



MINISTÉRIO DA CIÊNCIA E TECNOLOGIA
INSTITUTO NACIONAL DE PESQUISAS ESPACIAIS

INPE-7994-TDI/748

**RECONHECIMENTO DE PADRÕES INVARIANTE A ROTAÇÃO
UTILIZANDO UMA REDE NEURAL MORFOLÓGICA
NÃO-SUPERVISIONADA**

Erico Luiz Rempel

Dissertação de Mestrado em Computação Aplicada, orientada pelo Dr. Francisco Tavares Ferreira da Silva e pelo Dr. Lamartine Nogueira Frutuoso Guimarães, aprovada em 08 de março de 1999.

INPE
São José dos Campos
2000

Publicado por:

**Coordenação de Ensino, Documentação e
Programas Especiais - CEP**

Instituto Nacional de Pesquisas Espaciais - INPE

Caixa Postal 515

12201-970 - São José dos Campos - SP - Brasil

Fone: (012) 345.6911

Fax: (012) 345.6919

E-Mail: marciana@sid.inpe.br

- Solicita-se intercâmbio
- Exchange welcome
- Si sollecita intercambio
- Echange souhaité
- Mann bittet un Austausch
- Solicitamos intercambio
- Просим обмена
- 歡迎交換
- 出版物的交換

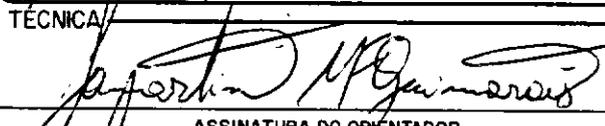
Publicação Externa - É permitida sua reprodução para interessados.

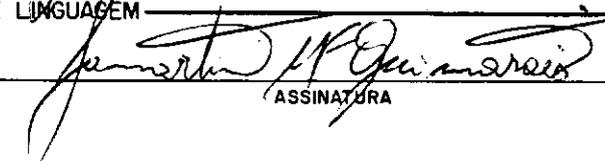
TÍTULO
Reconhecimento de Padrões Invariante a Rotação utilizando uma Rede Neural Morfológica Não-Supervisionada

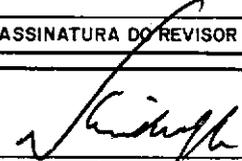
AUTOR
Erico Luiz Rempel

ORIENTADOR
Francisco de Assis Tavares Ferreira da Silva e Lamartine Nogueira Frutuoso Guimarães

ORIGEM SPG	CURSO CAP	SÉRIE	TIPO <input type="checkbox"/> TESE <input checked="" type="checkbox"/> DISSERTAÇÃO	APRESENTAÇÃO 8 / 3 / 99
Nº DE PÁGINAS 106	Nº DE FOTOS	Nº DE MAPAS	<input checked="" type="checkbox"/> LISTA DE DISTRIBUIÇÃO ANEXA	

REVISÃO TÉCNICA
CONCLUÍDA EM 21/09/00

ASSINATURA DO ORIENTADOR

REVISÃO DE LINGUAGEM
 SOLICITADA DISPENSADA
Nº _____
RECEBIDA ___/___/___ DEVOLVIDA ___/___/___

ASSINATURA DO REVISOR

AUTORIZAÇÃO FINAL
26/09/00

ASSINATURA

PALAVRAS-CHAVE
Reconhecimento de Padrões, Redes Neurais, Rotação de Imagens,
Processamento de Imagens, Invariância

PÓS-GRADUAÇÃO

___/___/___ 621.376.5 ENCAMINHADO POR	RECEBIDO ___/___/___ DEVOLVIDO ___/___/___ DEVOLVIDO POR
---	---

CONTROLE E DIVULGAÇÃO
RECEBIDO POR: Yolanda
DEVOLVIDO PARA: _____
PRONTO PARA PUBLICAÇÃO EM 25/09/00 Nº INPE 7994-TDI/748 QUANT. 9
___/___/___
ASSINATURA

OBSERVAÇÕES
Pattern Recognition, Neural Nets, Image Rotation, Image Processing, Invariance

368

INSTRUÇÕES PARA PREENCHIMENTO

- TÍTULO DA PUBLICAÇÃO: COMPLETO, EVITANDO ABREVIAR PALAVRAS.
- NOME DO AUTOR E DO ORIENTADOR: COMPLETO, POR EXTENSO, DA MESMA FORMA QUE IRÁ CONSTAR DA PUBLICAÇÃO.
- ORIGEM DA PUBLICAÇÃO: SIGLA DO ÓRGÃO DE PÓS-GRADUAÇÃO, SEGUIDA DA SIGLA DA UNIDADE, CONFORME A ESTRUTURA ORGANIZACIONAL DO INPE.
- CURSO: NOME OU SIGLA DO CURSO DE ACORDO COM A ESTRUTURA DE DIVISÃO DE TRABALHO DO INPE.
- SÉRIE: PROPOSTA PELO AUTOR, ENTRE AS APROVADAS PELO INSTITUTO.
- DATA DE APRESENTAÇÃO: COLOCAR A DATA DA APROVAÇÃO FINAL.
- NÚMERO DE PÁGINAS, FOTOS, MAPAS OU OUTROS MATERIAIS: NÚMERO EXATO DE CADA UM DELES.
- REVISÕES TÉCNICA E/OU DE LINGUAGEM: SOLICITADA OU DISPENSADA PELO ORIENTADOR OU CONFORME SUGESTÃO DA BANCA EXAMINADORA. DATAS E ASSINATURAS DE RECEBIMENTO E DEVOLUÇÃO DO ORIENTADOR E/OU DO REVISOR DE LINGUAGEM.
- AUTORIZAÇÃO FINAL: DATA E ASSINATURA DO TITULAR DA UNIDADE DIRETAMENTE SUBORDINADA À DIREÇÃO GERAL.
- PALAVRAS-CHAVE: INICIADAS PELOS CONCEITOS PRINCIPAIS, SEGUIDOS DOS TERMOS ESPECÍFICOS.
- PÓS-GRADUAÇÃO: NOME DA PESSOA QUE ENCAMINHA O TRABALHO, DATA E ASSINATURA; DATA DE DEVOLUÇÃO QUANDO HOVER.
- CONTROLE E DIVULGAÇÃO: NOME DA PESSOA QUE RECEBE O TRABALHO, DATA DE RECEBIMENTO, ASSINATURA E DATA DE DEVOLUÇÃO PARA CORREÇÕES, SE NECESSÁRIO. DATA, NÚMERO E QUANTIDADE DE CÓPIAS PARA REPRODUÇÃO.
- OBSERVAÇÕES: INFORMAÇÕES RELEVANTES.

