

Universidade Estadual de Campinas

Faculdade de Engenharia Elétrica e de Computação

Departamento de Computação e Automação

**Identificação de Pessoas por Reconhecimento de Íris Utilizando
Decomposição em Sub-bandas e uma Rede Neuro-Fuzzy**

autor: Denilson Palhares Ferreira

Orientador: Márcio Luiz de Andrade Netto

Co-Orientador: Max Henrique Machado Costa

Dissertação submetida à Faculdade de Engenharia Elétrica e de Computação da Universidade Estadual de Campinas, para preenchimento dos pré-requisitos parciais para a obtenção do Título de Mestre em Engenharia Elétrica

Este exemplar corresponde a redação final da tese defendida por Denilson Palhares Ferreira e aprovada pela Comissão julgada em 04/12/1998.

Márcio Luiz de Andrade Netto
Orientador

dezembro 1998

9915 056



UNIDADE	BC
N.º DE FOLHAS	
UNICAMP	
FAIZI	
V.	Ex.
TOMBO BC	38432
PROC.	229199
C	<input type="checkbox"/>
D	<input checked="" type="checkbox"/>
PREÇO	R\$ 11,00
DATA	24/08/99
N.º CPD	

CM-00134290-6

FICHA CATALOGRÁFICA ELABORADA PELA
BIBLIOTECA DA ÁREA DE ENGENHARIA - BAE - UNICAMP

F413i

Ferreira, Denilson Palhares

Identificação de pessoas por reconhecimento de íris utilizando decomposição em sub-bandas e uma rede neuro-fuzzy . / Denilson Palhares Ferreira.--Campinas, SP: [s.n.], 1998.

Orientadores: Márcio Luiz de Andrade Netto, Max Henrique Machado Costa

Dissertação (mestrado) - Universidade Estadual de Campinas, Faculdade de Engenharia Elétrica e de Computação.

1. Reconhecimento de padrões. 2. Redes neurais (Computação). 3. Biometria. 4. Processamento de imagens. I. Netto, Márcio Luiz de Andrade. II. Costa, Max Henrique Machado. III. Universidade Estadual de Campinas. Faculdade de Engenharia Elétrica e de Computação. IV. Título.

À minha esposa

Este trabalho teve suporte financeiro do Conselho
Nacional de Desenvolvimento Científico e
Tecnológico - **CNPq**

Agradecimentos

A Deus a quem tudo devo.

Ao CNPq que propiciou todas as condições para que este trabalho pudesse ser concluído

Ao Prof. Márcio Luiz Andrade Netto pela paciência, estímulo e orientação.

Ao Prof. Max Henrique Machado Costa pelo apoio e pelo exemplo.

À Dra. Cláudia Assis Lima pela cooperação.

Aos colegas do LCA

Aos amigos Geraldo Peres Caixeta, Paulo César Caixeta pela convivência e amizade.

Aos meus pais e meus irmãos pelo apoio aos meus projetos.

À minha esposa pela paciência e compreensão.

Resumo

Apresenta-se, neste trabalho, um sistema de identificação de pessoas baseado em seus padrões de íris. Estruturalmente este sistema pode ser dividido em dois módulos: o módulo de extração de características, no qual os padrões de íris são decompostos em sub-bandas e o módulo de identificação, em os padrões são identificados por uma rede “neuro-fuzzy”.

No módulo responsável pela extração de características, um determinado padrão de íris é decomposto em 67 sub-bandas. Inicialmente, os filtros em quadratura bidimensionais (2D QMF) são aplicados sobre toda a imagem gerando quatro sub-bandas. Em seguida, escolhe-se a sub-banda de baixa frequência nas duas direções (LL) e aplicam-se os filtros 2D QMF recursivamente sobre todas as sub-bandas até a obtenção de sub-bandas com (16x16) coeficientes. Neste ponto, a energia média de cada sub-banda é calculada e as sub-bandas são ordenadas em ordem decrescente de energia. A seguir, determinam-se quais sub-bandas serão utilizadas. Para tanto, aplica-se o procedimento descrito acima sobre um conjunto de dez padrões de íris diferentes e escolhem-se as 32 sub-bandas mais energéticas de cada padrão. Destas 320 sub-bandas, escolhem-se as 20 sub-bandas que apareceram com maior frequência.

Estas 20 sub-bandas de cada padrão são utilizadas para treinar a rede “neuro-fuzzy”. Após treinada, a rede “neuro-fuzzy” é utilizada pelo módulo responsável pela identificação. O desempenho do sistema de identificação baseia-se nas taxas de aceitação de autênticos e de rejeição de impostores. Este trabalho apresenta resultados de simulações do sistema em que as taxas de aceitação de autênticos foram, em média, superiores a 90% e as taxas de aceitação de impostores inferiores a 10%.

Abstract

In this work we develop a system that identifies a person by his or her iris pattern. Structurally this system can be divided in two modules: (1) the feature extraction one, performed via wavelet-based subband decomposition, and (2) the identification task module, performed by a neuro-fuzzy network.

In the feature extraction module, an iris pattern is decomposed in 67 subbands. Initially, 2D Quadrature Mirror Filters (2D QMF) are applied to the whole image and four subbands are generated. Then the low-low (LL) subband is further decomposed and the 2D QMF are applied recursively to all subbands until we obtain subbands with (16x16) coefficients. At this point, the energy of each subband is calculated and the subbands are ordered in decreasing mode. Then we determine which subbands should be applied to the identification module. To do so we use the procedure described above with a sample of ten different iris patterns. The 32 subbands with most energy are then chosen from each pattern. From this group of 320 subbands we choose the 20 subbands that occur most frequently.

These 20 subbands of each iris pattern are then used to train the neuro-fuzzy network. After the training phase, the neuro-fuzzy network is used by the identification module. This work presents results of simulations of the algorithms and estimates of the authentic recognition rate and the imposter rejection rate. The estimated authentic recognition rates were, on average, above 90% and the imposter recognition rates below 10%.

Capítulo 1 - Introdução	1
1.1 - Biométrica e sistemas biométricos	1
1.2 - Sistemas biométricos baseados na íris	3
Capítulo 2 - Reconhecimento de Padrões	9
2.1 - Visão Geral	9
2.1.1 - Estrutura de um sistema de reconhecimento de padrões	11
2.1.2 - Definições preliminares	11
2.2 - Abordagens ao reconhecimento de padrões	14
2.2.1 - Abordagem estatística	14
2.2.1.1 - Introdução	14
2.2.1.2 - Teoria da decisão de Bayes - caso contínuo	18
2.2.1.3 - Classificação por taxa de erro mínimo	20
2.2.1.4 - Classificadores, funções discriminantes e curvas de decisão	21
2.2.1.5 - Probabilidades de erro e integrais	23
2.2.1.6 - Densidade normal	25
2.2.1.6.1 - Distribuição normal monovariável	25
2.2.1.6.2 - Distribuição normal multivariável	26
2.2.1.7 - Funções discriminantes para a distribuição normal	27
2.2.1.8 - Conclusão	33
2.2.2 - Abordagem sintática	34
2.2.2.1 - Introdução	34
2.2.2.2 - Definições importantes	37
2.2.2.3 - Sistema sintático de reconhecimento de padrões	40
2.2.2.4 - Seleção de primitivas	42
2.2.2.5 - Gramática padrão	43
2.2.2.6 - Gramáticas de dimensão elevada	44
2.2.2.7 - Reconhecimento sintático	55
2.2.2.7.1 - Reconhecimento de linguagens de estado finito	55
2.2.2.7.2 - Reconhecimento de linguagens livres de contexto	58
2.2.2.8 - Conclusão	60
2.2.3 - Abordagem por redes neurais artificiais	62
2.2.3.1 - Introdução	62
2.2.3.2 - Características	63
2.2.3.3 - Arquiteturas de redes	65
2.2.3.3.1 - Redes “feedforward” de uma única camada	65
2.2.3.3.2 - Redes “feedforward” multicamadas	66
2.2.3.3.3 - Redes recorrentes	68
2.2.3.3.4 - Estruturas em “lattice”	69
2.2.3.4 - Processo de aprendizagem	69
2.2.3.4.1 - Aprendizado baseado na correção do erro de aprendizagem ...	71
2.2.3.4.2 - Aprendizado hebbiano	73
2.2.3.4.2.1 - Modelos matemáticos	74
2.2.3.4.3 - Aprendizado competitivo	77

2.2.3.4.4 - Aprendizado de Boltzmann	79
2.2.3.4.5 - Aprendizado supervisionado.....	81
2.2.3.4.6 - Aprendizado não-supervisionado.....	83
2.2.3.4.6.1 - Aprendizado supervisionado versus aprendizado auto-organizável.....	84

Capítulo 3 - Decomposição em Sub-bandas 87

3.1 - Introdução	87
3.2 - Processamento de sinais com múltiplas frequências.....	87
3.2.1 - Dizimação e interpolação.....	88
3.2.2 - Decomposição polifásica	95
3.3. - Sinais de banda passante, sinais modulados e filtros imagem	98
3.3.1 - Amostragem de banda inteira.....	99
3.3.2 - Modulação em quadratura	101
3.3.3 - Filtros imagem.....	105
3.4 - Banco de filtros de dois canais.....	105
3.4.1 - Banco de filtros PR-QMF de dois canais.....	106
3.4.2 - Sub-bandas geradas em forma de árvores binárias regulares.....	110
3.4.3 - Sub-bandas geradas em forma de árvores binárias irregulares.....	113
3.5 - Extensão para sinais bidimensionais.....	113

Capítulo 4 - Resolução do Problema 117

4.1 - Introdução	117
4.2 - Obtenção da imagem da íris.....	117
4.3 - Pré-processamento da imagem.....	118
4.4 - Extração de características	119
4.5 - Identificação e reconhecimento.....	122
4.5.1 - Introdução.....	122
4.5.2 - Rede Neural	122
4.5.2.1 - Neurônio nebuloso.....	122
4.5.2.2 - Classes de neurônios.....	123
4.5.2.2.1 - Neurônios de entrada (Ent)	123
4.5.2.2.2 - Neurônios de máximo (Max).....	124
4.5.2.2.3 - Neurônios de média (Med).....	124
4.5.2.2.4 - Neurônios competitivos (Comp).....	124
4.5.2.3 - Estrutura da rede	124
4.5.2.4 - Análise da rede	129

Capítulo 5 - Resultados 133

5.1 - Introdução	133
5.2 - Base de dados	133
5.3 - Fase de treinamento	133
5.4 - Fase de reconhecimento	134
5.5 - Análise dos resultados.....	134

Capítulo 6 - Conclusão Geral	145
6.1 - Conclusão	145
Bibliografia	147

Lista de Figuras

2.1 - Mapeamentos em uma representação abstrata de um sistema de geração/classificação/interpretação de padrões	10
2.2 - Estrutura de um sistema de reconhecimento de padrões típico	11
2.3 - Exemplos de regiões de decisão: (a) linear, (b) quadrática, (c) geral	13
2.4 - Regiões de decisão e funções discriminantes correspondentes: (a) Partição de R_2 , (b) Funções discriminantes correspondentes	14
2.5 - Classificador de padrões	22
2.6 - Componentes da probabilidade de erro	24
2.7 - Cena A e sua descrição hierárquica estrutural	35
2.8 - Losango e suas primitivas	36
2.9 - Grafo relacional da cena A	37
2.10 - Diagrama de blocos de um sistema sintático de reconhecimento de padrões	40
2.11 - Ilustração de (a) $A*B$, (b) $N(A+B+C)$, (c) ANB	49
2.12 - (a) Padrão, (b) Representação em forma de árvore	51
2.13 - Domínio de árvore universal	51
2.14 - Autômato de estado finito	56
2.15 - Representação gráfica de $d(q,a) = q'$	56
2.16 - Rede "feedforward" com uma única camada de neurônios	66
2.17 - Rede "feedforward" totalmente interconectada com uma camada de neurônios escondidos e uma camada de saída	67
2.18 - Rede "feedforward" parcialmente interconectada	68
2.19 - Rede recorrente sem auto-realimentação e sem neurônios escondidos	68
2.20 - "Lattice" unidimensional de três neurônios	69
2.21 - Grafo ilustrando um par de neurônios inserido em uma rede neural	70
2.22 - Taxonomia do processo de aprendizagem	71
2.23 - Ilustração da regra de Hebb	75
2.24 - Arquitetura de um rede de aprendizado competitivo com conexões excitatórias dos nós de entrada e conexões laterais inibitórias entre os neurônios	78
2.25 - Diagrama de blocos do aprendizado supervisionado	81
2.26 - Diagrama de blocos do aprendizado não supervisionado	83
3.1 - Representação esquemática das fases de análise e síntese para a decomposição em quatro sub-bandas	87
3.2 - Operação de dizimação: (a) filtro passa baixas e subamostrador, (b) seqüências típicas	88
3.3 - (a) Operação de superamostragem, (b) formas de onda de entrada e saída para $M=4$	91
3.4 - Compressão do eixo de freqüências devido a superamostragem para $M = 4$	92
3.5 - Sinais típicos em uma composição subamostrador-superamostrador para $M=4$	93
3.6 - Espectros dos sinais mostrados na Fig. 3.5	94
3.7 - Estruturas equivalentes	96
3.8 - (a) Filtro seguido por um subamostrador, (b) decomposição polifásica, (c) representação alternativa de uma rede polifásica	96

3.9 - Decomposição polifásica do interpolador	98
3.10 - Configuração da amostragem de banda inteira	100
3.11 - Espectro de um sinal de banda passante	102
3.12 - Modulação e demodulação em quadratura e redução de frequência	102
3.13 - Espectro de frequências dos sinais do modulador-demodulador em quadratura	104
3.14 - Sub-bandas geradas por um banco de filtros de dois canais	106
3.15 - Estrutura em árvore das fases de análise e síntese de uma decomposição em quatro sub-bandas	110
3.16 - (a) Sub-bandas geradas pela árvore binária regular em dois níveis, (b) bandas de frequência correspondentes às quatro sub-bandas com filtros ideais	112
3.17 - Estruturas equivalentes para filtros separados por subamostradores	112
3.18 - Estrutura de árvore regular para $L = 3$ e sua banda de frequência, assumindo que filtros ideais PR-QMFs sejam empregados	112
3.19 - Estrutura de árvore irregular e sua banda de frequência, assumindo que filtros PR-QMFs sejam empregados	113
3.20 - Fase de análise para um banco de filtros bidimensionais na geração de quatro sub-bandas	114
3.21 - Fase de síntese para um banco de filtros bidimensionais na geração de quatro sub-bandas	114
3.22 - Faixas de frequência das quatro sub-bandas geradas	115
4.1 - Desenho esquemático do sistema de identificação proposto	117
4.2 - Imagem em tons de cinza de uma íris esquerda	118
4.3 - Imagem da mesma íris da Fig. 4.2 obtida alguns segundos mais tarde	118
4.4 - Resultado do pré-processamento da Fig. 4.2	119
4.5 - Resultado do pré-processamento da Fig. 4.3	119
4.6 - Representação esquemática do processamento desenvolvido durante a fase de extração de características	120
4.7 - Seqüência de sub-bandas ordenadas de acordo com a sua energia média	121
4.8 - Neurônio "fuzzy"	123
4.9 - Arquitetura da rede "neuro-fuzzy"	125
4.10 - Função de "fuzzificação" ($\beta = 0.3$)	126
4.11 - Função de saída dos neurônios MAX-FN da segunda camada	127
5.1 - Imagem 1 da segunda íris do conjunto C_4	141
5.2 - Imagem 2 da segunda íris do conjunto C_4	141
5.3 - Imagem 1 da terceira íris do conjunto C_4	141
5.4 - Imagem 2 da terceira íris do conjunto C_4	141
5.5 - Imagem 1 da oitava íris do conjunto C_4	142
5.6 - Imagem 2 da oitava íris do conjunto C_4	142
5.7 - Imagem 1 da nona íris do conjunto C_4	142
5.8 - Imagem 2 da nona íris do conjunto C_4	142
5.9 - Imagem 1 da décima íris do conjunto C_4	143
5.10 - Imagem 2 da décima íris do conjunto C_4	143

Lista de Gráficos

5.1 - Curva de decisão Neyman-Person para o conjunto C_1	135
5.2 - Curva de decisão Neyman-Person para o conjunto C_1 com $T_f = 0.1$ (curva mais escura) e $T_f = 0.15$ (curva mais clara)	135
5.3 - Taxa de reconhecimento de autênticos e taxa de reconhecimento de impostores.....	137
5.4 - Curva de decisão Neyman-Person para o conjunto C_3	138
5.5 - Curva de decisão Neyman-Person para o conjunto C_3 com $T_f = 0.1$ (curva mais escura) e $T_f = 0.19$ (curva mais clara)	138
5.6 - Curva de decisão Neyman-Person para o conjunto C_4	139
5.7 - Curva de decisão Neyman-Person para o conjunto C_4 com $T_f = 0.1$ (curva mais escura) e $T_f = 0.16$ (curva mais clara)	139
5.8 - Curva de decisão Neyman-Person para o conjunto C_5	140
5.9 - Curva de decisão Neyman-Person para o conjunto C_5 com $T_f = 0.1$ (curva mais escura) e $T_f = 0.19$ (curva mais clara)	140

Capítulo 1 - Introdução

1.1 - Biométrica e sistemas biométricos

A palavra biométrica, à primeira vista, pode parecer estranha e nova e talvez muitos a associem a alguma tecnologia emergente e complexa. Porém, o conceito que ela encerra é bem simples e muito antigo - ela trata do reconhecimento de pessoas. Tecnicamente falando, a biométrica trata da mensuração automática de certas características físicas ou comportamentais de um determinado indivíduo e da comparação destas com a de outros com objetivo de identificá-lo ou reconhecê-lo.

A biométrica utiliza-se de características físicas, definidas como algo que possuímos, e de características comportamentais, algo que fazemos. Entre elas destacam-se:

1 - Características físicas

- . Composição química do odor corpóreo;
- . Características faciais;
- . Emissões de calor;
- . Características do olho - retina e íris;
- . Impressões digitais;
- . Geometria da mão;
- . Veias da mão e do punho.

2 - Características comportamentais

- . Assinatura;
- . Digitação;
- . Voz.

Destas características, apenas três características físicas (a retina, a íris e a impressão digital) e todas as características comportamentais, usadas atualmente nos sistemas biométricos disponíveis, podem ser consideradas realmente únicas.

Os sistemas biométricos atuais podem ser classificados, de acordo com sua funcionalidade, em dois grupos principais: sistemas de verificação e sistemas de reconhecimento. Nos sistemas de verificação, a pessoa que está sendo identificada afirma possuir determinada característica. Logo, o sistema deverá apenas aceitar ou rejeitar tal afirmação e, para tanto, ele deverá realizar uma comparação do tipo “um para muitos”. Já nos sistemas de reconhecimento, a identificação da pessoa é feita a partir de uma busca em uma base de dados cadastrada previamente. O sistema percorre esta base até encontrar um indivíduo que apresente um conjunto de características semelhantes àquelas apresentadas realizando, neste caso, uma comparação do tipo “um para um”.

Os sistemas biométricos podem ser usados em qualquer situação que requeira uma resposta rápida e correta para a questão “Quem é você?”. Uma das grandes vantagens dos sistemas biométricos sobre os sistemas de identificação tradicional baseados em senhas, cartões de acesso, R.G., etc. é que, nestes sistemas, o reconhecimento baseia-se em aspectos intrínsecos do ser humano. Os sistemas de reconhecimento que não são baseados nestes aspectos intrínsecos nem sempre são seguros. Por exemplo, chaves e cartões de acesso entre

outros podem ser perdidos, duplicados ou roubados. Senhas, códigos secretos e números de identificação pessoal podem ser facilmente esquecidos, compartilhados ou observados. Os sistemas biométricos, por outro lado, não são susceptíveis a estes problemas.

1.2 - Sistemas biométricos baseados na íris

A palavra íris provém do grego *ιρις* e, inicialmente, era usada para designar o arco-íris. No século XVI passou a ser aplicada também à porção externa e colorida do olho. Tecnicamente, a íris faz parte do *tracto uveal* sendo um fino diafragma que se estica sobre a porção anterior do olho suportada pelo cristalino. Por esta razão, ela adquire um aspecto de um cone tridimensional truncado. Em sua base, a íris esta conectada ao corpo ciliar; no lado oposto, ela termina na pupila, localizada um pouco abaixo da região central do olho. Sobre a íris, encontra-se a córnea, uma camada transparente, que lhe fornece proteção.

Para que se possa apreciar a riqueza da íris como um padrão de reconhecimento, faz-se necessário que sua estrutura seja analisada mais detalhadamente. Sua estrutura microscópica apresenta vários aspectos incomuns. Sua superfície anterior que forma o limite posterior da câmara anterior não é revestida por um epitélio distinto apesar de freqüentes afirmações em contrário; esta superfície é simplesmente uma “camada limitante anterior” modificada do *estroma* geral, que forma a massa da íris. O estroma contém os vasos e nervos da região e, perto da periferia da pupila, um conjunto de fibras musculares lisas forma uma estrutura contráctil anular, o *esfincter da pupila*. A face posterior da íris consiste de um prolongamento das mesmas duas camadas de epitélio que revestem o corpo ciliar. A mais profunda e por isso, na íris, a mais anterior destas camadas epiteliais é comumente, um pouco confusamente, denominada *epitélio anterior* da íris. Suas células são pigmentadas

como aquelas do epitélio ciliar; estreitamente associadas com elas estão as fibras lisas do dilatador da pupila, dispostas radialmente. Superficial e posteriormente a esta camada de células está o estrato de células densamente pigmentadas que formam o chamado *epitélio posterior*. Esta camada é contínua com a camada retínica *não pigmentada* do epitélio ciliar.

A camada limitante anterior ou superfície anterior da íris tem sido muito estudada com pequeno aumento por microscopia com lâmpada de fenda, com a qual ela mostra um aspecto um tanto felpudo exceto em íris muito pigmentadas. Depressões ou criptas, através das quais os vasos podem ser vistos no estroma, e várias pregas e estrias radiais e circulares podem ser observadas, porém detalhes destas formações podem ser procurados em autoridade adequada [10]. Os constituintes da lâmina limitante anterior são principalmente fibroblastos muito ramificados e melanócitos, sem vestígio do endotélio que a recobria ao nascimento e em extensão rapidamente decrescente durante o primeiro ano pós-natal [11]. Isto é confirmado por microscopia eletrônica [12]. Os fibroblastos formam na superfície uma camada única e contínua, com ramificações que não constituem junções verdadeiras [13]. Na periferia da íris, eles se misturam com o tecido conjuntivo da rede trabecular (ligamento pectinado) do ângulo iridocorneal. Na borda pupilar, eles entram em contato, mas ainda sem junções especializadas, com o epitélio pigmentado da face posterior da íris. Os melanócitos apresentam também intrincados processos ramificados e também não se observam junções especiais entre eles.

O *estroma da íris*, que é derivado, como a camada limitante anterior, do mesoderma entre a lente em desenvolvimento e o cálice óptico, é também formado por fibroblastos e melanócitos; mas nesta região há, além disso, uma considerável quantidade de tecido

colágeno frouxo, cujos espaços são preenchidos com fluido e por uma substância fundamental mucopolissacarídea. O estroma mesodérmico contém também não somente uma abundância de vasos sanguíneos e de nervos, mas também os músculos ectodérmicos, esfíncter e dilatador. As fibras colágenas que têm um diâmetro de cerca de 60 nm e uma periodicidade de 50 a 60 nm são arranjadas muito frouxamente, muitas descrevendo alças circunferenciais incompletas ao redor da pupila tomada como centro.

Afirmações de que a estrutura da íris é única para um indivíduo e que permanece estável com o passar dos anos provém de duas fontes principais. A primeira fonte de evidência é a observação clínica. Durante o período que examinaram grandes quantidades de olhos, os oftalmologistas e anatomistas notaram que o padrão detalhado de uma íris, mesmo a íris esquerda e direita de um determinado indivíduo, é único. Mais ainda, em repetidas observações, constatou-se que o padrão de uma determinada íris variava muito pouco. A segunda fonte de evidências é a biologia evolucionista. Nela, descobriu-se que, enquanto a estrutura geral da íris é determinada geneticamente, as particularidades do seu “*minutiae*” dependem basicamente de causas circunstanciais (e.g., as condições iniciais no precursor embrionário da íris). Portanto, é muito improvável que estas circunstâncias possam ser replicadas durante o curso natural destes eventos. Raramente o processo evolutivo transcorre de forma inadequada, dando origem a uma íris rudimentar (“*aniridia*”) ou a uma distorção na forma da pupila (“*colobloma*”). Evidências evolucionistas também suportam a estabilidade do padrão da íris no tempo. Determinadas partes da íris (e.g., “*vasculature*”) já se encontram desenvolvidas no nascimento, enquanto outras (e.g., *musculatura*) desenvolvem-se durante os primeiros dois anos de vida. De particular importância para o

reconhecimento, é o fato de que a pigmentação da íris continua até a adolescência. Assim como, o tamanho médio da pupila sofre pequenos acréscimos até esta fase. Após a adolescência, uma íris saudável varia muito pouco pelo resto da vida de um indivíduo, embora uma pequena despigmentação e redução do tamanho médio da abertura da pupila ocorram na velhice. Várias doenças de olhos podem alterar drasticamente a aparência da íris. Verificou-se, também, que a exposição intensiva a certos tipos de metais pode alterar a pigmentação da íris. Porém, estes casos são raros. Afirmações de que a íris sofre alterações de acordo com o estado de saúde do indivíduo (iridologia) tornaram-se desacreditadas [1, 9]. No geral, estas linhas de evidência sugerem que a íris é única e, após a infância, tipicamente estável. Porém, deve-se ressaltar que estudos mais aprofundados que tratem especificamente da unicidade e estabilidade da íris devem ser levados a cabo.

Um outro aspecto interessante da íris, sob o ponto de vista biométrico, está relacionado com a sua dinâmica. Devido a complexa interação dos músculos da íris, o diâmetro da pupila está em constante estado de oscilação, ainda que pequeno. Potencialmente, este movimento pode ser monitorado para garantir que um espécime vivo está sendo verificado. Mais ainda, como a íris reage rapidamente a luz, o monitoramento desta reação, quando ela estiver submetida a uma fonte de luz controlada, fornece evidência semelhante. Em contraste, em estado mórbido, a íris contrai-se e paralisa-se, o que pode servir de subsídios para a sua utilização na ciência forense.

Aparentemente, foram os parisienses os primeiros a utilizar o reconhecimento da íris como base para a identificação de pessoas. Eles distinguiam os condenados do seu sistema penal inspecionando visualmente suas íris, tendo como base a sua cor [2]. Mais

recentemente, o conceito de reconhecimento automático da íris foi proposto por Flom e Safir [4], porém este grupo não chegou a desenvolver e testar um sistema. Trabalhos mais recentes com o objetivo de construir um sistema de reconhecimento automático baseado na íris foram conduzidos em Los Alamos National Laboratories, CA [5]. Em seguida, dois grupos de pesquisa desenvolveram e documentaram um protótipo de um sistema biométrico utilizando a íris como característica básica [3, 8]. Outras pesquisas estão sendo desenvolvidas na América do Norte [7] e na Europa [6].

O trabalho aqui apresentado refere-se também a um sistema biométrico que utiliza a íris como característica básica. Ele é composto de duas partes principais: (1) a extração de características, realizada através de uma decomposição em sub-bandas e (2) a identificação do indivíduo, obtida a partir de uma rede neuro-fuzzy. Nos capítulos seguintes, apresentamos a base teórica necessária ao desenvolvimento deste sistema assim como os detalhes do mesmo. No capítulo 2, é apresentado o problema de reconhecimento de padrões e as abordagens atualmente aplicadas ao mesmo: (1) abordagem estatística, (2) abordagem sintática e (3) abordagem por redes neurais artificiais. No capítulo 3, é apresentada a técnica de decomposição em sub-bandas. No capítulo 4, apresenta-se o sistema biométrico desenvolvido e no capítulo 5, a bateria de testes aplicada ao mesmo assim como os resultados obtidos.



Capítulo 2 - Reconhecimento de Padrões

2.1 - Visão Geral

O problema de reconhecimento de padrões geralmente denota classificação ou descrição de um conjunto de processos ou eventos. Este conjunto pode ser de objetos físicos ou abstratos. Os processos com algumas propriedades similares são agrupados em uma classe. O número total de classes em um problema específico é geralmente determinado pela aplicação particular considerada. Em alguns problemas, o número exato de classes não é previamente conhecido sendo determinado a partir de observações dos vários padrões representativos. Neste caso, tem-se a possibilidade de detectar novas classes à medida que os padrões são apresentados.

Uma abordagem intuitiva do reconhecimento de padrões é dada pelo “casamento de padrões”. Neste caso, um conjunto de padrões ou protótipos, um para cada classe, é armazenado. O padrão de entrada (de classificação desconhecida) é comparado com o protótipo de cada classe e a classificação é feita segundo um critério de similaridade. Em outras palavras, se o padrão de entrada “casa-se” com o protótipo da i -ésima classe melhor do que com o protótipo de qualquer outra classe, então este padrão será classificado como um elemento desta classe. A desvantagem desta abordagem deve-se à dificuldade de se escolher um bom protótipo para cada classe como também de definir um critério próprio de “casamento”. Esta dificuldade é realçada quando grandes variações e distorções são esperadas nos padrões de uma classe.

Uma visão abstrata do problema de reconhecimento de padrões é dada pela Fig. 2.1

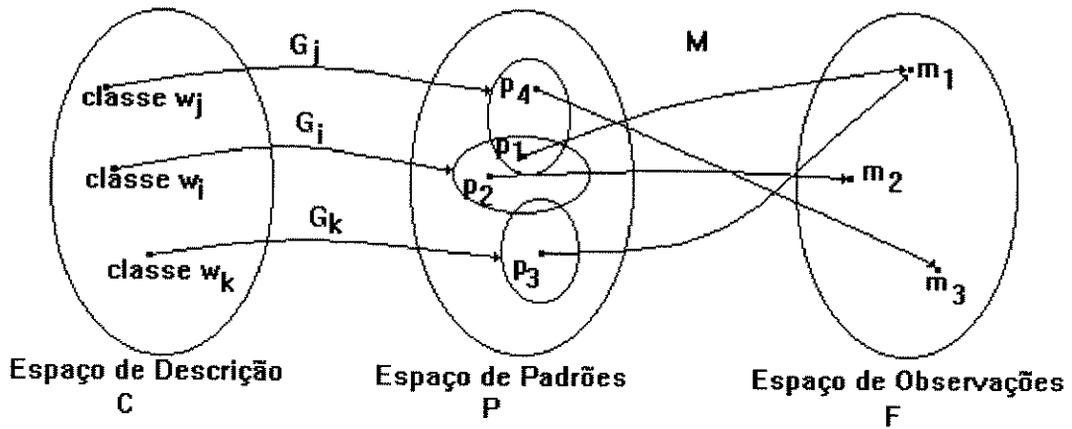


Fig. 2.1 - Mapeamentos em uma representação abstrata de um sistema de geração/classificação/interpretação de padrões.

Nesta figura, é apresentado um mapeamento entre o espaço de classes, C, e o espaço de padrões, P, representado pela relação G_i para cada classe. Cada classe w_i gera um subconjunto de padrões no espaço de padrões. Estes subconjuntos, como pode ser visto, podem se sobrepor permitindo que padrões de classes diferentes dividam os mesmos atributos. Uma outra relação, M, mapeia os padrões dos subespaços de P no espaço de características, F. Através desta visão, a caracterização de muitos problemas de reconhecimento de padrões resume-se em encontrar um meio de *identificar* e *inverter* os mapeamentos M e G_i para todo i dadas as características m_i . Infelizmente, na prática, estes mapeamentos não são funções e, mesmo que fossem, eles raramente são biunívocos, bijetores ou inversíveis. Na Fig. 2.1 observa-se que características idênticas podem resultar de diferentes p_i , que por sua vez correspondem a diferentes classes. Isso sugere um problema potencial de *ambigüidade*. Mas, é razoável modelar e tentar compreender estes processos na esperança de se obter melhores técnicas de classificação e descrição.

Pode-se considerar os conjuntos de padrões na Fig. 2.1 como dados básicos do ambiente que se está medindo. Logo, um outro aspecto importante desta figura é o mapeamento M. Este mapeamento reflete a escolha do sistema de medida. O projeto do

sistema de medida é um aspecto importante no projeto de um sistema de reconhecimento no sentido de que *características* ótimas provavelmente requerem medidas ótimas ou pelo menos adequadas. Finalmente, deve-se notar que padrões gerados pela mesma classe (p_4 e p_1 , de ω_i , por exemplo) e que estão próximos no espaço de padrões não necessariamente fornecerão medidas que também estejam próximas (neste caso, m_1 e m_3). Isto é um aspecto relevante quando a “clusterização” das medidas é utilizada como indicador da similaridade entre os padrões.

2.1.1- Estrutura de um sistema de reconhecimento de padrões

A estrutura de um sistema de reconhecimento de padrões típico é mostrada na Fig. 2.2.

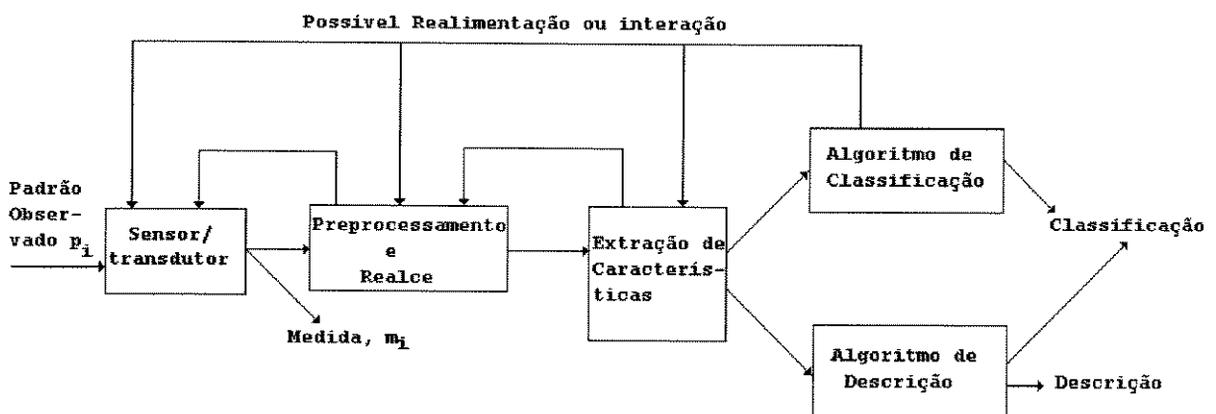


Fig. 2.2 - Estrutura de um sistema de reconhecimento de padrões típico.

O sistema é composto de um sensor (uma câmera, por exemplo), um mecanismo de extração de características e um algoritmo de classificação ou descrição (dependendo da abordagem). Usualmente, alguns dados que já foram classificados são utilizados para o treinamento do sistema (conjunto de treinamento).

2.1.2- Definições Preliminares

Classificação é o mapeamento que associa o dado de entrada à uma ou mais classes das c classes pré-especificadas tendo como base a extração de características significativas e o

processamento ou análise destes atributos. É comum recorrer-se a modelos probabilísticos ou gramaticais na classificação.

Reconhecimento é a habilidade de classificar. Geralmente, os problemas de reconhecimento de padrões são formulados com a $(c + 1)$ -ésima classe, que corresponde à “não-classificável” ou “não conheço” ou “não posso decidir”.

Descrição é uma alternativa à classificação onde uma descrição estrutural do padrão de entrada é desejável. É comum recorrer-se a modelos lingüísticos ou estruturais na descrição.

Classe de padrões é um conjunto de padrões conhecidos por originarem-se da mesma fonte em C . O problema principal, em muitas aplicações de reconhecimento de padrões, é identificar as características ideais, criar uma boa medida de similaridade e associá-las segundo um processo de “casamento”.

Pré-processamento é a transformação dos dados de entradas com o intuito de facilitar a extração de características e minimizar o ruído que porventura esteja presente.

Ruído é um conceito que, no reconhecimento de padrões, está associado a vários tipos de erro, incluindo:

- . Distorções ou erros nos sinais de entrada;
- . Erros de pré-processamento;
- . Erros de extração de características;
- . Erros nos dados de treinamento.

-Classificadores, Regiões de Decisão e Fronteiras, Funções Discriminantes

Um *classificador* particiona o espaço de características em classes denominadas *regiões de decisão*. Para que estas regiões possam ser utilizadas em uma atribuição de classes possível e única, é necessário que elas cubram totalmente o espaço \mathbf{R}^d e que sejam disjuntas

(não se sobreponham). A fronteira de cada região é uma *fronteira de decisão*. A partir desta visão, a classificação de um vetor de características \underline{x} torna-se bem simples:

Determina-se a região de decisão (em \mathbb{R}^d) que contém \underline{x} e associa-se \underline{x} à classe correspondente.

Embora esta estratégia de classificação seja bem simples, a determinação das regiões de decisão pode ser um desafio.

Os classificadores são então desenvolvidos de acordo com as funções que discriminam as classes presentes no espaço \mathbb{R}^d ; estas funções recebem a denominação de *funções discriminantes*. No caso de c classes, as funções discriminantes, denotadas por $g_i(\underline{x})$ com $i = 1, \dots, c$, são usadas para particionar \mathbb{R}^d como se segue:

Regra de Decisão: Associa-se \underline{x} à classe ω_m (região R_m), onde $g_m(\underline{x}) > g_i(\underline{x}) \forall i = 1, \dots, c$ e $i \neq m$.

Se $g_k(\underline{x}) = g_l(\underline{x})$, com $k \neq l$, tem-se uma fronteira de decisão. As Figs. 2.3 e 2.4 ilustram estes conceitos.

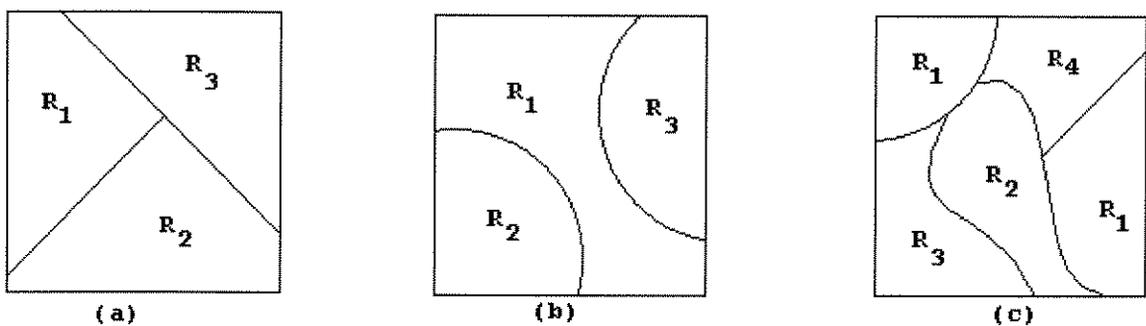


Fig. 2.3 - Exemplos de regiões de decisão: (a) linear, (b) quadrática, (c) geral.

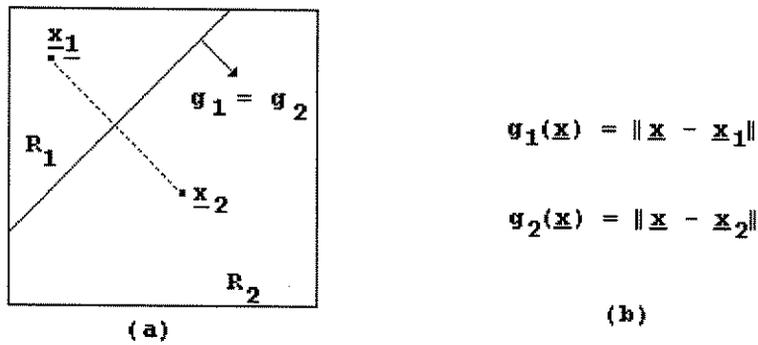


Fig. 2.4 - Regiões de decisão e funções discriminantes correspondentes (a) - Partição de R_2 ; (b) - Funções discriminantes correspondentes.

2.2 - Abordagens ao Reconhecimento de Padrões

As aplicações do reconhecimento de padrões surgem nas mais diversas formas. Em algumas, existe uma base estatística intrínseca e quantificável, responsável pela geração dos padrões. Em outras, é a estrutura intrínseca dos padrões que fornece as informações fundamentais para o reconhecimento. Noutras, porém, nenhuma das abordagens anteriores é adequada; todavia, existe a possibilidade de se desenvolver e “treinar” uma arquitetura que associe corretamente os padrões de entrada com as saídas esperadas. Nota-se, portanto, que dependendo do problema em questão, uma ou mais destas abordagens podem ser aplicadas para se obter a solução desejada.

2.2.1- Abordagem Estatística

2.2.1.1- Introdução

A teoria da de Decisão de Bayes é uma abordagem estatística fundamental do problema de classificação de padrões. Esta abordagem baseia-se na hipótese de que o problema de decisão é apresentado em termos probabilísticos, e que todos os valores das probabilidades relevantes são conhecidos. Neste Seção são apresentados os conceitos fundamentais desta teoria.

