais obtido com uma banda ser maior para o Parzen e com várias bandas para os classificadores Logístico e Bayesiano Linear pode ser explicado pelo fato das superfícies de decisão dos classificadores Logístico e Bayesiano Linear serem mais suaves, causando um maior erro no caso das bandas individuais, muito ruidosas, mas ignorando este ruído quando utilizamos de vários atributos. Já no caso do Parzen, a estimação de densidade é realizada diretamente pelos dados, o que pode ter privilegiado este classificador no caso das bandas ruidosas, porém, o nível do ruído por vezes é tão alto (principalmente na banda de 40 keV) devido à baixa exposição no procedimento de aquisição (detalhes na seção 5.1.3) que fica difícil analisar o comportamento do classificador. Ao tratar com várias bandas os parâmetros de classificação obtiveram superfícies irregulares, estimando a densidade levando em consideração o ruído existente em todos os atributos,

adicionando cada vez mais erros de classificação, o que contribuiu para o aumento na confusão do Parzen e k-vizinhos mais próximos com maior número de bandas, quando comparado com os outros classificadores.

Afirmamos acima que os classificadores escolhidos possuem características bem diferentes. Como examinado na seção 4.1, JAIN et. al. (2000) e DUIN e TAX (2000) afirmam que a atuação dos combinadores em situações deste tipo pode ampliar a performance. Como exemplo, verificamos os limiares de decisão calculados pelos combinadores produto e mínimo (respectivamente o que obteve a maior e menor performance) a partir dos classificadores para a imagem 40 e 85 keV (vistos na Figura 7.2). Neste caso, verifica-se uma redução considerável no erro, que, no melhor classificador individual, Bayesiano Linear, foi de 0.1563 e que para o combinador Produto foi de 0.0104. A melhoria obtida pelos combinadores pode ser constatada na Figura 6.2, onde o gráfico de barras mostra que o erro médio dos métodos de classificação claramente decai para os combinadores.

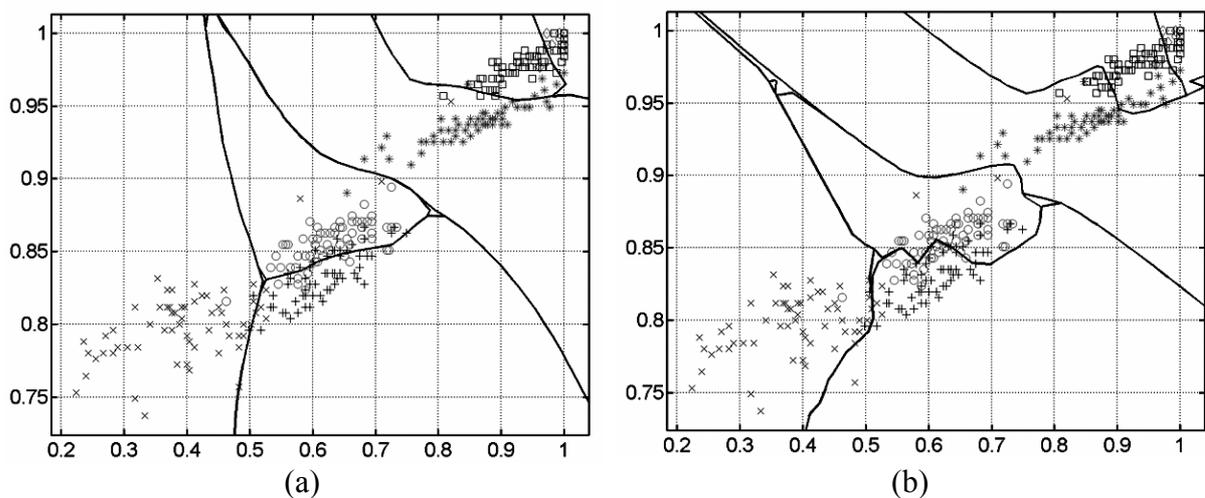


FIGURA 7.2 – Divisão do espaço de atributos pelos combinadores: (a) produto (maior performance), (b) mínimo (menor performance)

Uma observação importante a se fazer é que, conforme já comentado, os experimentos contavam com um pequeno número de amostras disponíveis para treinamento dos classificadores. Este fator, além de limitar a capacidade de classificação ainda aumenta a variância do estimador da probabilidade de erro. Segundo JAIN et. al. (2000), uma das propriedades interessantes dos métodos de combinação de classificadores é o fato de obterem uma menor variância na estimação do erro, gerando probabilidades de erro mais precisas. Nos experimentos realizados, notou-se que a combinação de classificadores alterou a variância

mostrada pelo coeficiente Kappa, gerando valores na ordem de 8 a 10 vezes menores do que nos experimentos com classificadores individuais.