PARA:	Sra. Marciana Leite Ribeiro	NOSSA REF. SPG-288/99	SUA REF. - 006
DE:	José Benedito dos Santos Novaes Martins	11/ 05 99	RAMAL 6845

ASSUNTO

Publicação de Dissertação

TEXTO

Estamos encaminhando, em consonância com o disposto na Resolução RE/DIR-261.1, Anexos I e II, a Dissertação do Sr. Érico Luiz Rempel, intitulada "*Reconhecimento de Padrões Invariante a Rotação Utilizando uma Rede Neural Morfológica Não Supervisionada*", por ele defendido em 08.03.99.

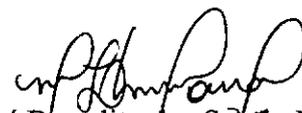
Assim, estou passando às suas mãos a versão definitiva da Dissertação, bem como disquete contendo o trabalho do Sr. Érico Luiz, que considera em condições de ser publicada, para as providências cabíveis.

Tendo em vista tratar-se de original que será, posteriormente, objeto de encadernação para depósito definitivo do SPG, solicito os seus bons officios no sentido de evitar, sempre que possível, anotações de qualquer natureza na dissertação, bem como a sua devolução a este Serviço tão logo haja sido concluído o processo de publicação.

Estou encaminhando ainda cópia de correspondência dirigida pelo Sr. Érico Luiz ao Serviço de Pós-Graduação, onde figura o seu endereço e/ou telefone, para o caso de haver necessidade de algum contato.

Colocando-nos ao seu dispôr para outros esclarecimentos que se fizerem necessários, subscrevo-me.

Atenciosamente,


José Benedito dos Santos Novaes Martins
Chefe do Serviço de Pós-Graduação

/caa.

FICHA DE ATUALIZAÇÃO DE ENDEREÇO

Nome: ERICO LUIZ REMPEL			Registro: 29003	
Curso: COMPUTAÇÃO APLICADA				
Rua: R. ANTARES		Nº 135	Complemento: 103 F	
Bairro: JDM. SATELITE	Cidade: S. J. CAMPOS			Estado: SP
CEP:	☎ (012) 331 1660	Recados: ☎ 345 6766	Fax:	

erico@pgrad.unp.br

ENDEREÇO DE CONTATO

Nome da pessoa a ser contatada(*):				
Relação de parentesco				
Rua:		Nº	Complemento:	
Bairro:	Cidade:			Estado:
CEP:	☎ ()	Recados: ☎	Fax:	

(* Preferentemente em endereço diferente do fornecido acima.

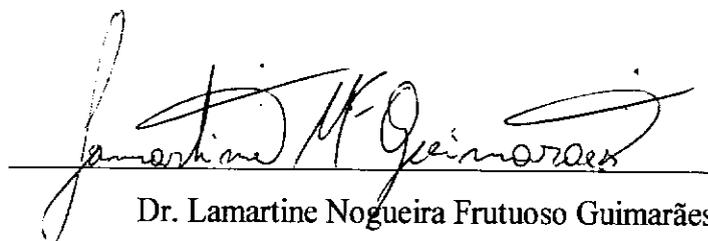
Lista de Distribuição de Cópias da Dissertação de Mestrado

Erico Luiz Kempel

Dr. Abraham Chian-Long Chian – DGE/INPE

Dr. Reinaldo Roberto Rosa – LAC/INPE

Dr. Gerald Jean Francis Banon – DPI/INPE



Dr. Lamartine Nogueira Frutuoso Guimarães

Orientador

São José dos Campos

Outubro de 2000



MINISTÉRIO DA CIÊNCIA E TECNOLOGIA
INSTITUTO NACIONAL DE PESQUISAS ESPACIAIS

INPE-7994-TDI/748

**RECONHECIMENTO DE PADRÕES INVARIANTE A ROTAÇÃO
UTILIZANDO UMA REDE NEURAL MORFOLÓGICA
NÃO-SUPERVISIONADA**

Erico Luiz Rempel

Dissertação de Mestrado em Computação Aplicada, orientada pelo Dr. Francisco Tavares Ferreira da Silva e pelo Dr. Lamartine Nogueira Frutuoso Guimarães, aprovada em 08 de março de 1999.

INPE
São José dos Campos
2000

Publicado por:

**Coordenação de Ensino, Documentação e
Programas Especiais - CEP**

Instituto Nacional de Pesquisas Espaciais - INPE

Caixa Postal 515

12201-970 - São José dos Campos - SP - Brasil

Fone: (012) 345.6911

Fax: (012) 345.6919

E-Mail: marciana@sid.inpe.br

- Solicita-se intercâmbio
- Exchange welcome
- Si sollecita intercambio
- Echange souhaité
- Mann bittet un Austausch
- Solicitamos intercambio
- Просим обмена
- 歡迎進行交換
- 出校為交換的申請

Publicação Externa - É permitida sua reprodução para interessados.

MINISTÉRIO DA CIÊNCIA E TECNOLOGIA
INSTITUTO NACIONAL DE PESQUISAS ESPACIAIS

INPE-7994-TDI/748

**RECONHECIMENTO DE PADRÕES INVARIANTE A ROTAÇÃO
UTILIZANDO UMA REDE NEURAL MORFOLÓGICA
NÃO-SUPERVISIONADA**

Erico Luiz Rempel

Dissertação de Mestrado em Computação Aplicada, orientada pelo Dr. Francisco Tavares Ferreira da Silva e pelo Dr. Lamartine Nogueira Frutuoso Guimarães, aprovada em 08 de março de 1999.

INPE
São José dos Campos
2000

621.376.5

REMPEL, E. L.

Reconhecimento de padrões invariante a rotação utilizando uma rede neural morfológica não-supervisionada /

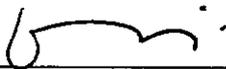
E. L. Rempel. - São José dos Campos: INPE, 1999.

106p. - (INPE-7994-TDI/748).

1.Reconhecimento de padrões. 2.Redes neurais. 3.Rotação de imagens. 4.Processamento de imagens. 5.Invariância. I.Título.

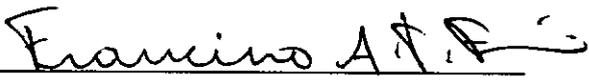
Aprovado pela Banca Examinadora em
cumprimento a requisito exigido para a
obtenção do Título de Mestre em
Computação Aplicada.

Dr. Atair Rios Neto



Presidente

Dr. Francisco de Assis Tavares F. da Silva



Orientador

Dr. Lamartine N.F. Guimarães



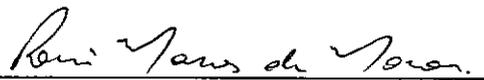
Orientador

Dr. Camilo Rodrigues Neto



Membro da Banca

Dr. Ronei Marcos de Moraes



Membro da Banca
Convidado

Candidato: Érico Luiz Rempel

AGRADECIMENTOS

Agradeço ao Dr. Francisco de A. Tavares F. da Silva pela orientação, apoio, incentivo e correções de grande valia durante todo o desenvolvimento deste trabalho;

Agradeço ao Dr. Lamartine N. F. Guimarães pela orientação, ajuda, conselhos e sugestões dadas;

Agradeço ao Dr. Gerald J. Francis Banon pela atenção dispensada, esclarecimentos e materiais concedidos;

Agradeço ao Dr. Celso L. Mendes pelo tempo e ajuda dispensados;

Agradeço ao CNPq pelo apoio financeiro;

Agradeço a Amândio e Eloísa pela manutenção e suporte no laboratório;

Agradeço aos colegas de curso, principalmente ao Cláudio Correa pela ajuda na elaboração do estilo L^AT_EX deste documento;

Agradeço à minha família pelo apoio em todos os momentos;

Agradeço a Deus pela saúde, direção e capacitação para terminar este trabalho.