Para melhor ilustrar os conceitos que serão apresentados nas Seções que se seguem, o exemplo a seguir será utilizado:

Deseja-se construir um classificador para separar dois tipos de metais preciosos, ouro e prata. Considera-se que um observador esteja monitorando o transporte de barras destes metais de um cofre que contém os dois tipos de metais (cofre “reservatório”) para cofres específicos (cofre “ouro” e cofre “prata”). Observando a saída do cofre reservatório, o observador conclui que é muito difícil prever que tipo de metal estará presente na próxima barra a sair. Portanto, a seqüência de aparecimento das barras de metais parece ser aleatória. Diz-se que à medida que cada barra de metal sai do cofre reservatório, a natureza está em um dos dois estados possíveis: ou a barra é de ouro ou ela é de prata. Denota-se w o estado da natureza com $w = w_1$ para o ouro e $w = w_2$ para a prata. Devido ao fato do estado da natureza ser tão imprevisível, considera-se w uma variável aleatória.

Se o cofre reservatório possuir a mesma quantidade de barras dos dois metais, pode-se dizer que a próxima barra a aparecer tem igual probabilidade de ser de ouro ou de prata. Generalizando, assume-se que exista uma probabilidade conhecida $P(w_1)$ que a próxima barra será de ouro e uma probabilidade $P(w_2)$ que será de prata. Estas probabilidades refletem o conhecimento de quão provável se verá ouro ou prata na próxima barra a sair.

Suponha agora que você precise fazer uma escolha sobre o tipo de barra de metal que irá aparecer a seguir. A única informação que você pode utilizar é o valor das probabilidades conhecidas de cada tipo de barra de metal. Se uma decisão deve ser tomada com tão pouca informação, a seguinte regra de decisão parece razoável:

Decide-se por w_1 se $P(w_1) > P(w_2)$; caso contrário, decide-se por w_2 .

Este procedimento pode parecer estranho, pois sempre é feita a mesma escolha apesar de se saber que os dois tipos de metais podem aparecer. O seu desempenho irá depender dos

valores das probabilidades conhecidas *a priori*. Se $P(w_1)$ for muito maior que $P(w_2)$, a decisão em favor de w_1 estará correta a maior parte do tempo. Se $P(w_1) = P(w_2)$ tem-se somente 50% de chance de se estar correto. Em geral, a probabilidade de erro é a menor das probabilidades ($P(w_1)$ e $P(w_2)$). Como mostrado a seguir, sob estas condições, nenhuma outra regra de decisão pode fornecer uma probabilidade de erro menor.

Em geral, ninguém é chamado a tomar decisões com tão poucas evidências. Neste exemplo, pode-se usar a intensidade do brilho dos metais, x , como uma característica relevante. Diferentes amostras de metais produzirão diferentes intensidades de brilho, e é natural expressar estas variações em termos probabilísticos. Considere x como uma variável aleatória contínua, representando a intensidade de brilho da barra, cuja distribuição densidade de probabilidade depende do estado da natureza. Seja $p(x|w_i)$ a função densidade de probabilidade condicional da variável x dado que o estado da natureza é w_i . Então a diferença entre $p(x|w_1)$ e $p(x|w_2)$ descreve a diferença de intensidade de brilho entre barras de ouro e barras de prata.

Suponha que as probabilidades $P(w_i)$ e as funções densidade de probabilidade condicionais $p(x|w_i)$ sejam conhecidas. Suponha a seguir que a intensidade de brilho de uma barra de metal seja medida. Com esta evidência, qual a decisão a tomar, ou seja, que tipo de metal deve-se escolher, ouro ou prata? Para solucionar este problema, é apresentada a **regra de Bayes**:

$$P(w_i|x) = \frac{p(x|w_i) \cdot P(w_i)}{p(x)}, \quad (2.1)$$

onde

$$p(x) = \sum_{i=1}^2 p(x|w_i) \cdot P(w_i), \quad (2.2)$$

A regra de Bayes mostra como, obtendo o valor de x , transformar a probabilidade conhecida *a priori* $P(w_i)$ na probabilidade *a posteriori* $P(w_i|x)$. Se para uma observação x a probabilidade $P(w_1|x)$ é maior que a probabilidade $P(w_2|x)$, toma-se a decisão em favor de w_1 . Da mesma forma, se $P(w_2|x)$ é maior que $P(w_1|x)$, a escolha é em favor de w_2 . Como justificativa para tal procedimento, calcula-se a probabilidade de erro na decisão. Dado um valor específico de x ,

$$P(\text{erro}|x) = \begin{cases} P(w_1|x) & \text{se a escolha foi } w_2 \\ P(w_2|x) & \text{se a escolha foi } w_1 \end{cases} \quad (2.3).$$

Mas ainda deve-se mostrar que a regra apresentada acima minimizará a probabilidade de erro.

A probabilidade de erro é dada por:

$$\begin{aligned} P(\text{erro}) &= \int_{-\infty}^{\infty} P(\text{erro}, x) dx \\ &= \int_{-\infty}^{\infty} P(\text{erro}|x) p(x) dx \end{aligned} \quad (2.4)$$

e se para todo x , $P(\text{erro}|x)$ for tão pequena quanto possível, a integral deverá ser tão pequena quanto possível. Logo conclui-se que a regra de decisão de Bayes,

Escolha w_1 , se $P(w_1|x) > P(w_2|x)$

Escolha w_2 , caso contrário.

minimiza a probabilidade de erro.

Da forma como está apresentada acima, esta regra de decisão realça o papel das probabilidades *a posteriori*. Usando a equação (2.1), esta regra pode ser expressa em termos das probabilidades condicionais e *a priori*. Deve-se notar que $p(x)$ na equação (2.1) não influi na tomada de decisão. Ele é apenas um fator de escala que garante que $P(w_1|x) + P(w_2|x) = 1$.

Ao se eliminar este fator de escala, a seguinte regra de decisão equivalente é obtida:

Escolha w_1 , se $p(x|w_1) \cdot P(w_1) > p(x|w_2) \cdot P(w_2)$

Escolha w_2 , caso contrário;

Em geral, os dois fatores, $p(x|w_i)$ e $P(w_i)$, são importantes para se fazer uma escolha e a regra de decisão de Bayes combina-os para obter a probabilidade de erro mínimo.

2.2.1.2- Teoria da Decisão de Bayes - Caso Contínuo

Nesta seção serão formalizadas e generalizadas as idéias consideradas na Seção anterior. As seguintes generalizações são feitas:

- .Uso de mais de uma característica;
- .Uso de um número maior de estados;
- .Consideração de ações diferentes daquelas onde ocorre apenas a escolha de um estado da natureza;
- .Generalização da probabilidade de erro com a noção da função perda.

Estas generalizações não devem obscurecer a idéia básica que foi ilustrada no exemplo da Seção 2.1.1. Ao permitir o uso de mais de uma característica, se está substituindo o escalar x por um vetor de características \underline{x} . Permitir ações diferentes daquelas essencialmente classificatórias abre espaço para a possibilidade de rejeição que é uma opção útil se a indecisão não for muito cara. A função de perda trata exatamente de quantificar o custo de cada ação. Teoricamente, isto poderia permitir o tratamento de situações nas quais alguns erros têm o custo maior que outros.

Seja $\Omega = \{w_1, \dots, w_n\}$ o conjunto finito de n estados da natureza e $A = \{a_1, \dots, a_k\}$ o conjunto de k ações possíveis. Seja $\lambda(a_j|w_i)$ o erro cometido ao tomar-se a ação a_j quando o estado da natureza é w_i . Seja o vetor de características \underline{x} um vetor de dimensão d cujos elementos são variáveis aleatórias e seja $p(\underline{x}|w_i)$ a função densidade de probabilidade

condicional para \underline{x} . Finalmente, seja $P(w_i)$ a probabilidade conhecida *a priori* de que a natureza está no estado w_i . Logo, a probabilidade *a posteriori* $P(w_i|\underline{x})$ pode ser calculada a partir de $p(\underline{x}|w_i)$ pela regra de Bayes:

$$P(w_i|\underline{x}) = \frac{p(\underline{x}|w_i) \cdot P(w_i)}{p(\underline{x})}, \quad (2.5)$$

onde

$$p(\underline{x}) = \sum_{i=1}^n p(\underline{x}|w_i) \cdot P(w_i), \quad (2.6)$$

Suponha que um dado \underline{x} seja observado e que a decisão a_j seja tomada. Se o estado da natureza é w_i , o erro cometido será $\lambda(a_j|w_i)$. Como $P(w_i|\underline{x})$ é a probabilidade de que o estado correto é w_i , o erro esperado associado à tomada da decisão a_j será dado por:

$$R(a_j|\underline{x}) = \sum_{i=1}^n \lambda(a_j|w_i) P(w_i|\underline{x}) \quad (2.7)$$

O erro esperado é denominado **risco** e $R(a_j|\underline{x})$ é conhecido como **risco condicional**. Pode-se minimizar o risco esperado escolhendo a ação que minimiza o risco condicional. Esta, na verdade, é a regra de decisão de Bayes ótima que é apresentada a seguir.

Formalmente falando, o problema que se tem aqui é o de encontrar uma regra de decisão que minimize o erro total. Uma regra de decisão é uma função $\alpha(\underline{x})$ que mostra qual decisão deve ser tomada em todas as observações possíveis de \underline{x} , ou seja, para cada \underline{x} a função de decisão assumirá um dos a valores a_1, \dots, a_k . O risco total é dado pelo erro esperado associado a uma regra de decisão. Como $R(a_j|\underline{x})$ é o erro condicional associado à ação a_j e como a regra de decisão especifica a ação, o risco total é dado por:

$$R = \int R(\alpha(\underline{x})|\underline{x}) p(\underline{x}) d\underline{x} \quad (2.8)$$

onde dx é a notação utilizada para um elemento de volume do espaço d -dimensional e onde a integral se estende por todo o espaço de características. Se $\alpha(\underline{x})$ é escolhido de tal forma que $R(\alpha(\underline{x})|\underline{x})$ seja tão pequeno quanto possível para todo \underline{x} , então o risco total será minimizado. Isto justifica a seguinte afirmativa:

. Para minimizar o risco total, calcule o risco condicional:

$$R(\alpha_j|\underline{x}) = \sum_{i=1}^n \lambda(\alpha_j|w_i)P(w_i|\underline{x}) \quad (2.9)$$

para todo $j = 1, \dots, k$ e escolha a ação α_j para a qual o risco $R(\alpha_j|\underline{x})$ seja mínimo. O erro total mínimo resultante é chamado de risco de Bayes.

2.2.1.3- Classificação por taxa de erro mínimo

Em problemas de classificação, cada estado da natureza geralmente está associado a uma das n classes distintas e a ação α_j é interpretada como a decisão de que o estado correto é o estado w_j . Se, após tomada a decisão α_j , constata-se que o estado correto é o estado w_i , então a decisão estará correta se $i = j$ e ter-se-á cometido um erro se $i \neq j$. Com o intuito de evitar estes erros, é natural que se desenvolva uma regra de decisão que minimize a probabilidade média de erro ou a taxa de erro. Uma função perda interessante, neste caso, é a chamada função perda simétrica ou zero-um:

$$\lambda(\alpha_j|w_i) = \begin{cases} 0 & \text{se } j = i \\ 1 & \text{se } j \neq i \end{cases} \quad j, i = 1, \dots, n \quad (2.10)$$

Esta função não atribui nenhuma perda à decisão correta e uma perda unitária a qualquer erro. Portanto, todos os erros apresentam o mesmo custo. O risco correspondente a esta função perda é precisamente a probabilidade média de erro pois o risco condicional é dado por:

$$\begin{aligned}
R(a_j|\underline{x}) &= \sum_{i=1}^n \lambda(a_j|w_i)P(w_i|\underline{x}) \\
&= \sum_{\substack{i=1 \\ i \neq j}}^n P(w_i|\underline{x}) \\
&= 1 - P(w_j|\underline{x})
\end{aligned}
\tag{2.11}$$

e $P(w_j|\underline{x})$ é a probabilidade condicional de que a ação a_j está correta. A regra de decisão de Bayes que minimiza o risco seleciona a ação que minimiza o risco condicional. Portanto, para minimizar a probabilidade média de erro, deve-se selecionar j que maximiza a probabilidade *a posteriori* $P(w_j|\underline{x})$. Portanto, para se obter uma taxa de erro mínima deve-se:

. Decidir por w_j se $P(w_j|\underline{x}) > P(w_i|\underline{x})$ para todo $i \neq j$.

2.2.1.4- Classificadores, Funções Discriminantes e Curvas de Decisão

Existem várias formas de se representar um classificador de padrões. Entre elas está o uso de funções discriminantes $g_i(\underline{x})$, $i = 1, \dots, n$. O classificador associa o vetor \underline{x} à classe w_i se:

$$g_i(\underline{x}) > g_j(\underline{x}) \quad \text{para todo } j \neq i \tag{2.12}$$

Ou seja, o classificador é visto como uma máquina que calcula n funções discriminantes e seleciona a categoria correspondente ao maior valor da função discriminante. Isto está ilustrado na Fig. 2.5

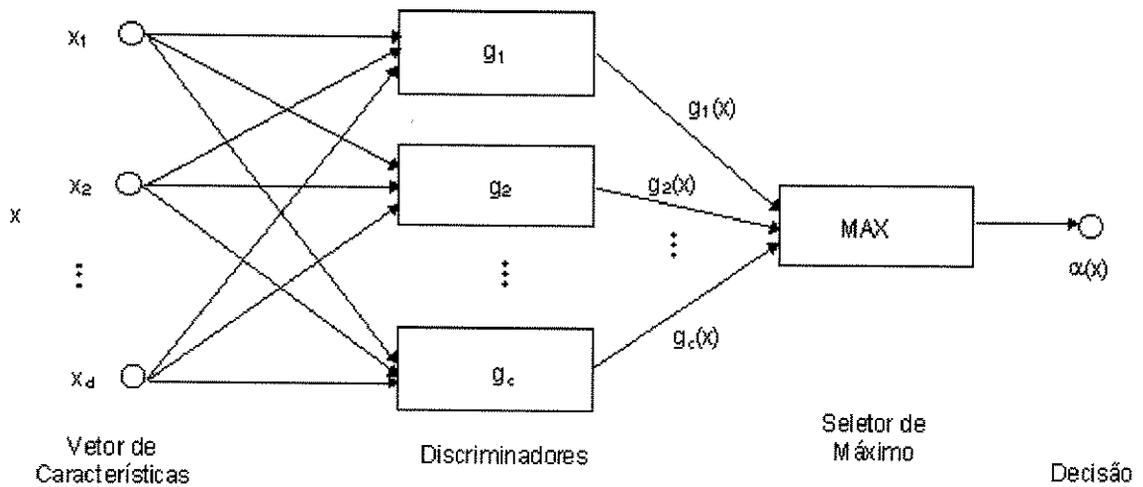


Fig. 2.5 - Classificador de Padrões.

Um classificador de Bayes pode ser facilmente representado deste modo. Neste caso, as funções $g_i(\underline{x})$ são da forma:

$$g_i(\underline{x}) = -R(\omega_i|\underline{x}) \quad (2.13)$$

onde a máxima função discriminante deve corresponder ao mínimo risco condicional. Para o caso da taxa de erro mínima apresentado na Seção 2.1.3, pode-se promover uma maior simplificação tomando como funções discriminantes as funções da forma $g_i(\underline{x}) = P(w_i|\underline{x})$, de forma que a máxima função discriminante corresponda à máxima probabilidade *a posteriori*.

Nota-se, portanto, que a escolha de funções discriminantes não é única. Pode-se multiplicá-las por uma constante positiva ou substituí-las por funções do tipo $f(\mathbf{g}_i(\mathbf{x}))$, onde f é uma função monotônica não decrescente, sem influenciar a decisão. Para o caso da classificação com taxa de erro mínimo qualquer um dos conjuntos de funções abaixo poderia ser utilizado:

$$g_i(\underline{x}) = P(w_i|\underline{x});$$

$$g_i(\underline{x}) = \frac{p(\underline{x}|w_i)P(w_i)}{\sum_{j=1}^n p(\underline{x}|w_j)P(w_j)};$$

$$g_i(\underline{x}) = p(\underline{x}|w_i)P(w_i);$$

$$g_i(\underline{x}) = \ln(p(\underline{x}|w_i)) + \ln(P(w_i));$$

Apesar de as funções discriminantes serem expressas numa variedade de formas, as regras de decisão são equivalentes. O efeito de qualquer regra de decisão é o de dividir o espaço de características em n regiões de decisão R_1, \dots, R_n . Se $g_i(\underline{x}) > g_j(\underline{x})$ para todo $i \neq j$, então \underline{x} está em R_i e a regra de decisão associa \underline{x} a w_i . As regiões são separadas por fronteiras de decisão, superfícies no espaço de características onde as funções discriminantes se encontram. Se R_i e R_j são contíguas a equação da fronteira de decisão que as separa é dada por:

$$g_i(\underline{x}) = g_j(\underline{x}) \quad (2.14)$$

Para pontos na fronteira de decisão, a classificação não é unicamente definida. Para um classificador de Bayes, o risco condicional associado a uma ou outra decisão é o mesmo.

2.2.1.5- Probabilidades de Erro e Integrais

Algumas informações interessantes podem ser obtidas sobre a operação do classificador de Bayes, se o este for considerado como um dispositivo que particiona o espaço de características em regiões de decisão. Para o caso onde são consideradas apenas duas características, suponha que o classificador tenha dividido o espaço em duas regiões, R_1 e R_2 . O erro de classificação pode ocorrer quando:

- . \underline{x} está na região R_2 e o estado da natureza correto é w_1 ;
- . \underline{x} está na região R_1 e o estado da natureza correto é w_2 .

Como estes eventos são mutuamente exclusivos e exaustivos, tem-se:

$$\begin{aligned}
 P(\text{erro}) &= P(x \in R_2, w_1) + P(x \in R_1, w_2) \\
 &= P(x \in R_2|w_1).P(w_1) + P(x \in R_1|w_2).P(w_2) \\
 &= \int_{R_2} p(x|w_1).P(w_1)dx + \int_{R_1} p(x|w_2).P(w_2)dx
 \end{aligned}
 \tag{2.15}$$

Esta probabilidade é ilustrada na Fig. 2.6.

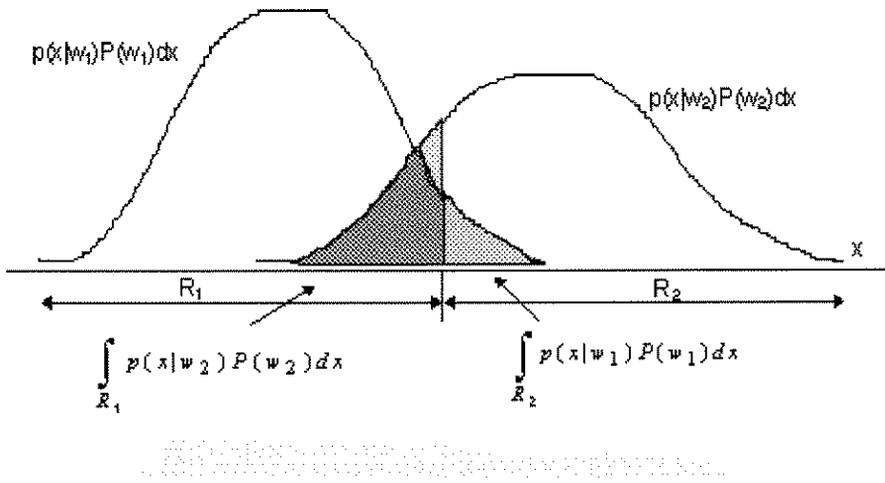


Fig. 2.6 - Componentes da probabilidade de erro.

Para o caso de várias classes como mostrado na Seção 2.1.4, existem mais formas de se estar errado do que de se estar correto sendo, portanto, mais fácil calcular a probabilidade de se estar correto. Logo:

$$\begin{aligned}
 P(\text{correto}) &= \sum_{i=1}^n P(x \in R_i, w_i) \\
 &= \sum_{i=1}^n P(x \in R_i|w_i).P(w_i) \\
 &= \sum_{i=1}^n \int_{R_i} p(x|w_i).P(w_i)dx
 \end{aligned}
 \tag{2.16}$$

Este resultado é válido independentemente da forma como o espaço de características é particionado em regiões de decisão. O classificador de Bayes maximiza esta probabilidade

escolhendo as regiões nas quais os integrandos são máximos; nenhuma outra partição fornece como resultado uma probabilidade de erro menor.

2.2.1.6- Densidade Normal

A estrutura de um classificador de Bayes é determinada inicialmente pelas funções densidade de probabilidade condicional $p(\underline{x}|w_i)$. Das várias funções densidade investigadas, nenhuma recebeu mais atenção que a função densidade normal multivariável. Nesta seção, será feita uma breve exposição das propriedades da função densidade normal multivariável, focalizando as propriedades de interesse para problemas de classificação.

2.2.1.6.1- Distribuição Normal Monovariável

A distribuição normal de uma variável é dada pela seguinte equação:

$$p(x) = \frac{1}{\sqrt{2\pi}} \exp\left[-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right] \quad (2.17)$$

na qual:

$$E[x] = \int_{-\infty}^{\infty} x \cdot p(x) dx = \mu \quad (2.18)$$

e

$$E[(x-\mu)^2] = \int_{-\infty}^{\infty} (x-\mu)^2 \cdot p(x) dx = \sigma^2 \quad (2.19)$$

A distribuição normal é completamente especificada por dois parâmetros, a média μ e a variância σ^2 . As amostras que apresentam uma distribuição normal tendem a se agrupar em torno da média, com espalhamento proporcional ao desvio padrão σ ; aproximadamente 95% das amostras de uma população com distribuição normal estão dentro do intervalo $|x - \mu| \leq 2\sigma$.

2.2.1.6.2- Distribuição Normal Multivariável

A distribuição normal multivariável é dada por:

$$p(\underline{x}) = \frac{1}{(2\pi)^{d/2} |\Sigma|^{1/2}} \exp\left[-\frac{1}{2}(\underline{x} - \underline{\mu})^t \Sigma^{-1}(\underline{x} - \underline{\mu})\right] \quad (2.20)$$

onde \underline{x} é um vetor de dimensão $dx1$, $\underline{\mu}$ é o vetor média cuja dimensão também é $dx1$, Σ é a matriz de covariância de dimensão $dx d$, $(\underline{x} - \underline{\mu})^t$ é o transposto de $(\underline{x} - \underline{\mu})$, Σ^{-1} é a inversa de Σ e $|\Sigma|$ é o determinante de Σ . Formalmente:

$$\underline{\mu} = E[\underline{x}] \quad (2.21)$$

e

$$\Sigma = E\left[(\underline{x} - \underline{\mu})(\underline{x} - \underline{\mu})^t\right] \quad (2.22)$$

A matriz de covariância Σ é sempre simétrica e semi-definida positiva. Apenas será tratado o caso em que Σ é definida positiva e, portanto, o seu determinante $|\Sigma|$ é estritamente positivo. O elemento da diagonal σ_{ii} é a variância de x_i e elemento que está fora da diagonal σ_{ij} é a covariância de x_i e x_j . Se x_i e x_j são estatisticamente independentes, $\sigma_{ij} = 0$. Se todos os elementos que não estão na diagonal principal são zero, $p(\underline{x})$ reduz-se ao produto de distribuições normais de uma variável para os componentes de \underline{x} .

A distribuição normal multivariável é completamente especificada por $d + \frac{d(d+1)}{2}$ parâmetros, os elementos do vetor de médias e os elementos independentes da matriz de covariância. As amostras escolhidas de uma população com distribuição normal tendem a pertencer a um grupamento simples. O centro do grupamento é determinado pelo vetor média e a forma do grupamento pela matriz de covariância. Segue da equação 2.20 que os loci dos pontos de densidade constante são hiper-elipsóides nas quais a quádrlica $(\underline{x} - \underline{\mu})^t \Sigma^{-1}(\underline{x} - \underline{\mu})$ é

constante. Os eixos principais destas hiper-elipsóides são dados pelos autovetores de Σ e os autovalores determinam os comprimentos destes eixos. A expressão:

$$r^2 = (\underline{x} - \underline{\mu})^t \Sigma^{-1} (\underline{x} - \underline{\mu}) \quad (2.23)$$

é geralmente denominada de **distância de Mahalanobis quadrada** de \underline{x} para $\underline{\mu}$. Portanto, os contornos de densidades constantes são hiper-elipsóides com distâncias de Mahalanobis a $\underline{\mu}$ constantes. O volume destas hiper-elipsóides é usado com medida do espalhamento destas amostras em torno da média. Pode-se mostrar que o volume da hiper-elipsóide correspondente à distância de Mahalanobis r é dado por:

$$V = V_d |\Sigma|^{1/2} r^d, \quad (2.24)$$

onde V_d é o volume de uma hiper-esfera unitária d-dimensional:

$$V_d = \begin{cases} \frac{\pi^{d/2}}{\left(\frac{d}{2}\right)!} & d \text{ par} \\ \frac{2^d \pi^{(d-1)/2} \left(\frac{d-1}{2}\right)!}{d!} & d \text{ impar} \end{cases} \quad (2.25)$$

Portanto, para uma dada dimensão, o espalhamento das amostras varia diretamente com $|\Sigma|^{1/2}$.

2.2.1.7- Funções Discriminantes para a Distribuição Normal

Na seção 2.1.4 foi mostrado que a classificação de acordo com a taxa de erro mínimo pode ser conseguida usando funções discriminantes do tipo:

$$g_i(\underline{x}) = \ln(p(\underline{x}|w_i)) + \ln(P(w_i)) \quad (2.26)$$

Considerando que $p(\underline{x}|w_i)$ é uma distribuição normal multivariável tem-se:

$$g_i(\underline{x}) = -\frac{1}{2} (\underline{x} - \underline{\mu})^t \Sigma^{-1} (\underline{x} - \underline{\mu}) - \frac{d}{2} \ln 2\pi - \frac{1}{2} \ln |\Sigma_i| + \ln P(w_i) \quad (2.27)$$

Com este resultado, passa-se, agora, à análise de alguns casos especiais.

2.2.1.7.1- Caso 1: $\Sigma_i = \sigma^2 I$

Este é o caso mais simples. Nele, as características são estatisticamente independentes e cada característica possui a mesma variância σ^2 . Neste caso, a matriz de covariância é uma matriz diagonal ($\sigma^2 I$). Geometricamente, isto corresponde à situação na qual as amostras pertencem a grupos hiper-esféricos de mesmo tamanho, o grupo correspondente à i -ésima classe está centrado em torno do vetor média μ_i . O cálculo do determinante e da matriz inversa de Σ_i é bem simples: $|\Sigma_i| = \sigma^{2d}$ e $\Sigma_i^{-1} = \left(\frac{1}{\sigma^2}\right)I$. Como, tanto $|\Sigma_i|$ como $\left(\frac{d}{2}\right)\ln(2\pi)$ são termos independentes de i na eq.(2.27), eles configuram apenas como constantes aditivas e podem ser ignorados. Logo, obtém-se a seguinte função discriminante:

$$g_i(\underline{x}) = -\frac{\|\underline{x} - \underline{\mu}_i\|^2}{2\sigma^2} + \ln(P(w_i)), \quad (2.28)$$

onde $\|\cdot\|$ é a norma Euclidiana, com:

$$\|\underline{x} - \underline{\mu}_i\| = \left(\underline{x} - \underline{\mu}_i\right)^t \left(\underline{x} - \underline{\mu}_i\right) \quad (2.29)$$

Se as probabilidades *a priori* $P(w_i)$ forem as mesmas para todas as n classes, então o termo $\ln(P(w_i))$ se torna uma constante aditiva que pode ser desprezada. Neste caso, a regra ótima de decisão pode ser definida da seguinte maneira:

Para classificar um vetor de características \underline{x} , calcule a distância Euclidiana $\|\underline{x} - \underline{\mu}_i\|$ do vetor \underline{x} a cada um dos n vetores média e associe \underline{x} à categoria com média mais próxima a \underline{x} .

Este classificador é denominado **classificador de mínima distância**.

Se as probabilidades *a priori* não forem iguais, então quando \underline{x} estiver igualmente próximo a dois vetores média diferentes, a decisão será em favor da categoria com probabilidade *a priori* maior. Percebe-se, porém, não ser necessário o cálculo das distâncias Euclidianas em nenhum dos casos. Expandindo a quádrlica $(\underline{x} - \underline{\mu}_i)^t (\underline{x} - \underline{\mu}_i)$ tem-se:

$$g_i(\underline{x}) = \frac{-1}{2\sigma^2} \left[\underline{x}^t \underline{x} - 2\underline{\mu}_i^t \underline{x} + \underline{\mu}_i^t \underline{\mu}_i \right] + \ln(P(\omega_i)), \quad (2.30)$$

que parece ser uma função quádrlica de \underline{x} . Entretanto, o termo quádrlico $\underline{x}^t \underline{x}$ é o mesmo para todo i tornando-o uma constante que pode ser ignorada. Logo, obtém-se o conjunto de funções discriminantes lineares equivalentes:

$$g_i(\underline{x}) = \underline{w}_i^t \underline{x} + w_{i0}, \quad (2.31)$$

onde:

$$\underline{w}_i = \frac{1}{\sigma^2} \underline{\mu}_i \quad (2.32)$$

e

$$w_{i0} = -\frac{1}{2\sigma^2} \underline{\mu}_i^t \underline{\mu}_i + \ln(P(\omega_i)) \quad (2.33)$$

Um classificador que utiliza um conjunto de funções lineares é denominado **máquina linear**.

Uma das propriedades de uma máquina linear é a caracterização das superfícies de decisão por hiperplanos definidos pelas equações lineares $g_i(\underline{x}) = g_j(\underline{x})$. Para o caso específico acima, esta equação pode ser escrita como:

$$\underline{w}^t (\underline{x} - \underline{x}_0) = 0, \quad (2.34)$$

onde:

$$\underline{w} = \underline{\mu}_i - \underline{\mu}_j \quad (2.35)$$

e

$$\underline{x}_0 = \frac{1}{2}(\underline{\mu}_i + \underline{\mu}_j) - \frac{\sigma^2}{\|\underline{\mu}_i - \underline{\mu}_j\|^2} \ln \frac{P(\omega_i)}{P(\omega_j)} (\underline{\mu}_i - \underline{\mu}_j) \quad (2.36)$$

Esta equação define um hiperplano que passa pelo ponto \underline{x}_0 e é ortogonal ao vetor \underline{w} . Como $\underline{w} = \underline{\mu}_i - \underline{\mu}_j$, o hiperplano que separa R_i de R_j é ortogonal à linha entre as duas médias. Se $P(\omega_i) = P(\omega_j)$, então o ponto \underline{x}_0 está exatamente na metade da linha que une as médias. Se $P(\omega_i) \neq P(\omega_j)$, o ponto \underline{x}_0 estará deslocado na direção da média mais provável. Note, no entanto, que se a variância σ^2 for pequena em relação à distância quadrada $\|\underline{\mu}_i - \underline{\mu}_j\|^2$, a posição da fronteira de decisão será relativamente insensível aos valores exatos das probabilidades *a priori*

2.2.1.7.2- Caso 2: $\Sigma_i = \Sigma$

Um outro caso que merece atenção, é o caso onde as matrizes de covariância são idênticas para todas as classes. Geometricamente isto corresponde à situação na qual as amostras estão em grupamentos hiper-elipsoidais de mesma forma e tamanho; o grupamento da i -ésima classe está centrado em torno do vetor de média $\underline{\mu}_i$. Como, tanto $|\Sigma_i|$ como $\left(\frac{d}{2}\right) \ln(2\pi)$, na eq. (2.27), são termos independentes de i , eles podem ser ignorados. Isto resulta no seguinte conjunto de funções discriminantes:

$$g_i(\underline{x}) = -\frac{1}{2}(\underline{x} - \underline{\mu}_i)^t \Sigma^{-1}(\underline{x} - \underline{\mu}_i) + \ln P(\omega_i) \quad (2.37)$$

Se as probabilidades *a priori* $P(\omega_i)$ forem iguais para todas as n classes, então o termo $\ln(P(\omega_i))$ pode ser ignorado. Neste caso, a regra ótima de decisão pode ser enunciada da seguinte maneira:

Para classificar um vetor de características \underline{x} , calcule a distância de Mahalanobis quadrada $(\underline{x} - \underline{\mu})^t \Sigma^{-1} (\underline{x} - \underline{\mu})$ de \underline{x} a cada um dos n vetores média e associe \underline{x} à categoria com média mais próxima.

A expansão da forma quádrlica $(\underline{x} - \underline{\mu})^t \Sigma^{-1} (\underline{x} - \underline{\mu})$ revela que o termo quádrlico $\underline{x}^t \Sigma^{-1} \underline{x}$ é independente de i . Eliminando-o, as funções discriminantes resultantes são novamente lineares:

$$g_i(\underline{x}) = \underline{w}_i^t \underline{x} + w_{i0}, \quad (2.38)$$

onde:

$$\underline{w}_i = \Sigma^{-1} \underline{\mu}_i \quad (2.39)$$

e

$$w_{i0} = -\frac{1}{2} \underline{\mu}_i^t \Sigma^{-1} \underline{\mu}_i + \ln(P(\omega_i)) \quad (2.40)$$

Como os discriminantes são lineares, as fronteiras de decisão são novamente hiperplanos. Se R_i e R_j são contíguos, a fronteira entre eles tem a seguinte equação:

$$\underline{w}^t (\underline{x} - \underline{x}_0) = 0, \quad (2.41)$$

onde:

$$\underline{w} = \Sigma^{-1} (\underline{\mu}_i - \underline{\mu}_j) \quad (2.42)$$

e

$$\underline{x}_0 = \frac{1}{2}(\underline{\mu}_i + \underline{\mu}_j) - \frac{\ln \frac{P(\omega_i)}{P(\omega_j)}}{(\underline{\mu}_i - \underline{\mu}_j)^t \Sigma^{-1}(\underline{\mu}_i - \underline{\mu}_j)} (\underline{\mu}_i - \underline{\mu}_j) \quad (2.43)$$

Como $\underline{w} = \Sigma^{-1}(\underline{\mu}_i - \underline{\mu}_j)$ geralmente não possui a direção de $(\underline{\mu}_i - \underline{\mu}_j)$, o hiperplano que separa R_i de R_j é geralmente não-ortogonal à linha entre as médias. Embora ele intercepte esta linha no ponto x_0 que está na metade desta linha se as probabilidades *a priori* forem iguais. Caso estas probabilidades sejam diferentes, o hiperplano fronteira será transladado em direção contrária à média mais provável.

2.2.1.7.3- Caso 3: Σ_i arbitrário

No caso mais geral as matrizes de covariância são diferentes para cada categoria. Logo, o único termo que pode ser eliminado na eq. 2.27 é o termo $\left(\frac{d}{2}\right) \ln(2\pi)$ e as funções discriminantes resultantes são inerentemente quádricas:

$$g_i(\underline{x}) = \underline{x}^t W_i \underline{x} + \underline{w}_i^t \underline{x} + w_{i0}, \quad (2.44)$$

onde:

$$W_i = -\frac{1}{2} \Sigma_i^{-1} \quad (2.45)$$

$$\underline{w}_i = \Sigma_i^{-1} \underline{\mu}_i \quad (2.46)$$

e

$$w_{i0} = -\frac{1}{2} \underline{\mu}_i^t \Sigma_i^{-1} \underline{\mu}_i - \frac{1}{2} \ln |\Sigma_i| + \ln(P(\omega_i)) \quad (2.47)$$

As superfícies de decisão são hiper-quádricas e podem assumir qualquer forma geral: pares de hiperplanos, hiper-esferas, hiper-elipsóides, hiper-parabolóides e hiper-hiperbolóides de vários tipos.

2.2.1.8- Conclusão

Nesta seção foi feita uma exposição da regra de decisão de Bayes. As idéias básicas são simples. Para minimizar o risco total, deve-se sempre escolher a ação que minimize o risco condicional:

$$R(a_j|\underline{x}) = \sum_{i=1}^n \lambda(a_j|w_i)P(w_i|\underline{x}) \quad (2.48)$$

Em particular, para minimizar a probabilidade de erro em um problema de classificação, deve-se sempre escolher o estado da natureza que maximize a probabilidade *a posteriori* $P(w_i|\underline{x})$. A regra de Bayes possibilita o cálculo destas probabilidades a partir das probabilidades *a priori* $P(w_i)$ e da função densidade de probabilidade $p(\underline{x}|w_i)$.

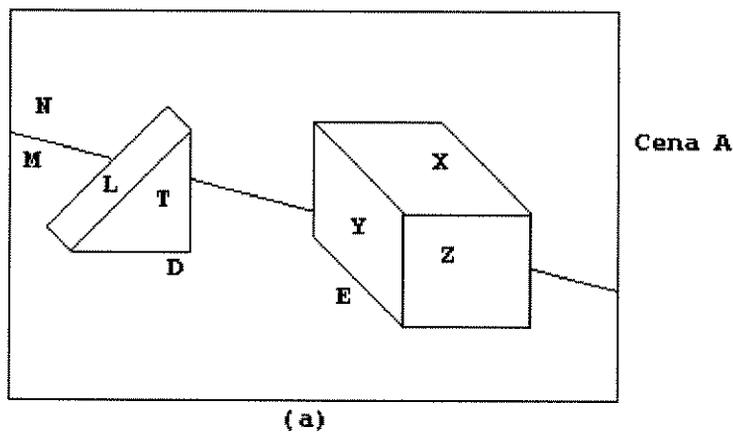
Para a maioria das aplicações de classificação de padrões, o problema principal em aplicar estes resultados é que as funções densidade de probabilidade $p(\underline{x}|w_i)$ não são conhecidas. Em alguns casos pode-se conhecer as formas que estas funções assumem, mas não os valores de parâmetros característicos. O caso clássico ocorre quando se sabe, ou pode-se assumir, que estas funções são distribuições normais multivariáveis porém não se conhecem o vetor de médias e nem as matrizes de covariância.

2.2.2 - Abordagem Sintática

2.2.2.1 - Introdução

Em alguns problemas de reconhecimento de padrões, a informação estrutural que descreve cada padrão é importante e o processo de reconhecimento abrange não apenas a capacidade de classificá-lo como também a de descrever aspectos do padrão que tornam a sua classificação em mais de uma classe algo improvável. Um exemplo típico desta classe de problemas de classificação é o de análise de uma cena. Nesta classe de problemas, os padrões considerados são geralmente complexos e o número de características extraídas é muito grande, o que torna a idéia, de descrever um padrão complexo em termos de uma composição hierárquica de subpadrões simples, muito atrativa. Observa-se, ainda, que, quando os padrões são complexos e o número de descrições possíveis muito grande, torna-se impraticável considerar cada descrição como a definição de uma classe.

Exemplo 2.1 - Os padrões mostrados na figura 2.7-a podem ser descritos em termos de uma estrutura hierárquica que é apresentada na figura 2.7-b.



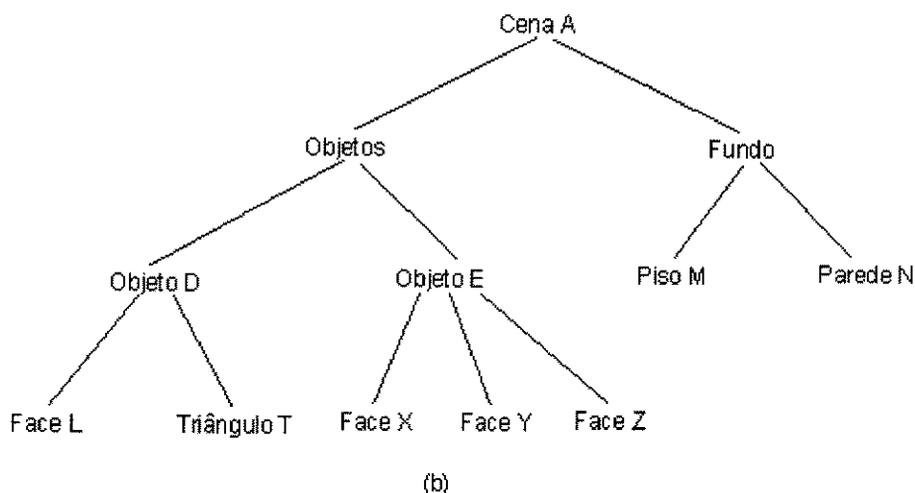


Fig. 2.7 - a e b: Padrão A e a sua descrição hierárquica estrutural.

A abordagem estrutural ou sintática foi proposta, em [14 - 19], com o objetivo de se representar a informação hierárquica estrutural de cada padrão, ou seja, um padrão é descrito em termos de subpadrões mais simples e cada subpadrão é novamente descrito em termos de subpadrões mais simples ainda. Esta abordagem apresenta uma analogia entre a estrutura hierárquica de padrões e a sintaxe das linguagens. Os padrões são especificados como sendo construídos a partir de subpadrões em várias formas de composição, assim como as frases e sentenças são construídas a partir da concatenação de palavras e estas, por sua vez, de caracteres. Evidentemente, os subpadrões mais simples que são escolhidos, *primitivas*, devem ser mais fáceis de se reconhecer do que os próprios padrões. A *linguagem* que fornece a descrição estrutural dos padrões a partir de um conjunto de primitivas e um conjunto de operações sobre estas é chamada de *Linguagem de Descrição de Padrões*. As regras que gerenciam a composição de primitivas para a criação de padrões são especificadas pela *gramática* desta linguagem. Após cada primitiva ser identificada no padrão, o processo de reconhecimento finaliza-se com a realização da análise sintática da *sentença* que descreve o padrão para determinar se o mesmo está sintaticamente (ou gramaticalmente) correto em

relação à gramática especificada. Concomitantemente, a análise sintática também produz uma descrição estrutural da sentença que representa o padrão dado.