Os experimentos demonstraram que, de maneira geral, os combinadores superaram a capacidade de generalização e são mais estáveis, além de melhorar a performance no caso das imagens aqui pesquisadas. No entanto algumas técnicas se diferenciam muito das outras, merecendo uma atenção especial, pois a literatura realiza diversas suposições acerca do uso destas técnicas (detalhes no Capítulo 4). Não é possível indicar uma técnica de combinação como sendo a melhor, pois como visto em COVER (1969), é preciso ser cuidadoso ao analisar o desempenho de uma técnica e recomendá-la, pois sua natureza teórica sempre dá margens a resultados dependentes de uma série de fatores. Vamos discutir o desempenho e as suposições teóricas realizadas para cada uma das técnicas e compará-las para as imagens em estudo.

Fundamentada por KITTLER et. al. (1998), as Regras do Produto e Soma dão origem a todas as outras. A Regra do Produto assume que as probabilidades *a posteriori* combinadas possuem aproximadamente erros independentes (diferentes para cada classificador). Apesar de ser uma situação improvável em muitos casos, a grande diferença entre as características dos classificadores nos experimentos demonstrou ser suficiente para satisfazer a aproximação em que se baseia a regra. Esta regra demonstrou o menor erro de maneira geral. Isto se deve provavelmente ao fato de que apenas uma suposição é feita para obter a aproximação que encerra em seu embasamento teórico.

A Regra da Soma, no entanto, que nos testes de KITTLER et. al. (1998) superou os outros métodos, apresentou performance mediana neste trabalho. Isto se deve ao fato de que suas suposições são muito rígidas. A principal delas é que as probabilidades *a posteriori* não se desviam muito das probabilidades *a priori*, o que não se pode afirmar ao certo em dados reais. Nos experimentos realizados neste trabalho, assumimos iguais probabilidades *a priori* para todas as classes o que torna a Regra da Soma uma espécie de Regra da Média. Outras características deste combinador segundo DUIN e TAX (2000), é que se assume teoricamente que os erros dos classificadores são correlacionados. KITTLER et. al. (1998) afirma entretanto que a principal suposição pode ser aproximada quando as informações disponíveis são ambíguas, como no caso da presença de alto nível de ruído, que é o caso das informações contidas nos atributos que utilizamos, o que possibilitou um desempenho razoável para esta técnica.

A Regra do Máximo se apresenta geralmente sensível ao ruído presente nos dados, segundo (DUIN; TAX, 2000), incorporando os erros dos classificadores causados pela presença do ruído. Não há um estudo aprofundado sobre o comportamento desta regra, mas ela é derivada a partir da regra da soma, herdando indiretamente algumas das características. Experimentalmente observamos que o comportamento desta técnica é similar ao da Soma (Figura 7.3), sendo que a Regra da Soma apresenta divisões do espaço de atributos geralmente mais suaves. O desempenho da Regra do Máximo também é comparável ao da Soma, tendo performance inferior apenas em alguns casos devido, provavelmente, ao ruído.