RESUMO

Neste trabalho é apresentado um modelo de reconhecimento de imagens digitais bidimensionais com capacidade de detecção invariante a rotação e translação usando uma rede neural morfológica não-supervisionada. Utilizando a característica de invariância a translação presente nos operadores da Morfologia Matemática, é sugerido um mecanismo de casamento morfológico de padrões baseado na comparação da imagem de entrada com um conjunto de elementos estruturantes rotacionados, permitindo a detecção invariante a rotação e translação sem necessidade de uma etapa de segmentação. Um conjunto de testes foram realizados utilizando imagens binárias e em escala de cinza.

ROTATION INVARIANT PATTERN RECOGNITION USING A MORPHOLOGICAL UNSUPERVISED NEURAL NETWORK

ABSTRACT

This work presents the development of a rotation and translation invariant 2-dimensional image detection system using an unsupervised morphological neural network. Utilizing the translation invariance property of the operators of Mathematical Morphology, it is suggested a morphological pattern-matching mechanism based on the comparison between the input image and a set of rotated structuring elements, making possible rotation and translation invariant image detection without the requirement of a segmentation stage. A series of experiments were done using binary and gray scale images.

SUMÁRIO

Pág.

LISTA DE FIGURAS

LISTA DE TABELAS

LISTA DE SIGLAS E/OU ABREVIATURAS

CAPÍTULO 1 INTRODUÇÃO	19
1.1 CONSIDERAÇÕES GERAIS	19
1.2 REVISÃO BIBLIOGRÁFICA	20
1.3 PROPOSTA DO TRABALHO	25
1.4 ESTRUTURA DO TEXTO	25
CAPÍTULO 2 CONCEITOS BÁSICOS	27
2.1 INTRODUÇÃO	27
2.2 REDES NEURAIS ARTIFICIAIS	27
2.2.1 Formas de Aprendizado	29
2.2.1.1 Aprendizado Competitivo	30
2.3 MORFOLOGIA MATEMÁTICA PARA IMAGENS BINÁRIAS ..	30
2.3.1 Conceitos Fundamentais	31
2.3.1.1 Translado por um vetor	31
2.3.1.2 Transposto (reflexão) em relação à origem	31
2.3.1.3 Complemento	31
2.3.2 Operadores Elementares	31
2.3.2.1 Dilatação	31
2.3.2.2 Erosão	32
2.3.3 Detecção de Padrões Binários por Morfologia Matemática	32
2.3.3.1 <i>Hit</i> ou <i>Miss</i>	33
2.4 MORFOLOGIA MATEMÁTICA EM NÍVEIS DE CINZA	34

2.4.1	Conceitos Básicos	34
2.4.1.1	Operadores Elementares da Morfologia Matemática	34
2.4.1.2	Janela	35
2.4.2	Casamento de Padrões por Morfologia Matemática	35
2.4.2.1	Operador Morfológico de Casamento de Padrões	35
2.4.2.2	Operador de Detecção com Folga	38
2.5	OPERAÇÕES MORFOLÓGICAS INVARIANTES A ROTAÇÃO	38

**CAPÍTULO 3 REDE MORFOLÓGICA NÃO-SUPERVISIONADA
(RMNS) 43**

3.1	INTRODUÇÃO	43
3.2	ARQUITETURA	44
3.3	OPERAÇÃO DA REDE MORFOLÓGICA NÃO-SUPERVISIONADA	46
3.3.1	Inicialização	46
3.3.2	Treinamento	46
3.3.3	Ativação	48

CAPÍTULO 4 O MODELO PROPOSTO 51

4.1	INTRODUÇÃO	51
4.2	DESCRIÇÃO DO MODELO	51
4.2.1	Detecção de Imagens Binárias	52
4.2.2	Detecção de Imagens em Níveis de Cinza	54
4.3	APLICAÇÃO DO MODELO À RMNS	56
4.3.1	Detalhes de Implementação	60
4.3.2	Insensibilidade ao Contexto	63
4.3.3	Comentários Sobre Possíveis Ambigüidades	65
4.3.4	Comparação entre RMNS e ART1 (<i>Adaptive Resonance Theory</i>) com Invariância a Rotação	66

CAPÍTULO 5 EXPERIMENTOS 69

5.1	INTRODUÇÃO	69
5.2	ESTUDO DE CASO PARA IMAGENS BINÁRIAS	69
5.2.1	Comparação com a rede ART1	69
5.2.2	Invariância ao Contexto	75
5.3	ESTUDO DE CASO PARA IMAGENS EM NÍVEIS DE CINZA	76
5.4	VARIAÇÃO DO LIMIAR DE RUÍDO	79
5.5	VARIAÇÃO DO INTERVALO DE ROTAÇÃO	82
CAPÍTULO 6 CONCLUSÕES		85
REFERÊNCIAS BIBLIOGRÁFICAS		89
APÊNDICE A ALGORITMO DA REDE ART1		99
APÊNDICE B ART1 INVARIANTE		103

LISTA DE FIGURAS

	Pág.
2.1	Erosão como mecanismo de detecção 33
2.2	Detecção de padrões por <i>Hit</i> ou <i>Miss</i> 34
2.3	Esquema para casamento de padrões 36
2.4	Abertura por múltiplos elementos estruturantes 39
3.1	Neurônio morfológico 45
3.2	Exemplo de RMNS 46
4.1	Diagrama morfológico do modelo para detecção de imagens binárias . 54
4.2	Diagrama morfológico do modelo para detecção em níveis de cinza . 56
4.3	Exemplo de mapa morfológico invariante a rotação. 57
4.4	Modelo de invariância a rotação aplicado à RMNS. 58
4.5	Modelo morfológico da RMNS com invariância a rotação 59
4.6	Padrão de referência e imagem de busca em níveis de cinza. 61
4.7	Rotação do padrão de referência. 61
4.8	Rotação de padrão de referência com forma circular. 62
4.9	Rotação de padrão de referência sem perda de informação 63
4.10	Caracter chinês 64
4.11	Padrão procurado inserido em outro objeto 64
4.12	Extração de objetos por múltiplas janelas 65
4.13	Exemplos de ambigüidades 65
5.1	Ativação da rede ART1 com entradas não rotacionadas 70
5.2	Ativação da rede ART1 com entradas rotacionadas 71
5.3	Efeito da redução da taxa de vigilância na rede ART1. 71
5.4	Problemas com ruído “sal” na rede ART1. 71
5.5	Pesos do Mapa Morfológico após o treinamento. 72
5.6	Saídas da RMNS para entradas ruidosas e rotacionadas. 73
5.7	Aumento do limiar de ruído na RMNS. 73