A abordagem sintática fornece os meios de se descrever um grande conjunto de padrões complexos usando pequenos conjuntos de primitivas e regras gramaticais. Um dos aspectos mais atrativos desta característica é o uso da natureza recursiva da gramática. Uma regra gramatical pode ser aplicada diversas vezes permitindo assim que algumas características estruturais básicas sejam expressas de forma bem compacta.

As várias relações ou operações de composição definidas entre os sub-padrões podem ser expressas em termos de operações lógicas e/ou matemáticas. Como exemplo, suponha que a concatenação seja escolhida como a única relação (operação de composição) usada na descrição dos padrões, então para as primitivas mostradas na Fig. 2.8-a o losango na Fig. 2.8-b seria representado pela seguinte seqüência: *bacdfc*.

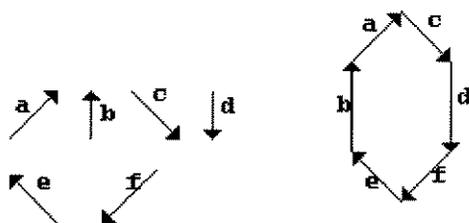


Fig. 2.8 - (a) e (b) - Losango e suas primitivas

Uma representação alternativa da informação estrutural de um padrão é dada por um *grafo relacional*. O grafo relacional da cena A na Fig. 2.7-a é mostrado na Fig. 2.9.

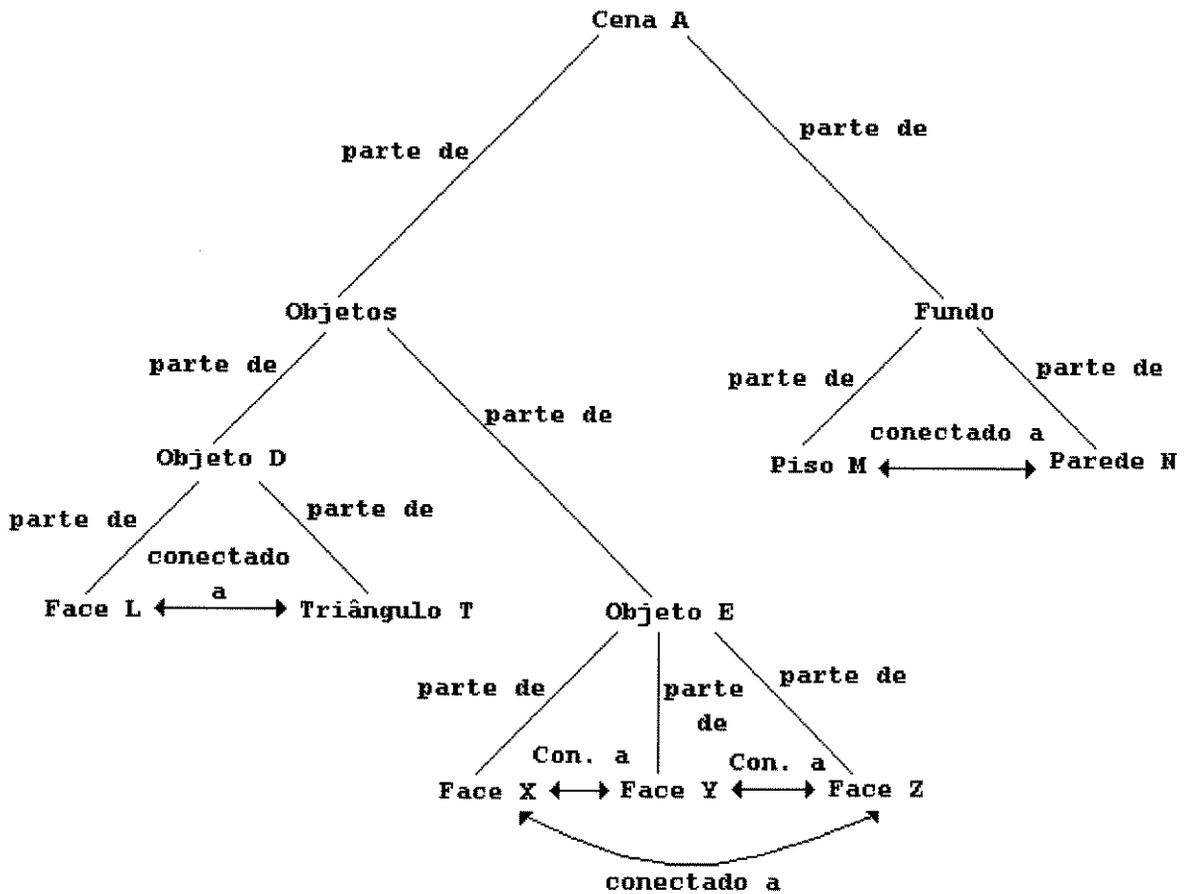


Fig. 2.9 - Grafo relacional da cena A

Como existe uma relação de um para um entre um grafo linear e uma matriz, um grafo relacional certamente pode ser expresso como uma *matriz relacional*. Com essas generalizações, pode-se expressar descrições mais detalhadas e ricas que aquelas feitas utilizando estruturas em árvore. Entretanto, o uso de estruturas em árvore fornece um canal direto para a adaptação das técnicas utilizadas na teoria de linguagens formais ao problema de representação compacta e de análise de padrões que possuem um conteúdo estrutural significativo.

2.2.2.2 - Definições Importantes

Nesta seção serão introduzidas algumas definições e notações importantes.

Definição 1.1: Uma gramática G é tupla da forma:

$$G = (V_N, V_T, P, S) \quad (2.49)$$

onde:

V_N é o conjunto finito de não-terminais,

V_T é o conjunto finito de terminais,

$S \in V_N$ é o símbolo inicial,

P é o conjunto finito de regras ou produções denotadas por:

$$\alpha \rightarrow \beta$$

α e β são cadeias da união entre V_T e V_N e α possui pelo menos um símbolo de V_N .

As seguintes notações também serão freqüentemente utilizadas:

- 1 - V^* é o conjunto de todas as cadeias de símbolos em V , incluindo ε , a seqüência de tamanho 0, $V^+ = V^* - \{\varepsilon\}$
- 2 - Se x é uma cadeia, x^n é a cadeia formada por n repetições de x .
- 3 - $|x|$ é o comprimento da cadeia x , ou o número de símbolos na cadeia x .
- 4 - $\eta_G \Rightarrow \gamma$, a cadeia η gera diretamente a cadeia γ se $\eta = \omega_1 \alpha \omega_2$, $\gamma = \omega_1 \beta \omega_2$ e $\alpha \rightarrow \beta$ for uma produção de P .
- 5 - $\eta_G^* \Rightarrow \gamma$, a cadeia η gera a cadeia γ se existir uma seqüência de cadeias $\xi_1, \xi_2, \dots, \xi_n$ de modo que $\eta = \xi_1$, $\gamma = \xi_n$, $\xi_i \Rightarrow \xi_{i+1}$, $i = 1, 2, \dots, n-1$. A seqüência de cadeias $\xi_1, \xi_2, \dots, \xi_n$ é chamada de derivação de γ a partir de η .

Definição 1.2 : A linguagem gerada pela gramática G é dada por:

$$L(G) = \{x / x \in V_T^* \text{ e } S_G^* \Rightarrow x\} \quad (2.50)$$

Ou seja, a linguagem consiste de todas as cadeias ou sentenças de terminais geradas a partir do símbolo inicial S.

Definição 2.3: Se, em 1.1, $|\alpha| \leq |\beta|$, a gramática será denominada gramática do tipo 1 ou gramática sensitiva ao contexto. Se $\alpha = A \in V_N$, a gramática será denominada gramática do tipo 2 ou gramática livre de contexto. Se, dado que $\alpha = A$, tem-se que $\beta = aB$ ou $\beta = a$, onde $a \in V_T$ e $B \in V_N$, a gramática será denominada gramática de estado finito ou gramática regular.

As linguagens geradas pelas gramáticas sensíveis ao contexto, livre de contexto e de estado finito são denominadas linguagens sensíveis ao contexto, livre de contexto e de estado finito, respectivamente.

As árvores de derivação constituem um método alternativo para se descrever qualquer derivação num gramática livre de contexto. Uma árvore de derivação para uma gramática livre de contexto pode ser constituída de acordo com o seguinte procedimento:

- i - Todo nó da árvore tem um rótulo que é um símbolo em V_N ou V_T .
- ii - A raiz da árvore recebe o rótulo S.
- iii - Se um nó possui pelo menos um descendente diferente de si, e possui ainda um rótulo A, então $A \in V_N$
- iv - Se os nós n_1, n_2, \dots, n_K são descendentes diretos do nó n (com rótulo A) e estão dispostos da esquerda para a direita na ordem A_1, A_2, \dots, A_K , então:

$$A \rightarrow A_1, A_2, \dots, A_K$$

é uma produção de P.

2.2.2.3 - Sistema Sintático de Reconhecimento de Padrões

Um sistema sintático de reconhecimento de padrões pode ser dividido em três grandes partes: pré-processamento, descrição do padrão e análise sintática. Um diagrama de blocos deste sistema é apresentado na Fig. 2.10.

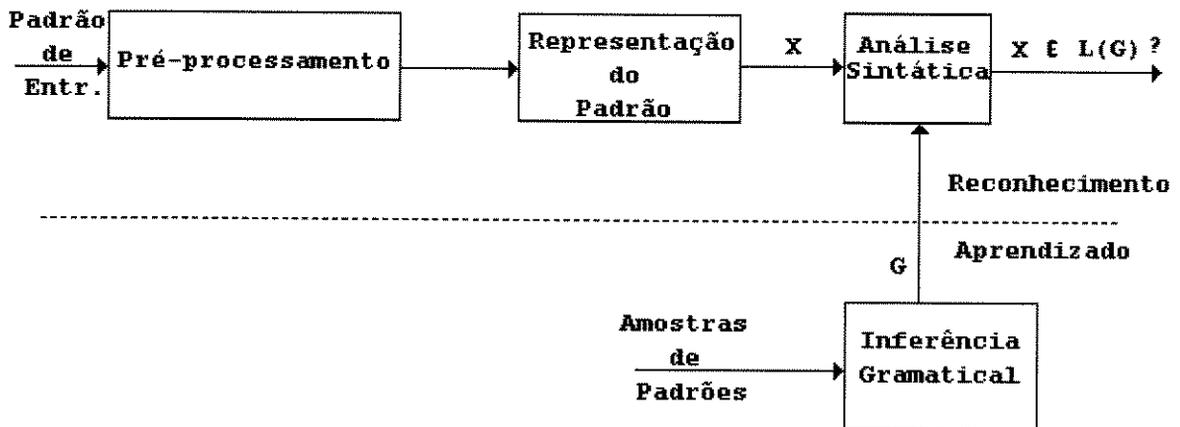


Fig. 2.10 - Diagrama de blocos de um sistema sintático de reconhecimento de padrões

No pré-processamento são realizadas as seguintes operações:

- 1 - codificação do padrão e aproximação;
- 2 - filtragem, restauração e realce;

O padrão de entrada é inicialmente codificado por uma forma conveniente para o processamento futuro. Geralmente, aplica-se, neste estágio, algum tipo de compressão de dados visando tornar mais eficiente o processamento nestes estágios iniciais. Posteriormente, técnicas de filtragem, restauração e/ou realce são aplicadas para retirar ruídos, restaurar regiões degradadas e melhorar a qualidade dos padrões codificados. Após estas operações espera-se que os padrões, na saída do pré-processador, sejam de “boa qualidade”. Cada padrão pré-processado é então representado por uma estrutura lingüística (uma cadeia, um grafo, etc.). Neste estágio, são realizadas as seguintes operações:

- 1 - segmentação do padrão;

2 - extração de características;

Para representar um padrão através dos seus subpadrões, deve-se promover uma segmentação deste padrão e, concomitantemente, suas primitivas devem ser identificadas assim como as relações existentes entre elas. Em outras palavras, cada padrão pré-processado é segmentado em subpadrões e primitivas de acordo com as operações sintáticas pré-estabelecidas; por sua vez, cada subpadrão é identificado por um conjunto de primitivas. Cada padrão é agora representado por um conjunto de primitivas com operações sintáticas específicas. Sistemas mais sofisticados também devem ser capazes de identificar outras relações sintáticas no padrão. Será do analisador sintático a palavra final sobre o fato de a representação do padrão estar sintaticamente correta ou não. Ao fazer a análise sintática o analisador pode produzir uma descrição sintática completa, em termos de uma árvore de análise, do padrão provando que ele está sintaticamente correto. Caso contrário, o padrão é rejeitado ou analisado de acordo com outras gramáticas presentes que descrevem outras classes de padrões possíveis.

Conceitualmente, a forma mais simples de reconhecimento é, provavelmente, o “casamento de padrões”. A cadeia de primitivas que representa um padrão de entrada é comparada com outras cadeias que representam um padrão de referência. Baseando-se em um critério de similaridade, o padrão de entrada é classificado na mesma classe do protótipo que melhor se aproxima de si. Por outro lado, a análise sintática completa da cadeia que representa o padrão de entrada explora a descrição estrutural hierárquica do padrão. Entre estas duas abordagens, existem várias outras intermediárias e a escolha da abordagem apropriada depende das características do problema. Se a descrição completa do padrão é exigida, a análise sintática será necessária; em restrições menos severas, a análise sintática pode ser evitada usando-se abordagens mais simples para melhorar a eficiência do processo de reconhecimento.

Para se ter uma gramática que descreva a informação estrutural da classe de padrões em estudo, deve-se ter uma máquina de inferência gramatical que a deduza a partir de um dado

conjunto de padrões de treinamento codificados numa representação lingüística. A descrição estrutural da classe de padrões em estudo é aprendida a partir das amostras de padrões desta classe. A descrição aprendida, na forma de uma gramática, é então usada para a descrição de padrões e análise sintática. Uma forma mais geral de aprendizagem deve também incluir a capacidade de aprender o melhor conjunto de primitivas e a descrição estrutural correspondente para a classe de padrões sob análise.

2.2.2.4 - Seleção de primitivas

Como discutido na seção 2.2.2.1, o primeiro passo na formulação de um modelo lingüístico para a descrição de padrões é a determinação de um conjunto de primitivas através das quais os padrões de interesse sejam descritos.

Este processo será muito influenciado pela natureza dos dados, pela aplicação em questão, pela tecnologia utilizada na implementação do sistema. Até o momento, não existe nenhuma solução geral para este problema. As características a seguir geralmente são observadas na escolha das primitivas:

- i - As primitivas devem servir como elementos básicos de padrões para que possam fornecer uma descrição compacta porém adequada dos dados em termos das relações estruturais especificadas.
- ii - As primitivas devem ser facilmente extraídas ou reconhecidas por métodos não-lingüísticos existentes, pois elas são consideradas padrões simples e compactos e suas informações estruturais não são importantes.

Em padrões que contenham linhas ou sejam descritos por esqueletos ou fronteiras, os segmentos de linha são geralmente usados como primitivas. Um segmento de linha reto por ser caracterizado pelas localizações de seu início e fim, seu tamanho e sua inclinação. Similarmente, um segmento curvo pode ser descrito em termos de seu início e fim e sua

curvatura. As informações que caracterizam as primitivas podem ser consideradas como suas informações semânticas associadas ou como características utilizadas para o reconhecimento das primitivas. Para a descrição de padrões em termos de regiões, usa-se geralmente algumas medidas que levam em consideração a forma e/ou textura das regiões.

2.2.2.5 - Gramática Padrão

Assumindo que uma solução satisfatória para o problema de “seleção de primitivas” tenha sido conseguida para uma determinada aplicação, o próximo passo será a construção de uma gramática (ou gramáticas) que gerará uma linguagem (ou linguagens) para descrever os padrões em estudo. Seria ótimo se existisse uma máquina de inferência gramatical que pudesse deduzir uma gramática a partir de um dado conjunto de cadeias que descrevem os padrões sob análise. Infelizmente, esta máquina não existe e não ser em algumas aplicações extremamente específicas [20]. Na maioria dos casos, o projetista constrói a gramática baseando-se em algum conhecimento disponível inicialmente e em sua experiência. É sabido que o poder descritivo de uma linguagem é alcançado com a complexidade do sistema de análise sintática. Autômatos de estado finito são capazes de reconhecer linguagens de estado finito, embora o poder descritivo destas linguagens seja inferior ao das linguagens livres de contexto e linguagens sensíveis ao contexto. A escolha de uma gramática específica para a descrição do padrão é afetada pelas primitivas escolhidas e pelo compromisso entre o poder descritivo da gramática e a eficiência da análise.

Deve-se ressaltar que uma gramática é mais apropriada para a descrição quando o padrão de interesse é construído a partir de um pequeno conjunto de primitivas através da aplicação recursiva de pequeno conjunto de regras. Tanto a “escolha das primitivas” quanto a “construção da gramática” devem ser tratadas simultaneamente e não em dois estágios diferentes. Não existe nenhuma dúvida de que uma escolha diferente de primitivas resultará em uma gramática diferente para a descrição do conjunto de padrões.

Um número de linguagens especiais foi proposto para a descrição de padrões tais como caracteres chineses, imagens de cromossomos, estruturas químicas e impressões digitais [21-26]. Com o propósito de realmente descrever padrões de imagem, gramáticas de uma maior dimensão tais como gramáticas web, gramáticas vetoriais, gramáticas em grafos, gramáticas em árvore e gramáticas de formas têm sido usadas.

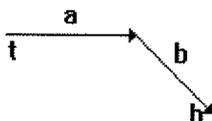
2.2.2.6 - Gramáticas de Dimensão Elevada

Na descrição de padrões através de uma gramática padrão, a única relação possível entre os subpadrões e/ou primitivas é a concatenação; neste caso, cada subpadrão pode ser conectado a um outro subpadrão somente pela esquerda ou pela direita. Esta relação unidimensional não é muito apropriada para a descrição de padrões bi ou tridimensionais. Uma generalização natural decorre do uso de um formalismo mais geral que inclua outras relações [21, 27-32]. Seja \mathbf{R} um conjunto de n relações ($n \geq 1$). Uma relação $r \in \mathbf{R}$ satisfeita pelos subpadrões X_1, \dots, X_n é denotada por $r(X_1, \dots, X_n)$. Por exemplo, **RETÂNGULO(a,b,c,d)** significa que a relação quaternária **RETÂNGULO** é satisfeita pelos segmentos **a**, **b**, **c** e **d**.

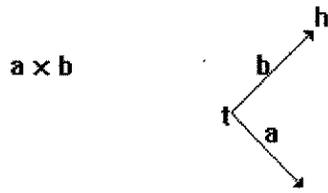
Uma generalização bidimensional simples das gramáticas padrão é obtida a partir da extensão das gramáticas de cadeias unidimensionais em gramáticas para vetores bidimensionais [33, 34]. As primitivas são os elementos dos vetores e a relação entre as primitivas é a concatenação bidimensional.

Shaw, atribuindo “head” (h) e “tail” (t) a cada primitiva, usou os quatro operadores binários $+$, \times , $-$ e $*$ para definir as relações binárias de concatenação entre as primitivas [35, 36]:

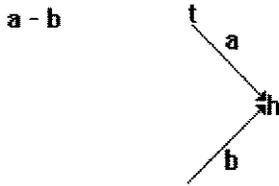
$\mathbf{a} + \mathbf{b}$



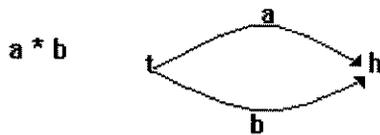
$$\begin{aligned} \mathbf{h}[\mathbf{a} + \mathbf{b}] &= \mathbf{h}[\mathbf{b}] \\ \mathbf{t}[\mathbf{a} + \mathbf{b}] &= \mathbf{t}[\mathbf{a}] \end{aligned}$$



$$\begin{aligned} h(a \times b) &= h(b) \\ t(a \times b) &= t(b) \end{aligned}$$



$$\begin{aligned} h(a - b) &= h(a) \\ t(a - b) &= t(a) \end{aligned}$$



$$\begin{aligned} h(a * b) &= h(a) \\ t(a * b) &= t(a) \end{aligned}$$

A estes operadores soma-se o operador unário \sim que atua como um inversor “tail/head” definido como:



$$\begin{aligned} h(\sim a) &= t(a) \\ t(\sim a) &= h(a) \end{aligned}$$

A gramática que gera sentenças em PDL (Linguagem de Descrição de Figuras) é uma gramática livre de contexto:

$$G = (V_N, V_T, P, S) \quad (2.51)$$

onde:

$$V_N = \{S, SL\}$$

$$V_T = \{b\} \cup \{+, \times, -, /, (\}, \cup \{\}$$

b pode ser qualquer primitiva (incluindo a “primitiva nula” e cujo head é igual a tail) e:

$S \rightarrow b$	$S \rightarrow (/SL)$	$\Phi_b \rightarrow +$
$S \rightarrow (S\Phi_b S)$	$SL \rightarrow SI$	$\Phi_b \rightarrow \times$
$S \rightarrow (\sim S)$	$SL \rightarrow (SL\Phi_b SL)$	$\Phi_b \rightarrow -$
$S \rightarrow SL$	$SL \rightarrow (/SL)$	$\Phi_b \rightarrow *$

l é um rótulo que é usado para permitir uma referência cruzada a expressões S com uma descrição. O operador / é usado para permitir que tail e head de uma expressão possam ser atribuídos arbitrariamente.

Baseando-se numa idéia de [37], Feder formalizou uma gramática “plex” que gera linguagens com terminais que possuem um número arbitrário de pontos de conexão para se conectarem a outras primitivas [24]. As primitivas da gramática plex são chamadas de entidades com N pontos de conexão (**N-Attaching Point Entity - NAPE**). Cada produção da gramática plex está numa forma livre de contexto na qual a conectividade das primitivas é descrita através de listas explícitas de pontos de concatenação rotulados (listas de junção).

Pfaltz e Rosenfeld estenderam o conceito de gramáticas de cadeias para gramáticas de grafos rotulados denominados **webs**. Grafos rotulados orientados são explicitamente usados nas produções. Cada produção descreve como transformar um grafo α num grafo β e também contém uma regra de fixação E que especifica a conexão de β com seus grafos vizinhos no web hospedeiro quando α sofre a transformação. Uma gramática web G é uma tupla:

$$G = (V_N, V_T, P, S) \quad (2.52)$$

onde V_N é um conjunto de não-terminais, V_T é um conjunto de terminais, S é um conjunto de webs “iniciais” e P é o conjunto de produções web. Uma produção web é definida como:

$$\alpha \rightarrow \beta, E \quad (2.53)$$

onde α e β são webs, E é uma regra de fixação de β . Quando se deseja substituir a subweb α na web ω por outra subweb β , é necessário que se especifique como “fixar” β em ω no lugar

de α . A definição de uma regra de fixação não deve depender da web hospedeira ω pois deseja-se substituir α por β em qualquer web que contenha α como subweb. Geralmente, E é constituído de um conjunto de funções lógicas que especificam se cada vértice de $\omega-\alpha$ está ou não conectado a cada vértice de β .

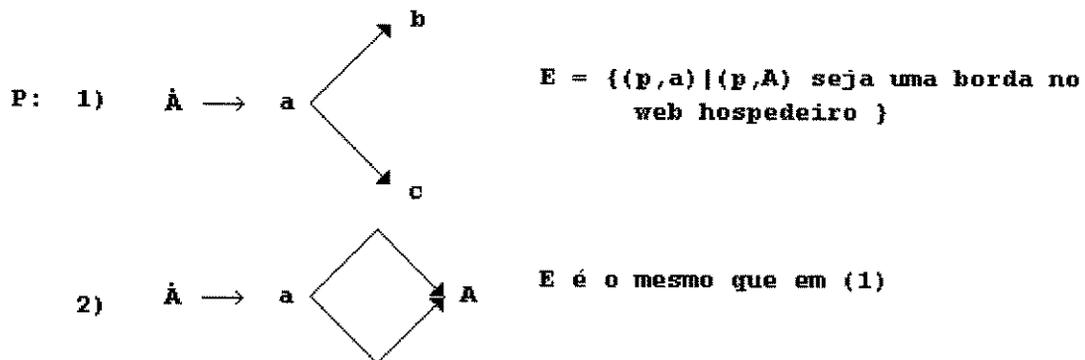
Exemplo 2.2: Considere a gramática web

$$G = (V_N, V_T, P, S)$$

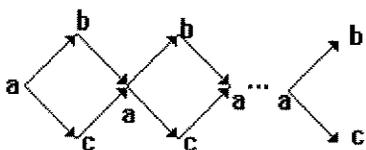
onde:

$$V_N = \{A\}, \quad V_T = \{a, b, c\}, \quad S = \{\dot{A}\}$$

e



A linguagem desta gramática é dada pelo conjunto de todas as webs da forma:



Um caso particular, porém importante, das gramáticas web é aquele no qual o conjunto de terminais V_T consiste de apenas um símbolo. Neste caso, todos os pontos de todas as webs

na linguagem possuem o mesmo rótulo, logo estes rótulos podem ser ignorados e as webs são identificadas apenas por seus grafos. Este tipo de gramática web é chamada de “gramática de grafos” e sua linguagem, linguagem de grafos [38]. O formalismo das gramáticas web foi extensivamente analisado em [39,40]. As relações entre as gramáticas PDL, gramáticas plex e gramáticas web foram discutidas por Shaw [41] e Rosenfeld [42].

Pavlidis [38,43] generalizou as gramáticas de cadeias em gramáticas de grafos incluindo símbolos não-terminais diferentes dos simples ramos ou nós. Uma estrutura não terminal de ordem m é definida como uma entidade que está conectada ao restante do grafo através de m nós. Particularmente, uma estrutura de segunda ordem é denominada ramo e a de primeira, nó. Logo, uma gramática de grafo livre de contexto de ordem m é uma tupla G_g :

$$G_g = (V_N, V_T, P, S) \quad (2.53)$$

onde:

- . V_N é um conjunto de estruturas não-terminais de ordem m : nós, ramos, triângulos, ..., polígonos com m vértices;
- . V_T é um conjunto de terminais: nós e ramos;
- . P é um conjunto finito de produções da forma $A \rightarrow \alpha$, onde A é uma estrutura não-terminal e α , um grafo contendo tanto estruturas terminais como não-terminais. α está conectado ao restante do grafo exatamente pelos mesmos nós que A ;
- . S é o conjunto de grafos iniciais.

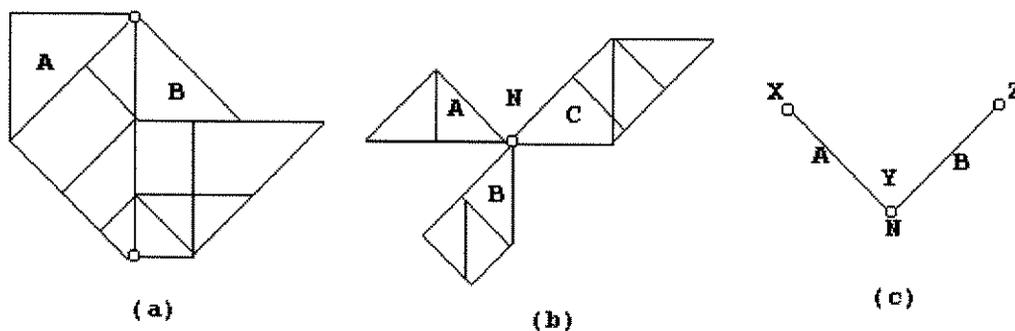


Fig. 2.11 - Ilustração de (a) $A*B$; (b) $N(A+B+C)$; e (c) ANB

A expressão $A*B$ denota que dois grafos A e B estão conectados por um par de nós (Fig. 2.11a) e $N(A+B+C)$ denota que os grafos A, B e C estão conectados por um nó comum N (Fig. 2.11b). Portanto a produção $A \rightarrow B*C$ onde A, B e C são ramos deve ser interpretada assim:

Substitua o ramo A pelas estruturas B e C conectadas ao grafo pelos mesmos nós que A. Não existe nenhuma outra conexão entre B e C.

De modo similar deve ser interpretada a produção $N \rightarrow M(A+B)$:

Substitua o nó N pelo nó M e duas outras estruturas A e B conectadas ao resto do grafo pelo mesmo nó como N.

Quando não existe nenhuma ambigüidade, as concatenações simples podem ser usadas:

ANB denota um subgrafo não-terminal constituído de um ramo A com nós X e Y conectado ao nó N através de Y e um ramo B com nós Y e Z conectado a N também através de Y (Fig. 2.11c). O subgrafo está conectado ao restante do grafo através dos nós X e Z.

O seguinte exemplo ilustra o uso de gramáticas de grafos para a descrição de padrões:

Exemplo 2.3 : A seguinte gramática descreve grafos que representam redes série-paralelas entre dois terminais:

$$G_g = (V_N, V_T, P, S)$$

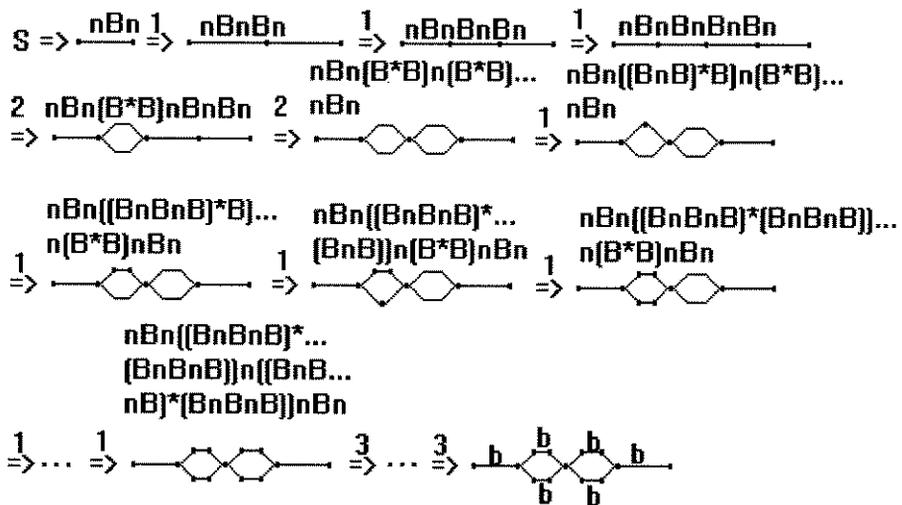
onde:

$$V_N = \{B, S\}, \quad V_T = \left\{ \begin{matrix} b & n \\ -, & \cdot \end{matrix} \right\}, \quad S = \{nBn\}$$

e

- P : (1) $B \rightarrow BnB$
 (2) $B \rightarrow B^*B$
 (3) $B \rightarrow b$

Uma geração típica seria:



Extendendo a concatenação unidimensional para a concatenação multidimensional, as cadeias transformam-se em árvores. Gramáticas de árvores e os correspondentes reconhecedores, autômatos de árvores, foram estudados por diversos autores [44,45]. Naturalmente, se um padrão pode ser convenientemente descrito por uma árvore, ele será facilmente gerado por uma gramática de árvore. Por exemplo, na Fig. 2.12, tem-se um padrão e sua respectiva representação em forma de árvore tirado de [46,47].

Definição 2.6 - Uma árvore em Σ é uma função

$$\alpha : D \rightarrow \Sigma$$

tal que D é o domínio da árvore e

$$r[\alpha(a)] = \max\{i / a \ i \in D\}$$

O domínio de uma árvore α é denotado por $D(\alpha)$ e T_Σ é o conjunto de todas as árvores em Σ .

Definição 2.7 - Seja α uma árvore e a um membro de $D(\alpha)$; α/a , uma subárvore de α em a , é definida como:

$$\alpha/a = \{(b, x) / (a \cdot b, x) \in \alpha\}$$

Definição 2.8 - Uma gramática de árvore regular em $\langle V_T, r \rangle$ é uma tupla:

$$G_t = (G, r', P, S)$$

que satisfaz as seguintes condições:

- (a) - $\langle V, r' \rangle$ é um alfabeto hierárquico finito, $V_T \subseteq V$ e $r' / V_T = r$; $V - V_T = V_N$ é o conjunto de não-terminais;
- (b) - P é um conjunto finito de produções da forma $\Phi \rightarrow \Psi$, onde Φ e Ψ são árvores em $\langle V, r' \rangle$.
- (c) - S é um subconjunto finito de T_V , onde T_V é o conjunto de árvores no alfabeto V .

Definição 2.9 $\alpha \Rightarrow^a \beta$ está em G_t se e somente se existe uma produção $\Phi \rightarrow \Psi$ em P tal que Φ seja uma subárvore de α em a e β seja obtido substituindo a ocorrência de Φ em a por Ψ .

Escreve-se $\alpha \Rightarrow \beta$ em G_t se e somente se existe $a \in D(\alpha)$ tal que $\alpha \Rightarrow^a \beta$.

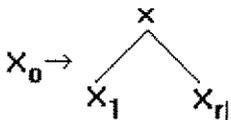
Definição 2.10 $\alpha \Rightarrow^* \beta$ está em G_t se e somente se existe $\alpha_0, \alpha_1, \dots, \alpha_m$ ($m > 0$) tal que

$$\alpha = \alpha_0 \Rightarrow \alpha_1 \Rightarrow \dots \Rightarrow \alpha_m = \beta$$

em G_t . A seqüência $\alpha_0, \alpha_1, \dots, \alpha_m$ é chamada de derivação ou dedução de β a partir de α , e m é o comprimento da derivação.

Definição 2.11 - $L(G_t) = \left\{ \alpha \in T_{V_T} / \text{existe } Y \in S \text{ tal que } Y \Rightarrow^* \alpha \text{ em } G_t \right\}$ é chamado de linguagem de árvore gerada por G_t .

Definição 2.12 - Uma gramática de árvore é expansível se e somente se cada produção em P é da forma:



ou

$$X_j \rightarrow x$$

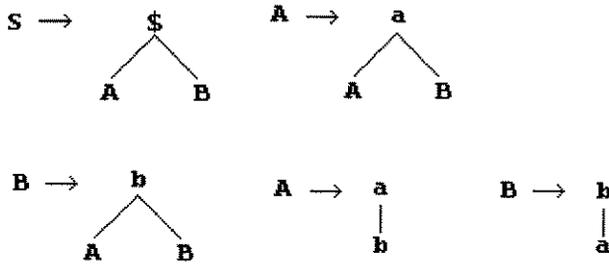
onde $x \in V_T$ e $X_0, X_1, \dots, X_{r(x)}$ são símbolos não-terminais. Para cada gramática de árvore regular G_t , pode-se efetivamente construir uma gramática expansível equivalente G_t' tal que:

$$L(G_t) = L(G_t')$$

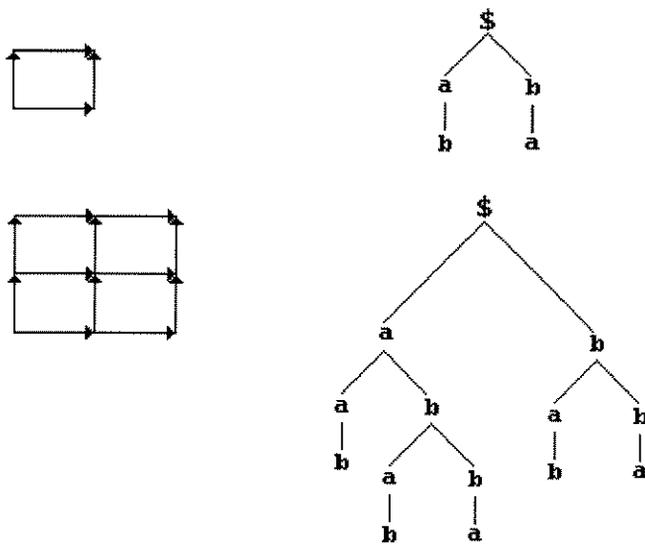
Exemplo 2.4: A gramática de árvore:

$$G_t = (G, r', P, S)$$

onde $V = (S, a, b, \$, A, B)$; $V_T = \left\{ \begin{matrix} a \\ \rightarrow, \uparrow \\ b, \cdot \\ \$ \end{matrix} \right\}$ e P :



gerando os padrões tais como:



Definição 2.13 - Um autômato de árvore em Σ é uma $(k+2)$ -tupla:

$$M_t = (Q, f_1, \dots, f_K, F)$$

onde:

i - Q é um conjunto finito de estados;

ii - Para cada i , $1 \leq i \leq K$, f_i é uma relação em $Q^{r(\sigma_i)} \times Q$, $\sigma_i \in \Sigma$, ou seja,

$$f_i : Q^{r(\sigma_i)} \rightarrow Q$$

iii - $F \subseteq Q$ é o conjunto de estados finais.

Definição 2.14 - A relação resposta ρ de um autômato de árvore M_t é definida como:

i - se $\sigma \in \Sigma_0$, $\rho(\sigma) \sim X$ se e somente se $f_\sigma \sim X$, ou seja, $\rho(\sigma) = f_\sigma$;

ii - se $\sigma \in \Sigma_n$, $n > 0$, $\rho(\sigma, x_0, \dots, x_{n-1}) \sim X$ se e somente se existir x_0, \dots, x_{n-1} tal que $f_\sigma(x_0, \dots, x_{n-1}) \sim X$ e $\rho(x_i) \sim X_i$, $1 \leq i \leq n$, ou seja:

$$\rho(\sigma, x_0, \dots, x_{n-1}) = f_\sigma(\rho(x_{n-1}))$$

Definição 2.15 $T(M_t) = \{\alpha \in T_\Sigma / \text{existe } x \in F \text{ tal que } \rho(\alpha) \sim x\}$ é chamado de conjunto de árvores aceitas por M_t . Para cada gramática de árvore regular G_t , pode-se efetivamente construir um autômato M_t , tal que $T(M_t) = L(G_t)$. A construção deste autômato é resumida a seguir:

i - Obtenha uma gramática de árvore expansível $G_t = (V', r, P', S)$ para a gramática de árvore regular $G_t = (V, r, P, S)$ em V_T .

ii - O autômato de árvore (não determinístico) equivalente é dado por:

$$M_t = (V' - V_T, f_1, \dots, f_K, \{S\})$$

onde

$$f_x = (x_1, \dots, x_n) \sim X_0 \text{ se } X_0 \rightarrow xX_1, \dots, X_n \text{ está em } P'.$$

2.2.2.7 - Reconhecimento Sintático

Nesta seção é apresentada a análise sintática para linguagens de estado finito e linguagens livres de contexto [48, 49].

2.2.2.7.1 - Reconhecimento de Linguagens de Estado Finito

Se uma classe de padrões pode ser descrita por uma linguagem de estado finito, um autômato de estado finito pode ser construído para reconhecer as cadeias ou sentenças que descrevem esta classe de padrões.

Definição 2.16 - Um autômato de estado finito não determinístico é uma tupla $(\Sigma, Q, \delta, q_0, F)$, onde Σ é um conjunto finito de símbolos de entrada (alfabeto), Q é um conjunto finito de estados, δ é um mapeamento de $Q \times \Sigma$ em subconjuntos de Q , $q_0 \in Q$ é o estado inicial e $F \subseteq Q$ é o conjunto de estados finais.

Uma representação conveniente de um autômato de estado finito é dada pela Fig. 2.14 abaixo:

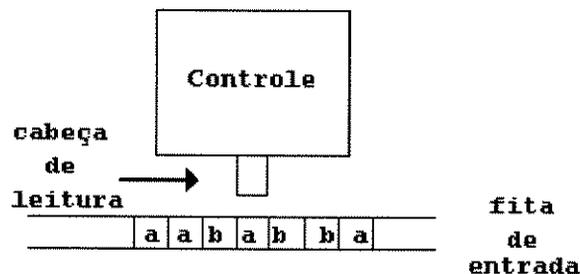


Fig. 2.14 - Autômato de estado finito

A interpretação de:

$$\delta(q, a) = q' \quad q, q' \in Q \quad e \quad a \in \Sigma$$

é a seguinte:

O autômato A , no estado q , “observa” o símbolo de entrada a , vai para o estado q' deslocando a cabeça de entrada de uma célula para a direita. Uma forma conveniente de representar este mapeamento é através do uso do diagrama de transição de estados mostrado na figura abaixo:



Fig. 2.15 - Representação gráfica de $d(q, a) = q'$

O mapeamento δ pode ser estendido a uma cadeia de símbolos de entrada:

$$\delta(q, \lambda) = q' \quad \delta(q, xa) = \delta(\delta(q, x), a) \quad x \in \Sigma^* \text{ e } a \in \Sigma$$

cuja interpretação é a seguinte:

O autômato A , no estado q , “observa” a cadeia x na fita de entrada e vai para o estado q' deslocando a cabeça de entrada de x células para a direita.

Uma cadeia ou seqüência x é dita ser aceita por A se:

$$\delta(q_0, x) = p \quad \text{para algum } p \in F$$

O conjunto de cadeias aceitas por A é definido como:

$$T(A) = \{x / \delta(q_0, x) \in F\}$$

Quando $\delta(q, a)$ consistir de apenas um único estado, o autômato será então chamado de autômato de estado finito determinístico.