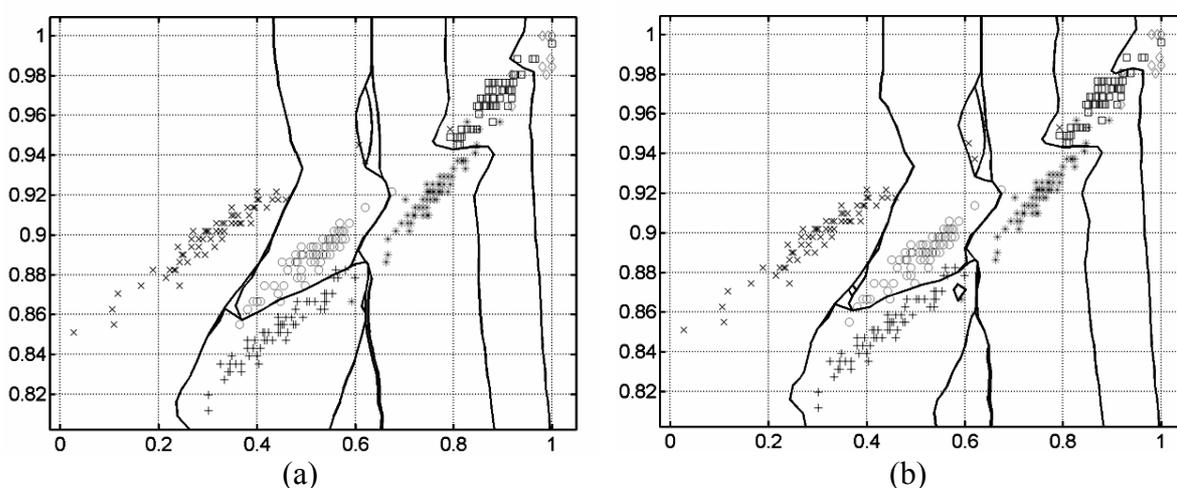


FIGURA 7.3 – Divisão do espaço de atributos da imagem de 60 e 662 keV pelos combinadores: (a) soma, (b) máximo

A Regra do Mínimo baseia-se na Regra do Produto, tomando por base o máximo produto das probabilidades *a posteriori* para cada classe, escolhendo a máxima probabilidade entre as classes das mínimas probabilidades *a posteriori* dos classificadores. É uma opção de combinação pessimista, supondo que os classificadores em geral cometem erros. Seu desempenho é muito variável e apresentou-se melhor quando a confusão dos classificadores foi maior. Na maioria dos casos, sua performance não apresentou melhoria significativa perante os classificadores individuais logístico e Bayesiano linear.

A Regra da Mediana é uma modificação da Regra da Média – que, por sua vez representa uma variação da Regra da Soma. Esta modificação foi realizada, pois, no caso de alguma saída de um determinado classificador divergir muito das saídas dos outros classificadores a média seria extremamente afetada, podendo gerar erros grosseiros. Para obter-se uma estimaco mais robusta da tendncia central da probabilidade, KITTLER et. al (1998) estabeleceu e sugeriu a Regra da Mediana. Em testes realizados por DUIN e TAX

(2000), foi verificado que esta técnica é boa quando os classificadores individuais estimam a mesma quantidade, pois ela é capaz de corrigir erros de estimação e suavizar a classificação. Esta vantagem foi observada nos experimentos deste, em que a Regra da Mediana obteve bons resultados.

A Votação por Maioria foi bastante estudada nos últimos anos por ser uma das primeiras técnicas a serem desenvolvidas para o problema de combinação de classificadores, provavelmente por resolver o problema de uma forma bem intuitiva. Em comparações realizadas na literatura, há casos em que se apresentava melhor (DUIN e TAX, 2000) ou pior (KITTLER et. al., 1998) que as demais regras de combinação. Realmente, como todas as outras técnicas, a Votação baseia-se teoricamente em uma série de suposições. Segundo DUIN e TAX (2000), assim como a Regra da Soma, a Votação assume teoricamente que os erros dos classificadores são correlacionados, ou seja, que os classificadores cometem os mesmos tipos de erros. Em KITTLER e ALKOOT (2003) verifica-se que esta teoria supõe erros identicamente distribuídos. Estes autores ainda demonstram que o desempenho desta técnica só é maximizado quando, além de satisfeitas as suposições, a diferença entre as probabilidades *a posteriori* é pequena, no caso de duas classes. Estas suposições são teoricamente muito diferentes das suposições da Regra do Produto que obteve os melhores resultados. Apesar de todas as adversidades deste método para o escopo deste trabalho, comparando os resultados com os classificadores individuais, ainda assim temos um pequeno ganho na performance, mas, assim como Regra do Mínimo, não foi um ganho estatisticamente significativo.