5.8	Comportamento da RMNS para entradas com ruído “sal”	74
5.9	Comportamento da RMNS com limiar muito grande.	74
5.10	Comportamento da ART1 com taxa de vigilância pequena	74
5.11	Ferramentas usadas para o treinamento da RMNS.	75
5.12	Saída da RMNS treinada com uma chave de fenda.	75
5.13	Saída da RMNS treinada com uma chave de boca.	75
5.14	Saída da RMNS treinada com um alicate.	76
5.15	Conjunto de padrões de referência extraídos do MIT Face DataBase .	77
5.16	Conjunto de testes.	77
5.17	Ativação com <i>joel3</i>	78
5.18	Ativação com <i>rob2</i>	78
5.19	Ativação com <i>thad3</i>	78
5.20	Ativação com <i>david2</i>	79
5.21	Ativação com <i>wave</i>	79
5.22	Conjunto de padrões de referência.	80
5.23	Ativação com <i>woman1</i> rotacionado por 15°	81
5.24	Quantidade de erros em relação à variação do limiar	82
5.25	Quantidade de erros em relação à variação do intervalo de rotação . .	83
A.1	Arquitetura da rede ART1	101
B.1	Rede ART1 invariante a rotação e translação	106

LISTA DE TABELAS

	Pág.
5.1 Variação do limiar	81
5.2 Variação do intervalo de rotação na RMNS	82

LISTA DE SIGLAS E/OU ABREVIATURAS

ADALINE - *Adaptive Linear Neuron*

ART1 - *Adaptive Resonance Theory*

GCF - *Geometrical Correlation Function*

MAT - *Morphological Autocorrelation Transform*

MIT - *Massachusetts Institute of Technology*

OMCP - Operador Morfológico de Casamento de Padrões

RMNS - Rede Morfológica Não-Supervisionada

RNA - Rede Neural Artificial

CAPÍTULO 1

INTRODUÇÃO

1.1 CONSIDERAÇÕES GERAIS

O processamento de imagens digitais é uma área de pesquisa que tem experimentado um grande crescimento na última década. Dentro desta área, grandes esforços têm sido aplicados na tarefa de reconhecimento de padrões visuais.

O propósito do reconhecimento de padrões, conforme [1], é a realização do mapeamento de um espaço de representação para um espaço de interpretação, e um aspecto importante é a determinação do operador que fará este mapeamento. As técnicas de reconhecimento de padrões atribuem um objeto ou evento a uma dentre várias categorias pré-estabelecidas. Jain [2] caracteriza um sistema de reconhecimento de padrões como sendo uma regra de decisão automática que transforma medições de um padrão em atribuições de classe. Os padrões são reconhecidos com base nas características extraídas destas medições.

Na tarefa de reconhecimento de padrões, várias fontes de ruído podem distorcer os dados de entrada, podendo haver significativa variação dentre os padrões pertencentes a uma mesma categoria, implicando em diferentes características ou medições sobre estes padrões. Esta diferença deve ser pequena comparada com a diferença existente entre padrões pertencentes a categorias diferentes [2].

Dentro da área de reconhecimento de imagens, uma série de dificuldades são encontradas. Além de os padrões de entrada a serem detectados poderem estar corrompidos por ruído, muitas vezes se encontram conectados com outros objetos. Além disso, os objetos podem se encontrar em diversos tamanhos e posições, podendo haver necessidade de o reconhecimento ser invariante a transformações de escala, translação e rotação. Dentre estes problemas, o mais complicado de se resolver, segundo [3], é o reconhecimento invariante a rotação.

A invariância a transformações em escala, rotação e translação tem recebido grande atenção nas últimas décadas. Conforme [4], o problema é geralmente abordado através de dois processos. Primeiro, um conjunto de características com as propriedades invariantes desejadas é extraído da imagem; em seguida, as características extraídas são submetidas a um classificador que determina a qual classe estes dados pertencem. Uma forma alternativa de se lidar com o problema é construir um classificador que seja invariante a estas transformações, sendo capaz de reconhecer os padrões independente de assistência externa. Recentes trabalhos demonstram que a área permanece aberta a pesquisas [5].

A Morfologia Matemática pode ser empregada com eficiência em tarefas de processamento de imagens. As operações são realizadas por meio de uma sondagem e transformações de um objeto por padrões de formas pré-definidas. Esta abordagem é bastante adequada ao reconhecimento por casamento de padrões [6], sendo computacionalmente vantajosa pelo fato de que seu formalismo permite simplificações algébricas que possibilitam aumentos sensíveis de velocidade de processamento [7].

1.2 REVISÃO BIBLIOGRÁFICA

A seguir é feita uma breve discussão a respeito das dificuldades encontradas no reconhecimento de padrões invariante a transformações de translação, rotação e escala. No levantamento bibliográfico apresentado, são destacadas algumas das técnicas de reconhecimento invariante de padrões encontradas na literatura consultada.

Muitas abordagens para detecção de imagens invariante a transformações baseiam-se em técnicas de casamento de padrões. Casamento de padrões é o processo de se determinar a posição de uma máscara ou modelo dentro de uma imagem, comparando a diferença entre duas janelas de casamento por meio de medidas de similaridade, tais como a soma das diferenças absolutas de pontos ou a correlação [8]. Em vários métodos, a invariância é atingida através do casamento de características invariantes da figura, em vez da própria figura. Exemplo disso são as expansões circulares harmônicas [9]; [10] e [11]. Neste caso, os pontos de uma imagem de entrada são

representados em coordenadas polares $f(r, \theta)$ e os coeficientes de Fourier dos círculos de raio r são calculados, sendo usada uma das componentes harmônicas da imagem teste para a correlação com a imagem de referência. O módulo da correlação entre uma imagem e sua n – ésima componente harmônica é invariante a rotação.

Invariância a escala, translação e rotação pode ser atingida através da transformada Fourier-Mellin. A imagem de entrada é mapeada em coordenadas polares e convertida para escala logarítmica (*complex-log mapping*), o que transforma variações na rotação e na escala em translações. A magnitude da transformada de Fourier é invariante a translação e pode ser usada para gerar a saída invariante às três transformações [12]; [13] e [14]. Esta forma conjunta de distorção logarítmica com transformada de Fourier é equivalente à transformada Mellin. No entanto, por este procedimento, informações sobre a posição, tamanho e orientação do objeto são perdidas [15].

Segundo [16] e [17], as características invariantes mencionadas podem não descrever uma imagem de maneira singular, podendo não haver uma correspondência um-para-um entre as imagens e estas características. O uso exclusivo da magnitude do espectro de Fourier, com ausência da informação sobre fase, pode gerar falsos reconhecimentos.