A relação entre autômatos de estado finito determinístico e não determinístico assim como a relação entre linguagens de estado finito e os conjuntos de cadeias aceitas pelos autômatos de estado finito são expressas pelos seguintes teoremas:

Teorema 2.1 - Seja L um conjunto de cadeias aceitas por um autômato de estado finito não-determinístico $A = (\Sigma, Q, \delta, q_0, F)$. Então existe um autômato de estado finito determinístico $A' = (\Sigma', Q', \delta', q'_0, F')$ que aceita L . Os estados de A' são todos subconjuntos de Q . F' é o conjunto de todos os estados de Q' que contém um estado de F . Um estado de A' será denotado por $[q_1, q_2, \dots, q_i] \in Q'$ onde $q_1, q_2, \dots, q_i \in Q$. $q_0 = [q_0]$.

$$\delta^r([q_1, q_2, \dots, q_i], a) = [p_1, p_2, \dots, p_j] \quad \text{se e somente se}$$

$$\delta(\{q_1, q_2, \dots, q_i\}, a) = \bigcup_{k=1}^i \delta(q_k, a) = \{p_1, p_2, \dots, p_j\}$$

Teorema 2.2 - Seja $G = (V_N, V_T, P, S)$ uma gramática de estado finito. Existe um autômato $A = (\Sigma, Q, \delta, q_0, F)$ com $T(A) = L(G)$, onde:

- 1 - $\Sigma = V_T$;
- 2 - $Q = V_N \cup \{T\}$;
- 3 - $q_0 = S$;
- 4 - Se P contém a produção $S \rightarrow \lambda$, então $F = \{S, T\}$, caso contrário, $F = \{T\}$;
- 5 - O estado T está em $\delta(B, a)$ se $B \rightarrow a$, $B \in V_N$, $a \in V_T$ for uma produção de P ;
- 6 - $\delta(B, a)$ contém todos $C \in V_N$ tal que $B \rightarrow aC$ é uma produção de P e $\delta(T, a) = \Phi$ para cada $a \in V_T$

Teorema 2.3 - Dado um autômato de estado finito $A = (\Sigma, Q, \delta, q_0, F)$, existe uma gramática de estado finito $G = (V_N, V_T, P, S)$ com $L(G) = T(A)$, onde:

- 1 - $V_N = Q$;
- 2 - $V_T = \Sigma$;
- 3 - $S = q_0$;
- 4- $B \rightarrow aC$ é uma produção de P se $\delta(B, a) = C$, $B, C \in Q$ e $a \in \Sigma$.
- 5 - $B \rightarrow a$ é uma produção de P se $\delta(B, a) = C$, $C \in F$

2.2.2.7.2 - Reconhecimento de Linguagens Livres de Contexto

Quando uma linguagem livre de contexto é usada para descrever uma classe de padrões, o sistema de reconhecimento correspondente consiste, em geral, de um autômato

não-determinístico. Nem todo autômato não-determinístico tem um autômato determinístico equivalente. Portanto, o processo que realiza o reconhecimento é, em geral, um processo não-determinístico. A saída do analisador sintático geralmente contém a decisão sobre a aceitação ou não da cadeia gerada pela gramática dada assim como a árvore de derivação da cadeia que, por sua vez, possibilita uma descrição estrutural completa do padrão.

Muitos algoritmos de análise sintática foram propostos para as linguagens livres de contexto. Destes, apresenta-se abaixo, o mais comumente usado: Algoritmo de Earley.

Algoritmo de Análise Sintática de Earley

Entrada: Uma gramática livre de contexto $G = (N, \Sigma, P, S)$ e uma cadeia de entrada $x = a_1 a_2 \dots a_n$ em Σ^* .

Saída: Listas de análise sintática I_0, I_1, \dots, I_n .

Método:

(A) Construção de I_0

Passo1 - Se $S \rightarrow \alpha$ é uma produção de P , acrescente $[S \rightarrow \cdot \alpha, 0]$ a I_0 . Execute os passos 2 e 3 até que nenhum novo item possa ser adicionado a I_0 .

Passo2 - Se $[B \rightarrow \gamma \cdot, 0]$ está em I_0 , acrescente $[A \rightarrow \alpha B \cdot \beta, 0]$ para todo $[A \rightarrow \alpha \cdot B \beta, 0]$ em I_0 .

Passo3 - Suponha que $[A \rightarrow \alpha \cdot B \beta, 0]$ é um item de I_0 . Acrescente a I_0 , para todas as produções em P da forma $B \rightarrow \gamma$, o item $[B \rightarrow \cdot \gamma, 0]$, se este item ainda não estiver presente em I_0 .

(B) Construção de I_j a partir de I_0, I_1, \dots, I_{j-1} :

Passo 4 - Para cada $[B \rightarrow \alpha \cdot a\beta, i]$ em I_{j-1} tal que $a = a_j$, acrescente $[B \rightarrow \alpha a \cdot \beta, i]$ a I_j .

Execute os passos 5 e 6 até que nenhum novo item possa ser adicionado.

Passo 5 - Seja $[A \rightarrow \alpha \cdot, i]$ um item em I_j . Examine I_i procurando itens da forma

$[B \rightarrow \alpha \cdot A\beta, k]$. Para cada item encontrado, acrescente $[B \rightarrow \alpha A \cdot \beta, k]$ a I_j .

Passo 6 - Seja $[A \rightarrow \alpha \cdot B\beta, i]$ um item em I_j . Para todo $B \rightarrow \gamma$ em P , acrescente

$[B \rightarrow \cdot \gamma, j]$ a I_j .

2.2.2.8 - Conclusões

Demonstrou-se nesta seção que as linguagens podem ser usadas para descrever padrões complexos. Logo, procedimentos que realizam a análise sintática podem ser usados para implementar o processo de reconhecimento de padrões. Deve-se notar que em muitas aplicações práticas, tanto a abordagem sintática quanto a abordagem estatística são usadas [15, 50], neste caso, a abordagem estatística é usada no reconhecimento das primitivas dos padrões. Isto é feito porque a informação estrutural das primitivas não é considerada importante e também porque as medidas locais das primitivas são sensíveis a ruídos e distorções. Já no reconhecimento de subpadrões e padrões propriamente ditos, onde a informação estrutural é abundante, a abordagem sintática é mais apropriada.

Em algumas aplicações práticas, existem certas incertezas no processo em estudo que ocorrem devido a existência de ruído e variações nas medidas dos padrões ocasionando ambiguidades nas linguagens que descrevem os padrões. Para descrever e reconhecer padrões ruidosos sob possíveis situações ambíguas, surgiram as linguagens estocásticas [51 - 57].

Uma gramática estocástica é uma tupla $G_S = (V_N, V_T, P_S, S)$ onde P_S é um conjunto finito de produções estocásticas e todos os outros símbolos são semelhantes aos definidos na

seção 2.2.2.2. Para uma gramática estocástica livre de contexto, uma produção P_S tem a forma:

$$A_i \xrightarrow{p_{ij}} \alpha_j, \quad A_i \in V_N, \quad \alpha_j \in (V_N \cup V_T)^*$$

onde p_{ij} é denominado probabilidade da produção. A probabilidade de se gerar uma cadeia x , denominada probabilidade da cadeia $p(x)$, é o produto de todas as probabilidades de produção associadas às produções usadas na geração de x . A linguagem gerada por uma gramática estocástica consiste das cadeias geradas pela gramática e as probabilidades a elas associadas.

Ao se associar às cadeias valores de probabilidades, impõem-se à linguagem de descrição uma estrutura probabilística. A distribuição de probabilidades que caracteriza os padrões em uma classe pode ser interpretada como a distribuição de probabilidades das cadeias em uma linguagem. Portanto, as regras de decisão estatística podem ser aplicadas para a classificação de um padrão sob situações ambíguas. Mais ainda, devido a existência de informações sobre as probabilidades das produções, a rapidez com que a análise sintática é feita pode ser melhorada [57,58].

Outras abordagens para o reconhecimento de padrões ruidosos ou distorcidos, usando métodos sintáticos, incluem a utilização de gramáticas transformacionais [59] e a aplicação de técnicas de correção de erro de análise [60]. Para usar as técnicas de correção de erro de análise como sistema de reconhecimento, deve-se introduzir diferentes tipos de erro na extração de características (substituição, deleção, inserção). A gramática de padrões original é modificada observando-se estes erros. O processo de reconhecimento é então construído com base no analisador desenvolvido de acordo com a gramática modificada.

Na especificação de uma primitiva, um conjunto de atributos é geralmente necessário. Uma primitiva com diferentes propriedades pode ser expressa em termos dos valores dos seus atributos. Este conjunto de atributos pode ser considerado como uma informação semântica da

primitiva. Cada atributo pode ser expresso por funções numéricas ou predicados lógicos. A informação semântica de um subpadrão é, em geral, calculada tanto a partir da informação semântica das primitivas, compostas de acordo com as relações sintáticas ou operadores e regras semânticas associadas com cada produção da gramática, como com base num conjunto separado de funções ou regras que não estão necessariamente definidas em conjunto com as produções da gramática [15]. Essa classe de gramáticas é algumas vezes chamada de classe das gramáticas de atribuição ou “gramáticas com coordenadas”.

2.2.3 - Abordagem por Redes Neurais Artificiais

2.2.3.1 - Introdução

A motivação para o trabalho desenvolvido na área de redes neurais artificiais provém da observação do modo como o cérebro humano trabalha. O seu funcionamento é totalmente diferente do funcionamento dos computadores digitais convencionais. Pode-se considerá-lo como um computador altamente complexo, não-linear e paralelo. Ele é capaz de organizar os neurônios (elementos estruturais) de forma a executar determinadas tarefas (reconhecimento de padrões, percepção, controle motor, entre outras) numa velocidade bem superior à dos supercomputadores atuais. A visão humana é um bom exemplo [61 - 63]. O sistema visual fornece uma representação do ambiente onde a pessoa se encontra assim como as informações necessárias para que esta possa interagir com este ambiente. Tarefas como estas são executadas pelo cérebro em tempos da ordem de 100 a 200 ms, enquanto tarefas menos complexas exigem dias nos atuais supercomputadores [64].

Em sua forma mais geral, uma rede neural artificial, ou simplesmente rede neural, é uma máquina desenvolvida para “modelar” o modo como o cérebro executa uma determinada

tarefa ou função de interesse. Para ter um bom desempenho, a rede neural utiliza uma interconexão maciça de elementos simples denominados “neurônios” ou “unidades de processamento”. A definição a seguir congrega estes conceitos [65]

“Uma rede neural é um processador distribuído, paralelo com interconexões maciças que possui uma propensão natural em armazenar o conhecimento adquirido através de experiências e torná-lo disponível para o uso. Ela é semelhante ao cérebro em dois aspectos:

- 1 - O conhecimento é adquirido através de um processo de aprendizagem;
- 2 - A intensidade das conexões entre os neurônios conhecida como **peso sináptico** é usada para armazenar o conhecimento.”

O procedimento usado no processo de aprendizagem é chamado de **algoritmo de aprendizagem** cuja função é modificar os pesos sinápticos da rede neural de forma organizada para se atingir o objetivo desejado.

2.2.3.2 - Características

A partir do que foi apresentado na Seção anterior, parece razoável afirmar que o poder computacional de uma rede neural emana de sua estrutura paralela, distribuída e maciçamente interconectada e de sua habilidade de aprender e generalizar; a generalização se deve ao fato da rede neural produzir respostas razoáveis para entradas que não foram apresentadas durante o treinamento. Estas são as duas características básicas de uma rede neural. A partir delas surgem outras entre as quais destacam-se:

- 1 - Não-Linearidade - um neurônio é basicamente um dispositivo não-linear. Logo, uma rede neural, construída a partir da interconexão de neurônios, é um sistema não-linear cuja não-linearidade está distribuída através da rede.

2 - Mapeamento Entrada-Saída - um paradigma de aprendizagem chamado de **aprendizagem supervisionada** baseia-se na modificação dos pesos sinápticos a partir da apresentação de um conjunto de amostras de treinamento. Cada amostra consiste de um único sinal de entrada e a correspondente resposta esperada. À rede é apresentada uma amostra, escolhida aleatoriamente dentro do conjunto, e os pesos sinápticos são modificados de forma a minimizar a diferença entre a resposta desejada e a resposta produzida pela rede, tomando-se como base um critério estatístico apropriado. O treinamento da rede é repetido para várias amostras do conjunto até que a rede atinja um estado estacionário onde não ocorra variações nos pesos sinápticos. Logo, a rede aprende através de um conjunto de amostras, construindo um mapeamento entrada-saída.

3 - Adaptação - as redes neurais possuem uma capacidade natural de adaptar os seus pesos sinápticos a mudanças no ambiente que as circunda. Por exemplo, para operar em um ambiente não-estacionário, uma rede pode ser desenvolvida de modo a permitir a alteração, em tempo real, de seus pesos sinápticos.

4 - Resposta Confiável - no contexto da classificação de padrões, uma rede neural pode ser desenvolvida para fornecer informações não somente sobre qual padrão escolher, mas também sobre a confiabilidade da decisão tomada. Esta última informação pode ser usada para rejeitar padrões ambíguos melhorando o desempenho da rede.

5 - Informação Contextual - o conhecimento é representado pela estrutura e o estado de ativação de uma rede neural. Cada neurônio, na rede, é potencialmente afetado pela atividade global de todos os outros neurônios. Logo, a informação contextual é tratada com naturalidade pela rede.

6 - Tolerância a Falhas - uma rede neural, implementada em “hardware”, tem a característica de ser tolerante a falhas no sentido de que seu desempenho sofre uma degradação suave quando opera em condições adversas [66].

7 - Analogia Neurobiológica - o desenvolvimento de uma rede neural é motivado por sua analogia com o cérebro, que é uma prova viva de que o processamento paralelo tolerante a falhas não é apenas fisicamente possível, mas também é rápido e poderoso. Os neurobiologistas consideram as redes neurais artificiais como instrumentos de pesquisa úteis na interpretação dos fenômenos neurobiológicos. Por outro lado, os engenheiros buscam na neurobiologia novas idéias para resolver problemas complexos que não são satisfatoriamente solucionados por técnicas convencionais.

2.2.3.3 - Arquiteturas de Redes

O modo como os neurônios, em uma rede neural, estão estruturados está intimamente ligado ao algoritmo de aprendizagem usado em seu treinamento. Em geral, existem quatro diferentes classes de arquiteturas.

2.2.3.3.1 - Redes “Feedforward” de uma única camada

Em sua forma mais simples, uma rede neural em camadas é constituída por uma camada de entrada que projeta os valores de entrada em uma camada de neurônios de saída, não ocorrendo o inverso. Ou seja, nesta rede a informação flui em uma única direção (“feedforward”), não existindo realimentação. A Fig. 2.16 ilustra uma rede deste tipo. Esta rede é denominada rede de uma única camada (devido à camada de neurônios de saída) pois a camada de entrada não é considerada porque nenhum processamento é realizado nela.

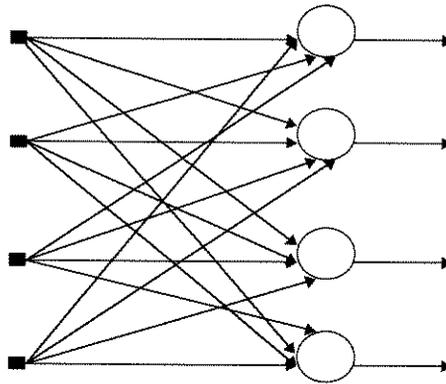


Fig. 2.16 - Rede Feedforward com uma única camada de neurônios.

2.2.3.3.2 - Redes “Feedforward” Multicamadas

A segunda classe de redes “feedforward” distingue-se pela presença de uma ou mais **camadas escondidas** cujos elementos computacionais são chamados de *neurônios escondidos* ou *unidades escondidas*. A função dos neurônios escondidos é a de intervir entre a entrada externa e a saída da rede. Adicionando uma ou mais camadas escondidas, a rede é capaz de extrair estatísticas de ordem elevada sendo particularmente interessante quando a camada de entrada é extensa.

Os elementos da camada de entrada fornecem os valores do padrão de ativação (vetor de entrada), que constituem os sinais de entrada aplicados aos neurônios da segunda camada (primeira camada escondida). Os sinais de saída da segunda camada são usados como entradas na terceira camada e assim sucessivamente até o final da rede. O conjunto dos sinais de saída dos neurônios na camada de saída (camada final) constitui a resposta da rede ao padrão de ativação fornecido pelos elementos da camada de entrada. O grafo da Fig. 2.17 ilustra a arquitetura de uma rede “feedforward” com apenas uma camada escondida.

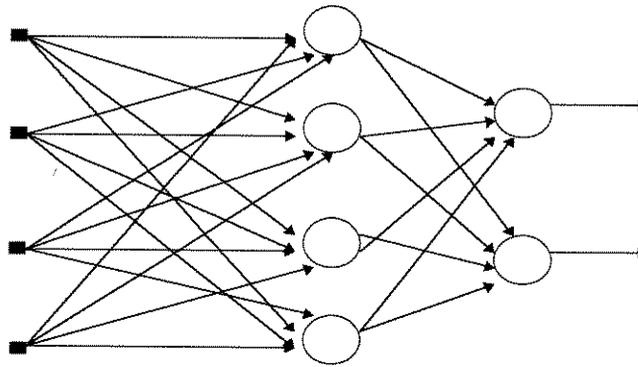


Fig. 2.17 - Rede Feedforward totalmente interconectada com uma camada de neurônios escondidos e uma camada de saída.

A rede apresentada na Fig. 2.17 é dita ser totalmente conectada no sentido de que cada elemento em cada camada está conectado a todos os outros elementos na camada seguinte. Se, por acaso, algumas conexões não existirem, a rede será dita parcialmente conectada. Uma forma de rede multicamada parcialmente conectada de certo interesse é a rede localmente conectada. Um exemplo desta rede com apenas uma camada escondida é mostrado na Fig. 2.18. Cada neurônio da camada escondida está conectado a um conjunto local de elementos de entrada; tal conjunto de elementos localizados que alimenta um neurônio constitui *o campo receptivo* deste neurônio. Da mesma forma, cada neurônio da camada de saída está conectado a um conjunto local de neurônios escondidos. Comparando as redes ilustradas pelas Figs. 2.17 e 2.18, observa-se que a rede com conexões locais (Fig. 2.18) apresenta uma estrutura especializada. Na prática, a estrutura especializada construída no desenvolvimento da rede reflete as informações obtidas a priori sobre as características do padrão de ativação a ser classificado.

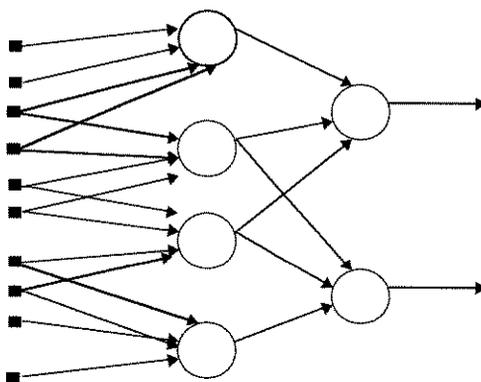


Fig. 2.18 - Rede Feedforward parcialmente interconectada.

2.2.3.3.3 - Redes Recorrentes

Uma rede neural recorrente diferencia-se de uma rede neural “feedforward” pelo fato dela possuir pelo menos uma realimentação. Como exemplo, pode-se ter uma rede recorrente com apenas uma única camada de neurônios onde cada neurônio realimenta o seu sinal de saída na entrada de todos os outros neurônios, como mostra a Fig. 2.19. A presença da realimentação tem um impacto profundo na capacidade de aprendizagem da rede assim como no seu desempenho. Mais ainda, a realimentação se utiliza de ramos particulares compostos de unidades de atraso (denotados por z^{-1}), que resulta em um comportamento dinâmico não-linear em virtude da natureza não-linear dos neurônios.

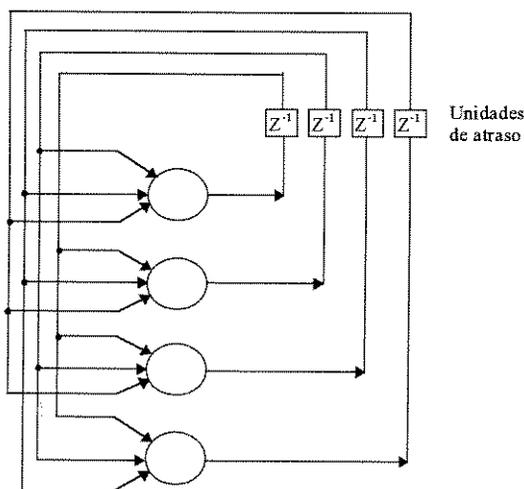


Fig. 2.19 - Rede recorrente sem auto realimentação e sem neurônios escondidos.

2.2.3.3.4 - Estruturas em “Lattice”

Um “lattice” consiste de um vetor unidimensional, bidimensional ou de dimensões mais elevadas de neurônios com um conjunto de nós de entrada correspondente que fornece os sinais de entrada para o vetor; a dimensão do “lattice” está relacionada ao número de dimensões do espaço no qual o grafo se encontra. O grafo da Fig. 2.20 apresenta um “lattice” unidimensional de 3 neurônios alimentado por 3 nós de entrada. Uma rede em “lattice” é na verdade uma rede “feedforward” com os neurônios de saída arranjados em linhas e colunas.

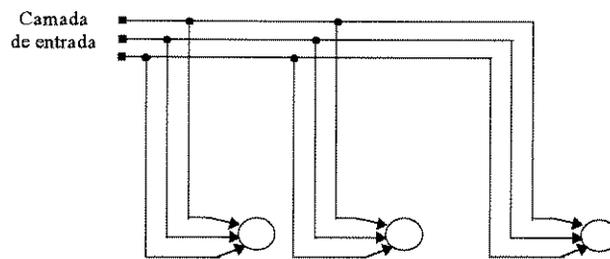


Fig. 2.20 - Lattice unidimensional de três neurônios.

2.2.3.4 - Processo de Aprendizagem

Entre as diversas propriedades de uma rede neural, a propriedade de maior relevância é a habilidade da rede em *aprender* com o ambiente onde está inserida como também em melhorar o seu desempenho através do aprendizado. A rede aprende sobre o ambiente através de um processo iterativo de ajuste dos pesos sinápticos e dos limiares. Idealmente, a rede apresenta um conhecimento maior sobre o seu ambiente após cada iteração do processo de aprendizagem.

O aprendizado, no contexto de redes neurais, pode ser definido da seguinte forma [67]

“Aprendizado é um processo pelo qual os parâmetros de uma rede neural são ajustados de acordo com um processo contínuo de estimulação desta pelo ambiente no qual se insere. O tipo de aprendizado da forma como os parâmetros são ajustados”

Exemplificando, considere os sinais x_j e v_k e o peso sináptico w_{kj} mostrados na Fig. 2.21.

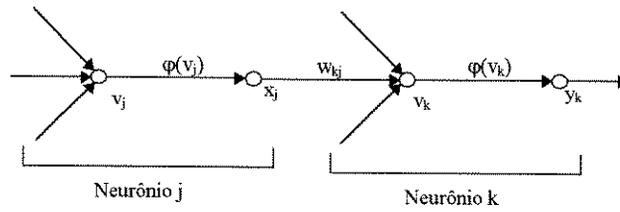


Fig. 2.21 - Grafo ilustrando um par de neurônios inserido em uma rede neural.

O sinal x_j representa a saída do neurônio j e o sinal v_k , a atividade interna do neurônio k. No contexto do peso sináptico w_{kj} , os sinais x_j e v_k são comumente chamados de atividades pré-sináptica e pós-sináptica, respectivamente. Seja $w_{kj}(n)$ o peso sináptico no instante n. Neste instante, um ajuste $\Delta w_{kj}(n)$ é aplicado ao peso sináptico $w_{kj}(n)$, fornecendo o valor atualizado $w_{kj}(n+1)$. Logo,

$$w_{kj}(n+1) = w_{kj}(n) + \Delta w_{kj}(n) \quad (2.1)$$

onde $w_{kj}(n)$ e $w_{kj}(n+1)$ podem ser vistos como os valores antigo e novo do peso sináptico w_{kj} , respectivamente. Em particular, o ajuste $\Delta w_{kj}(n)$ é calculado como resultado do estímulo do ambiente e o valor atualizado $w_{kj}(n+1)$ define a mudança na rede como resultado deste estímulo.

Um conjunto de regras bem definidas para a solução do problema de aprendizado é denominado algoritmo de aprendizagem. Na verdade, não existe um único algoritmo de aprendizagem para o desenvolvimento de redes neurais. Basicamente, estes algoritmos diferem um do outro na forma como o ajuste Δw_{kj} do peso sináptico w_{kj} é formulado. Um outro fator que deve ser considerado é o modo como a rede neural se relaciona com o ambiente. Neste último contexto, fala-se do paradigma de aprendizagem referindo-se a um modelo do ambiente

no qual a rede neural atua. A taxonomia da aprendizagem é descrita na Fig. 2.22 e seus elementos são explicados a seguir.

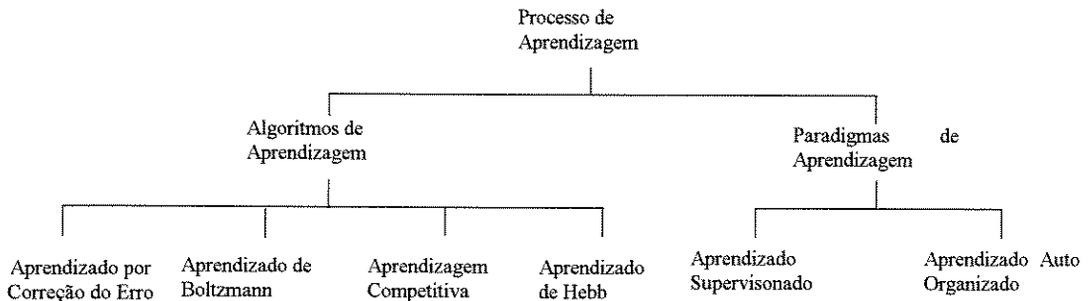


Fig. 2.22 - Taxonomia do processo de aprendizagem.

2.2.3.4.1 - *Aprendizado Baseado na Correção do Erro de Aprendizagem*

Suponha que $d_k(n)$ denote uma resposta esperada ou desejada para o neurônio k no instante n . Seja $y_k(n)$ a resposta fornecida por este neurônio. Esta resposta é produzida por um estímulo $\underline{x}(n)$ aplicado à entrada da rede da qual o neurônio k faz parte. O vetor de entrada $\underline{x}(n)$ e a resposta desejada $d_k(n)$ para o neurônio k constituem um exemplo particular apresentado à rede no instante n . Assume-se que este exemplo, assim como todos os outros apresentados, são gerados por um ambiente de natureza probabilística cuja distribuição de probabilidade é desconhecida.

Tipicamente, a resposta $y_k(n)$ fornecida pelo neurônio k é diferente da resposta desejada $d_k(n)$ gerando um sinal de erro dado pela diferença entre a resposta esperada $d_k(n)$ e a resposta obtida $y_k(n)$, como mostra,

$$e_k(n) = d_k(n) - y_k(n) \quad (2.2)$$

O propósito principal deste algoritmo de aprendizagem é minimizar a função custo associada ao sinal de erro $e_k(n)$, de tal forma que a resposta de cada neurônio de saída da rede se aproxime da resposta desejada para estes neurônios. Um critério comumente usado como

função custo é o *erro quadrático médio*, definido como o valor quadrático médio da soma dos erros quadráticos,

$$J = E \left[\frac{1}{2} \sum_k e_k^2(n) \right] \quad (2.3)$$

onde E é o operador estatístico **esperança** e a soma é calculada usando-se todos os neurônios da camada de saída da rede. A minimização da função de custo J em relação aos parâmetros da rede é conhecida como **método do gradiente descendente** [68, 85]. A desvantagem deste procedimento de otimização é que ele requer o conhecimento das características estatísticas dos processos sob análise. Para superar este problema prático, assume-se uma solução aproximada para a otimização. Especificamente, usa-se como critério o valor instantâneo da soma dos erros quadráticos,

$$E(n) = \frac{1}{2} \sum_k e_k^2(n) \quad (2.4)$$

A rede é então otimizada minimizando-se E(n) em relação aos pesos sinápticos da rede. Logo, de acordo com esta regra de aprendizagem (algumas vezes chamada de **regra delta**), o ajuste $\Delta w_{kj}(n)$ feito aos pesos sinápticos w_{kj} no instante n é dado por [69],

$$\Delta w_{kj}(n) = \eta e_k(n) x_j(n) \quad (2.5)$$

onde η é uma constante positiva que determina a taxa de aprendizagem.

Plotando-se a função custo J em relação aos pesos sinápticos que caracterizam a rede neural, obtém-se uma superfície multidimensional referida como superfície de erro. Dependendo do tipo dos neurônios usados na construção da rede, duas situações distintas podem ocorrer:

1 - A rede neural constitui-se apenas de neurônios lineares e neste caso, a superfície de erro é uma função quádrlica tendo portanto um único ponto de mínimo;

2 - A rede neural é constituída de neurônios não-lineares resultando numa superfície de erro com um mínimo global assim como vários mínimos locais.

Em ambos os casos, o objetivo deste algoritmo de aprendizagem é, partindo de um ponto arbitrário na superfície de erro, chegar a um mínimo global. No primeiro caso, este objetivo é possível de ser alcançado. No segundo, nem sempre; o algoritmo pode “cair” em um mínimo local e, portanto, não conseguirá atingir o mínimo global da superfície de erro.

2.2.3.4.2 - Aprendizagem Hebbiana

O postulado de aprendizagem de Hebb é o mais antigo e famoso de todos os algoritmos de aprendizagem,

“Quando um axônio de uma célula A está próximo o bastante para excitar uma célula B e repetidamente ou persistentemente participa de sua ativação, mudanças metabólicas ocorrem em uma ou nas duas células de tal forma que a eficiência de A, como uma das células em ativar B, é aumentada.”

Este postulado foi concebido num contexto neurobiológico. Expandindo e parafraseando-o tem-se a seguinte regra [70, 86],

1 - Se dois neurônios em ambos os lados de uma sinapse (conexão) são ativados simultaneamente, a força desta ligação sofre um aumento;

2 - Se dois neurônios em ambos os lados de uma sinapse são ativados assincronamente, a sinapse é enfraquecida ou eliminada.

Tal sinapse é chamada **sinapse Hebbiana**. Mais precisamente, uma sinapse hebbiana é uma sinapse que usa um mecanismo com dependência temporal, é altamente local e fortemente

interativo para aumentar a eficiência sináptica como uma função da correlação entre as atividades pré e pós-sináptica. Desta definição, pode-se destacar quatro propriedades que caracterizam uma sinapse hebbiana [71],

1 - Mecanismo com dependência temporal - este mecanismo se refere ao fato de que as mudanças em uma sinapse hebbiana dependem do tempo exato de ocorrência das atividades pré e pós-sináptica.

2 - Mecanismo local - por natureza, uma sinapse é um local de transmissão onde os sinais que “carregam” a informação são espaço-temporalmente contíguos. Esta informação disponível localmente é usada pela sinapse hebbiana para produzir uma modificação sináptica local que depende da entrada. É este mecanismo local que permite a uma neural, composta por sinapses hebbianas, realizar o aprendizado não-supervisionado.

3 - Mecanismo Interativo - nota-se aqui que a ocorrência de mudanças na sinapse hebbiana depende dos níveis de atividade em ambos os lados da sinapse. Esta dependência pode ser de natureza determinística ou probabilística.

4 - Mecanismo Conjuncional - uma interpretação do postulado de aprendizagem de Hebb mostra que para que ocorra uma mudança na eficiência sináptica é necessário que haja conjunção entre as atividades pré e pós-sináptica. Portanto, de acordo com esta interpretação, a co-ocorrência das atividades pré e pós-sináptica é suficiente para produzir uma modificação sináptica.

2.2.3.4.2.1 - Modelos Matemáticos

Considere novamente a situação apresentada na Fig. 2.21. Ela mostra um peso sináptico w_{kj} com atividades pré e pós-sináptica denotadas por x_j e y_k , respectivamente. De acordo com o postulado de Hebb, o ajuste do peso sináptico w_{kj} no instante n é dado por,

$$\Delta w_{kj}(n) = F(y_k(n), x_j(n)) \quad (2.6)$$

onde $F(\cdot, \cdot)$ é função das atividades pré e pós sinápticas.

Como um caso especial da Eq. 2.6, pode-se escrever,

$$\Delta w_{kj}(n) = \eta y_k(n) x_j(n) \quad (2.7)$$

onde η é uma constante positiva que determina a taxa de aprendizagem. Esta regra enfatiza claramente a natureza correlacional da sinapse hebbiana. A curva superior da Fig. 2.23 apresenta uma representação gráfica da Eq. (2.7).

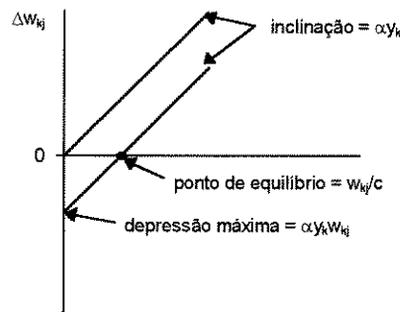


Fig. 2.23 - Ilustração da regra de Hebb.

Desta representação gráfica, nota-se que a aplicação repetida do sinal de entrada (atividade pré-sináptica) x_j leva a um crescimento exponencial que promove a saturação do peso sináptico w_{kj} . Para evitar tal situação, deve-se impor um limite ao crescimento dos peso sinápticos. Para se fazer isto, basta introduzir um fator de esquecimento não-linear na fórmula do ajuste sináptico Δw_{kj} apresentada na Eq. (2.7). Especificamente, Δw_{kj} é redefinido como [72],

$$\Delta w_{kj}(n) = \eta y_k(n) x_j(n) - \alpha y_k(n) w_{kj}(n) \quad (2.8)$$

onde α é uma nova constante positiva e $w_{kj}(n)$, o peso sináptico no instante n . Equivalentemente,

$$\Delta w_{kj}(n) = \alpha y_k(n) [cx_j(n) - w_{kj}(n)] \quad (2.9)$$

onde $c = \eta/\alpha$. Esta equação implica que para entradas onde $x_j(n) < w_{kj}(n)/c$, o peso sináptico modificado $w_{kj}(n+1)$ no instante $n+1$ diminuirá de uma quantidade proporcional à atividade pós-sináptica $y_k(n)$, como mostrado pela curva inferior da Fig. 2.23. Por outro lado, quando $x_j(n) > w_{kj}(n)/c$, o peso sináptico modificado $w_{kj}(n+1)$ sofrerá um aumento proporcional a $y_k(n)$. Nota-se, portanto, que o ponto de equilíbrio da atividade para a modificação do peso sináptico no instante $n+1$ é uma variável, igual a w_{kj}/c , proporcional ao valor w_{kj} no instante da ativação pré-sináptica.

Pode-se ainda formular o postulado de Hebb em termos estatísticos, considerando que as mudanças nos pesos sinápticos são proporcionais à covariância entre as atividades pré e pós sinápticas [73, 77]. Especificamente,

$$\begin{aligned} \Delta w_{kj}(n) &= \eta \text{cov}[y_k(n), x_j(n)] \\ &= \eta E\left[\left(y_k(n) - \bar{y}_k\right) \left(x_j(n) - \bar{x}_j\right)\right] \end{aligned} \quad (2.10)$$

onde η é uma constante de proporcionalidade que determina a taxa de aprendizagem; E é o operador estatístico esperança; \bar{x}_j e \bar{y}_k são os valores médios das atividades pré e pós sináptica, respectivamente. De acordo com a Eq. 2.10, na média, a força da sinapse deve aumentar se as atividades pré e pós sináptica estiverem correlacionadas positivamente, diminuir se elas estiverem correlacionadas negativamente, e manter-se constante se elas não estiverem correlacionadas.

2.2.3.4.3 - *Aprendizado Competitivo*

No aprendizado competitivo, os neurônios de saída de uma rede neural competem entre si para ser o escolhido e, portanto, ativado. Logo, enquanto numa rede baseada no aprendizado hebbiano vários neurônios de saída podem ser ativados simultaneamente, no caso do aprendizado competitivo apenas um único neurônio de saída será ativado em um instante de tempo.

Existem três elementos básicos em uma regra de aprendizagem competitiva [74]:

- 1 - Um conjunto de neurônios iguais a não ser por alguns pesos sinápticos distribuídos aleatoriamente fazendo com que respondam de forma diferente a um dado conjunto de padrões de entrada;
- 2 - Um limite imposto à “força” de cada neurônio;
- 3 - Um mecanismo que permita aos neurônios competirem pelo direito de responder a um dado conjunto de entradas, de forma que apenas um neurônio de saída, ou apenas um neurônio por grupo, fique ativo em um determinado instante.

Na forma mais simples de aprendizado competitivo, a rede neural é constituída de uma única camada de neurônios de saída que, por sua vez, estão totalmente conectados aos nós de entrada. A rede pode conter conexões laterais entre os neurônios, como mostra a Fig. 2.24. Nesta arquitetura de rede, as conexões laterais realizam a inibição lateral, onde cada neurônio tenta inibir o neurônio com o qual esteja lateralmente conectado. O restante das conexões são sinapses excitatórias.

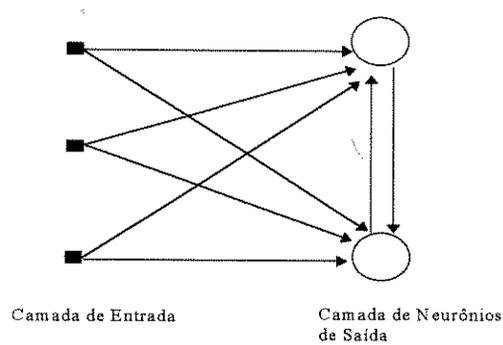


Fig. 2.24 - Arquitetura de uma rede de aprendizado competitivo com conexões excitatórias dos nós de entrada e conexões laterais inibitórias entre os neurônios.

Para que um neurônio j seja o neurônio vencedor, é necessário que o seu nível de atividade interna, v_j , para um determinado padrão de entrada x , seja o maior entre todos os neurônios na rede. O sinal de saída y_j do neurônio vencedor j é igualado a 1; os sinais de saída de todos os neurônios que perderam a competição são igualados a 0.

Seja w_{ji} o peso sináptico que conecta o nó de entrada i ao neurônio j . A cada neurônio é designado uma quantidade fixa de peso sináptico que é distribuída entre seus nós de entrada; ou seja,

$$\sum_i w_{ji} = 1 \quad \text{para todo } j \quad (2.11)$$

Um neurônio aprende deslocando os pesos sinápticos dos seus nós de entrada inativos para os ativos. Se um neurônio não responde a um determinado padrão de entrada, nenhum aprendizado é realizado neste neurônio. Se um determinado neurônio ganha a competição, então cada nó de entrada deste neurônio cede alguma porção de seu peso sináptico e o peso sináptico cedido é então distribuído igualmente entre os nós de entrada ativados. De acordo com a regra de aprendizagem competitiva padrão, a variação Δw_{ji} aplicada ao peso sináptico w_{ji} é definida por,

$$\Delta w_{ji} = \begin{cases} \eta(x_i - w_{ji}) & \text{se o neurônio } j \text{ ganhar a competição} \\ 0 & \text{se o neurônio } j \text{ perder a competição} \end{cases}$$

onde η é a taxa de aprendizagem. O efeito geral desta regra é mover o vetor de pesos sinápticos w_j do neurônio vencedor j em direção ao padrão de entrada x .

2.2.3.4.4 - *Aprendizado de Boltzmann*

A regra de aprendizado de Boltzmann é um algoritmo de aprendizagem estocástico derivado de considerações da Teoria da Informação e da Termodinâmica [75, 76].

Em uma máquina de Boltzmann, os neurônios constituem uma estrutura recorrente e operam em um modo binário. A máquina é caracterizada por uma função de energia E cujo valor é determinado pelos estados de cada neurônio da máquina,

$$E = \frac{1}{2} \sum_i \sum_j w_{ji} s_j s_i \quad i \neq j \quad (2.12)$$

onde s_i é o estado do neurônio i e w_{ji} é o peso sináptico que conecta o neurônio i ao neurônio j . O fato de i ser diferente de j apenas indica que nenhum dos neurônios na máquina possui auto-realimentação. A máquina opera escolhendo um neurônio aleatoriamente em algum passo do processo de aprendizagem e troca o estado do neurônio j de s_j para $-s_j$ em alguma temperatura T com probabilidade

$$W(s_j \rightarrow -s_j) = \frac{1}{1 + \exp(-\Delta E_j / T)} \quad (2.13)$$

onde ΔE_j é a variação da energia resultante de tal troca. A temperatura T não é uma temperatura física e sim uma pseudotemperatura. Se esta regra for aplicada repetidamente, a máquina atingirá um estado de equilíbrio termal.

Os neurônios de uma máquina de Boltzmann podem ser divididos em dois grupos funcionais: visíveis e escondidos. Os neurônios visíveis são responsáveis pela interface entre a

rede e o ambiente no qual ela opera, enquanto os neurônios escondidos sempre operam livremente. Existem dois modos de operação destes neurônios:

- . Condição “Amarrada” - na qual os neurônios visíveis estão todos “presos” a estados específicos determinados pelo ambiente.
- . Condição de Execução Livre - na qual todos os neurônios (visíveis e escondidos) operam livremente.