Finalmente, a Votação dos Combinadores, é ainda uma técnica sem um embasamento teórico apropriado. Não foi encontrado nenhum trabalho que utilizasse mais de um estágio de combinação, apenas casos em que se utilizam algoritmos iterativos de combinação, que convergem em uma determinada situação. Este procedimento foi proposto e realizado neste trabalho pela disponibilidade de dados dos combinadores. Os resultados obtidos foram muito interessantes, muito próximos dos obtidos com a Regra do Produto. Obteve-se a melhor média para o coeficiente Kappa, mas a diferença não foi significativa para diferenciá-la dos demais métodos de combinação. Uma característica interessante observada foi a constância dos bons resultados, mas não houve uma melhora muito grande na performance. É interessante notar que a boa performance da votação dos combinadores caracteriza a boa performance geral dos outros combinadores, pois quer dizer que a maioria dos combinadores classificou corretamente os objetos. Um estudo mais apurado, talvez com dados simulados, de forma a

ter-se conhecimento *a priori* das distribuições, poderia revelar o comportamento e as características desta técnica, revelando então a situação adequada para o uso de métodos de segundo estágio de combinação.

8. Conclusões

Após experimentos, extensivos estudos e análises acerca dos resultados, apresentaremos as conclusões, focando os métodos de reconhecimento de padrões, especialmente na combinação de classificadores.

Aplicações de reconhecimento de padrões baseadas no uso de múltiplas energias em tomografia computadorizada apresentam-se viáveis e abrem-se possibilidades para a identificação de materiais em diversas áreas. Utilizando-se como atributos os coeficientes de atenuação linear e realizando a escolha das energias segundo o material que se objetiva identificar, obtém-se boa discriminação entre as classes. Alguma dificuldade de reconhecimento devido ao ruído ou outros problemas, pode ser superada pelo uso de múltiplas energias.

A estimação do erro por validação cruzada *leave-one-out* revelou-se mais precisa e adequada para casos de poucas amostras disponíveis, obtendo uma estimação aproximadamente sem viés, porém com variância um pouco maior do que a estimada pelo coeficiente Kappa, e alto custo computacional.

O uso de combinadores em problemas de reconhecimento de padrões pode ser muito interessante, desde que se tenham disponíveis diferentes conjuntos de atributos ou diferentes classificadores treinados sobre um mesmo conjunto de atributos. A estratégia de “dividir para conquistar” funciona no caso dos classificadores que, mesmo possuindo baixo desempenho, fornece informação relevante para melhores resultados dos combinadores.

A Regra do Produto possui o menor número de suposições acerca dos dados e dos erros cometidos pelos classificadores o que a torna vantajosa em casos em que os classificadores disponíveis são bem diferentes.

A Regra da Mediana apresenta-se a melhor técnica que assume erros correlacionados de classificação entre as probabilidades *a posteriori* combinadas.

O uso da Votação dos Combinadores, em um segundo estágio de combinação, obteve uma das melhores médias de performance. Esta técnica é vantajosa quando não é possível

definir um combinador ótimo para determinada aplicação, aproveitando-se os resultados dos diferentes combinadores.

A Votação dos Combinadores, como técnica de segundo estágio de combinação, apresentou resultados interessantes e obteve um dos melhores resultados. Não é possível, porém, definir o seu comportamento pois não há estudos teóricos ou experimentais na literatura. Além disso, foram realizados aqui apenas experimentos com dados específicos e utilizando todos os combinadores. Experimentos com dados controlados (simulados) e classificadores bem estabelecidos teoricamente podem ser realizados para ter uma idéia melhor do potencial desta técnica.

A Regra do Produto e a Votação dos Combinadores obtiveram os melhores resultados. Este resultado está ligado à característica dos métodos de combinação e das probabilidades *a posteriori* fornecidas por cada classificador combinado. Não há como definir uma melhor regra de combinação de classificadores para todos os casos, pois cada qual é desenvolvida de forma a atuar em situações diferentes.

A combinação de classificadores com características bem diferentes, treinados com conjuntos de atributos iguais, apresentou-se muito útil. Neste trabalho utilizamos imagens de ciência do solo, mas outras áreas podem ser exploradas visando a identificação de materiais. Para qualquer aplicação difícil, as regras de combinação geralmente melhoram a performance e certamente proporcionam maior capacidade de generalização, que é uma das características mais desejáveis em um sistema de reconhecimento de padrões. No entanto, como observado por COVER (1969), deve-se ter cuidado ao dizer que um sistema de classificação é melhor do que outro, pois fatores como alguns bons resultados experimentais e embasamento teórico não implicam em uma comparação “melhor/pior” definitiva. Segundo LAM e SUEN (1997), a combinação de classificadores não substitui a construção de um classificador especializado para uma determinada aplicação. Mais do que a aplicação direta de combinadores para estudos experimentais, outras áreas vem sendo estudadas e possuem vasto campo de estudos, como as transformações de espaços de atributos, pré-seleção de classificadores e outras. O desenvolvimento de técnicas em reconhecimento de padrões continua a desafiar os pesquisadores, e isto não é diferente com a combinação de múltiplos classificadores.