A transformada Radon tem sido utilizada para reconhecimento invariante a rotação e translação [18] ou em conjunto com outras técnicas para extração de características invariantes a rotação, translação e escala [19]. A transformada Radon consiste em um conjunto de projeções unidimensionais de uma função bidimensional. A projeção de uma imagem em um ângulo θ é a soma de todas as intensidades que se encontram ao longo de um conjunto de linhas paralelas com ângulo $\theta + \pi/2$ com relação ao eixo x . Os dados extraídos da transformada podem ser apresentados na forma de um sinograma, um diagrama onde cada linha contém os valores da projeção da imagem em um determinado ângulo. Translações da imagem são refletidas como deslocamentos horizontais de cada projeção no sinograma. Rotações ocasionam mudanças cíclicas na posição das linhas. Uma correlação invariante a translação e

rotação pode ser obtida através de uma normalização dos dados do sinograma [18]. Métodos como descritores Fourier [20]; [21] e abordagens sintáticas [22] e [23] detectam padrões baseados nos contornos das imagens. Os descritores Fourier trabalham com um conjunto de pontos extraídos da silhueta de um objeto. Mapeando-se os pontos da imagem em coordenadas polares, tem-se uma silhueta cujos pontos formam um sinal periódico unidimensional. É possível a obtenção de invariância a rotação, translação e escala através de normalizações dos coeficientes de Fourier obtidos. O reconhecimento sintático representa a imagem através de segmentos de reta e curvas. Em [22], a entrada do sistema consiste de uma codificação linear do contorno de uma figura e a saída é uma sentença com descrições de alto nível produzidas por uma gramática.

O uso de técnicas de casamento estrutural permite o reconhecimento de padrões insensível a transformações geométricas e radiométricas. As imagens são representadas por um conjunto de atributos que descrevem simbolicamente os objetos contidos na imagem e como eles se encontram relacionados espacialmente [24]. Em [25] é descrito um método para casamento estrutural de imagens de satélite. Primeiramente as imagens são segmentadas e os objetos extraídos são representados por atributos como descritores de forma e tons de cinza. Para cada imagem é feita uma descrição na forma de grafos direcionados com base nas relações espaciais entre os objetos. Um grafo de associação é gerado a partir das correspondências entre nós e arestas de duas descrições e o casamento é calculado através de técnicas de relaxação.

Momentos e funções de momentos geométricos têm sido usados com muita frequência para extração de características invariantes a escala, rotação e translação [26]; [27]; [8], [21]; [28]; [29]; [30]; [31] e [32]. Reeves et al. [21] apresentam os resultados da comparação entre o uso de descritores Fourier e momentos para o reconhecimento de aeronaves. Wang et al. [30] propõem o uso dos dois métodos em conjunto para a tarefa de classificação de caracteres chineses. Varella et al. [29] utilizam os valores de funções de momentos invariantes no treinamento de redes neurais para a classificação de caracteres, alertando ao fato de que as propriedades de invariância dos momentos

apenas se confirmam quando aplicados a funções contínuas. Para imagens digitais os resultados mostram variações expressivas.

Dentre as técnicas conexionistas encontram-se as redes de terceira ordem, que codificam invariância a escala, rotação e translação em sua própria estrutura [33]; [34]; [35]; [4]; [36] e [37]. Nestas redes, em vez de se ter um peso para cada entrada, *triplas* de *pixels* recebem um peso. Para uma imagem representada pelos valores X de seus *pixels*, a função de ativação de um neurônio i , denotada a seguir por O_i , opera sobre produtos de valores de X ,

$$O_i = f \left(\sum_j \sum_k \sum_l W_{ijkl} X_j X_k X_l \right); \quad (1.1)$$

onde f é uma função de ativação não linear e W_i é um peso do nó i para cada uma das *triplas*. A invariância é resultado da configuração dos pesos sinápticos. Os valores de quaisquer dois pesos W_{ijkl} e W_{ipqr} são igualados sempre que os triângulos com aqueles *pixels* como vértices possuírem os mesmos ângulos internos [37] e [36]. A mais severa limitação deste método é que, conforme aumenta o tamanho da entrada, o número de possíveis combinações de *triplas* também aumenta. Fukumi et al. [5] afirmam que em aplicações práticas, o método causa uma explosão combinatorial de conexões. Kwon et al. [38] sugerem o uso de mapeamento *complex-log* e redes de segunda ordem invariates a translação para obtenção de invariância a escala e rotação com redução no número de conexões.

A rede neocognitron [39] foi desenvolvida com o intuito de ser um sistema com inspiração biológica para reconhecimento de padrões binários invariante a distorções, incluindo variações de escala e translação. Trata-se de uma rede com múltiplas camadas em que os neurônios são treinados para representar pequenas regiões da imagem de treinamento. Estas regiões contém traços característicos do padrão de entrada que são replicados em diferentes posições nos nós da rede. Em [40] é feita uma expansão da rede neocognitron, para a detecção invariante a rotação por meio de um conjunto de camadas responsáveis pela representação rotacionada da infor-

mação contida nos neurônios da rede. Isto confere à rede a capacidade de detecção invariante a escala, distorções, translações e rotações de aproximadamente 60° . Para rotações maiores, pode-se treinar a rede com padrões rotacionados.

Abordagens conexionistas que usam extratores de características invariantes mapeiam as imagens em vetores que são reconhecidos por redes neurais com aprendizado supervisionado. Tais características podem ser obtidas por métodos já mencionados, como momentos [29] e [36], ou transformada Fourier-Mellin [12] e [41]; ou através de uma rede invariante como pré-processamento. Widrow et al. [42] apresentam um modelo que utiliza uma rede invariante baseada em agrupamentos de neurônios do tipo *Adaptive Linear Neuron* (ADALINE) para gerar uma saída invariante a translação e rotação que serve de entrada para uma rede multi-camadas de reconhecimento. Na rede de pré-processamento, cada agrupamento de ADALINE's possui um conjunto de pesos inicializados aleatoriamente que são replicados em diferentes orientações e translações. Cada agrupamento produz um *pixel* de saída invariante. Srinivasa e Jouaneh [43] aplicaram o modelo na rede *Adaptive Resonance Theory* (ART1) de Carpenter e Grossberg [44], verificando, contudo, a sua dificuldade em responder corretamente a entradas ruidosas. Em [45] os autores apresentaram uma variação do sistema, em que os pesos da camada invariante são inicializados com os valores treinados pela rede ART1, sendo obtidos resultados satisfatórios. Hosokawa et al. [46] e Fukumi et al. [47] descrevem uma variação do sistema de Widrow [42], usando neurônios com funções sigmóides. Fukumi et al. [47] utilizam o modelo em uma aplicação de reconhecimento de moedas.