Seja ρ_{ji}^+ a correlação entre os estados dos neurônios i e j , condicionada ao fato da rede operar no estado amarrado. Seja ρ_{ji}^- a correlação não-condicional entre os estados dos neurônios i e j (a rede opera na condição de execução livre). As correlações ρ_{ji}^+ e ρ_{ji}^- são definidas como:

$$\rho_{ji}^+ = \sum_{\alpha} \sum_{\beta} P_{\alpha\beta}^+ s_{j/\alpha\beta} s_{i/\alpha\beta} \quad (2.14)$$

$$\rho_{ji}^- = \sum_{\alpha} \sum_{\beta} P_{\alpha\beta}^- s_{j/\alpha\beta} s_{i/\alpha\beta} \quad (2.15)$$

onde o estado $s_{i/\alpha\beta}$ denota o estado do neurônio i dado que os neurônios visíveis da máquina estejam no estado α e os neurônios escondidos, no estado β . O fator $P_{\alpha\beta}^+$ é a probabilidade condicional que os neurônios visíveis estejam no estado α e que, concomitantemente, os neurônios escondidos estejam no estado β , dado que as máquinas estejam operando na condição “amarrada”; e $P_{\alpha\beta}^-$ é a probabilidade condicional de que os neurônios visíveis estejam no estado α e os neurônios escondidos no estado β , dado que as máquinas estejam operando na condição de execução livre. Portanto, de acordo com a regra de aprendizagem de Boltzmann, a variação Δw_{ji} do peso sináptico w_{ji} que conecta o neurônio i ao neurônio j é definida por [75]

$$\Delta w_{ji} = \eta (\rho_{ji}^+ - \rho_{ji}^-) \quad i \neq j \quad (2.16)$$

onde η é a taxa de aprendizagem.

Uma característica própria do aprendizado de Boltzmann é que ele usa apenas observações disponíveis localmente sob dois modos de operação: “amarrada” e execução livre.

2.3.4.5 - *Aprendizado Supervisionado*

Um ingrediente essencial do aprendizado supervisionado é a disponibilidade de um “professor” externo como é mostrado na Fig. 2.25.

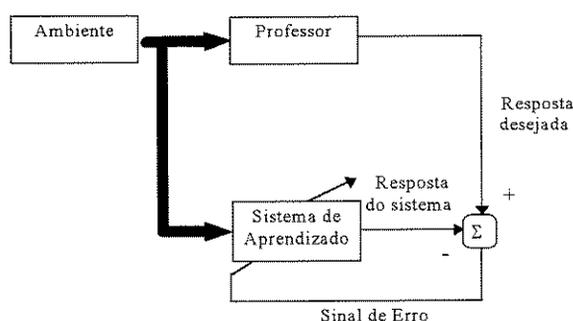


Fig. 2.25 - Diagrama de Blocos do Aprendizado Supervisionado.

Em termos conceituais, o professor pode ser encarado como um elemento do sistema que tem conhecimento do ambiente representado por um conjunto de pares entrada-saída. Entretanto, o ambiente é desconhecido para a rede neural em questão. Suponha que o professor e a rede sejam expostos a um vetor de treinamento com dados do ambiente. Devido ao seu conhecimento intrínseco, o professor é capaz de fornecer à rede a resposta desejada para este vetor de treinamento. Na verdade, a resposta desejada representa a ação ótima a ser realizada pela rede. Os parâmetros da rede são ajustados sob a influência conjunta do vetor de treinamento e do sinal de erro; o sinal de erro é definido como a diferença entre a resposta fornecida pela rede e a resposta desejada. O ajuste é feito iterativamente, passo a passo, com o objetivo de fazer com que a rede eventualmente emule o professor; presume-se que a emulação seja ótima de acordo com algum critério estatístico. Em outras palavras, o conhecimento do

ambiente disponível ao professor é transferido ao máximo para a rede. Quando se atinge esta condição, o professor é dispensado e a rede passa a interagir sozinha com o ambiente.

A forma do aprendizado supervisionado apresentada acima é, na verdade, o aprendizado por correção de erro discutido anteriormente na Seção 2.2.3.4.1. Como uma medida do desempenho do sistema, pode-se pensar na medida do erro quadrático médio definido em função dos parâmetros livres do sistema. Esta função pode ser visualizada como uma superfície de erro multidimensional tendo os parâmetros livres como coordenadas. Qualquer operação do sistema sob a supervisão do professor é representada como um ponto na superfície de erro. Para que o sistema possa melhorar seu desempenho com o tempo, é necessário que o ponto de operação caminhe sucessivamente em direção a um ponto de mínimo da superfície de erro; este ponto de mínimo pode ser local ou global. Um sistema com treinamento supervisionado é capaz de realizar isto devido a algumas informações úteis que ele possui sobre o gradiente da superfície de erro correspondente ao comportamento atual do sistema. O gradiente de uma superfície de erro em qualquer ponto é um vetor que aponta na direção decrescente. De fato, no caso do aprendizado supervisionado, o sistema usa uma estimativa instantânea do vetor gradiente que resulta em um movimento de “caminhada aleatória” do ponto de operação na superfície de erro. Porém, dado um algoritmo desenvolvido com o objetivo de minimizar a função custo, um conjunto adequado de pares entrada-saída e um tempo suficiente para se realizar o treinamento, um sistema com aprendizado supervisionado realiza satisfatoriamente tarefas como a classificação de padrões e aproximação de funções.

Os exemplos de algoritmos de aprendizado supervisionado incluem o algoritmo LMS (Least Mean Square) [69] e sua generalização conhecida como algoritmo Backpropagation (BP) [78]. O algoritmo LMS envolve apenas um neurônio, enquanto o algoritmo

Backpropagation várias camadas de neurônios interconectados. Naturalmente, o algoritmo Backpropagation é mais poderoso que o algoritmo LMS, incluindo-o como um caso especial.

O aprendizado supervisionado pode ser realizado de um modo “on-line” ou “off-line”. No caso “off-line”, um sistema computacional separado é usado para desenvolver o sistema de aprendizado supervisionado. Quando o desempenho desejado é obtido, a estrutura da rede é “congelada”, significando que a rede opera em um modo estático. Por outro lado, no aprendizado “on-line” o processo de aprendizagem é implementado junto ao sistema, ou seja, o aprendizado é feito em tempo real resultando em uma rede neural dinâmica.

A grande desvantagem do aprendizado supervisionado, tanto “on-line” como “off-line”, é que, sem um professor, a rede neural não consegue aprender novas estratégias para situações particulares que não são abordadas pelo conjunto de exemplos usados no treinamento.

2.2.3.4.6 - Aprendizado Não-Supervisionado

No aprendizado não supervisionado ou auto organizável não existe nenhum professor externo para supervisionar o processo de aprendizagem, como mostra a Fig. 2.26.

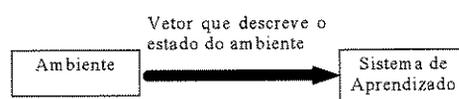


Fig. 2.26 - Diagrama de Blocos do Aprendizado Não Supervisionado.

O objetivo de um algoritmo de aprendizagem não supervisionada é descobrir características ou padrões significativos nos dados de entrada, e deve fazê-lo sem o auxílio de um professor. Para tanto, é fornecido ao algoritmo um conjunto de regras de natureza local permitindo que ele aprenda o mapeamento entrada-saída com algumas propriedades específicas desejáveis. Ao aprender estas características, a rede sintoniza-se à regularidades estatísticas dos dados de

entrada desenvolvendo representações internas para as características da entrada e criando novas classes automaticamente [79].

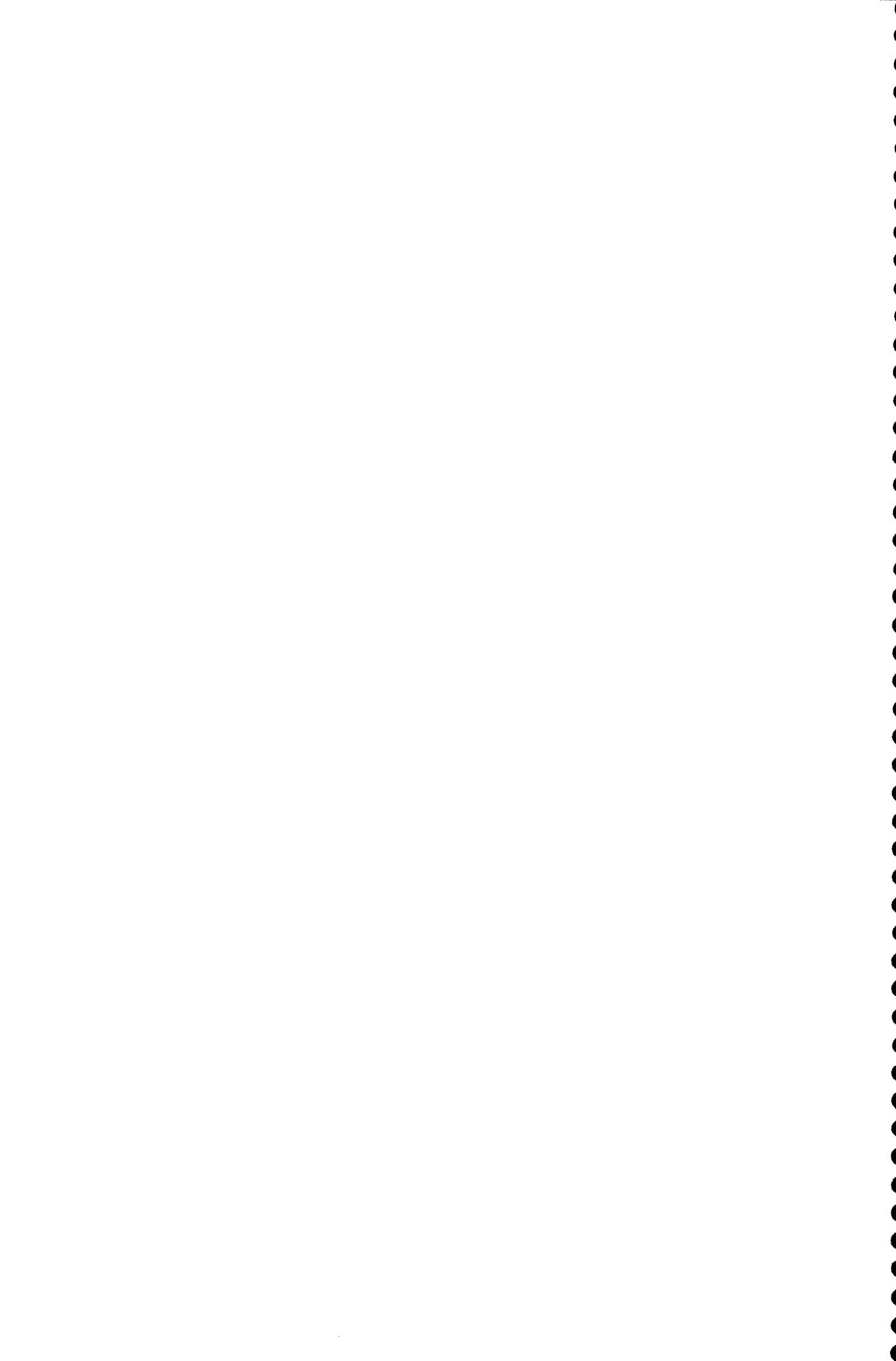
Para que o aprendizado não supervisionado possa ser feito, deve-se utilizar uma regra de aprendizagem competitiva. Como exemplo, pode-se usar uma rede neural consistindo de duas camadas, uma camada de entrada e uma camada de neurônios competitivos. A camada de entrada recebe os dados disponíveis. Na camada de neurônios competitivos, estes competem entre si pela “oportunidade” de responder às características presentes nos dados de entrada. Na forma mais simples, a rede opera de acordo com a estratégia “winner-takes-all”. Como descrito na Seção 2.2.3.4.3, nesta estratégia o neurônio com um conjunto maior de entradas “ganha” a competição e é acionado; todos os outros neurônios são “desligados”.

2.2.3.4.6.1 - Aprendizado Supervisionado versus Aprendizado Auto-Organizável

Entre os algoritmos usados no aprendizado supervisionado, o algoritmo Backpropagation emergiu como o mais usado e bem sucedido para o treinamento de redes “feedforward” com múltiplas camadas. Existem duas fases distintas na operação do algoritmo BackPropagation: a fase “forward” e a fase “backward”. Na fase “forward”, os sinais de entrada se propagam pela rede, camada por camada, produzindo, eventualmente, alguma resposta na saída da rede. A resposta produzida é então comparada com a resposta desejada, gerando sinais de erro que são propagados em direção contrária à do fluxo inicial. Na fase “backward”, os parâmetros da rede são ajustados de forma a minimizar a soma dos erros quadráticos. O algoritmo BackPropagation foi aplicado com sucesso na solução de alguns problemas difíceis tais como: reconhecimento de sinais de fala [80], reconhecimento de dígitos manuscritos [81] e controle adaptativo [82]. Infelizmente, o algoritmo BackPropagation assim como outros de aprendizagem supervisionada podem ser limitados pelo seu comportamento em escala. Considere, por exemplo, uma rede multicamadas “feedforward” com L camadas. O efeito de

um peso sináptico na primeira camada sobre a saída da rede depende de suas interações com aproximadamente L outros F_i^L pesos sinápticos, onde F_i é definido como a média das ligações de entrada dos neurônios na rede. Logo, à medida que o tamanho da rede cresce, ela se torna, computacionalmente, mais intensa fazendo com que o tempo gasto em seu treinamento cresça exponencialmente tornando-o, assim, mais lento.

Uma solução possível para o problema de escala descrito acima, é usar um procedimento de aprendizado auto-organizável. Em particular, se for possível aplicar um processo não supervisionado de modo sequencial, o treinamento de redes deste tipo torna-se factível. Mais ainda, com a habilidade das redes auto-organizáveis em formar representações internas que modelam a estrutura básica dos dados de entrada de forma mais explícita e simples, espera-se que a versão transformada da entrada sensorial possa ser mais fácil de ser interpretada de forma que as respostas corretas possam ser associadas mais rapidamente às representações internas que a rede faz do ambiente [79]. Em outras palavras, o uso híbrido de procedimentos de aprendizado supervisionado e não supervisionado pode fornecer uma solução mais aceitável que as soluções apresentadas unicamente pelo aprendizado supervisionado, principalmente se a dimensão do problema for elevada [83, 84, 87].



Capítulo 3 - Decomposição em Sub-bandas

3.1 - Introdução

Na decomposição em sub-bandas, a imagem é inicialmente filtrada criando um conjunto de imagens cada qual contendo uma faixa limitada de freqüências espaciais. Essas imagens são denominadas *sub-bandas*. Como cada sub-banda possui uma largura de faixa reduzida em comparação com a largura de faixa da imagem original, faz-se necessário a *subamostragem* destas sub-bandas. Este processo de filtragem e subamostragem é denominado *fase de análise*. A reconstrução é obtida através da *superamostragem* das sub-bandas, do uso de uma filtragem apropriada e da soma das sub-bandas reconstruídas. Esta fase é denominada *fase de síntese*. A Fig. 3.1 ilustra estes conceitos .

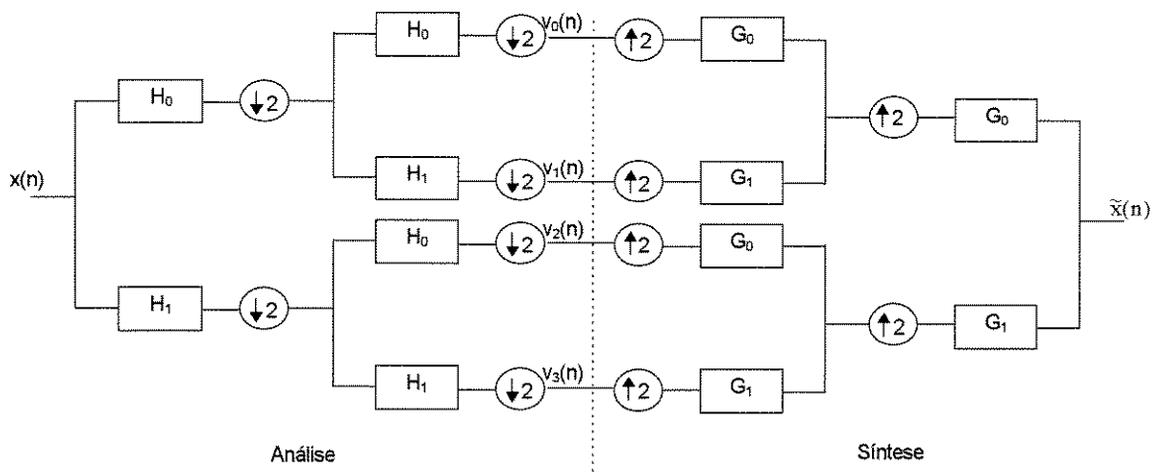


Fig. 3.1 - Representação esquemática das fases de análise e síntese para a decomposição em quatro sub-bandas.

3.2 - Processamento de Sinais com Múltiplas Freqüências

Em um sistema com múltiplas freqüências, as amostras do sinal são processadas e manipuladas em diferentes freqüências em vários pontos na configuração. Usualmente, um sinal analógico de banda limitada é amostrado segundo a freqüência de Nyquist gerando o sinal $\{x(n)\}$, com conteúdo espectral variando de zero à metade da freqüência de amostragem. Estas

amostras do sinal podem ser manipuladas tanto em frequências altas como em baixas através da interpolação e dizimação, respectivamente. O sinal deve ser adequadamente condicionado por filtros antes ou após a alteração da frequência de amostragem.

3.2.1 - Dizimação e Interpolação

Os operadores de dizimação e interpolação são mostrados nas Figs. 3.2 e 3.3 respectivamente. A dizimação consiste em reduzir a frequência de amostragem de um sinal por um fator inteiro M . Este processo é feito passando-se o sinal $\mathbf{x(n)}$ através de um filtro $\mathbf{h(n)}$ “anti-alias” (tipicamente passa-baixa) e subamostrando o sinal filtrado, como ilustrado na Fig. 3.2(a).

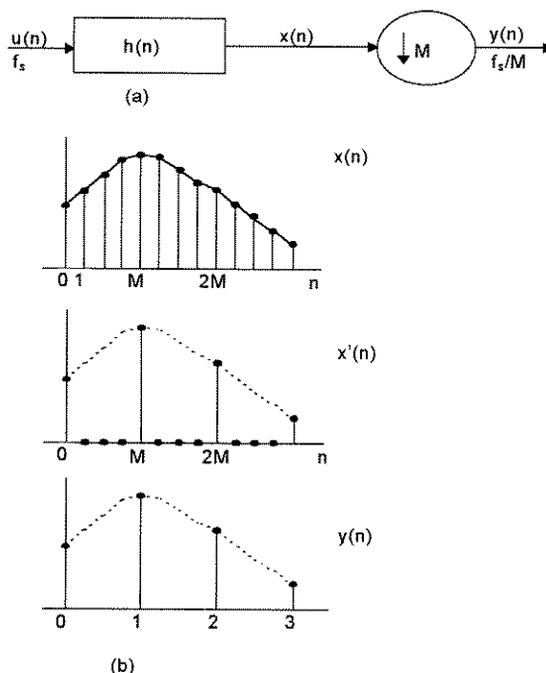


Fig. 3.2 - Operação de dizimação: (a) filtro passa baixas e subamostrador; (b) seqüências típicas.

O processo de subamostragem consiste em reter toda amostra múltipla de M do sinal $\mathbf{x(n)}$ e reindexar o eixo de índices como mostrado na Fig. 3.2(b). A Fig. 3.2(b) mostra ainda um sinal intermediário $\mathbf{x'(n)}$, do qual o sinal subamostrado $\mathbf{y(n)}$ é obtido.

$$x'(n) = \begin{cases} x(n) & n = 0, \pm M, \pm 2M, \dots \\ 0 & \text{caso contrario} \end{cases} \quad (3.1)$$

$$y(n) = x'(Mn) = x(Mn) \quad (3.2)$$

O sinal intermediário $x'(n)$ operando na mesma frequência que $x(n)$ pode ser expresso como o produto de $x(n)$ com uma função de amostragem, o trem de impulsos periódico $i(n)$:

$$x'(n) = i(n) \cdot x(n) = \left[\sum_{r=-\infty}^{\infty} \delta(n - rM) \right] \cdot x(n) \quad (3.3)$$

Mas, expandindo $i(n)$ em série discreta de Fourier [88], tem-se:

$$i(n) = \sum_r \delta(n - rM) = \frac{1}{M} \sum_{k=0}^{M-1} e^{j \frac{2\pi}{M} nk} \quad (3.4)$$

Portanto,

$$x'(n) = \frac{1}{M} \sum_{k=0}^{M-1} x(n) \cdot e^{j \frac{2\pi}{M} nk}$$

Logo a transformada Z será dada por:

$$X'(z) = \frac{1}{M} \sum_k Z \left\{ x(n) \left(e^{j \frac{2\pi}{M} k} \right)^n \right\} = \frac{1}{M} \sum_{k=0}^{M-1} x \left(z \cdot e^{-j \frac{2\pi}{M} k} \right).$$

Usando $w = e^{-j2\pi/M}$, tem-se:

$$X'(z) = \frac{1}{M} \sum_{k=0}^{M-1} X(z \cdot w^k). \quad (3.5)$$

No círculo unitário $z = e^{j\omega}$, a resposta em frequência será dada por:

$$X'(e^{j\omega}) = \frac{1}{M} \sum_{k=0}^{M-1} X \left(e^{j \left(\omega - \frac{2\pi}{M} k \right)} \right). \quad (3.6)$$

Esta última equação mostra que a transformada discreta de Fourier de $x'(n)$ é dada pela soma de M réplicas da resposta em frequência do sinal original espaçadas de $2\pi/M$.

Posteriormente, o eixo dos tempos é reindexado através da eq. (3.2) comprimindo a escala de tempo de M . Logo segue-se que:

$$Y(z) = \sum_{n=-\infty}^{\infty} X'(Mn)z^{-n} = \sum_{k=-\infty}^{\infty} X'(k) \left(z \frac{1}{M} \right)^{-k}$$

ou

$$Y(z) = X' \left(z \frac{1}{M} \right) \quad (3.7)$$

e

$$Y(e^{j\omega}) = X' \left(e^{j \frac{\omega}{M}} \right). \quad (3.8)$$

Usando a eq.(3.5), a transformada do subamostrador M será dada por:

$$Y(z) = \frac{1}{M} \sum_{k=0}^{M-1} X \left(z^{1/M} \cdot W^k \right) \quad (3.9)$$

$$Y(e^{j\omega}) = \frac{1}{M} \sum_{k=0}^{M-1} X \left(e^{j \left(\frac{\omega - 2\pi k}{M} \right)} \right). \quad (3.10)$$

Logo, a compressão do eixo dos tempos implícita na Eq. (3.2) é acompanhada por uma expansão no domínio da frequência de forma que o intervalo $[0, \pi/M]$ cobre agora a banda de 0 a π . É evidente que ao descartar algumas amostras, este processo pode levar a uma perda de informação. No domínio da frequência isto gera o “aliasing” como indica a Eq. (3.6). Para evitar isto, a largura da banda do sinal original deve ser reduzida a $\pm\pi/M$ antes de ser feita a subamostragem pelo fator M. Esta é a função do filtro “antialiasing” $h(n)$.

A interpolação consiste em aumentar a frequência de amostragem de um sinal por um fator inteiro M. Como mostrado na Fig. 3.3(a), este processo é feito combinando-se um superamostrador e um filtro passa baixa $g(n)$.

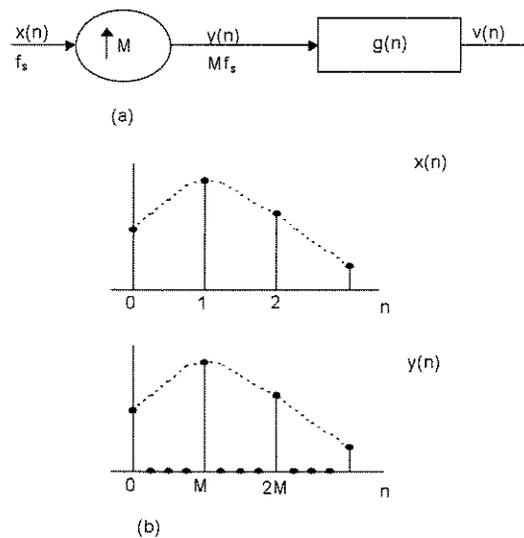


Fig. 3.3 - (a) Operação de superamostragem; (b) formas de onda de entrada e saída para $M = 4$.

O superamostrador é definido por:

$$y(n) = \begin{cases} x(n/M) & n = 0, \pm M, \pm 2M, \dots \\ 0 & \text{caso contrario} \end{cases} \quad (3.11)$$

Este operador insere $(M-1)$ zeros entre os valores das amostras e reindexa a escala de tempo como mostra a Fig. 3.3(b). A frequência efetivamente aumenta de um fator M.

A superamostragem tem dois efeitos. Primeiro, a expansão do eixo dos tempos induz a compressão em frequência; segundo, ao “forçar” a presença de zeros entre as amostras de $x(n)$, gera-se sinais de altas frequências. Estes efeitos são demonstrados no domínio transformado por:

$$Y(z) = \sum_{n=-\infty}^{\infty} y(n) \cdot z^{-n} = \sum_{n=-\infty}^{\infty} X\left(\frac{n}{M}\right) \cdot z^{-n} = \sum_{k=-\infty}^{\infty} X(k) \cdot (z^M)^{-k}$$

ou

$$Y(z) = X(z^M), \quad Y(e^{j\omega}) = X(e^{j\omega M}). \quad (3.12)$$

A Fig. 3.4 ilustra a compressão em frequência para $M = 4$. Observa-se que o eixo de frequência de 0 a 2π teve sua escala alterada para 0 a $2\pi/M$ e foi periodicamente repetido. A finalidade do filtro passa baixa $g(n)$ é suavizar o sinal superamostrado.

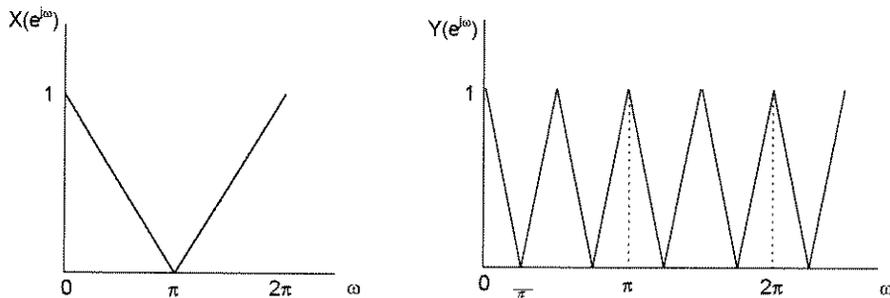


Fig. 3.4 - Compressão do eixo de frequências devido a superamostragem para $M = 4$.

As representações no domínio do tempo do dizimador e do interpolador das Figs. 3.2 e 3.3 são dadas, respectivamente, por:

$$y(n) = \sum_k h(Mn - k) \cdot u(k)$$

$$v(n) = \sum_k g(n - Mk) \cdot x(k) \quad (3.13)$$

Considere agora o que acontece quando se posiciona o subamostrador e o superamostrador como mostrado na Fig. 3.5. A saída do interpolador $v(n)$ é reconhecida como o sinal intermediário $x'(n)$ na Fig. (3.2). Logo $V(z)$ reduz-se, neste caso, a $X'(z)$ nas Eqs. (3.5) e (3.6). Algumas representações do espectro destes sinais são mostradas na Fig. 3.6 . O espectro do sinal de entrada na Fig. 3.6(a) possui uma largura de faixa maior que π/M para $M = 4$, logo deve ocorrer “aliasing”. Três dos quatro termos na Eq. (3.10) são mostrados em (b), (c) e (d); quando todos os quatro são somados, obtém-se o espectro dizimado em (e), que apresenta “aliasing” devido a sobreposição das faixas de frequência. A superamostragem comprime o eixo de frequência, como na Fig. (3.6)(f) e induz as imagens. Portanto, o espectro do sinal após a subamostragem e posterior superamostragem exibe tanto o “aliasing” (a faixa de frequência original é muito grande para o parâmetro de dizimação usado) quanto as imagens que resultam como consequência da superamostragem.

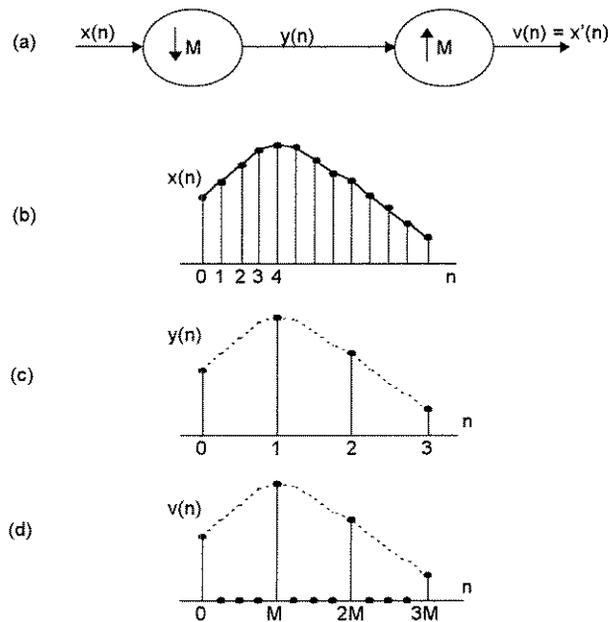
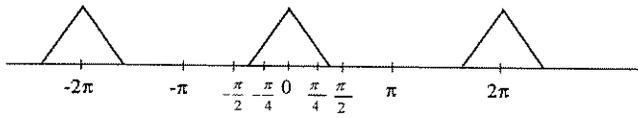


Fig. 3.5 - Sinais típicos em uma composição subamostrador-superamostrador para $M = 4$.

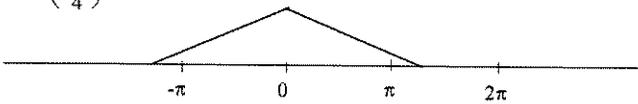
$$X(e^{j\omega}) = F(\omega)$$

(a)



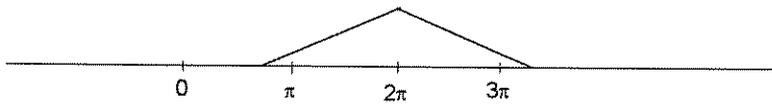
(b)

$$F\left(\frac{\omega}{4}\right)$$



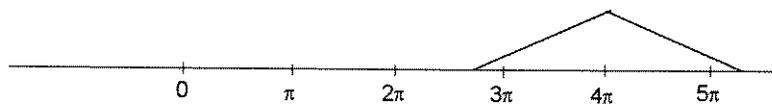
(c)

$$F\left(\frac{\omega - 2\pi}{4}\right)$$



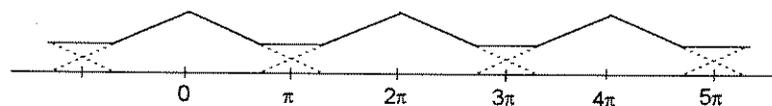
(d)

$$F\left(\frac{\omega - 4\pi}{4}\right)$$



(e)

$$Y(e^{j\omega}) = \sum_{k=0}^3 F\left(\frac{\omega - 2k\pi}{4}\right)$$



(f)

$$V(e^{j\omega}) = Y\left(e^{j\frac{\omega}{4}}\right)$$

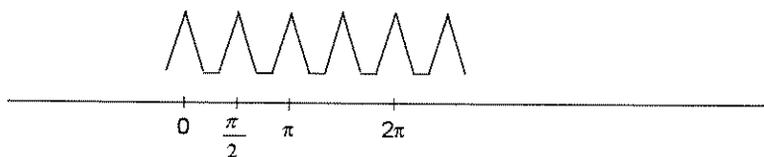


Fig. 3.6 - Espectros dos sinais mostrados na Fig. 3.5.

3.2.2 - Decomposição Polifásica

Para prevenir ou mesmo reduzir o “aliasing” devido ao subamostrador, um filtro “antialiasing” - tipicamente passa-baixa - é geralmente colocado à frente do subamostrador como na Fig. 3.8(a). Esta combinação pode ser representada por uma decomposição polifásica mostrada na Fig. 3.8(b) dada por:

$$H(z) = \sum_{k=0}^{M-1} z^{-k} G_k(z^M) \quad (3.14)$$

onde

$$G_k(z) = h(k) + h(k+M)z^{-1} + h(k+2M)z^{-2} + \dots \quad (3.15)$$

A resposta impulsiva do k-ésimo filtro polifásico é simplesmente a subamostragem de $h(n+k)$,

$$g_k(n) = h(Mn+k) \quad k = 0, 1, \dots, M-1. \quad (3.16)$$

Expandindo $H(z)$ e agrupando os termos:

$$\begin{aligned} H(z) &= h_0 + h_1 z^{-1} + h_2 z^{-2} + \dots \\ &= (h_0 + h_M z^{-M} + h_{2M} z^{-2M} + \dots) + (h_1 z^{-1} + h_{M+1} z^{-(M+1)} + \dots) + \dots \\ &\quad + (h_{M-1} z^{-(M-1)} + h_{2M-1} z^{-(2M-1)} + \dots) \end{aligned}$$

Através desta expansão, nota-se que $H(z)$ pode ser escrito como:

$$H(z) = G_0(z^M) + z^{-1} G_1(z^M) + \dots + z^{-(M-1)} G_{M-1}(z^M)$$

com $G_k(z)$ dado pela Eq. 3.15. Com outras palavras pode-se dizer que: $g_k(n)$ é apenas uma subamostragem de $h(n)$ deslocado por k .

Similarmente, pode-se mostrar que a decomposição polifásica do superamostrador é a ilustrada na Fig. 3.9. Como primeiro passo, substitui-se $H(z)$, na Fig. 3.9(a), por um banco de filtros polifásicos do ponto 1 ao ponto 2 da Fig. 3.9(c). Deslocando o superamostrador para a direita e usando a equivalência sugerida pela Fig. 3.7(b) tem-se a estrutura mostrada na Fig. 3.9(b).

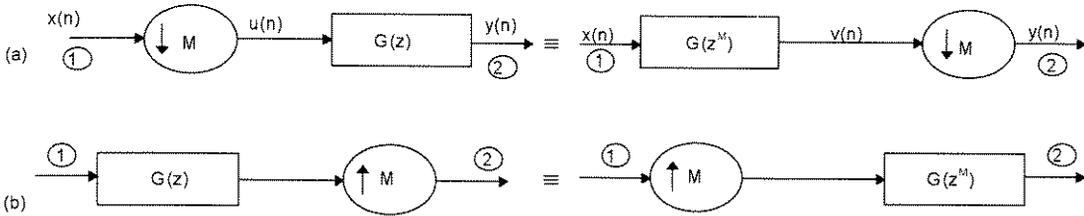


Fig. 3.7 - Estruturas equivalentes.

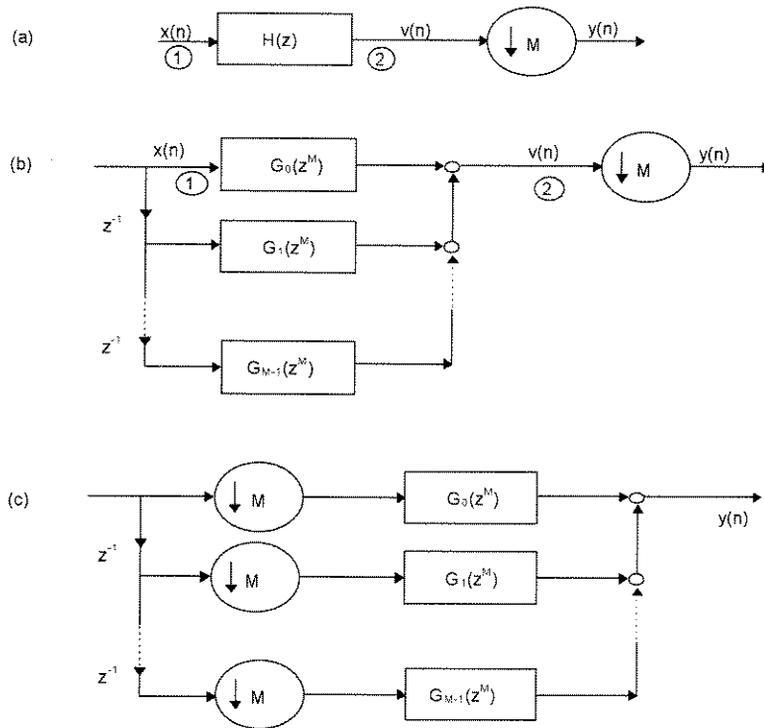


Fig. 3.8 - (a) Filtro seguido por um subamostrador; (b) decomposição polifásica; (c) representação alternativa de uma rede polifásica.

A Eq. 3.15 representa os componentes polifásicos $G_k(z)$ em termos das amostras dizimadas $\{h(k+M)\}$. Para isto possa ser representado no domínio transformado, deve-se notar que $g_r(n)$ é uma dizimação de $h(n+r)$ de M vezes. Portanto:

$$h(n+r) \leftrightarrow z^r H(z) \triangleq F_r(z)$$

e usando a Eq. 3.9 para a transformada de um sinal dizimado, obtém-se:

$$\begin{aligned} G_r(z) &= \frac{1}{M} \sum_{k=0}^{M-1} F_r(z^{1/M} W^k) \\ &= \frac{z^{r/M}}{M} \sum_{k=0}^{M-1} W^{kr} H(z^{1/M} W^k) \end{aligned} \quad (3.17)$$

e, no círculo unitário

$$G_r(e^{j\omega}) = \frac{e^{jr\omega/M}}{M} \sum_{k=0}^{M-1} W^{kr} H(e^{j(\omega-2k\pi)/M}) \quad (3.18)$$

As equações 3.14 a 3.18 definem completamente a análise polifásica e as equações de síntese.

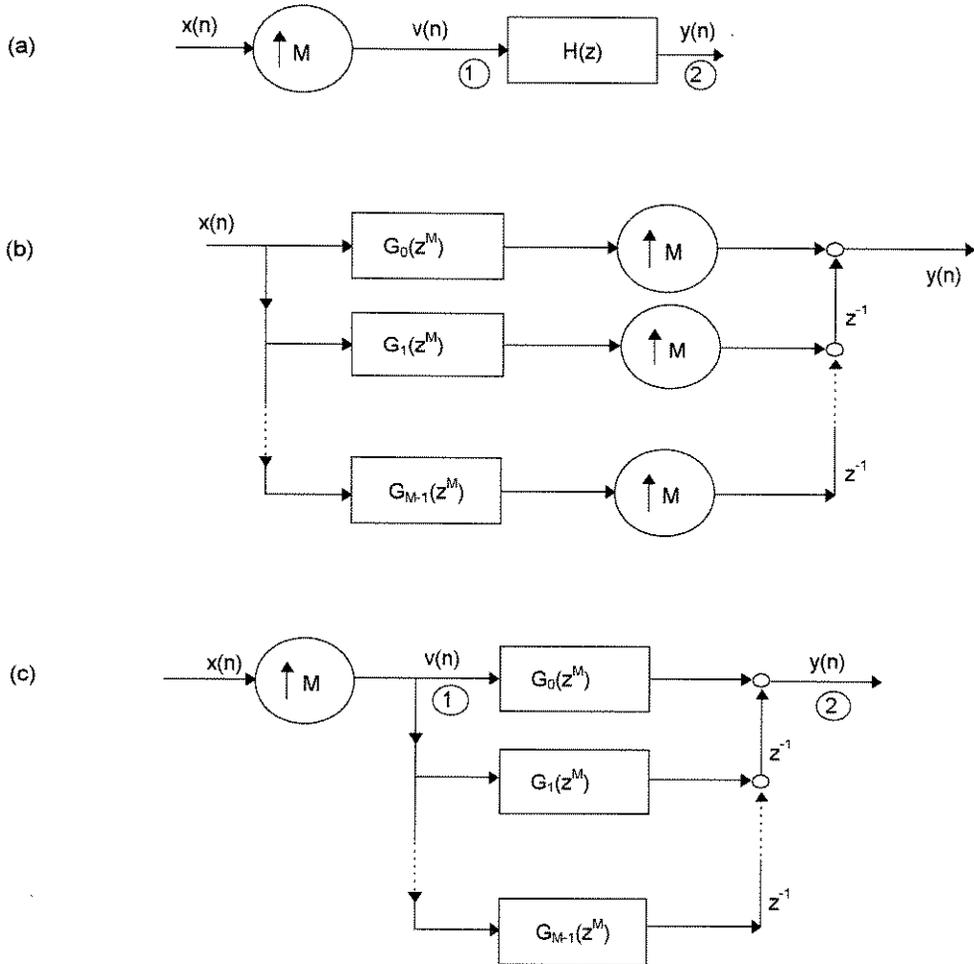


Fig. 3.9 - Decomposição polifásica do interpolador.

3.3 - Sinais de Banda Passante , Sinais Modulados e Filtros Imagem

Nesta seção serão apresentados os sinais de banda passante assim como os meios de se obter a modulação e demodulação através da dizimação e interpolação. O primeiro caso a ser considerado é o da amostragem de banda inteira. Posteriormente, será apresentada a modulação em quadratura de sinais de banda passante. Finalmente, na terceira parte desta Seção, será apresentada a definição de um filtro imagem assim como suas principais propriedades.

3.3.1 - Amostragem de Banda Inteira

Seja um sinal com espectro $X(e^{j\omega})$ como o mostrado na Fig. 3.10-b. Divida a banda de frequência de 0 a π em M bandas iguais ($M = 4$). Como mostrado na Fig. 3.10-a, o sinal é apresentado a um filtro passa-faixa ideal H_2 isolando a banda 2. Esta banda é então subamostrada por um fator $M = 4$ “esticando-se” para ocupar toda a faixa de frequência $[0, \pi]$

No processo de reconstrução da informação na banda 2 ocorre exatamente o inverso.