8.1 TRABALHOS FUTUROS

Diversos trabalhos podem ser desenvolvidos como extensão deste, na área de combinação de múltiplos classificadores:

- Seleção de classificadores para combinação - uma etapa de pré-seleção de classificadores - de forma a utilizar apenas classificadores relevantes para o processo de combinação;
- Transformações do espaço de atributos dos classificadores a serem combinados para produzir diferenças maiores entre as classificações individuais, resultando em melhor performance na combinação;
- Outros estágios de combinação de classificadores para melhoria da classificação, como por exemplo, a votação dos combinadores testada neste trabalho;
- Utilização de redes neurais para combinar resultados de classificação;
- Combinação de diferentes redes neurais ou combinação de uma mesma rede utilizando diferentes inicializações ou kernel diferente em vários treinamentos;
- Estudo de teorias de incerteza para resolver problemas de combinação de classificadores.

Referências Bibliográficas

ANDERSON, J. A., Logistic Discrimination. In: KRISHNAIAH, P. R.; KANAL, L.N (Eds). **Classification, Pattern Recognition and Reduction of Dimensionality**: v. 2 of Handbook of Statistics. Amsterdam: North-Holland, p. 169-191, 1982.

BISHOP, C. M., **Neural Networks for Pattern Recognition**, New York: Oxford, 1995.

BREIMAN, L. Bagging Predictors, **Machine Learning**, v. 24, n. 2, p. 123-140, 1996.

COVER, T. Learning in Pattern Recognition. In: WATANABE, S (Ed). **Methodologies of Pattern Recognition**. New York: Academic Press, 1969.

CONGALTON, R. G., A Review of Assessing the Accuracy of Classifications of Remotely Sensed Data. **Remote Sensing Environment**, v. 37, p. 35-46, 1991.

CONGALTON, R. G.; GREEN, K. **Assessing the Accuracy of Remotely Sensed Data: principles and practices**. Boca Raton: CRC/Lewis Press, 1999.

CRUVINEL, P. E.; CESAREO, R.; MASCARENHAS, S. X and γ - Rays Computerized Minitomograph Scanner for Soil Science, **IEEE Transactions on Instrumentation and Measurements**. v. 39, n. 5, p. 745-750, 1990.

DEVJVER, P. A.; KITTLER, J. **Pattern Recognition: a statistical approach**. New York: Prentice-Hall, 1982.

DUDA, R. O.; HART, P. E. **Pattern Classification and Scene Analysis**, New York: Wiley, 1973.

DUDA, R. O.; HART, P. E., STORK, D. G. **Pattern Classification**, 2.ed. New York: Wiley, 2000.

DUIN, R. P. W. **PRTools 3.2: a matlab toolbox for pattern recognition (manual)**. Delft University of Technology: Delft, 39 p., 2003.

DUIN, R. P. W.; TAX, D. M. J. Experiments With Classifier Combining Rules. In: KITTLER, J.; ROLI, F., (Eds). **Multiple Classifier Systems 2000**: v. 1897 of Lecture Notes in Computer Science. Berlin Heidelberg: Springer-Verlag, pp. 16-29, 2000.

FUKUNAGA, K. **Introduction to Statistical Pattern Recognition**, New York: Academic Press, 1990.

FUKUNAGA, K.; HAYES, R. R. The Reduced Parzen Classifier, **IEEE Transactions on Pattern Analysis and Machine Intelligence**, v. 11, p. 423-425, 1989.