Algumas soluções utilizam redes neurais para aprender um conjunto de exemplos de dados rotacionados em diversos ângulos, aproveitando a capacidade de generalização das redes para realizar as aproximações corretas. Ranganath e Arun [48] e Howell e Buxton [49] treinaram redes supervisionadas de função de base radial com diversas imagens rotacionadas e com variações de escala para a tarefa de reconhecimento de faces.

Algumas das abordagens conexionistas mencionadas requerem que se especifique o

intervalo de inclinação para os padrões de treinamento serem classificados de maneira invariante à rotação. Em geral, um padrão é treinado com diversas inclinações. Quanto menor o intervalo de inclinação, melhor a rede se comporta com entradas rotacionadas; porém, mais complexa ela se torna, havendo aumento no número de neurônios ou redes envolvidas no processo.

Em [50] é apresentado um sistema de reconhecimento invariante de imagens baseado em Morfologia Matemática, utilizando um descritor de formas chamado de *morphological autocorrelation transform* (MAT). A MAT de uma imagem é formada por um conjunto de funções de correlação geométrica (GCF's) que definem sua covariância morfológica em uma direção específica. O sistema invariante a rotação, translação e escala apresentado envolve uma etapa de pré-processamento para segmentação, localização da área de interesse e normalização da escala da imagem de entrada. Os valores da MAT da entrada são obtidos e comparados com os valores extraídos do padrão de referência.

1.3 PROPOSTA DO TRABALHO

O presente trabalho tem por objetivos:

- Implementar um modelo de reconhecimento morfológico de imagens bidimensionais invariante a rotação e translação baseado no conceito de replicação de imagens apresentado por [45], usando a Rede Morfológica Não-Supervisionada (RMNS) [7].
- Comparar os resultados obtidos pelo modelo aplicado à RMNS com os obtidos pelo sistema descrito em [45].
- Verificar o comportamento do sistema com entradas em níveis de cinza.

1.4 ESTRUTURA DO TEXTO

Este trabalho está dividido em 6 capítulos. O capítulo 2 contém uma introdução sobre Redes Neurais Artificiais e apresenta conceitos básicos de Morfologia Matemática, além de apresentar alguns trabalhos sobre o uso de Morfologia Matemática

em tarefas que exigem invariância a rotação. Neste capítulo também é descrito o Operador de Detecção com Folga. Este operador é utilizado pela Rede Morfológica Não Supervisionada, descrita no capítulo 3. O capítulo 4 apresenta o modelo proposto para reconhecimento morfológico invariante a rotação e translação, mostrando a forma de implementação deste modelo na RMNS. O capítulo 5 contém os resultados de uma série de experimentos realizados com o modelo implementado. As conclusões do trabalho se encontram no capítulo 6.

Os Apêndices A e B contêm a descrição da rede neural ART1, e de uma versão da rede alterada de maneira a englobar invariância a rotação e translação. Os algoritmos extraídos de [51] e [45], respectivamente, foram implementados em Matlab para a comparação de resultados com o modelo implementado sobre a RMNS.

CAPÍTULO 2

CONCEITOS BÁSICOS

2.1 INTRODUÇÃO

Neste capítulo é feita uma introdução sobre Redes Neurais Artificiais e sobre Morfologia Matemática, sendo dada ênfase na descrição do Operador de Detecção com Folga, que é utilizado pela Rede Morfológica Não-Supervisionada, descrita no capítulo 3. Ao final do capítulo são apresentados alguns trabalhos referentes ao uso de Morfologia Matemática para processamento de imagens invariante a rotação.

2.2 REDES NEURAIS ARTIFICIAIS

Redes Neurais Artificiais (RNA's) podem ser concebidas como uma arquitetura composta por uma grande quantidade de unidades simples ativas interconectadas, as quais são caracterizadas por uma pequena memória local, e processamento paralelo distribuído, cujas bases de arquitetura e funcionamento se assemelham a estruturas de redes dos tecidos nervosos cerebrais [52]. Pham e Liu [53] definem as RNA's como sendo modelos computacionais do cérebro. Embora ainda muito distantes da complexidade de um modelo real do cérebro, este conceito é aceitável uma vez que as RNA's são projetadas com base nos modelos de neurônios biológicos apresentados pela neurociência. A neurocomputação envolve o desenho de máquinas com circuitos inspirados em redes neurais cerebrais, mas que não precisam emulá-las fielmente [54].

Existem diferentes tipos de neurônios no cérebro humano, sendo combinados em redes responsáveis por tarefas distintas. McClelland e Plunkett [55] propõem que o cérebro seja visto como uma rede de redes, em vez de um único sistema homogêneo. Cada rede pode possuir uma “arquitetura” diferente, baseada em um conjunto de neurônios com propriedades distintas, características de uma parte específica do cérebro. Esta diversidade de modelos foi o que inspirou a elaboração de vários

tipos de neurônios artificiais, sendo que o comportamento de uma rede depende grandemente da maneira como os neurônios se conectam e interagem entre si.

Na última década a atividade no campo de Redes Neurais Artificiais tem aumentado consideravelmente [56], [53] e [57]. Este aumento de interesse tem sido impulsionado por sucessos obtidos tanto no campo teórico como no prático [58]. RNA's constituem modelos computacionais adequados ao processamento paralelo distribuído, e são capazes de aprender a partir de exemplos. Seu uso permite que o comportamento de sistemas complexos seja modelado e previsto, e que um controle apurado seja alcançado através do treinamento, sem que haja necessidade de informações iniciais sobre a estrutura ou parâmetros do sistema [53].

Esta capacidade de aprender através de exemplos, podendo modificar seu comportamento em resposta ao ambiente, é a principal característica das redes neurais [58], sendo a maior vantagem em relação a técnicas de programação convencional. Para se resolver um problema usando computação programada é necessária a criação de algoritmos e a codificação detalhada dos procedimentos, o que em geral consome tempo e é uma tarefa altamente sujeita a erros. Por meio de RNA's, torna-se possível a resolução de uma série de classes de problemas, possivelmente ainda sem algoritmos existentes, através de modelos conexionistas já conhecidos, que resultam em redução do código gerado [59].

Os diversos modelos de redes neurais artificiais possuem diferentes características que se aplicam à resolução de problemas que incluem reconhecimento de padrões, controle, previsão, otimização, aproximação de funções, dentre outros [60], [57] e [7]. O desenvolvimento atual de *hardwares* específicos para o processamento neural em paralelo permite um alto desempenho, necessário em aplicações que exijam respostas em tempo real [60].

Uma vez treinada, a resposta de uma rede pode ser insensível a um certo grau de variação na sua entrada. Esta capacidade de generalização, ou habilidade para lidar com ruído e distorções é essencial para o reconhecimento de padrões [58].

2.2.1 Formas de Aprendizado

Nos modelos de neurônios artificiais, os impulsos de entrada são operados por um conjunto de pesos sinápticos e o resultado é submetido a uma função que determina qual será a saída do neurônio. O aprendizado em redes neurais consiste na forma como os pesos sinápticos são ajustados de maneira a se alcançar a configuração necessária para o desempenho da tarefa desejada. Durante o treinamento, um conjunto de dados é apresentado à rede para que seus pesos sejam adaptados segundo uma regra de aprendizado.