O sinal de baixa frequência $Y_2(e^{j\omega})$ é superamostrado por um fator $M = 4$ resultando em $Y_2(e^{j\omega}) = Y_2(e^{jM\omega})$ como mostrado na Fig. 3.10-a. Nota-se que uma réplica da banda 2 ocupa a banda original $[\pi/2, 3\pi/4]$. O segundo filtro passa-faixa remove as imagens e retém somente a banda 2 em sua localização em frequência original. O sinal no ponto 5 é, portanto, igual ao sinal no ponto 2.

Um banco de filtros passa-faixa paralelos pode ser usado para separar as M bandas. A subamostragem faz a heterodinagem destas na região de baixas frequências. Estes sinais são quantizados, transmitidos e reconstruídos em um receptor por uma superamostragem e filtragem passa-faixa.

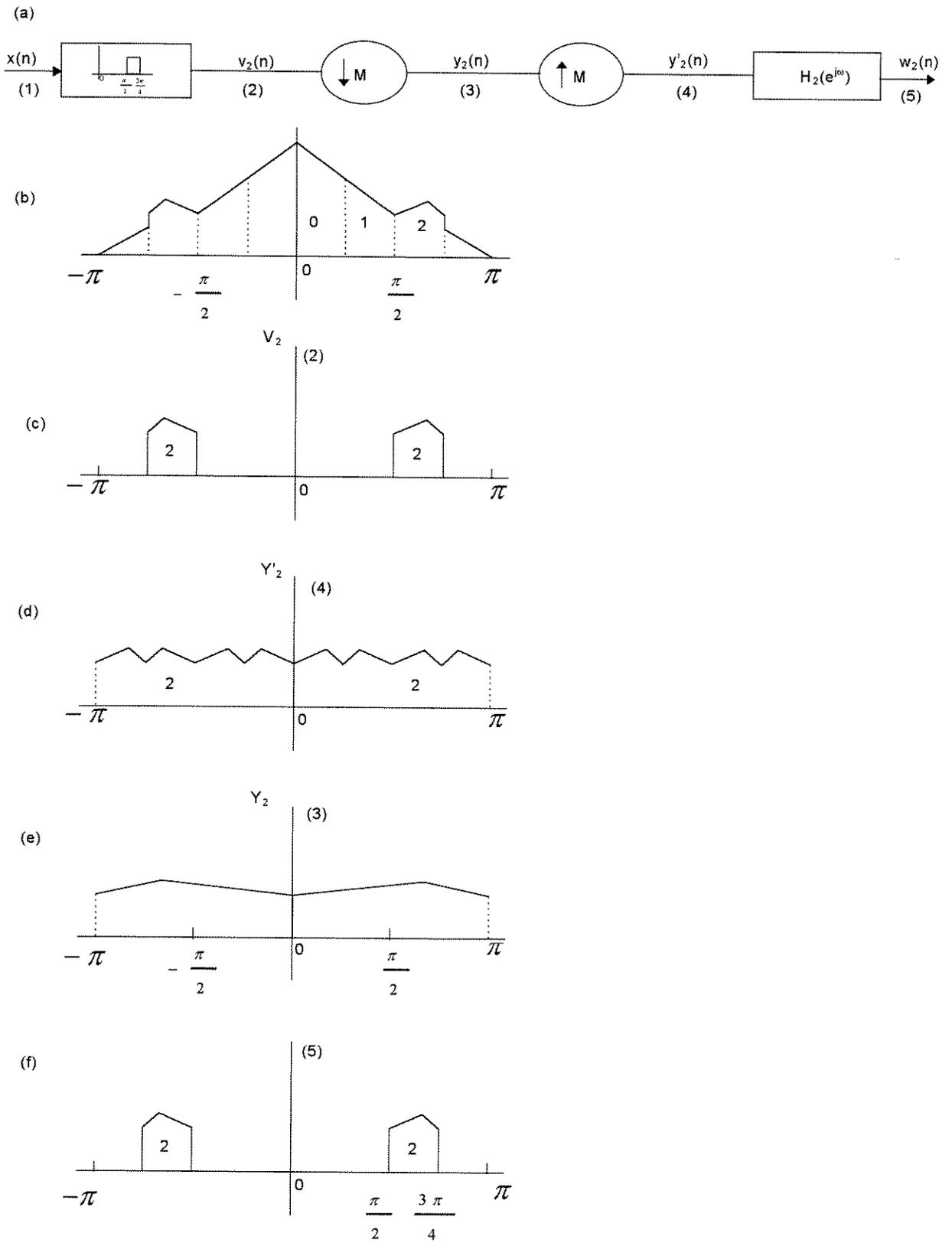


Fig. 3.10 - Configuração da amostragem de banda inteira.

3.3.2 - Modulação em Quadratura

Uma alternativa à amostragem de banda inteira é uma abordagem mais convencional à heterodinagem denominada modulação senoidal. Suponha que um sinal $x(n)$ seja um sinal de banda passante com o espectro mostrado na Fig. 3.11. Este sinal pode ter o seu espectro de frequência deslocado para o nível DC e ser subamostrado sem “aliasing” empregando-se a configuração mostrada na Fig. 3.12.

O espectro dos sinais em vários pontos no modulador-demodulador são mostrados na Fig. 3.13. Para $x(n)$ real tem-se:

$$\begin{aligned} x(n) &\leftrightarrow X(e^{j\omega}) \\ x(n)e^{j\omega_0 n} &\leftrightarrow X(e^{j(\omega - \omega_0)}) \\ X^*(e^{-j\omega}) &= X(e^{j\omega}) \end{aligned}$$

Nos pontos (1) e (2) a modulação em quadratura gera os espectros

$$\begin{aligned} x(n) \cos(\omega_0 n) &\leftrightarrow \frac{1}{2} \left[X(e^{j(\omega - \omega_0)}) + X(e^{j(\omega + \omega_0)}) \right] \\ x(n) \sin(\omega_0 n) &\leftrightarrow \frac{1}{2j} \left[X(e^{j(\omega - \omega_0)}) - X(e^{j(\omega + \omega_0)}) \right] \end{aligned}$$

O filtro passa baixa, na fase de análise, remove as imagens em $\pm 2\omega_0$, deixando apenas os espectros da soma e da diferença nos pontos (3) e (4) respectivamente com larguras de banda $\pm \Delta/2$. A subamostragem “estica” estes espectros por um fator M . Se a largura de banda original for dada por $\Delta/2 = \pi/M$, obtém-se os espectros mostrados nos pontos (5) e (6). O superamostrador “comprime” o eixo de frequências como nos pontos (7) e (8). Nota-se, portanto, que o filtro passa-baixas, na fase de síntese, remove todos os lóbulos exceto o lóbulo em torno do DC. A modulação senoidal em quadratura desloca novamente estes espectros

para $\pm\omega_0$. Finalmente, após a adição dos dois componentes do sinal obtém-se o sinal reconstruído mostrado no ponto (13).

Este método tende ser mais robusto e menos limitado que a amostragem de banda inteira pois nele a frequência central ω_0 pode ser escolhida.

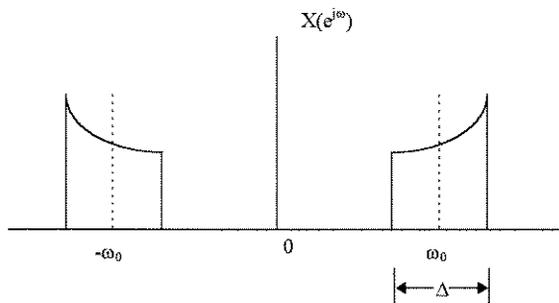


Fig. 3.11 - Espectro de um sinal de banda passante.

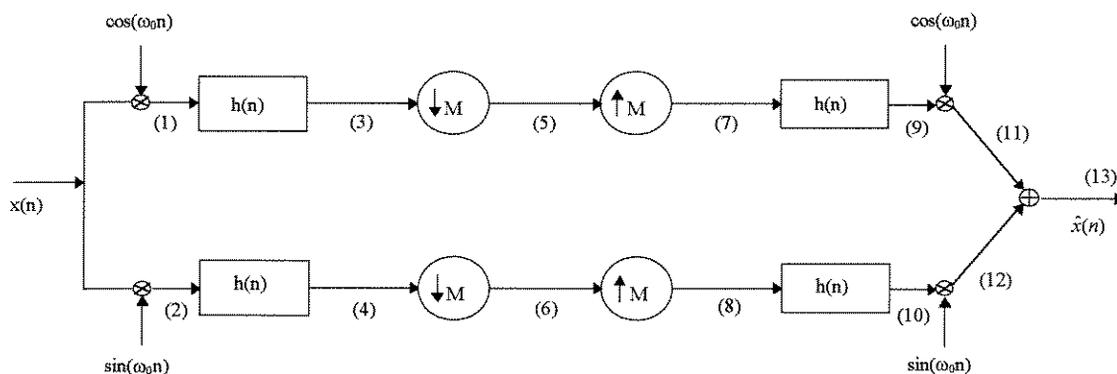
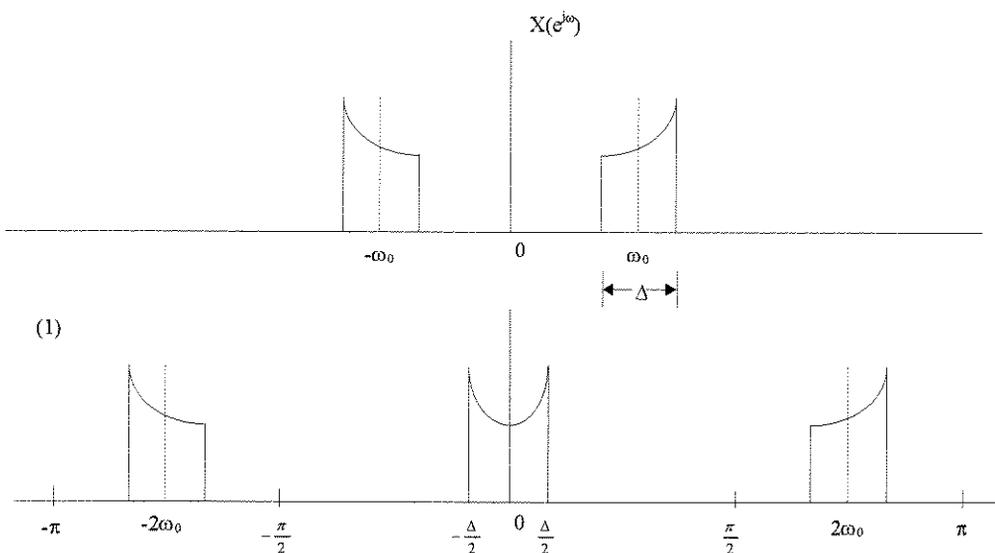
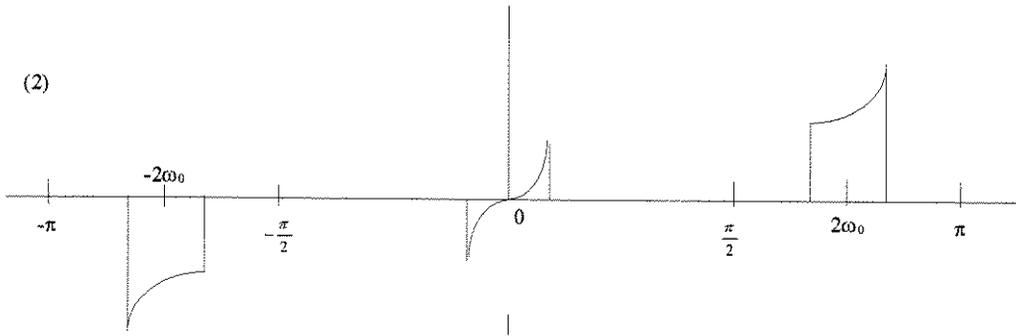


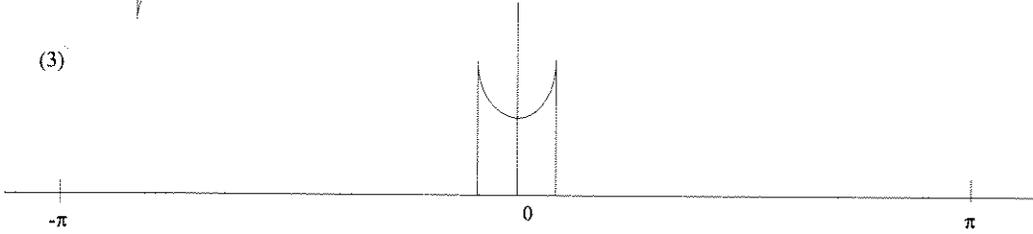
Fig. 3.12 - Modulação e Demodulação em Quadratura, e redução de frequência.



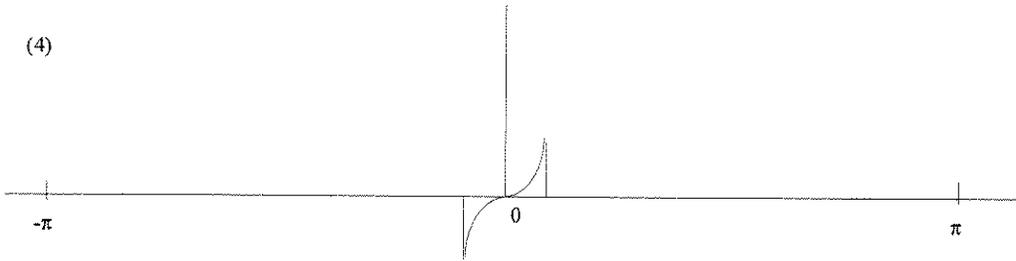
(2)



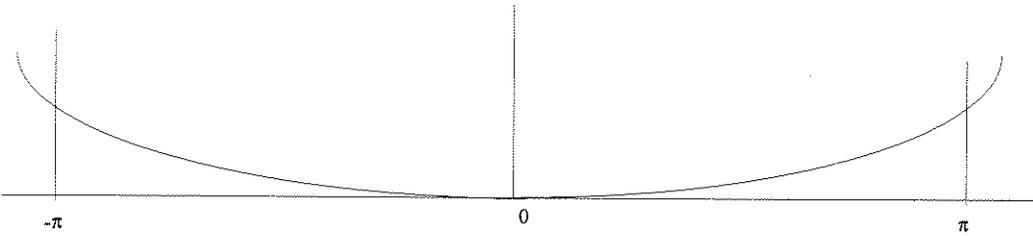
(3)



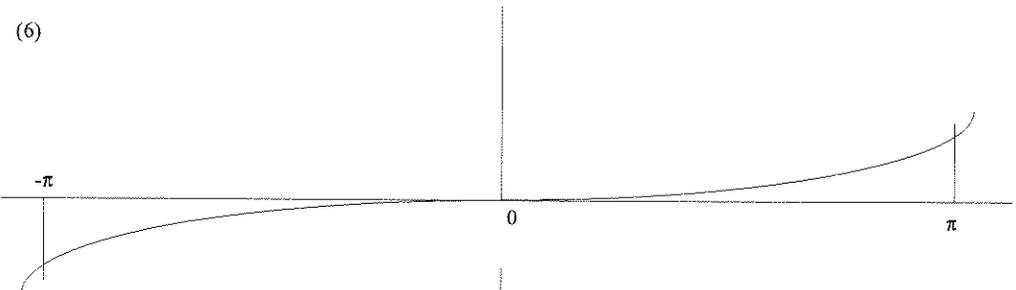
(4)



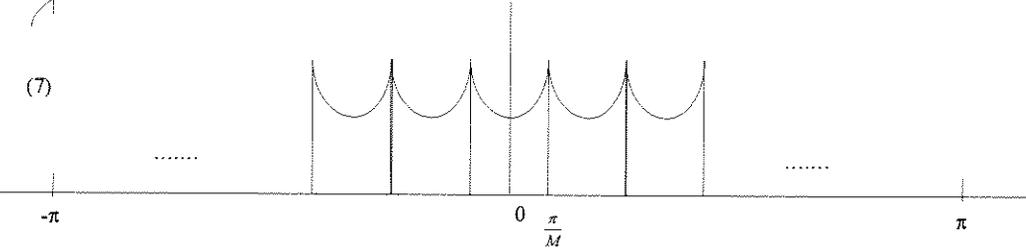
(5)



(6)



(7)



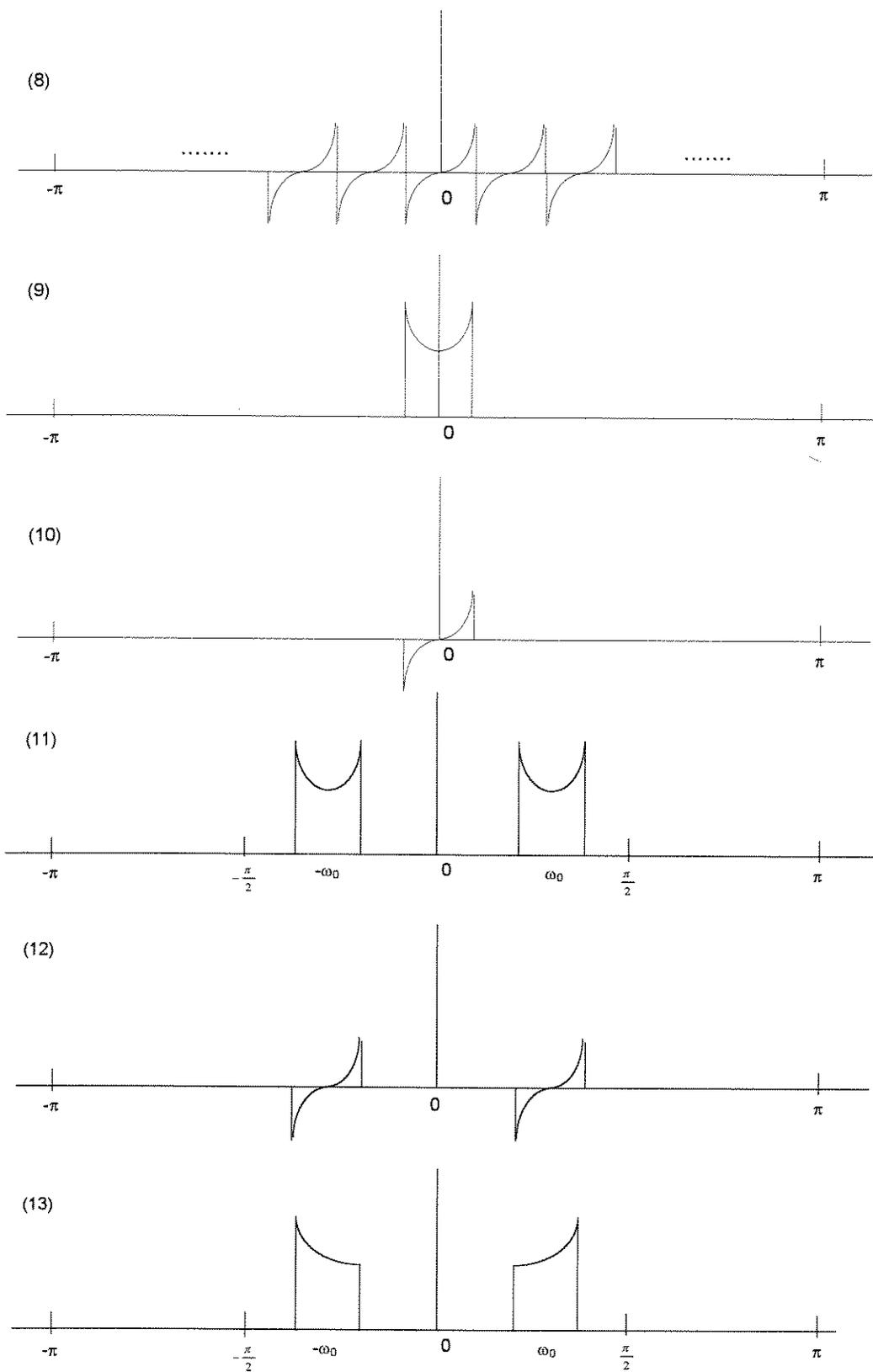


Fig. 3.13 - Espectro de freqüências dos sinais do modulador-demodulador em quadratura.

3.3.3 - Filtros Imagem

Seja $h_0(n)$ um filtro passa baixa de resposta finita ao impulso (FIR - passa baixa) com coeficientes reais. O filtro imagem é definido como

$$h_1(n) = (-1)^n h_0(n) \quad (3.19)$$

ou, no domínio transformado,

$$\begin{aligned} H_1(z) &= H_0(-z) \\ H_1(e^{j\omega}) &= H_0(e^{j(\omega - \pi)}) \end{aligned} \quad (3.20)$$

Substituindo ω por $\frac{\pi}{2} - \omega$ e observando-se que a magnitude é uma função par de ω , obtém-se

$$\left| H_1\left(e^{j\left(\frac{\pi}{2} - \omega\right)}\right) \right| = \left| H_0\left(e^{j\left(\frac{\pi}{2} + \omega\right)}\right) \right| \quad (3.21)$$

Esta última equação demonstra a propriedade de imagem de H_0 e H_1 em $\omega = \frac{\pi}{2}$; estes filtros são chamados de filtros imagem em quadratura (“quadrature mirror filters - QMF”). Os pólos e zeros também se refletem através do eixo imaginário do plano Z, como exigido pela Eq. 3.20.

3.4 - Banco de Filtros de Dois Canais

O propósito de um sistema de filtragem em subbandas é separar o sinal em bandas de frequência e posteriormente codificá-las de acordo com a sua energia. Nesta Seção serão apresentadas as propriedades de um banco de filtros de dois canais que realiza esta tarefa permitindo a reconstrução perfeita do sinal.

O banco de filtros de dois canais é mostrado a Fig. 3.14. O espectro de entrada $X(e^{j\omega})$, $0 \leq \omega \leq \pi$, é dividido em duas subbandas iguais. Os filtros de análise $H_0(z)$ e $H_1(z)$ funcionam como filtros anti “aliasing” dividindo o espectro em duas bandas iguais. Logo, de acordo com o teorema de Nyquist $\theta_0(z)$ e $\theta_1(z)$ são subamostradas por um fator de 2 gerando os sinais de subbanda $v_0(n)$ e $v_1(n)$ como saídas do estágio de análise. Em seguida, estes sinais são passados ao estágio de síntese onde cada sinal é superamostrado por um fator de 2 gerando sinais $f_0(n)$ e $f_1(n)$ entrelaçados de zeros. Estes sinais são então processados pelos filtros de interpolação $G_0(z)$ e $G_1(z)$ e somados na saída para gerar o sinal reconstruído $\hat{x}(n)$.

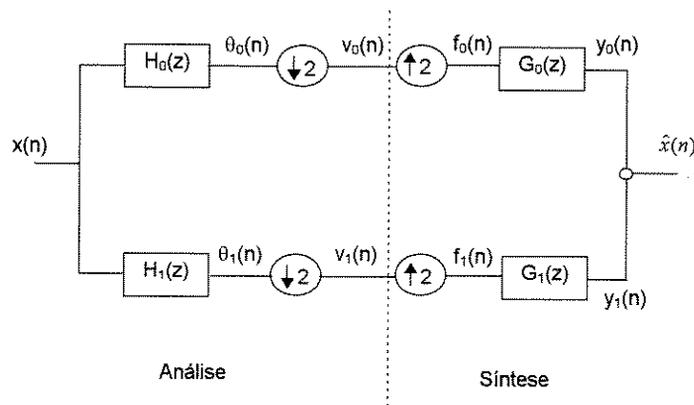


Fig. 3.14 - Subbandas geradas por um Banco de Filtros de Dois Canais.

3.4.1 - Banco de Filtros PR-QMF de Dois Canais

As condições para a reconstrução perfeita do sinal usando um banco de filtros FIR-QMF de dois canais foram inicialmente obtidas por Smith e Barnwell (1986) e tratadas por Vaidyanathan (1987). Estas condições serão aqui apresentadas tanto no domínio da frequência quanto no domínio do tempo.

Observando o ramo superior da Fig. 3.14, tem-se:

$$\Theta_0(z) = H_0(z) \cdot X(z)$$

$$Y_0(z) = G_0(z) \cdot F_0(z)$$

(3.22)

que representam as saídas dos filtros dizimador e interpolador, respectivamente. O subamostrador e o superamostrador impõem que:

$$V_0(z) = \frac{1}{2} [\Theta_0(z^{1/2}) + \Theta_0(-z^{1/2})]$$

$$F_0(z) = V_0(z^2)$$

Combinando estes resultados, obtém-se:

$$Y_0(z) = \frac{1}{2} G_0(z) [H_0(z) \cdot X(z) + H_0(-z) \cdot X(-z)] \quad (3.23)$$

Similarmente,

$$Y_1(z) = \frac{1}{2} G_1(z) [H_1(z) \cdot X(z) + H_1(-z) \cdot X(-z)]$$

Logo, a transformada Z do sinal reconstruído é dada por:

$$\begin{aligned} \hat{X}(z) &= \frac{1}{2} [H_0(z) \cdot G_0(z) + H_1(z) \cdot G_1(z)] X(z) \\ &\quad + \frac{1}{2} [H_0(-z) \cdot G_0(z) + H_1(-z) \cdot G_1(z)] X(-z) \\ &= T(z) \cdot X(z) + S(z) \cdot X(-z) \end{aligned} \quad (3.24)$$

A reconstrução perfeita do sinal exige que:

1 - $S(z) = 0$ para todo o z .

2 - $T(z) = c \cdot z^{-n_0}$ onde c é uma constante.

Para eliminar o “aliasing” e forçar $S(z) = 0$, faz-se necessário que:

$$\frac{G_0(z)}{G_1(z)} = -\frac{H_1(-z)}{H_0(-z)}$$

que pode ser obtido selecionando:

$$\begin{aligned} G_0(z) &= -H_1(-z) \\ G_1(z) &= H_0(-z) \end{aligned} \quad (3.25)$$

resultando em,

$$T(z) = \frac{1}{2} [H_0(-z) \cdot H_1(z) - H_0(z) \cdot H_1(-z)]. \quad (3.26)$$

Várias escolhas podem ser feitas para forçar $T(z) = c \cdot z^{-n_0}$ e entre elas, apresenta-se a seguir a solução FIR paraunitária.

Seja $H_0(z)$ e $H_1(z)$ dois filtros FIR de N-taps, onde N é um número par; seja ainda:

$$H_1(z) = z^{-(N-1)} H_0(-z^{-1}) \quad (3.27)$$

Esta escolha faz com que

$$H_1(-z) = G_0(z)$$

logo,

$$T(z) = \frac{1}{2} z^{-(N-1)} [H_0(z) \cdot H_0(z^{-1}) + H_0(-z) \cdot H_1(-z^{-1})] \quad (3.28)$$

Portanto, as exigências para se fazer a reconstrução perfeita reduzem-se a encontrar $H(z) = H_0(z)$ de forma que

$$\begin{aligned} Q(z) &= H(z) \cdot H(z^{-1}) + H(-z) \cdot H(-z^{-1}) = \text{constante} \\ &= R(z) + R(-z) \end{aligned} \quad (3.29)$$

Esta escolha faz com que todos os quatro filtros sejam causais quando $H_0(z)$ for causal.

A Eq. 3.29 também pode ser expressa no domínio do tempo. Observa-se que $F(z)$ é uma função densidade espectral sendo, portanto, representada por uma série finita da forma:

$$R(z) = y_{N-1}z^{N-1} + y_{N-2}z^{N-2} + \dots + y_0z^0 + \dots + y_{N-1}z^{-(N-1)} \quad (3.30)$$

Logo,

$$R(-z) = -y_{N-1}z^{N-1} + y_{N-2}z^{N-2} - \dots + y_0z^0 - \dots - y_{N-1}z^{-(N-1)} \quad (3.31)$$

Portanto, $Q(z)$ é composto apenas por potências pares de z . Para fazer com que $Q(z) =$ constante, basta que todos os coeficientes pares em $R(z)$ sejam iguais a zero com exceção de y_0 . Mas os y_n coeficientes em $R(z)$ são as amostras da autocorrelação $\rho(n)$ dada por:

$$\begin{aligned} \rho(n) &= \sum_{k=0}^{N-1} h(k)h(k+n) = \rho(-n) \\ &\stackrel{\Delta}{=} h(n) \otimes h(n) \end{aligned} \quad (3.32)$$

onde \otimes indica a operação de correlação. Isto provém de uma propriedade da transformada Z,

$$R(z) = H(z) \cdot H(z^{-1}) \leftrightarrow h(n) * h(-n) = \rho(n) \quad (3.33)$$

onde $\rho(n)$ é a convolução de $h(n)$ com $h(-n)$, ou equivalentemente, a autocorrelação temporal.

Portanto, $\rho(n)$ deve ser igual a 0 para n par e $n \neq 0$. Logo,

$$\rho(2n) = \sum_{k=0}^{N-1} h(k) \cdot h(k+2n) = 0 \quad n \neq 0 \quad (3.34)$$

Se a normalização,

$$\sum_{k=0}^{N-1} |h(k)|^2 = 1 \quad (3.35)$$

for imposta, a exigência para uma reconstrução perfeita no domínio do tempo será dada por:

$$\rho(2n) = \sum_{k=0}^{N-1} h(k) \cdot h(k + 2n) = \delta_n \tag{3.36}$$

Em resumo, a estrutura de duas bandas de filtros PR-FIR paraunitários satisfaz:

$$\begin{aligned} G_0(z) &= -H_1(-z). \\ G_1(z) &= H_0(-z). \\ H_1(z) &= z^{-(N-1)}H_0(-z^{-1}). \\ H_0(z) &= H(z). \end{aligned}$$

com

$$\begin{aligned} R(z) &= H(z)H(z^{-1}) \leftrightarrow \rho(n) = h(n)*h(-n). \\ R(z) + R(-z) &= 1 \leftrightarrow \rho(2n) = \delta(n). \end{aligned} \tag{3.37}$$

3.4.2 - Subbandas geradas em forma de árvores binárias regulares

As técnicas multifrequenciais fornecem a ferramenta básica para a análise espectral em multiresolução e o banco de filtros PR QMF fornece a mais eficiente decomposição em blocos. Estes filtros dividem o espectro em duas subbandas iguais, bandas de baixa (L) e de alta (H) frequência. Esta decomposição pode ser novamente aplicada nas duas novas bandas (L) e (H) gerando outras quatro bandas: (LL), (LH), (HL) e (HH).

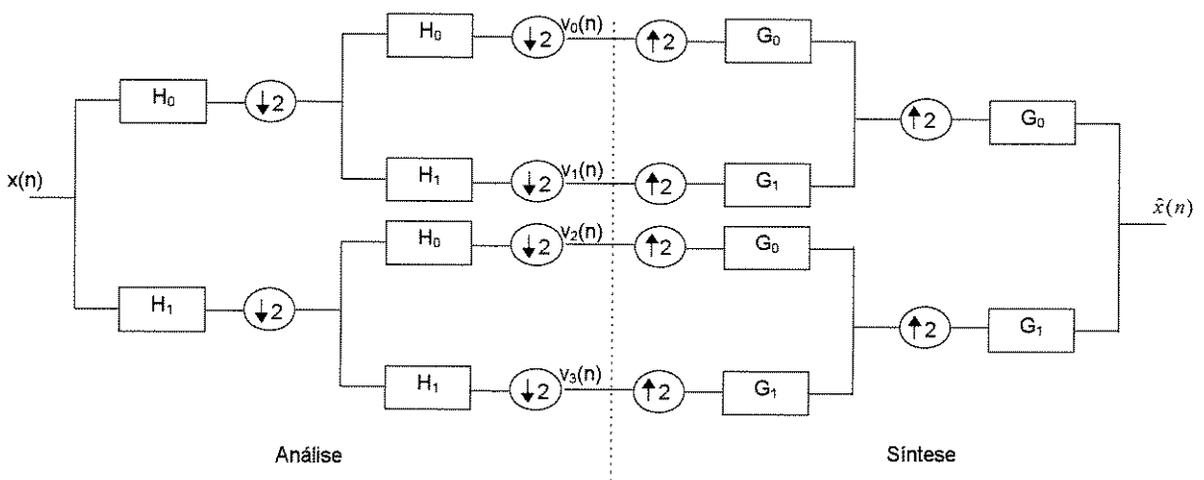


Fig. 3.15 - Estrutura em árvore das fases de análise e síntese de uma decomposição em quatro.

Esta decomposição em dois níveis é mostrada na Fig. 3.15 onde o sinal original, operando numa frequência f_s , é decomposto em quatro subbandas $v_0(n)$, ..., $v_3(n)$, cada uma operando na frequência $f_s/4$. Smith e Barnwell mostraram que a árvore de análise-síntese mostrada na Fig. 3.21 produz uma reconstrução perfeita se a estrutura progenitora análise-síntese for PR, com $\{H_0, H_1, G_0, G_1\}$ satisfazendo a Eq. 3.37. Portanto, esta estrutura pode sofrer várias iterações com a garantia de que a reconstrução perfeita será obtida caso não haja erro de codificação.

A seção de análise da Fig 3.15 pode ser representada por um banco de quatro filtros de análise como na Fig. 3.16. Considere, agora, a estrutura em cascata apresentada na Fig. 3.17 mostrando três filtros separados por dizimadores com taxas M_1 e M_2 . Usando as estruturas equivalentes da Fig. 3.7, pode-se permutar o filtro e o subamostrador substituindo z por z^{M_1} . Logo, a decomposição em duas bandas e dois níveis é equivalente à decomposição hierárquica de duas bandas se:

$$\begin{aligned}
 H_0'(z) &= H_0(z)H_0(z^2). \\
 H_1'(z) &= H_0(z)H_1(z^2). \\
 H_2'(z) &= H_1(z)H_0(z^2). \\
 H_3'(z) &= H_1(z)H_1(z^2).
 \end{aligned} \tag{3.38}$$

Quando a decomposição é aplicada em L níveis, são obtidas 2^L subbandas de mesma largura. Esta abordagem fornece a frequência máxima possível de $\pi/2^L$ com L níveis. Esta estrutura de análise espectral é chamada de árvore binária regular de L níveis ou árvore de subbandas completa. Para $L = 3$, a estrutura em árvore binária regular e a correspondente divisão de bandas de frequência são mostradas na Fig. 3.18.

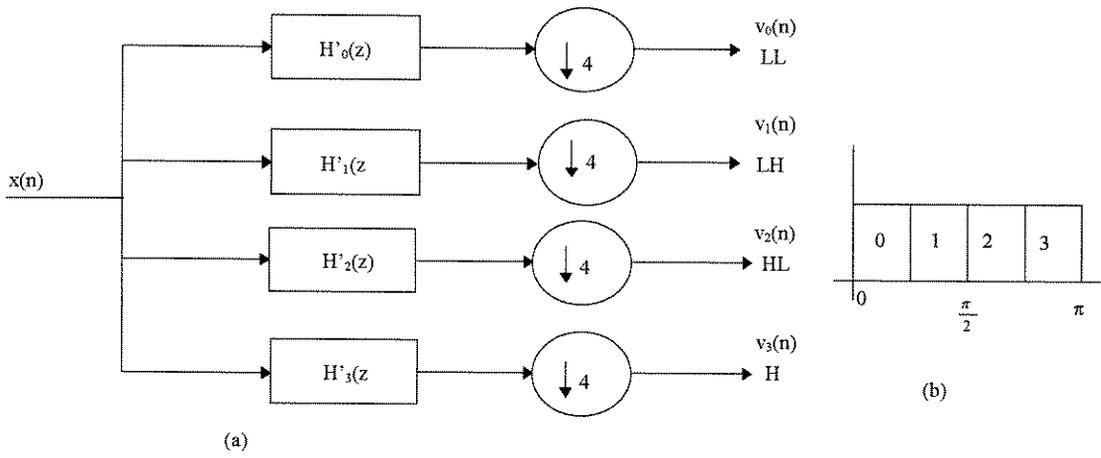


Fig. 3.16 - (a) Subbandas geradas pela árvore binária regular em dois níveis; (b) bandas de frequência correspondentes às quatro subbandas com filtros ideais.

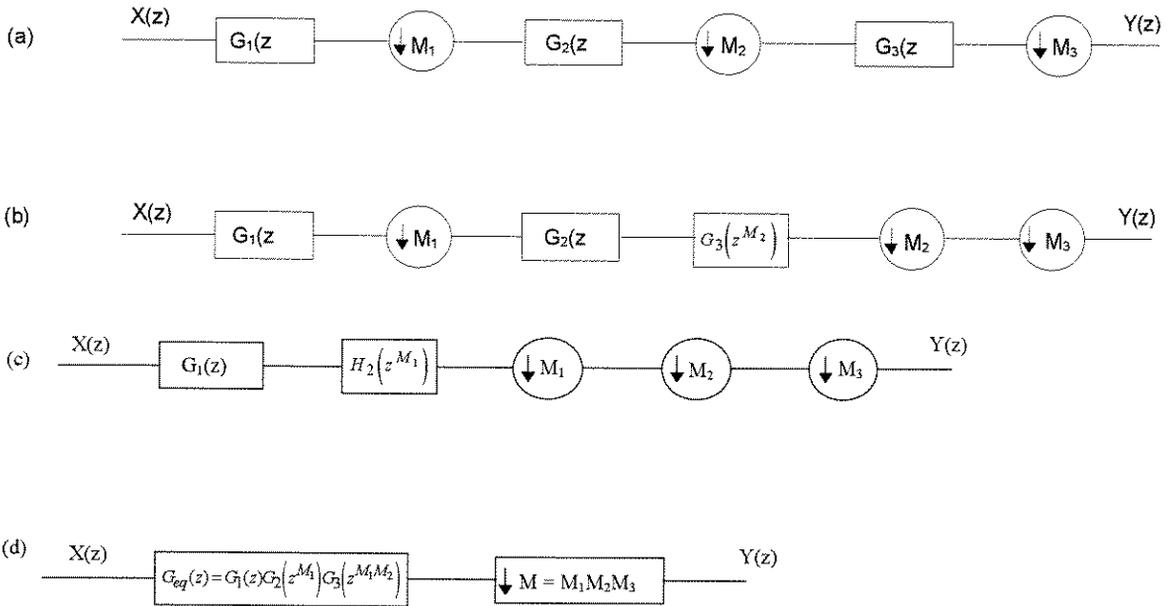


Fig. 3.17 - Estruturas equivalentes para filtros separados por subamostradores.

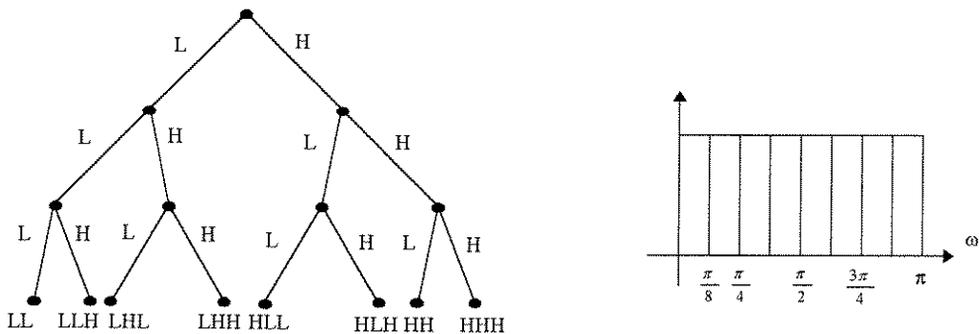


Fig. 3.18 - Estrutura de árvore regular para $L = 3$ e sua banda de frequência, assumindo que filtros ideais PR-QMFs sejam empregados.

3.4.3 - Subbandas geradas em forma de árvores binárias irregulares

Quase todos os sinais reais concentram porções significativas de suas energias em subregiões de seus espectros. Isto indica que alguns intervalos do espectro são mais significativos e importantes que outros. Logo, todas as subbandas de uma árvore binária regular não serão necessárias e algumas subbandas poderão ser combinadas para gerar subbandas com largura de faixa maior. Com isso haverá terminações irregulares dos ramos da árvore. A Fig. 3.19 mostra uma árvore irregular arbitrária com nível máximo L igual a 3 e sua correspondente divisão de bandas de frequência.

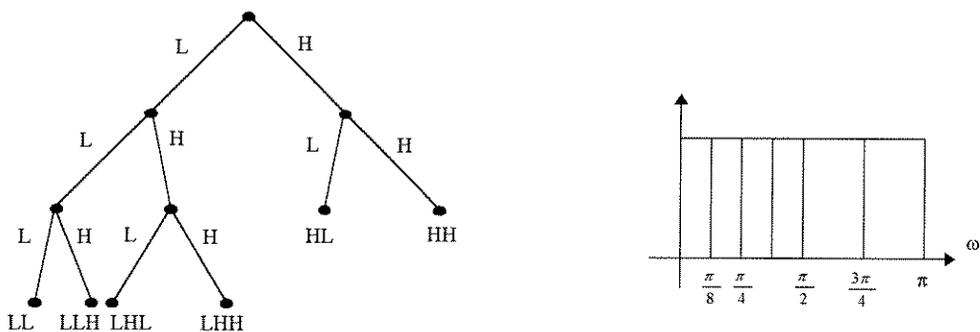


Fig. 3.19 - Estrutura de árvore irregular e sua banda de frequência, assumindo que filtros PR-QMFs sejam empregados.