- FUKUNAGA, K; HUMMELS, D. M. Leave-one-out Procedures for Nonparametric Error Estimates. **IEEE Transactions on Pattern Analysis and Machine Intelligence**. v. 11, n. 4, p. 421-423, 1989.
- HASTIE, T.; TIBSHIRAMI, R.; FRIEDMAN, J., **The Elements of Statistical Learning: data mining, inference and prediction**, New York: Springer Verlag, 2001.
- HO, T. K.; HULL, J. J.; SRIHARI, S. N. Decision Combination in Multiple Classifier Systems. **IEEE Transactions on Pattern Analysis and Machine Intelligence**, v. 16, n. 1, p. 66-75, 1994.
- HOMEM, M. R. P. **Técnicas de Reconhecimento de Padrões Aplicadas à Imagens Multiespectrais Obtidas com Múltiplas Energias**. Tese de Dissertação: Universidade Federal de São Carlos. EDUFSCar: São Carlos, 1998.
- HOMEM, M. R. P.; MASCARENHAS, N. D. A.; CRUVINEL, P. E. The Linear Attenuation Coefficients as Features of Multiple Energy CT Image Classification, **Nuclear Instruments and Methods in Physics Research**. v. 452, p. 351-360, 2000.
- HOUNSFIELD, G. N. Computerized Transverse Axial Scanning (tomography). Part I: description of system. Part II: clinical applications. **British Journal of Radiology**. v. 46, p. 1016-1022, 1973.
- JAIN, A. K.; DUIN, R. P. W.; MAO, J. Statistical Pattern Recognition: a review, **IEEE Transactions on Pattern Analysis and Machine Intelligence**. v. 22, n. 1, p. 4-37, 2000.
- KAK, A. C.; SLANEY, M. **Principles of Computerized Tomographic Imaging**. New York: Society of Industrial and Applied Mathematics, 1988.
- KANDEL, E. R.; SCHWARTZ, J. H. **Principles of Neural Science**. New York: Elsevier, 1987.
- KUNCHEVA, L. I. **Combining Pattern Classifiers: methods and algorithms**. New York: Wiley-Interscience, 2004.
- KITTLER, J.; ALKOOT, F. M. Sum Versus Vote in Multiple Classifier Systems. **IEEE Transactions on Pattern Analysis and Machine Intelligence**. v. 25, n. 1, p. 110-115, 2003.
- KITTLER, J.; HATEF, M.; DUIN, R. P. W.; MATAS, J. On Combining Classifiers, **IEEE Transactions on Pattern Analysis and Machine Intelligence**. v. 20, n. 3, p. 226-239, 1998.
- LAM, L., SUEN, C. Y. Application of Majority Voting to Pattern Recognition: an analysis of its behavior and performance. **IEEE Transactions on Systems, Man and Cybernetics**, v. 27, n. 5, p. 553-568, 1997.
- LIDE, R. D. **CRC Handbook of Chemistry and Physics: a ready-reference book of chemical and physical data**, 83.ed. CRC, 2002.
- MACOVSKI, A. **Medical Imaging Systems**. New York: Prentice-Hall, 1983.

- NADAL, C.; LEGAUT, R.; SUEN, C. Y. Complementary Algorithms for the Recognition of Totally Unconstrained Handwritten Numerals. **10th International Conference on Pattern Recognition**, Atlantic City - New Jersey, p. 443-449, 1990.
- PARKER, J. A. **Image Reconstruction in Radiology**, New York: CRC Press, 1990.
- PATRICK, E. A.; FISCHER, F. P., III. A Generalized k-Nearest Neighbor Rule. In: DASARATHY, B. V. (Ed). **Nearest Neighbour (NN) Norms: nn pattern classification techniques**. Los Alamitos: IEEE Computer Society, p. 64-83, 1970.
- ROUX, C; UDUPA, J. K. Scanning the Issue: Special Issue on Emerging Medical Imaging Technology. **Proceedings of IEEE**. v. 91, n. 10, p. 1479-1482, 2003.
- SCHAPIRE, R. E. The Strenght of Weak Learnability. **Machine Learning**. v. 5, p. 197-227, 1990.
- SHUMWAY-COOK, A.; WOOLLACOTT, M. H. **Controle Motor: teoria e aplicações práticas**. 2.ed. Manole: Barueri, 2003.
- STONE, M. Cross-Validatory Choice and Assessment of Statistical Predictions. **Journal of the Royal Statistical Society**. v. 36, p. 111-147, 1974.
- WEBB, A. **Statistical Pattern Recognition**. London: Arnold, 1999.
- WOLPERT, D. Stacked Generalization. **Neural Networks**. v. 5, p. 241-259, 1992.
- XU, L.; KRZYZAK, A.; SUEN, C.Y. Methods of Combining Multiple Classifier and Their Applications to Handwriting Recognition. **IEEE Transactions on Systems, Man and Cybernetics**. v. 22, n. 3, p. 418-435, 1992.