De acordo com a forma de treinamento, as RNA's podem ser divididas em duas categorias principais: treinamento supervisionado e não-supervisionado (auto-organização) [61]. Alguns autores, como Hecht-Nielsen [59], incluem ainda o treinamento graduado (reforço).

No treinamento supervisionado, a rede recebe uma seqüência de exemplos de pares de entradas e saídas desejadas. Para cada entrada X_k , uma saída desejada Y_k é também fornecida. A rede busca minimizar a diferença entre a saída obtida e a saída desejada para uma dada entrada.

O aprendizado graduado ou por reforço é uma forma semi-supervisionada de treinamento, em que a saída desejada é substituída por uma avaliação ou nota do desempenho da rede em uma dada situação. Em determinados intervalos de tempo durante o treinamento, a rede recebe uma nota numérica que indica sua medida de performance. Desta forma não é necessário conhecer a saída correta para cada entrada apresentada. Destacam-se, para esta forma de aprendizado, as aplicações na área de controle e otimização, onde não há meios de se saber as saídas desejáveis [7].

No aprendizado não-supervisionado a rede se auto modifica em resposta à entrada x , criando sua própria representação do padrão de entrada, sem haver necessidade de auxílio externo. A rede pode aprender a se adaptar baseada na experiência adquirida com padrões de treinamento anteriores.

2.2.1.1 Aprendizado Competitivo

O aprendizado competitivo é comentado a seguir por ser esta a forma de treinamento empregada na Rede Morfológica Não-Supervisionada, usada neste trabalho.

As redes neurais não supervisionadas possuem a capacidade de auto-organizar a informação recebida durante o aprendizado. Certas redes podem criar representações das características dos dados de entrada através de mecanismos de auto-organização baseados em *aprendizado competitivo*, uma forma de aprendizado não-supervisionado em que cada padrão de entrada é associado a um neurônio, ou a um pequeno grupo de neurônios, levando à formação de agrupamentos (*clusters*) [62].

O objetivo da competição é a determinação do neurônio que melhor representa o padrão de entrada, ou seja, que possui o vetor de pesos mais próximo do vetor de entrada, segundo uma forma de medida de distância (usualmente a distância Euclidiana).

A competição se dá por meio de conexões laterais entre neurônios de uma camada. Através destas conexões, um neurônio inibe a atividade dos outros nós da camada. O neurônio vencedor é aquele que consegue inibir todos os demais. Somente os pesos deste elemento e, em alguns casos, dos vizinhos ao seu redor, são atualizados segundo uma regra de aprendizado para se aproximarem mais do padrão de entrada. Isto gera uma representação esparsa do conjunto de padrões de treinamento. Detalhes deste modelo e uma aplicação em rede neural morfológica são apresentados em [7].

2.3 MORFOLOGIA MATEMÁTICA PARA IMAGENS BINÁRIAS

Desenvolvida por George Matheron e Jean Serra por volta de 1964 na *École Nationale Supérieure des Mines de Paris*, em Fontainebleau, a Morfologia Matemática consiste em uma abordagem para o tratamento de problemas de análise de imagem baseada na extração de informações a partir de transformações de formas, através de dois operadores elementares denominados dilatação e erosão. Estes operadores realizam transformações sobre imagens de acordo com elementos estruturantes

pré-estabelecidos, que as sondam localmente [63].

2.3.1 Conceitos Fundamentais

Para as definições apresentadas nas próximas seções serão utilizados os seguintes conceitos, usando a notação apresentada por [63], onde Z é o conjunto dos inteiros, Z^2 o produto cartesiano $Z \times Z$ e B é um subconjunto de Z^2 , chamado elemento estruturante.

2.3.1.1 Translado por um vetor

O translado de B por um vetor h em Z^2 , denotado por $B + h$, é definido por [64] como

$$B + h = \{x + h | x \in B\}. \quad (2.1)$$

2.3.1.2 Transposto (reflexão) em relação à origem

O transposto de B , denotado por B^t , é definido por [64] como

$$B^t = \{-x | x \in B\}. \quad (2.2)$$

2.3.1.3 Complemento

O complemento de B , denotado por B^c , é definido por [14] como

$$B^c = \{x | x \notin B\}. \quad (2.3)$$

2.3.2 Operadores Elementares

2.3.2.1 Dilatação

A dilatação de um conjunto A por um elemento estruturante B em Z^2 , denotada por $\delta_B(A)$, pode ser definida conforme a expressão:

$$\delta_B(A) = \{x | (B^t + x) \cap A \neq \emptyset\}. \quad (2.4)$$

A dilatação de A por B consiste no conjunto de pontos x nos quais a intersecção entre A e o transposto de B transladado por x é diferente de nulo. Ou ainda, o conjunto dos centros de B^t que tocam o conjunto A [65].

A dilatação também pode ser vista como sendo a união de todos os pontos de uma matriz deslocada sobre os pontos do elemento estruturante ou vice-versa, como mostrado em [65]:

$$\delta_B(A) = \bigcup_{x \in B} (A + x) \quad (2.5)$$

$$\delta_B(A) = \bigcup_{x \in A} (B + x), \quad (2.6)$$

sendo a dilatação uma operação comutativa.

2.3.2.2 Erosão

A erosão de um conjunto A por um elemento estruturante B , denotada por $\varepsilon_B(A)$, pode ser definida conforme a expressão

$$\varepsilon_B(A) = \{x | (B + x) \subset A\}. \quad (2.7)$$

Assim, a erosão de A pelo elemento estruturante B é o conjunto de pontos x tais que B , transladado por x , está contido em A [65].

A erosão pode também ser definida conforme [63] pela expressão:

$$\varepsilon_B(A) = \bigcap_{b \in B} (A - b). \quad (2.8)$$

2.3.3 Detecção de Padrões Binários por Morfologia Matemática

O reconhecimento de figuras binárias por meio de Morfologia Matemática é possível através de um mecanismo baseado no operador de erosão. Uma figura particular de uma imagem pode ser detectada erodindo-se a imagem por um elemento estruturante com a mesma forma e tamanho da figura a ser detectada. Objetos menores do que

o elemento estruturante são erodidos por completo. A erosão de objetos maiores ou iguais ao elemento estruturante geram uma mancha na posição de sua origem.

2.3.3.1 *Hit ou Miss*

O uso de uma simples erosão pode produzir resultados insatisfatórios no reconhecimento de imagens. Na Figura 2.1, o objeto superior esquerdo foi erroneamente detectado como sendo a figura procurada.