O número de bandas desta decomposição espectral irregular é menor que o gerado pelas árvores regulares, $M \leq 2^L$. A árvore regular fornece a melhor resolução em frequência possível para um determinado nível L . A árvore regular também possui bandas de frequência de mesma largura enquanto a árvore irregular fornece bandas de larguras diferentes. Porém, as duas estruturas dividem o espectro em um número de subbandas que é potência de 2 pois elas empregam um algoritmo de divisão baseado em duas bandas de frequência.

3.5 - Extensão para Sinais Bidimensionais

A extensão dos resultados apresentados nas Seções precedentes é relativamente simples devido ao fato dos filtros bidimensionais em quadratura (2D-QMF) poderem ser desenvolvidos como filtros separáveis [89, 90]. Logo, os filtros unidimensionais são aplicados primeiro em uma dimensão e posteriormente, na outra para gerar as subbandas. Convém

destacar que a subamostragem pode ser feita logo após a filtragem na primeira dimensão visando reduzir o número de operações necessárias na filtragem na outra dimensão. Este processo é mostrado na Fig. 3.20 para a fase de análise e Fig. 3.21 para a fase de síntese; observa-se que são geradas quatro subbandas classificadas do seguinte modo: baixa/baixa, baixa/alta, alta/baixa e alta/alta nas direções horizontal e vertical respectivamente. A faixa de frequência aproximada de cada subbanda é mostrada na Fig. 3.22 onde LL designa a subbanda baixa/baixa, LH a subbanda baixa/alta, HL a subbanda alta/baixa e HH a subbanda alta/alta. Para um conjunto maior que quatro subbandas, os filtros QMF são aplicados repetidamente em uma ou mais subbandas.

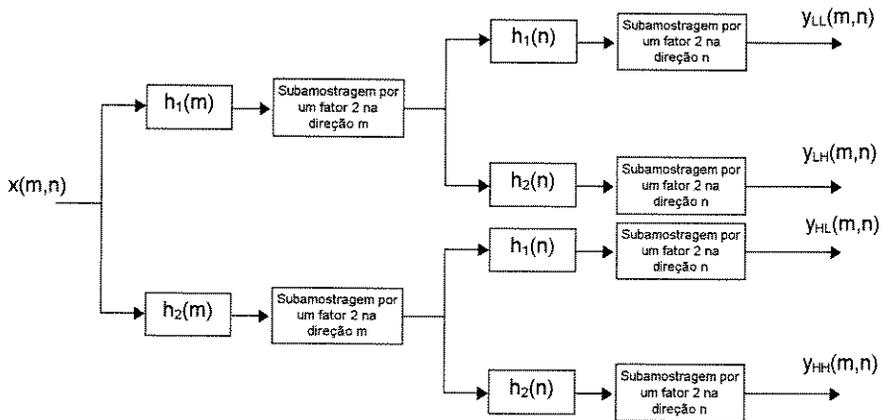


Fig. 3.20 - Fase de análise para um banco de filtros bidimensionais na geração de quatro subbandas.

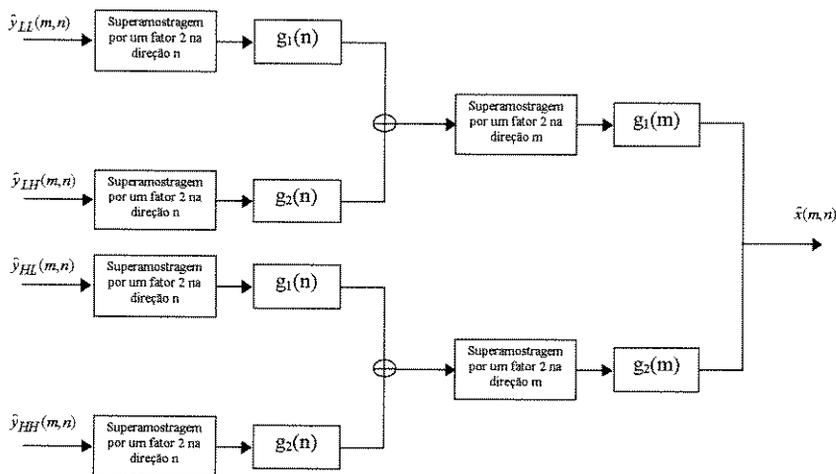


Fig. 3.21 - Fase de síntese para um banco de filtros bidimensionais na geração de quatro subbandas.

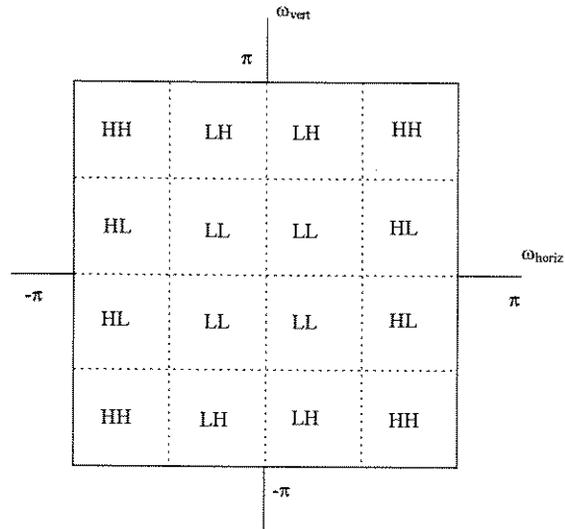
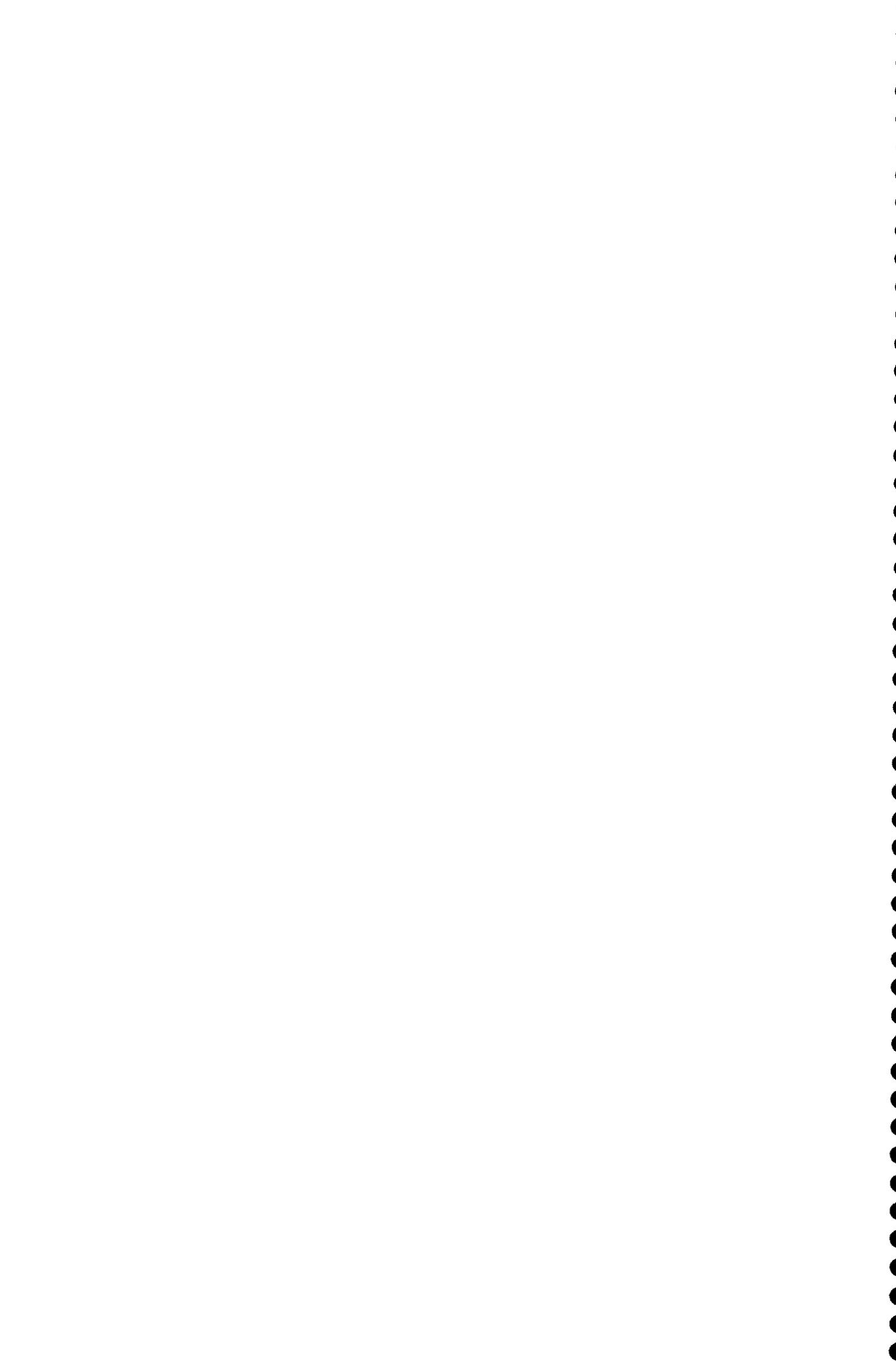


Fig. 3.22 - Faixas de frequência das quatro subbandas geradas.



Capítulo 4 - Resolução do Problema

4.1 - Introdução

Neste capítulo é apresentado o sistema de identificação desenvolvido. Este sistema pode ser dividido em quatro módulos principais como mostra a Fig. 4.1. Destes, procurou-se estudar com maior atenção, neste trabalho, o módulo responsável pela extração de características e o módulo que realiza a identificação do indivíduo. Para tanto, assumiu-se que os problemas presentes na obtenção da imagem da íris e os problemas de registro (rotação, escala e translação da imagem), tratados pelo módulo de pré-processamento, estavam resolvidos.

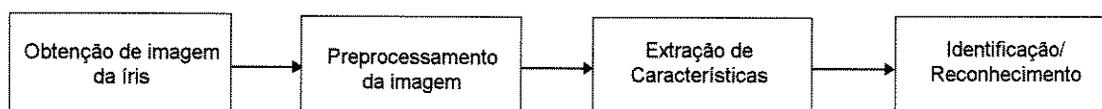


Fig. 4.1 - Esquema do Sistema de Identificação Proposto.

Passa-se a seguir à descrição das operações realizadas em cada módulo, enfatizando os dois últimos, como mencionado anteriormente.

4.2 - Obtenção da Imagem da Íris

As imagens aqui utilizadas foram obtidas durante consultas oftalmológicas. A Dra. Cláudia Assis Lima, do Departamento Oftalmo-Otorrino da Faculdade de Ciências Médicas da Universidade Estadual de Campinas, fotografou, com a autorização de seus respectivos pacientes, 80 pares de olhos. Para tanto, ela utilizou uma máquina fotográfica com uma lente de 105 mm acoplada a uma lâmpada de fenda. Para evitar que o paciente movimentasse a cabeça enquanto seus olhos estivessem sendo fotografados, ele foi instruído a apoiá-la sobre um suporte. Este suporte é geralmente utilizado pelos oftalmologistas durante as consultas.

Foram feitas duas fotografias de cada íris com o intuito de utilizar uma delas na fase de treinamento do módulo de identificação e a outra na fase de reconhecimento neste mesmo módulo. Porém, estas fotos não são idênticas devido a uma série de fatores dentre os quais convém destacar: a variação na dimensão da pupila e os problemas de registro. Estes problemas são tratados no módulo de pré-processamento da imagem.

Posteriormente, estas fotografias, inicialmente coloridas, são “scaneadas” originando imagens com 256x256 pixels e resolução de 200 pixels/pol.. Como a cor não é um aspecto relevante para este sistema pois as imagens são analisadas de acordo com a sua textura e não a sua cor, estas, após serem “scaneadas”, são convertidas em imagens em nível de cinza, com os tons de cinza variando de “0” (preto) a “255” (branco). As Figs. 4.2 e 4.3 são exemplos de imagens de saída deste módulo.

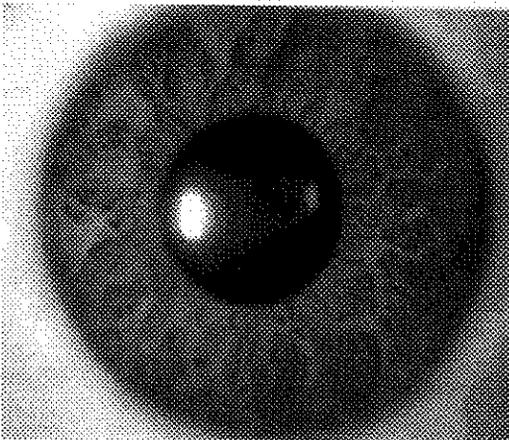


Fig. 4.2 - Imagem em tons de cinza de uma íris esquerda.

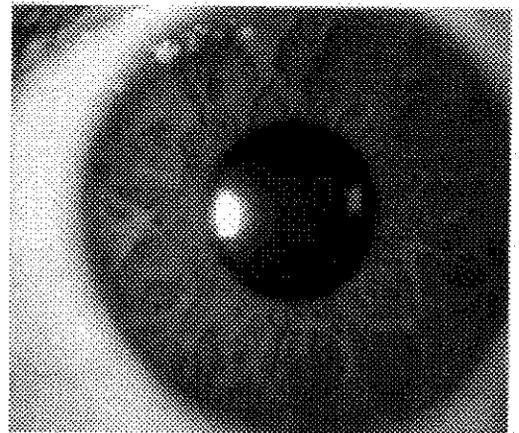


Fig. 4.3 - Imagem da mesma íris da Fig. 4.2 obtida alguns segundos mais tarde.

4.3 - Pré-processamento da Imagem

Antes de serem apresentadas ao sistema de reconhecimento, as imagens geradas no módulo anterior devem ser trabalhadas por este módulo. O processamento realizado aqui é

constituído das seguintes atividades: determinação da região da imagem correspondente à íris e à pupila seguida pela determinação da mesma região íris-pupila nas duas imagens. O objetivo deste trabalho é evitar que sejam utilizadas regiões diferentes da mesma imagem na fase de treinamento e reconhecimento. Para tanto, foi utilizado o software Aldus PhotoStyler 2.0a Special Edition. As Figs. 4.4 e 4.5 ilustram a saída deste módulo.

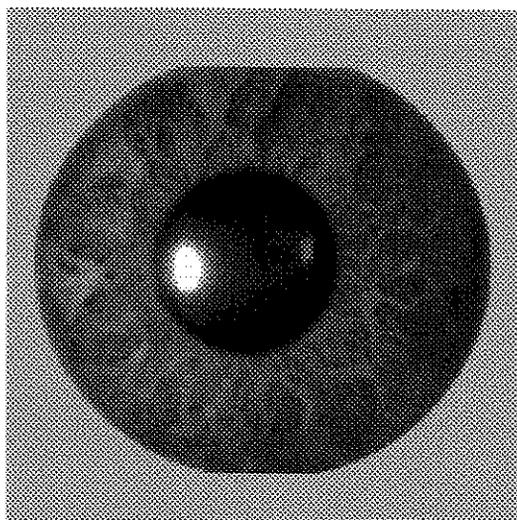


Fig. 4.4 - Resultado do pré-processamento da Fig. 4.2 .

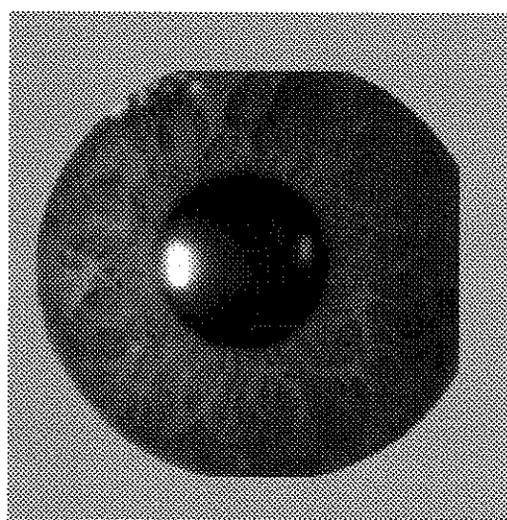


Fig. 4.5 - Resultado do pré-processamento da Fig. 4.3 .

4.4 - Extração de Características

A extração de características é feita por meio de uma decomposição em subbandas. Nesta decomposição utiliza-se a estrutura de árvore binária irregular apresentada no Cap. 3. Esta estrutura foi adotada neste trabalho pois verificou-se que, nos padrões de íris, grande parte da informação relevante está concentrada em determinadas regiões de frequências, especificamente nas regiões de baixas frequências nas direções x e y.

Os filtros 2D-QMF de Symmlet operam sobre a imagem original com 256x256 pixels e geram quatro subbandas com 128x128 coeficientes cada:

- . LL - “Low-Low” - subbanda de baixa frequência na direção X e Y;

- . LH - “Low-High” - subbanda de baixa frequência na direção X e alta frequência na direção Y;
- . HL - “High-Low” - subbanda de alta frequência na direção X e baixa frequência na direção Y;
- . HH - “High-High” - subbanda de alta frequência na direção X e Y.

Após uma série de testes, verificou-se que a subbanda LL concentrava a maior parte da informação da imagem e, por esta razão, ela foi novamente submetida aos filtros 2D-QMF originando mais quatro subbandas com 64x64 coeficientes. Os filtros 2D-QMF são aplicados então sobre todas as subbandas de forma recursiva até que se obtenham subbandas com 16x16 coeficientes. Após efetuada a decomposição, 64 subbandas com 16x16 coeficientes e 3 subbandas com 128x128 coeficientes são geradas. A Fig. 4.6 ilustra todo este processo.

0	1	2	3	4	5	6	7	64
8	9	10	11	12	13	14	15	
16	17	18	19	20	21	22	23	
24	25	26	27	28	29	30	31	
32	33	34	35	36	37	38	39	
40	41	42	43	44	45	46	47	
48	49	50	51	52	53	54	55	
56	57	58	59	60	61	62	63	
65								66

Fig. 4.6 - Representação esquemática do processamento desenvolvido durante a fase de extração de características.

As 64 subbandas com 16x16 coeficientes, por concentrarem a maior parte da informação relevante, são rotuladas como mostra a Fig. 4.6 e suas energias médias são calculadas. Posteriormente, de acordo com o valor de suas energias médias, estas 64

subbandas são dispostas em ordem decrescente gerando-se uma seqüência semelhante à mostrada na Fig. 4.7.

Seqüência A =	0	1	8	16	...	48
---------------	---	---	---	----	-----	----

Fig. 4.7 - Seqüência de 64 subbandas ordenadas de acordo com a sua energia média..

Neste ponto, todos os padrões de íris passam a ser representados por uma seqüência ordenada de rótulos. Posteriormente, estas seqüências são agrupadas em conjuntos. Como cada padrão de íris, de um determinado conjunto de N padrões, possui uma seqüência de rótulos específica, faz-se necessário definir uma seqüência de rótulos que caracterize o conjunto. Para tanto, foi aplicada a seguinte heurística, obtida após vários testes e simulações:

1. Escolha os 32 primeiros rótulos de cada seqüência. Estes rótulos representam as 32 subbandas mais energéticas de cada padrão de íris.
2. Do conjunto resultante, N (padrões) x 32 (rótulos), escolha os 20 rótulos que aparecem com maior freqüência. Esta seqüência de 20 rótulos passa agora a caracterizar o conjunto de N padrões.

Nota-se que esta seqüência é a mesma para todos os padrões presentes neste conjunto, sendo totalmente diferente da seqüência de um outro conjunto de padrões. As subbandas caracterizadas por estas seqüências constituem o conjunto de características que será utilizado tanto no treinamento da rede neural como na identificação de um determinado padrão dentro de um conjunto de padrões aprendidos.

4.5 - Identificação e Reconhecimento

4.5.1 - Introdução

Neste módulo foi utilizada a rede neural proposta por Kwan e Cai [91]. Porém, a estrutura desta rede não era totalmente adequada para o tratamento dos padrões fornecidos pelo módulo de extração de características. Para tanto, foram feitas algumas modificações nesta estrutura principalmente na terceira camada onde foram utilizados os neurônios do tipo MED (apresentado na seção 4.5.2.2.3) no lugar dos neurônios do tipo MIN [91] utilizados na estrutura original. A estrutura da rede neural utilizada neste módulo é descrita a seguir.

4.5.2 - Rede Neural

4.5.2.1 - Neurônio Nebuloso

O neurônio nebuloso possui N entradas x_i ponderadas por pesos w_i com $i = 1, \dots, N$ e M saídas y_j com $j = 1, \dots, M$. As entradas x_i , ponderadas por w_i , são agregadas por uma função $h[\cdot]$ e armazenadas em uma variável interna z que representa a entrada deste neurônio,

$$z = h [w_1x_1, w_2x_2, \dots, w_nx_n].$$

Esta entrada, submetida à ação de uma função de ativação $f[\cdot]$ e um limiar de ativação T , constitui o estado s do neurônio nebuloso,

$$s = f [z - T].$$

As saídas y_j estão associadas a M funções de saída $\{g_j, j= 1, \dots, M\}$ que representam as funções de pertinência dos M conjuntos nebulosos,

$$y_j = g_j (s), \quad j = 1, \dots, M.$$

A Fig. 4.8 apresenta um diagrama esquemático do neurônio nebuloso,

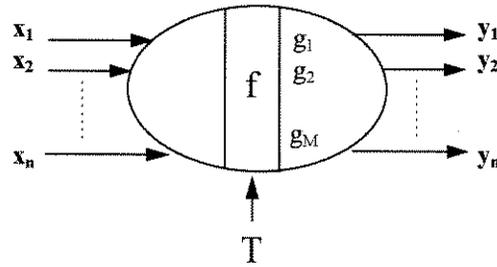


Fig. 4.8 - Neurônio fuzzy.

Geralmente, os pesos w_i , o limiar de ativação T e as funções de saída g_i são ajustados durante o processo de aprendizado. Logo, o neurônio nebuloso é adaptativo e uma rede constituída por um conjunto de neurônios deste tipo é capaz de aprender com o ambiente em que está inserida. A função de agregação e a função de ativação são características intrínsecas do neurônio nebuloso. Usando-se diferentes funções $h[.]$ e $f[.]$, obtém-se neurônios com diferentes características e propriedades que definem uma classe de neurônio nebuloso. A definição das quatro classes de neurônios usados na estrutura de rede apresentada é feita na Seção que se segue.

4.5.2.2 - Classes de Neurônios

Na construção da rede neural, foram utilizadas quatro classes de neurônios. Cada classe possui certas características desempenhando um papel específico dentro da estrutura de rede idealizada. Estas classes de neurônios, assim como suas principais características, são descritas a seguir.

4.5.2.2.1 - Neurônios de entrada (Ent)

Esta classe é constituída por neurônios nebulosos que possuem apenas uma entrada x tal que,

$$z = x.$$

4.5.2.2.2 - Neurônios de máximo (Max)

Estes neurônios são assim denominados porque usam como função de agregação uma função de máximo, ou seja,

$$z = \max_{i=1}^N (w_i x_i)$$

4.5.2.2.3 - Neurônios de média (Med)

Esta classe de neurônios apresenta como função de agregação uma função que calcula a média dos valores de entrada, ou seja,

$$z = \frac{\sum_{i=1}^N w_i x_i}{N}$$

4.5.2.2.4 - Neurônios competitivos (Comp)

Esta classe de neurônios é caracterizada pelo fato de o neurônio possuir um limiar variável e apenas uma única saída tal que,

$$y = g[s - T] = \begin{cases} 0, & \text{se } s < T \\ 1, & \text{se } s \geq T \end{cases}$$

$$T = t [c_1, c_2, \dots, c_K],$$

onde s representa o estado do neurônio nebuloso; $t[\cdot]$, a função de limiar e c_k ($k=1, \dots, K$), as variáveis competitivas.

4.5.2.3. Estrutura da Rede

A rede neural proposta em [91] é uma rede feedforward composta de 4 camadas de neurônios nebulosos representada pela Fig. 4.9,

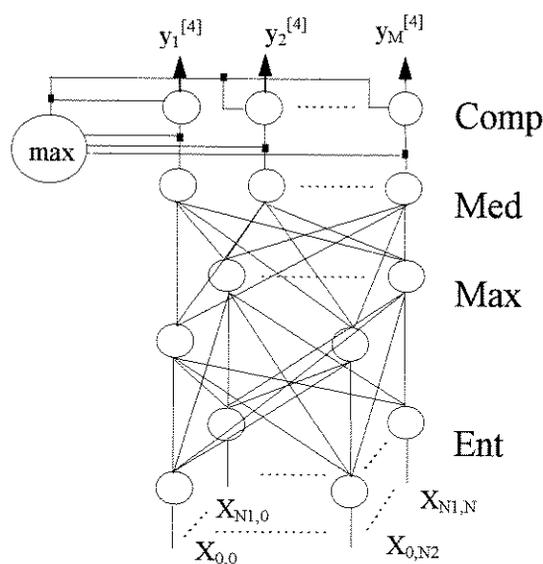


Fig. 4.9 - Arquitetura da Rede Neuro-Fuzzy.

Primeira Camada

Esta camada é composta por neurônios de entrada e sua principal função é admitir os padrões que serão analisados. Cada neurônio, nesta camada, recebe uma componente do padrão de entrada. Esta camada é uma camada bidimensional pois os padrões de entrada estão organizados na forma de matrizes com $N_1 \times N_2$ componentes. Além de permitir a entrada das componentes dos padrões de entrada, os neurônios desta camada promovem a normalização destas no intervalo $[0,1]$.

$$y_{ij}^{[1]} = x_{ij}^{[1]} \quad , \text{ para } i = 1, \dots, N_1 \text{ e } j = 1, \dots, N_2.$$

Segunda camada

A segunda camada também é uma camada bidimensional composta por neurônios de máximo. O objetivo principal desta camada é promover a “fuzzificação” dos padrões de entrada utilizando para tanto uma função de pesos $w[m,n]$.

O estado s do (p,q) -ésimo neurônio desta camada é dado por:

$$S_{pq}^{[2]} = \max_{i=1}^{N_1} \left(\max_{j=1}^{N_2} \left(w[p-i, q-j] y_{ij}^{[1]} \right) \right), \quad \text{para } p = 1, \dots, N_1 \text{ e } q = 1, \dots, N_2,$$

onde $w[p-i, q-j]$ é o peso da conexão entre o neurônio (i, j) da primeira camada e o neurônio (p, q) da segunda sendo definido por:

$$w[m, n] = \exp\left(-\beta^2(m^2 + n^2)\right)$$

para $m = -(N_1 - 1)$ a $(N_1 - 1)$

$n = -(N_2 - 1)$ a $(N_2 - 1)$.

Um mostra da função $w[m, n]$ para $\beta = 0,3$ é dada pela Fig. 4.10 .

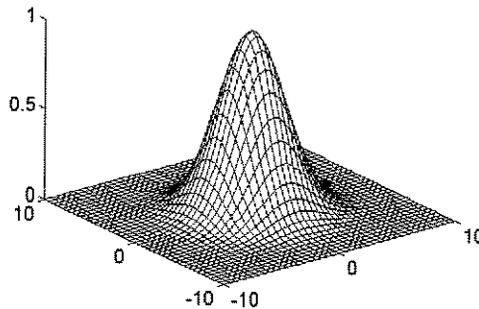


Fig. 4.10 - Função de "Fuzzificação" ($\beta = 0.3$)

Observando a Fig. 4.10, pode-se concluir que os neurônios desta camada são semelhantes à uma lente, ou seja, eles focalizam uma determinada componente do padrão de entrada e observam, concomitantemente, o comportamento das componentes vizinhas a esta. A quantidade de componentes vizinhas observadas depende do valor de β e é determinado pelo algoritmo de aprendizagem.

Cada neurônio desta camada possui M saídas, uma para cada neurônio da terceira camada. Estas saídas, para o (p, q) -ésimo neurônio, são dadas por,

$$y_{pqm}^{[2]} = g_{pqm} [s_{pq}^{[2]}],$$

para $p = 1, \dots, N_1$, $q = 1, \dots, N_2$, $m = 1, \dots, M$.

onde $y_{pqm}^{[2]}$ é a m -ésima saída do neurônio (p, q) , ou seja, o peso que conecta este neurônio ao m -ésimo neurônio na terceira camada. A função $g_{pqm}[s_{pq}^{[2]}]$ deve ser determinada pelo algoritmo de aprendizagem mas, por simplicidade, adotou-se uma função definida por um triângulo isóceles de altura 1 e base α como mostrada pela Fig. 4.11 ,

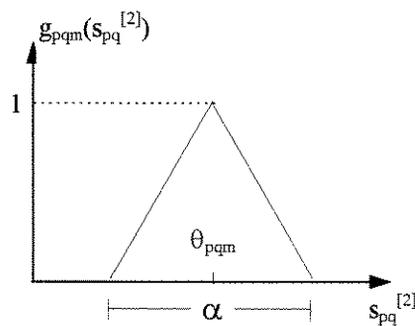


Fig. 4.11 - Função de saída dos MAX-FN da segunda camada.

Logo:

$$y_{pqm}^{[2]} = g_{pqm} [s_{pq}^{[2]}] = \begin{cases} 1 - \frac{2}{\alpha} |s_{pq}^{[2]} - \theta_{pqm}|, & \text{se } \alpha/2 \geq |s_{pq}^{[2]} - \theta_{pqm}| \geq 0 \\ 0 & \text{, fora do intervalo acima} \end{cases}$$

para $\alpha \geq 0$, $p = 1, \dots, N$, $q = 1, \dots, N_2$ e $m = 1, \dots, M$.

θ_{pqm} é o ponto central da base do triângulo e representa o padrão a ser reconhecido. Quanto mais próximo de θ_{pqm} estiver $s_{pq}^{[2]}$, mais significativa será a percepção que este neurônio tem da componente (p, q) do padrão de entrada m . Os valores de α , θ_{pqm} , p , q e m são determinados pelo algoritmo de aprendizagem.

Terceira camada

Nesta camada, são usados neurônios de média onde cada neurônio representa um padrão aprendido. Portanto, o número de neurônios nesta camada só poderá ser estabelecido quando o processo de aprendizagem estiver finalizado. A saída do m-ésimo neurônio desta camada é dada por:

$$y_m^{[3]} = s_m^{[3]} = \sum_{p=1}^{N_1} \left(\sum_{q=1}^{N_2} y_{pqm}^{[2]} \right) / (N_1 \times N_2), \quad \text{para } m = 1, \dots, M$$

onde $s_m^{[3]}$ representa o estado do m-ésimo neurônio desta camada.

Quarta camada

Esta camada é a camada de saída. Nela são usados neurônios competitivos (um para cada um dos M padrões aprendidos), proporcionando uma saída não nebulosa. Nesta camada, apenas um único neurônio tem a saída $y_m^{[4]}$ igual a 1 enquanto os outros têm suas saídas iguais a 0, pois a função de saída, que rege o comportamento dos neurônios desta camada, é dada por,

$$y_m^{[4]} = g[s_m^{[4]} - T] = \begin{cases} 0, & \text{se } s_m^{[4]} < T \\ 1, & \text{se } s_m^{[4]} = T, \end{cases} \quad m = 1, \dots, M$$

onde:

$$T = \max_{m=1}^M (y_m^{[3]})$$

$$s_m^{[4]} = z_m^{[4]} = y_m^{[3]}$$

Se um padrão de entrada é mais semelhante ao m-ésimo padrão aprendido, então o m-ésimo neurônio desta camada é o vencedor e o padrão de entrada é reconhecido como sendo o padrão aprendido m.

Algoritmo de Aprendizagem

Antes da apresentação do algoritmo de aprendizagem, deve-se definir os parâmetros de entrada T_f e K . O parâmetro T_f refere-se ao erro de aprendizagem da rede variando no intervalo $[0,1]$; o parâmetro K especifica o número total de padrões de treinamento ($k = 1, \dots, K$). Feito isto, passa-se ao algoritmo de aprendizagem:

1 - Crie $N_1 \times N_2$ neurônios de entrada na primeira camada e $N_1 \times N_2$ neurônios de máximo na segunda camada. Escolha um valor para α ($\alpha \geq 0$) e um valor para β .

2 - Faça $M = 0$ e $k = 1$.

3 - Faça $M = M+1$. Crie o m -ésimo neurônio de média na terceira camada e o m -ésimo neurônio competitivo na quarta camada. Faça:

$$\Theta_{pqM} = s_{pqM}^{[2]} = \max_{i=1}^{N_1} \left(\max_{j=1}^{N_2} (w[p-i, q-j] y_{ijk}^{[1]}) \right), \quad \text{para } p = 1, \dots, N_1 \text{ e } q = 1, \dots, N_2.$$

4 - Faça $k = k+1$. Se $k > K$, então a aprendizagem está finalizada. Caso contrário, entre com o k -ésimo padrão de treinamento na rede e obtenha a saída da rede corrente (com M neurônios nas terceira e quarta camadas). Faça:

$$\sigma = 1 - \max_{j=1}^M (y_{jk}^{[3]})$$

onde $y_{jk}^{[3]}$ é a saída do j -ésimo neurônio de média da terceira camada para o k -ésimo padrão de treinamento X_k . Se $\sigma \leq T_f$, volte ao 4º passo. Caso contrário, volte ao 3º passo.

4.5.2.4 . Análise da rede

Na Seção 4.5.2.3, foi apresentada a estrutura e o algoritmo de aprendizagem de uma rede de quatro camadas “feedforward”. Nesta Seção, serão analisadas as características

marcantes desta rede. A primeira delas diz respeito aos seus elementos constituintes. Como mostrado na Seção 4.5.2.3, esta rede é composta por quatro tipos de neurônios nebulosos: neurônios de entrada, neurônios de máximo, neurônios de média e neurônios competitivos. Estes neurônios, organizados de forma adequada, permitem a manipulação e o processamento de informações incompletas e dúbias. A manipulação de informações desta natureza é também uma característica relevante desta estrutura. Uma outra característica que também deve ser destacada refere-se à forma como esta rede é construída. Como pôde ser observado na Seção anterior, a construção desta rede ocorre durante o processo de aprendizagem e não se restringe à uma única etapa. Isto significa que após P padrões terem sido aprendidos e desejando-se que N padrões diferentes de P sejam ensinados à rede, deve-se iniciar novamente o algoritmo de aprendizagem a partir do segundo passo tendo o cuidado de atribuir à variável M o número de padrões já aprendidos, ou seja, P . Após terem sido citadas as principais características deste sistema, passa-se à descrição das ações realizadas em cada camada desta rede.

A primeira camada de neurônios recebe os dados do padrão de entrada. Os neurônios nebulosos desta camada transformam os valores das componentes do padrão em valores normalizados no intervalo $[0,1]$.

A segunda camada é responsável pela “fuzzificação” do padrão de entrada. Cada neurônio de máximo desta camada está conectado a todos os neurônios de entrada da camada anterior por intermédio de uma matriz de pesos $W[m,n]$ e têm como estado de ativação o máximo valor das diversas entradas ponderadas. O grau de “fuzzificação” dado a um padrão de entrada depende do valor de β . Se β for muito pequeno, a rede não consegue separar alguns padrões de treinamento distintos. Se β for muito grande, a rede perde a capacidade de

reconhecer padrões um pouco distorcidos ou um pouco deslocados. Portanto, o valor de β deve ser escolhido visando minimizar estes problemas.

A m -ésima saída do (p,q) -ésimo neurônio de máximo na segunda camada, $y_{pqm}^{[2]}$, expressa o conceito nebuloso com relação à extensão do grau de similaridade entre os valores das componentes em torno da (p,q) -ésima componente do padrão de entrada e os valores das componentes em torno da (p,q) -ésima componente do m -ésimo padrão aprendido. A função de saída desta camada é, de fato, uma função de pertinência deste conjunto nebuloso e contém informações sobre os valores das componentes em torno da (p,q) -ésimo componente do m -ésimo padrão aprendido. Por esta razão, usa-se a matriz Θ_{pqM} para armazenar esta informação, no terceiro passo do algoritmo de aprendizagem. Tal procedimento assegura que a rede “lembrar-se-á” de todos os padrões aprendidos.

Os neurônios nebulosos na terceira camada fornecem as similaridades entre o padrão de entrada e todos os demais padrões aprendidos. Como foram usados neurônios de média nesta camada, a similaridade entre o padrão de entrada $X=\{x_{ij}\}$ e o m -ésimo padrão aprendido é calculada pela rede como a saída do m -ésimo neurônio de média na terceira camada, ou seja:

$$y_m^{[3]} = \begin{cases} \sum_{p,q} \left(1 - (2 / \alpha) \times \left| s_{pq}^{[2]} - \Theta_{pqM} \right| \right) / (N_1 \times N_2), & \text{se } \max_{p,q} \left(\left| s_{pq}^{[2]} - \Theta_{pqM} \right| \right) \leq \alpha / 2 \\ 0, & \text{caso contrario} \end{cases},$$

para $m = 1, \dots, M$

onde $s_{pq}^{[2]}$ é o estado do (p,q) -ésimo neurônio de máximo na segunda camada quando o padrão de entrada é X . Da equação acima, pode-se observar que α é um parâmetro de escopo e seu valor afeta o cálculo das similaridades. Quando um padrão de entrada X é um dos padrões aprendidos, haverá uma similaridade $y_m^{[3]}$ (onde $1 \leq m \leq M$) igual a 1. Quando o padrão de entrada não é algum dos padrões aprendidos todas as M similaridades serão menores que 1.

A camada de saída da rede é usada na “defuzzificação” e fornece saídas as não nebulosas da rede. Ela escolhe a similaridade máxima como fronteira de ativação de todos os neurônios competitivos na quarta camada. Se $y_m^{[3]}$ for o máximo entre todas as saídas dos neurônios da terceira camada, a saída do m -ésimo neurônio competitivo na quarta camada será igual a 1 e as saídas de todos os outros neurônios desta camada serão iguais a 0

Novos padrões de treinamento podem ser aprendidos a qualquer momento bastando somente reiniciar o algoritmo de aprendizagem a partir do segundo passo e atribuir à variável M o número de padrões aprendidos anteriormente. Logo, novos neurônios serão adicionados à terceira e quarta camadas quando padrões distintos dos já existentes forem usados para treinamento adicional. Se um padrão aprendido ou um padrão similar aos padrões aprendidos for apresentado à rede, ele será tratado como padrão previamente aprendido e a rede não irá reaprendê-lo. A similaridade entre um padrão adicional e todos os padrões aprendidos, bem como os parâmetros de aprendizado α , β e T_f são determinantes em decidir se o novo padrão será classificado como um padrão aprendido ou um padrão distinto. Se uma das similaridades de um padrão de entrada for maior ou igual a $1 - T_f$, então o padrão será considerado como padrão aprendido. Caso contrário, será tratado como um novo padrão. Percebe-se, assim, que o aprendizado desta rede é contínuo, ou seja, a rede não necessita estar completamente treinada para ser usada apresentando um comportamento semelhante ao do cérebro humano!

Capítulo 5 - Resultados

5.1 - Introdução

No capítulo 4 apresentou-se o sistema biométrico desenvolvido neste trabalho. Neste capítulo, apresenta-se a análise de desempenho do mesmo. Para tanto, fez-se necessário a construção de uma base de dados consistente e a definição de uma estratégia de testes adequada. A seção 5.2 trata especificamente da base de dados utilizada e as seções 5.3 e 5.4 da estratégia criada para avaliar o sistema desenvolvido. Na seção 5.5 é feita a análise dos resultados obtidos durante a bateria de testes a que este sistema foi exposto.

5.2 - Base de Dados

A base de dados utilizada para validar o sistema biométrico proposto foi construída durante os últimos dez meses com o auxílio da Dra. Cláudia (Seção 4.2). Foram fotografados, durante este período, 160 diferentes olhos de 80 pessoas. De cada olho foram retiradas, em média, duas fotografias. Algumas destas fotografias foram rejeitadas à primeira vista por estarem desfocadas ou pelo fato de as pálpebras encobrirem grande parte da íris. As fotografias aprovadas (320 ao todo) foram então digitalizadas e processadas como descrito no Cap. 4.

5.3 - Fase de Treinamento

Durante a fase de treinamento, foram apresentadas à rede neuro-fuzzy 160 imagens de íris diferentes divididas em subconjuntos de dez e vinte imagens cada:

. Conjuntos $C - c_i$, ($i = 1 \dots 16$)- conjuntos com 10 imagens;

. Conjuntos $D - d_j$, ($j = 1 \cdot \cdot 8$) - conjuntos com 20 imagens.

Esta partição teve por objetivo analisar o comportamento da rede neuro-fuzzy à medida que o número de imagens treinadas aumentasse.

5.4 - Fase de Reconhecimento

Na fase de reconhecimento foi apresentada à rede a segunda imagem de cada íris presente nos conjuntos formados pela partição descrita anteriormente. Esta fase foi dividida em dois momentos complementares: (1) determinação da taxa de aceitação de autênticos e (2) determinação da taxa de aceitação de impostores. No primeiro caso, apresenta-se à rede a segunda imagem de cada um dos conjuntos em que foi dividido o conjunto de 160 imagens. No segundo, a primeira imagem dos conjuntos c_i e d_j é apresentada à rede.

Após a determinação das taxas de aceitação de autênticos e de impostores para cada conjunto, foram gerados diversos gráficos em função destes valores. Estes gráficos, apresentados na Seção a seguir, ilustram o desempenho do sistema biométrico proposto.

5.5 - Análise dos Resultados

Nesta seção, são apresentados alguns dos gráficos gerados. Em sua maioria, estes gráficos assemelham-se ao Gráfico 5.1 abaixo e serão analisados de forma semelhante. Os casos destonantes serão analisados detalhadamente.

Para cada conjunto foram gerados dois gráficos. O primeiro apresenta uma série de curvas obtidas a partir da variação da tolerância T_f da rede (durante o treinamento da rede a tolerância T_f variou de 0.1 a 0.16 em média). O segundo apresenta apenas o caso mais restritivo (T_f menor) e o caso mais liberal (T_f maior).

1 - Erro Tipo II

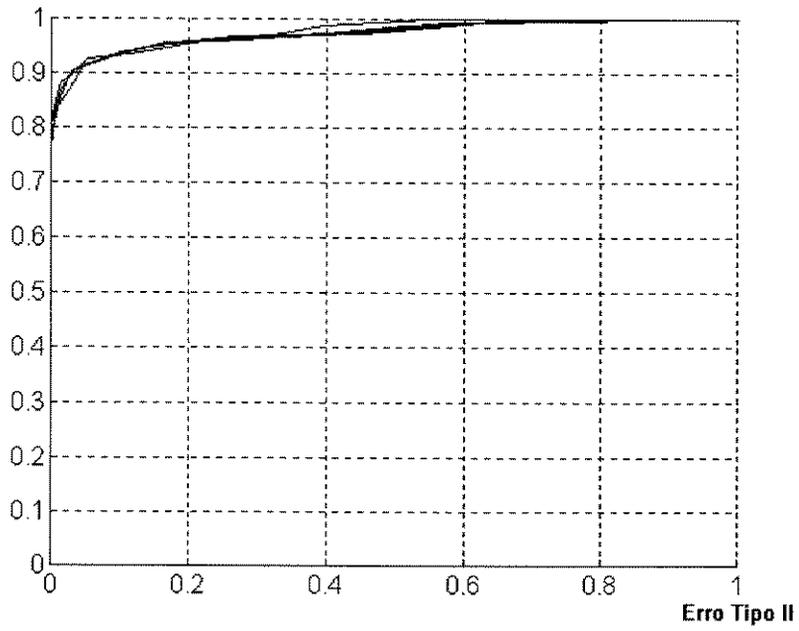


Gráfico 5.1 - Curva de Decisão Neyman-Person para o conjunto C_1

1 - Erro Tipo II

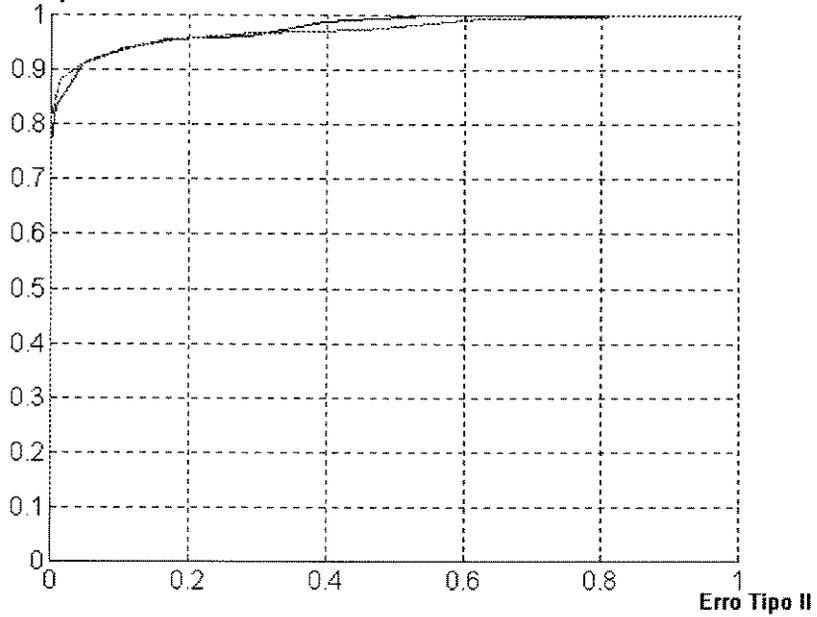


Gráfico 5.2 - Curva de Decisão Neyman-Person para o conjunto C_1 com $T_f = 0.1$ (curva mais escura) e $T_f = 0.15$ (curva mais clara).

Um dado importante que deve ser destacado nos gráficos obtidos é a taxa de reconhecimento de autênticos para a taxa de aceitação de impostores igual a zero. No Gráfico 5.1 esta taxa é, em média, igual a 80%

Também foram gerados alguns gráficos da taxa de aceitação de autênticos e da taxa de rejeição de impostores em função do número de subbandas reconhecidas pela rede neuro-fuzzy (ver Gráfico 5.3 abaixo). Após a análise de outros gráficos semelhantes a este verifica-se que o reconhecimento de 10 subbandas associadas a cada íris é suficiente para determinar se esta pertence ou não a um determinado conjunto de íris cadastradas.

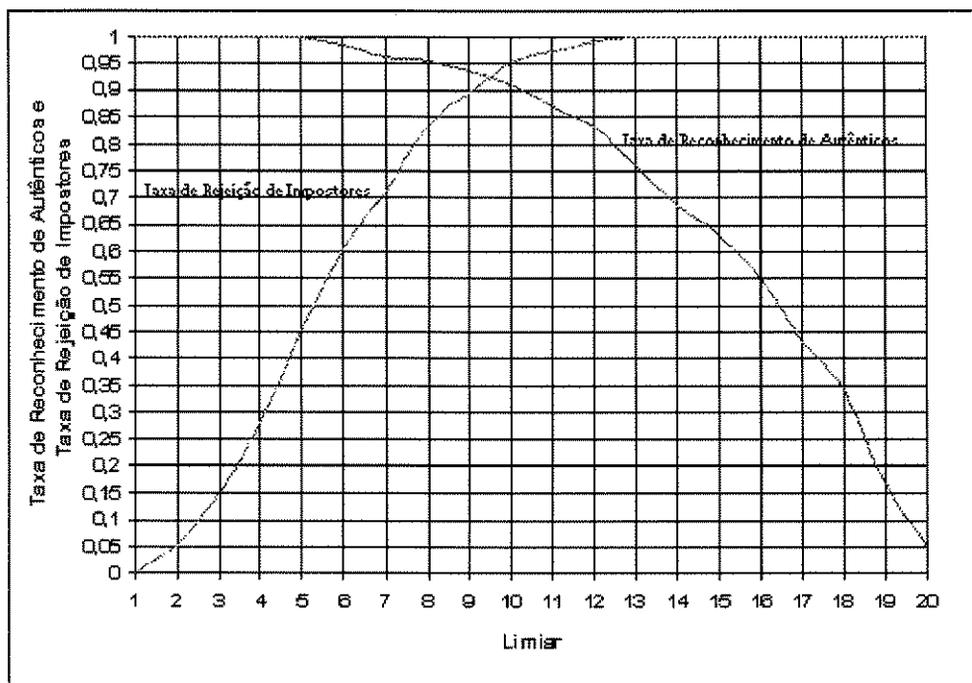


Gráfico 5.3 - Taxa de Reconhecimento de Autênticos e Taxa de Rejeição de Impostores

Casos Destonantes

Como foi dito anteriormente, alguns gráficos da curva de decisão de Neyman-Person fogem ao padrão apresentado pelos Gráficos 5.1 e 5.2. Os Gráficos 5.4 e 5.5 são um bom exemplo deste desvio.

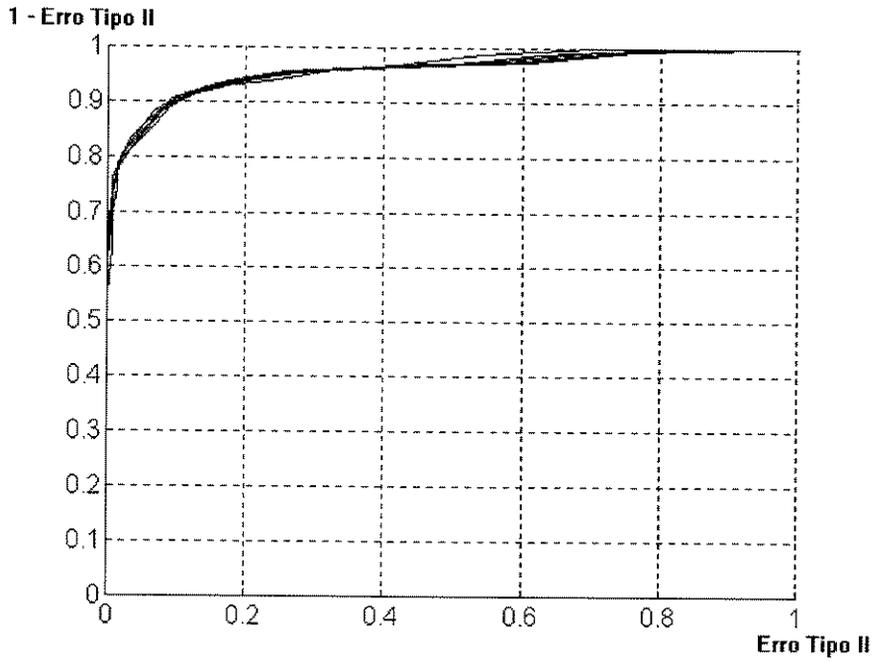


Gráfico 5.4 - Curva de Neyman-Person para conjunto C_3

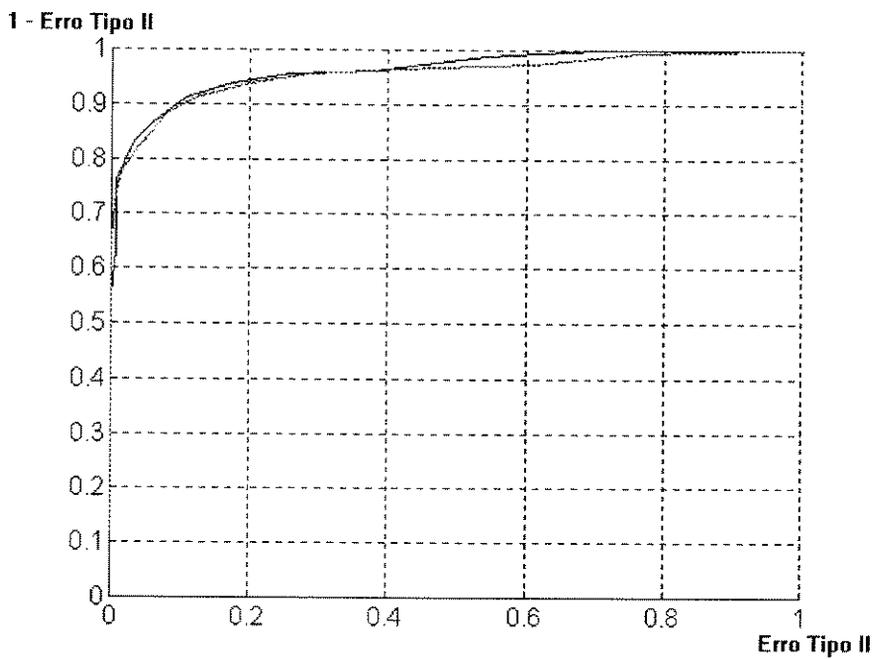


Gráfico 5.5 - Curva de Decisão Neyman-Person para o conjunto C_3 com $T_f = 0.1$ (curva mais escura) e $T_f = 0.19$ (curva mais clara).

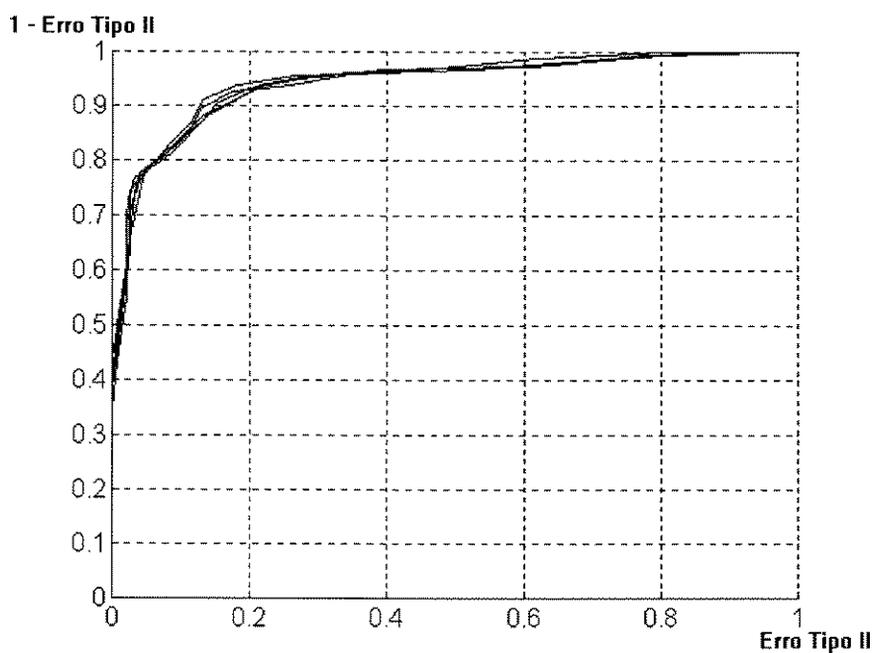


Gráfico 5.6 - Curva de Decisão Neyman-Person para o conjunto C_4

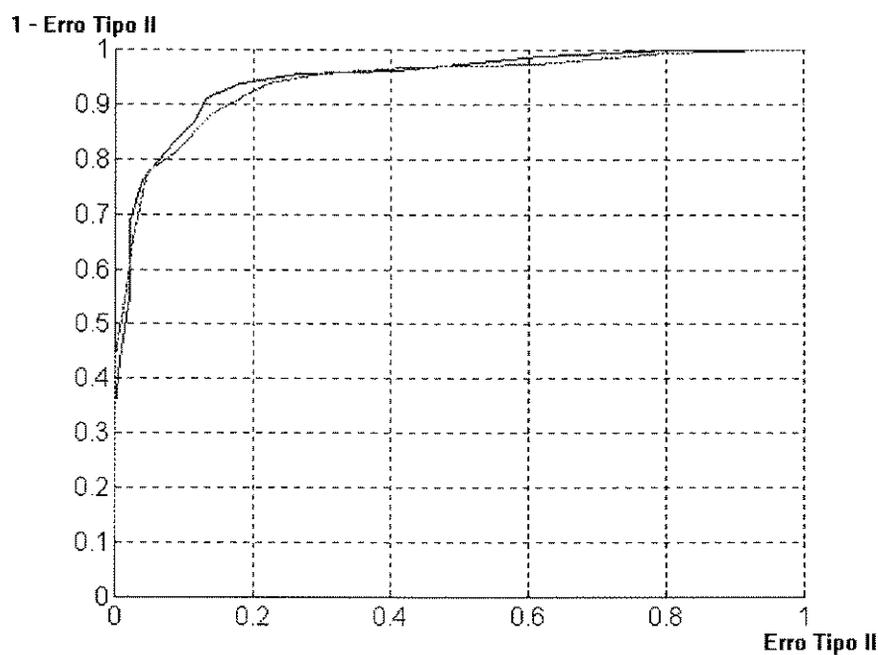


Gráfico 5.7 - Curva de Decisão Neyman-Person para o conjunto C_4 com $T_f = 0.1$ (curva mais escura) e $T_f = 0.16$ (curva mais clara).

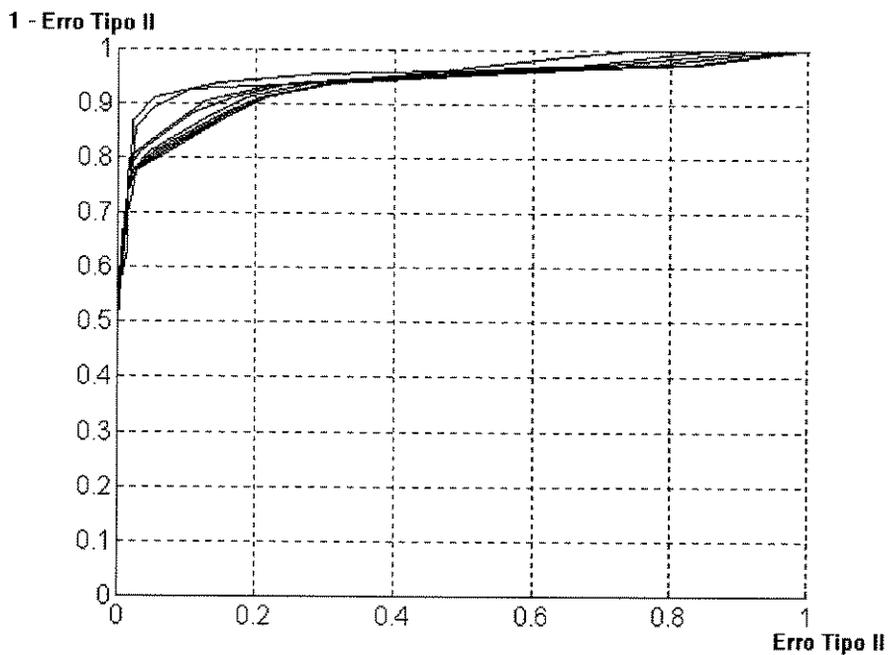


Gráfico 5.8 - Curva de Decisão Neyman-Person para o conjunto C_5

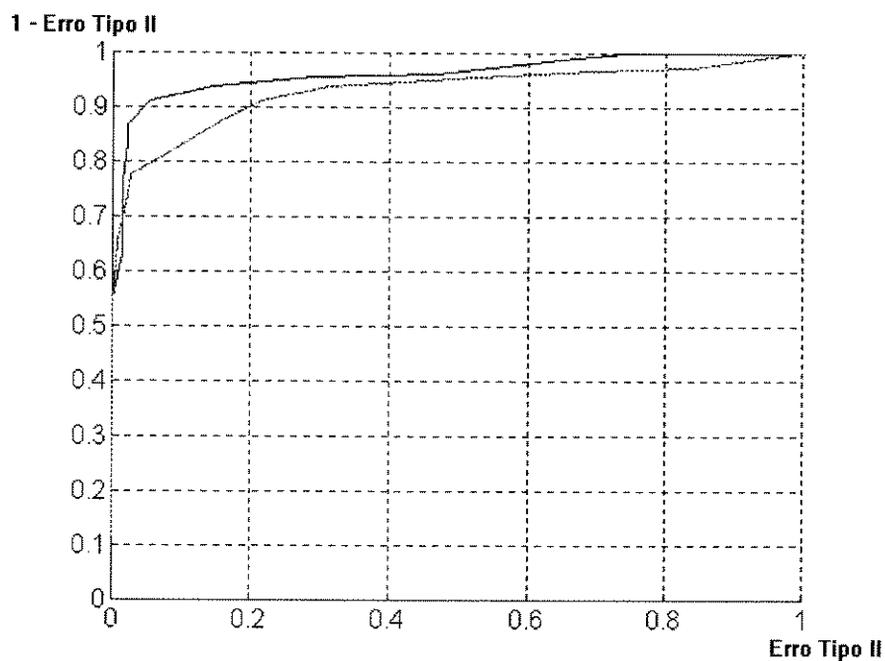


Gráfico 5.9 - Curva de Decisão Neyman-Person para o conjunto C_5 com $T_f = 0.1$ (curva mais escura) e $T_f = 0.19$ (curva mais clara).

Analisando as imagens presentes nos conjuntos associados a estes gráficos, encontra-se o porquê deste desvio.

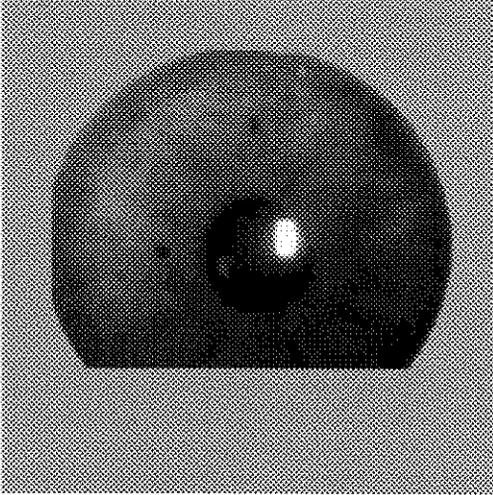


Fig. 5.1 - Imagem 1 da segunda íris do conjunto C_4

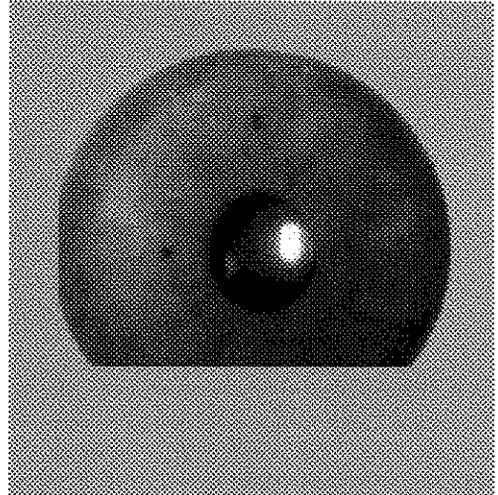


Fig. 5.2 - Imagem 2 da segunda íris do conjunto C_4

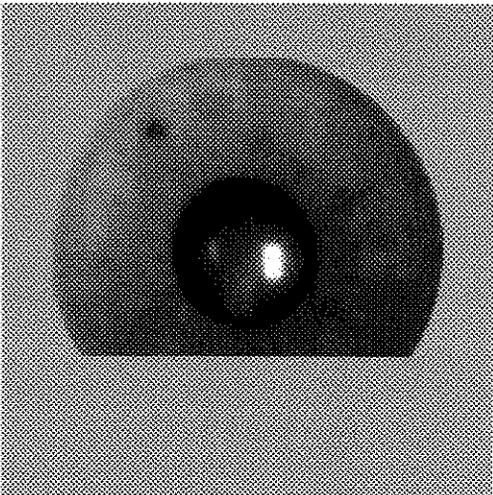


Fig. 5.3 - Imagem 1 da terceira íris do conjunto C_4

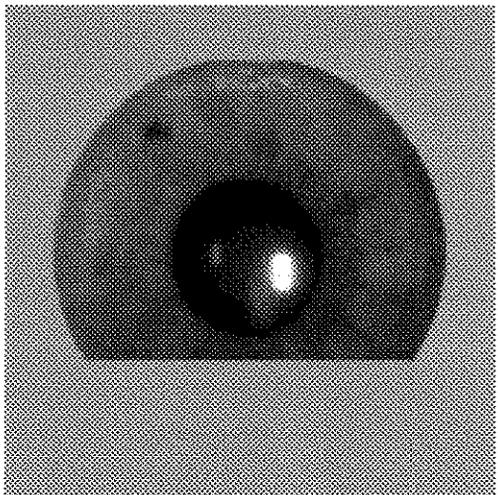


Fig. 5.4 - Imagem 2 da terceira íris do conjunto C_4

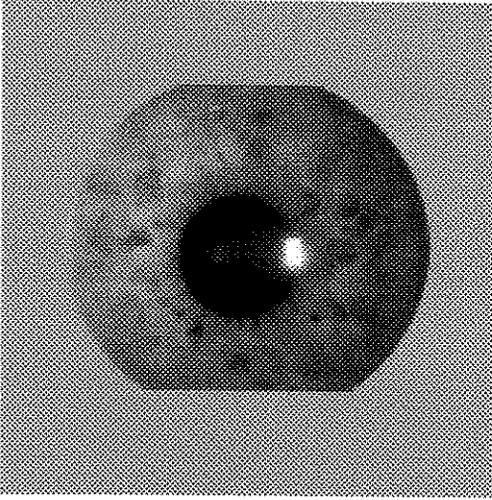


Fig. 5.5 - Imagem 1 da oitava íris do conjunto C_4

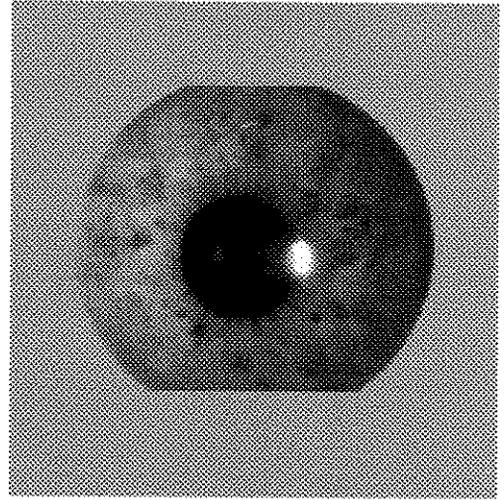


Fig. 5.6 - Imagem 2 da oitava íris do conjunto C_4

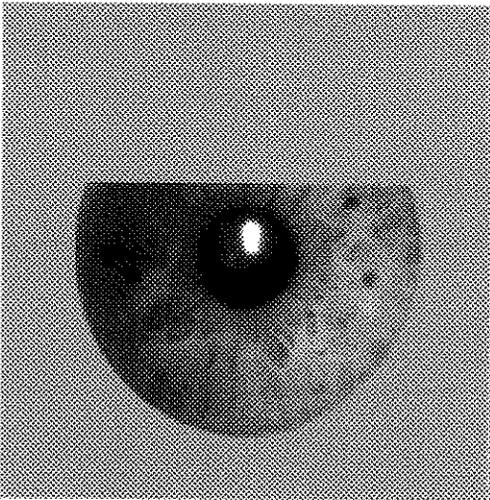


Fig. 5.7 - Imagem 1 da nona íris do conjunto C_4

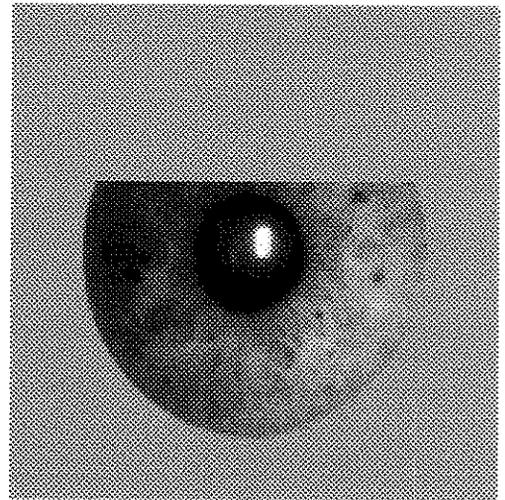


Fig. 5.8 - Imagem 2 da nona íris do conjunto C_4

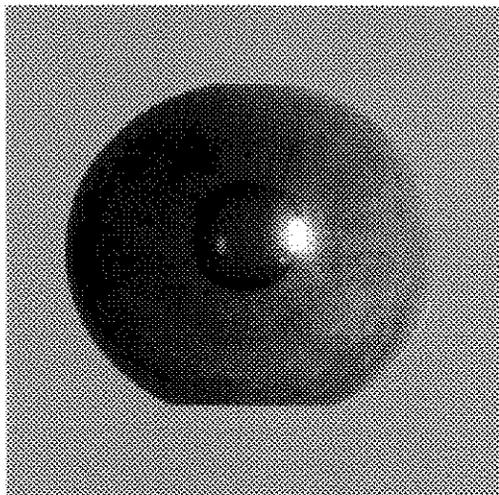


Fig. 5.9 - Imagem 1 da décima íris do conjunto C₄

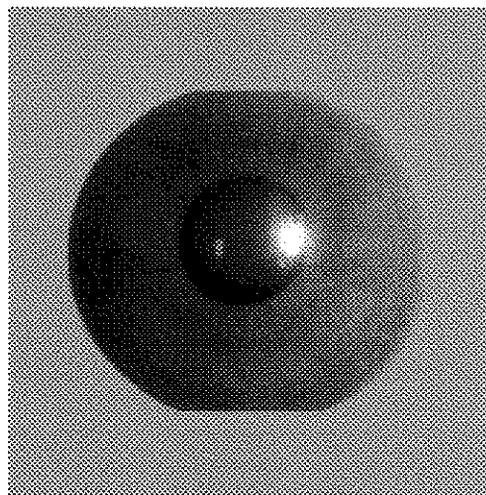


Fig. 5.10 - Imagem 2 da décima íris do conjunto C₄

As imagens acima estão presente no conjunto associado aos Gráficos 5.6 e 5.7 . Analisando-as cuidadosamente, identificam-se diferenças significativas entre as duas imagens de uma mesma íris. Entre os fatores, responsáveis por estas diferenças, destacam-se: a luminosidade, a posição do foco luminoso na imagem e problemas de registro. Estes fatores fazem com que as duas imagens sejam consideradas imagens de íris diferentes e por esta razão podem comprometer o trabalho de identificação. Porém, como pode ser observado pela série de gráficos apresentados, mesmo diante de “ruídos” como estes, a rede neuro-fuzzy se mostrou robusta, atingindo uma taxa de reconhecimento de autênticos acima de 90% para uma taxa de aceitação de impostores de 20%.

Capítulo 6 - Conclusão Geral

6.1 - Conclusão

Neste trabalho foi apresentado um sistema biométrico capaz de identificar uma pessoa a partir da imagem de sua íris. Este sistema apresenta duas diferenças básicas em relação aos sistemas propostos por Daugman [3] e Wildes [8]: (1) a extração de características é realizada através de uma decomposição em sub-bandas e (2) a identificação do indivíduo é feita por uma rede neuro-fuzzy. O sistema apresentou resultados satisfatórios, semelhantes àqueles obtidos pelos dois pesquisadores citados.

Este sistema também se mostrou bastante robusto pois foi capaz de identificar determinados indivíduos cujas imagens de treinamento e de reconhecimento apresentavam pequenas diferenças devido ao processamento manual da imagem como foi descrito no Cap. 4. A automatização deste processo, além de permitir a eliminação de uma etapa sujeita à subjetividade do administrador do sistema, poderá propiciar um ganho de desempenho. Pelo que foi mostrado no Cap. 5, este ganho poderá estar relacionado à forma como for tratado, neste processo, os problemas de registro (escala, rotação e translação), problemas de luminosidade e problemas de localização da íris na imagem. A automatização deste processo será um grande passo em direção ao aperfeiçoamento deste sistema.

Um outro fator que influenciou o desempenho deste sistema foi a escolha das sub-bandas apresentadas à rede neuro-fuzzy. De acordo com o procedimento de escolha das sub-bandas descrito no Cap. 4, priorizou-se a primeira sub-banda (LL) de baixa frequência nas direções X e Y de dimensão 128x128 pois a mesma continha a maior parte das

informações do padrão analisado. Como sugestão para um trabalho futuro, pode-se incluir no procedimento dados das outras sub-bandas e em especial na sub-banda (HH) de alta frequência em ambas as direções.

Bibliografía

- [1] - Berggren, L., 1985. "Iridology: A critical review", *Acta Ophthalmologica*, vol. 63, págs. 1-8.
- [2] - Bertillon, A., 1885. "La couleur de l'iris", *Rev. Sci.*, vol. 36, n° 3, págs. 65-73.
- [3] - Daugman, J.G., 1993. "High confidence visual recognition of persons by a test of statistical independence", *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 5, n° 11, págs. 1148-1161.
- [4] - Flom, L., and A. Safir, 1987. "Iris recognition system", U.S. Patent 4641349.
- [5] - Johnson, R.G., 1991. "Can iris patterns be used to identify people", Los Alamos National Laboratory, CA, Chemical and Laser Sciences Division, Rep. LA-12331-PR.
- [6] - Malickas, A., 1994. *Personal Communication*.
- [7] - Stotz, J., 1994. *Personal Communication*.
- [8] - Wildes, R.P., J.C. Asmuth, G. L. Green, S. C. Hsu, R. J. Kolczynski, J.R. Matey, and S. E. McBride, 1996. "A machine vision system for iris recognition", *Mach. Vision Applicat.*, vol. 9, págs. 1-8.
- [9] - Worrall, R.S., 1983. "Iridology: Diagnosis or delusion", *The Skeptica Inquirer*, págs. 23-35.
- [10] - Vogt, A., 1942, *Lehrbuch und Atlas der Spaltlampenmikroskopie des lebenden Auges. Teil 3. Iris, Glaskörper, Bindehaut*, Ente. Stuttgart.
- [11] - Vrabec, F., 1952. *Ophthalmologica*, **123**.
- [12] - Tousimis, A.J., and B.S. Fine, 1959. *Archs Ophthalmologica, New York*, **62**.
- [13] - Smelser, G.K., and T. Ishikawa, 1966. *Acta XIX Conc. Opthal. India*.
- [14] - Young, T.Y., T.W. Calvert, 1973 *Classification, Estimation, and Pattern Recognition* New York: American Elsevier.
- [15] - Fu, K. S., 1974 *Syntatic Methods in Pattern Recognition* New York: Academic Press.
- [16] - Miller, W.F., A.C. Shaw, 1968 Proc. AFIPS Fall Joint Computer Conf.
- [17] - Narasimhan, R. 1962 Report 121 (Digital Computer Laboratory, University of Illinois, Urbana, Illinois).
- [18] - Special Issues of *Pattern Recognition* on Syntatic Pattern Recognition, vol. 3, No. 4, 1971 and vol. 4, No 1, 1972.
- [19] - Zavalishin, N.V., I.B. Muchnik, 1969 *Automatika i Telemekhanika*.
- [20] - Fu, K.S., T.L. Booth, 1975 *IEEE Trans. SMC-5* (Jan. and July).
- [21] - Narasimhan, R. 1969 "On the description, generation and recognition of classes of pictures" in *Automatic Interpretation and Classification of Images* ed. by Grasselli, A. New York: Academic Press.
- [22] - Ledley, R.S., L.S. Rotolo, T.J. Golab, J.D. Jacobson, M.D. Ginsburg, J.B. Wilson, 1965 in *Optical and Electro-Optical Information Processing*, ed. by J.T. Tippett, D. Beckowitz, L. Clapp, C. Koester and A. Vanderburgh, Cambridge, MA. MIT Press.
- [23] - Shaw, A.C., 1968 - Rept. SLAC-84, Stanford Linear Accelerator Center, Stanford, CA.
- [24] - Feder, J., 1971 - *Inform. Sci.* 3.

- [50] - Blackwell, F.W. 1974 "Combining Mathematical and Structural Pattern Recognition" Proc. Second Intern. Joint Conf. Pattern Recognition, Copenhagen, Denmark.
- [51] - Grenander, V. 1967 "Syntax-Controlled Probabilities"; Tech. Rept., Division of Applied Math., Brown University, Providence, Rhode Island.
- [52] - Fu, K.S. 1972 "Syntactic pattern recognition and stochastic languages" in *Frontiers of Pattern Recognition*, ed. by S. Watanabe, New York: Academic Press.
- [53] - Kovalevsky, V.A., 1968 "Sequential Optimization in Pattern Recognition and Pattern Description", *Proc. IFIP Congress*, Amsterdam, Holland
- [54] - Fu, K.S. 1973 - "Stochastic Languages for Picture Analysis" U.S. - Japan Seminar on Picture and Scene Analysis, Kyoto, Japan.
- [55] - Fung, L.W., K.S. Fu, 1974 "Stochastic Syntactic Classification of Noisy Patterns", Proc. Second Intern. Joint Conf. Pattern Recognition, Copenhagen, Denmark.
- [56] - Dimitrov, V., 1973 "Multilayered Stochastic Languages for Pattern Recognition", Proc. First Intern. Joint Conf. Pattern Recognition, Washington, D.C.
- [57] - Lee, H.C., K.S. Fu, 1972 IEEE Trans. C-21, 660.
- [58] - Huang, T., K.S. Fu, 1972 Computer Graphics and Image Processing 1, 257.
- [59] - Clowes, M.C., 1969 "Transformational grammars and the organization of pictures" in *Automatic Interpretation and Classification of Images*, ed. by Grasselli New York: Academic Press
- [60] - Aho, A.V., T.G. Peterson, 1972 SIAM J. Comp. 1, 305.
- [61] - Churchland, P.S., and T.J. Sejnowski, 1992. *The Computational Brain*. Cambridge, MA: MIT Press.
- [62] - Levine, M., 1985. *Man and Machine Vision*. New York: McGraw-Hill.
- [63] - Marr, D., 1982. *Vision*. New York: Freeman.
- [64] - Churchland, P.S., 1986. *Neurophilosophy: Toward a Unified Science of the Mind/Brain*. Cambridge, MA: MIT Press.
- [65] - Aleksander, I., and H. Morton, 1990. *An Introduction to Neural Computing*. London: Chapman & Hall.
- [66] - Bolt, G.R., 1992. "Fault Tolerance in Artificial Neural Networks." D. Phil. Thesis, York University, Ontario. (Bolt, 1992)
- [67] - Haykin, S., 1994. *Neural Networks - A Comprehensive Foundation*. Macmillan College Publishing Company, Inc..
- [68] - Haykin, S., 1991. *Adaptive Filter Theory*, 2^a Ed. Englewood Cliffs, NJ: Prentice-Hall.
- [69] - Widrow, B., and M. E. Hoff, Jr., 1960. "Adaptive switching circuits." IRE WESCON Convention Record, págs. 96-104.
- [70] - Stent, G. S., 1973. "A physiological mechanism for Hebb's postulate of learning." *Proceedings of the National Academy of Sciences of the U.S.A.* 70, págs. 997-1001.
- [71] - Brown, T.H., E.W. Kairiss, and C.L. Keenan, 1990. "Hebbian synapses: Biophysical mechanisms and algorithms." *Annual Review of Neuroscience* 13, págs. 475-511.
- [72] - Kohonen, T., 1988. *Self-Organization and Associative Memory*, 3^a Ed.. New York: Springer-Verlag.
- [73] - Sejnowski, T.J., 1977a. "Strong covariance with nonlinearly interacting neurons." *Journal of Mathematical Biology* 4, págs. 303-321.

- [74] - Rumelhart, D.E. and D. Zipser, 1985. "Feature discovery by competitive learning." *Cognitive Science* **9**, págs. 75-112.
- [75] - Hinton, G.E., and T.J. Sejnowski, 1986. "Learning and relearning in Boltzmann machines." Em *Parallel Distributed Processing: Explorations in Microstructure of Cognition* (D.E. Rumelhart and J.L. McClelland, eds.), Cambridge, MA: MIT Press.
- [76] - Ackley, D.H., G.E. Hinton, and T.J. Sejnowski, 1985. "A learning algorithm for Boltzmann machines." *Cognitive Science* **9**, págs. 147-169.
- [77] - Sejnowski, T.J., 1977b. "Statistical constraints on synaptic plasticity." *Journal of Theoretical Biology* **69**, págs. 385-389.
- [78] - Werbos, P.J., 1974. "Beyond regression: New tools for prediction and analysis in the behavioral sciences." Ph. D. Thesis, Havard University, Cambridge, MA.
- [79] - Becker, S., 1991. "Unsupervised learning procedures for neural networks." *International Journal of Neural Systems* **2**, págs. 17-33.
- [80] - Sejnowski, T.J., and C.R. Rosenberg, 1987. "Parallel networks that learn to pronounce English text." *Complex Systems* **1**, págs 145-168.
- [81] - LeCun, Y., B. Boser, J.S. Denker, D. Henderson, R.E. Howard, W. Hubbard, and L.D. Jackel, 1990a. "Handwritten digit recognition with back-propagation network." Em *Advances in Neural Information Processing Systems 2* (D.S. Touretzky, ed.), págs. 396-404. San Mateo, CA: Morgan Kaufmann.
- [82] - Narendra, K.S., and K. Parthasarathy, 1990. "Identification and control of dynamical systems using neural networks." *IEEE Transactions on Neural Networks* **1**, págs. 4-27.
- [83] - Jacobs, R.A., and M.I. Jordan, 1991. "A competitive modular connectionist architecture." Em *Advances in Neural Information Processing Systems 3* (R.P. Lippmann, J.E. Moody, and D.S. Touretzky, eds.), págs. 767-773. San Mateo, CA: Morgan Kaufmann.
- [84] - Nowlan, S.J., and G.E. Hinton, 1991. "Evaluation of adaptive mixtures of competing experts." Em *Advances in Neural Information Processing Systems 3* (R.P. Lippmann, J.E. Moody, and D.S. Touretzky, eds.), págs. 774-780. San Mateo, CA: Morgan Kaufmann.
- [85] - Widrow, B., and S.D. Stearns, 1985. *Adaptive Signal Processing*. Englewood Cliffs, NJ: Prentice-Hall.
- [86] - Changeux, J.P., and A. Danchin, 1976. "Selective stabilization of developing synapses as a mechanism for the specification of neural networks." *Nature (London)* **264**, págs. 705-712.
- [87] - deSa, V., and D. Ballard, 1992. "Top-down teaching enables task-relevant classification with competitive learning." *International Joint Conference on Neural Networks*, vol. 3, págs. 364-371, Baltimore, MD.
- [88] - Haddad, R.A., and A.N. Akansu, 1991. "A class of Fast Gaussian Binomial Filters for Speech and Image Processing." *IEEE Trans. on Signal Processing*, vol. 39, págs. 723-727.
- [89] - Viscito, E. and J.P. Allebach, 1991. "The Analysis and Design of Multidimensional FIR Perfect Reconstruction Filter Banks for Arbitrary Sampling Lattices." *IEEE Trans. Circuits and Systems*, vol. CAS-38, págs. 29-41.

- [90] - Karlsson, G. and M. Vetterli, 1990. "Theory of Two-Dimensional Multirate Filter Banks." *IEEE Trans. ASSP*, vol. 38, págs. 925-937.
- [91] - Kwan, H.K., and Y. Cai, 1994. "A Fuzzy Neural Network and its Application on Pattern Recognition." *IEEE Trans. on Fuzzy Systems*, vol. 2, págs. 185-193.