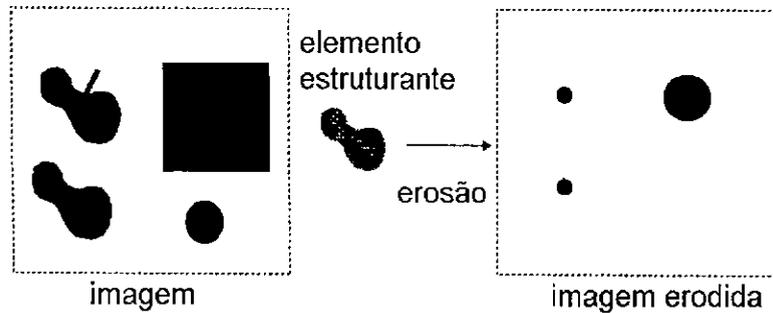


Fig. 2.1 - Utilização do operador de erosão como mecanismo de detecção de padrões binários.

FONTE: ([66], p. 3).

Uma forma de se evitar tais problemas é usando-se a operação *Hit* ou *Miss*, definida por [65] como sendo a transformação de um conjunto X por um elemento estruturante B , composto por dois conjuntos H e M . Sendo B centrado no ponto x , denotando-se (H_x, M_x) , tal ponto pertence à transformação *Hit* ou *Miss* $X \otimes B$ de X somente se H está incluído em X e M está incluído no complemento X^c de X :

$$X \otimes B = \{x | H_x \subset X; M_x \subset X^c\}. \quad (2.9)$$

Esta transformação pode ser expressa por meio de erosões, como mostrado na expressão a seguir:

$$X \otimes B = \varepsilon_H(X) \cap \varepsilon_M(X^c). \quad (2.10)$$

A Figura 2.2 mostra o resultado da aplicação de uma operação de *Hit* ou *Miss*. Como resultado, a imagem de saída contém somente os pontos onde o elemento estruturante *Hit* está contido em um objeto da imagem e o elemento estruturante *Miss* está contido no complemento deste objeto.

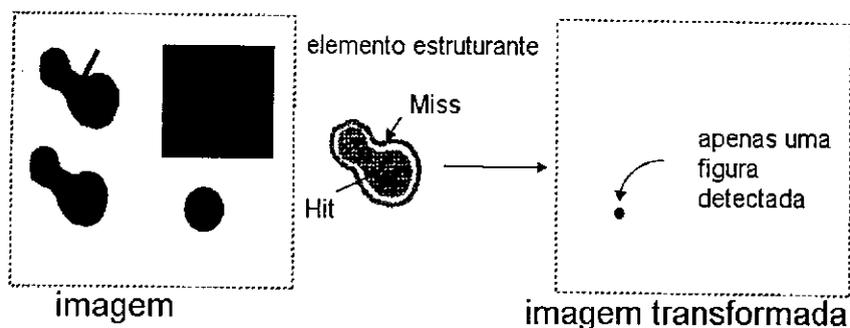


Fig. 2.2 - Utilização do operador *Hit* ou *Miss* para detecção de padrões binários.

FONTE: ([66], p. 3).

2.4 MORFOLOGIA MATEMÁTICA EM NÍVEIS DE CINZA

Nas próximas seções são apresentados alguns conceitos da Morfologia Matemática em níveis de cinza, segundo o formalismo desenvolvido por Gerald Banon descrito em [6] e [67].

2.4.1 Conceitos Básicos

Seja Z o conjunto dos números inteiros e Z^2 o produto cartesiano $Z \times Z$. Sejam D e E dois subconjuntos de Z^2 (dois retângulos finitos de Z^2 [67]), e seja K_m o intervalo $[0, m]$ de Z . O conjunto de mapeamentos de E em K_m , denotado por K_m^E , representa o conjunto de imagens em níveis de cinza com domínio E (conjunto das posições dos *pixels*) e escala de cinza K_m (conjunto dos possíveis valores dos *pixels*).

2.4.1.1 Operadores Elementares da Morfologia Matemática

Seja E um subconjunto não-vazio de Z^2 e l um número inteiro entre 0 e m . Os operadores δ_l , ε_l , δ_l^a e ε_l^a , respectivamente, dilatação, erosão, anti-dilatação e anti-

erosão, podem ser definidos conforme [6]:

$$\delta_l(f)(x) = \begin{cases} 1 & \text{se } f(x) > l \\ 0 & \text{caso contrário} \end{cases} \quad (2.11)$$

$$\varepsilon_l(f)(x) = \begin{cases} 1 & \text{se } f(x) \geq l \\ 0 & \text{caso contrário} \end{cases} \quad (2.12)$$

$$\delta_l^a(f)(x) = \begin{cases} 1 & \text{se } f(x) \leq l \\ 0 & \text{caso contrário} \end{cases} \quad (2.13)$$

$$\varepsilon_l^a(f)(x) = \begin{cases} 1 & \text{se } f(x) < l \\ 0 & \text{caso contrário} \end{cases} \quad (2.14)$$

para todo $f \in K_m^E$ e $x \in E$.

2.4.1.2 Janela

Sejam w_1 e w_2 dois números naturais. Denomina-se janela o subconjunto de Z^2 , denotado por W , dado por:

$$W = [-w_1, w_1] \times [-w_2, w_2] \subset Z^2. \quad (2.15)$$

2.4.2 Casamento de Padrões por Morfologia Matemática

Esta seção contém uma descrição do Operador Morfológico de Casamento de Padrões (OMCP), apresentado em [6] e [67], e do Operador de Detecção com Folga, apresentado em [7].

2.4.2.1 Operador Morfológico de Casamento de Padrões

O algoritmo para casamento de padrões, ilustrado na Figura 2.3, baseia-se na busca de uma imagem de referência f_W em uma imagem de busca g . A máscara f_W é uma subimagem de f , delimitada pela janela W .

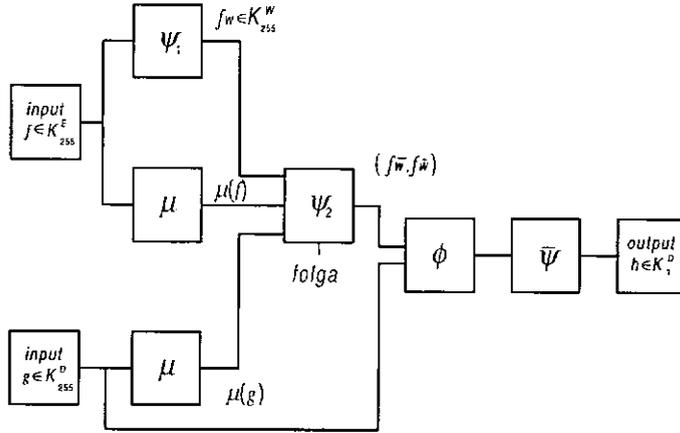


Fig. 2.3 - Representação esquemática para casamento de padrões utilizando o OMCP.

FONTE: ([6], p.35).

O operador ψ_1 seleciona uma janela de f , gerando f_W . Esta seleção pode ser feita manualmente [67].

O operador ψ_2 gera um par de imagens, f_W^- e f_W^+ , a partir de f_W em K_m^W , sendo

$$f_W^-(x) = \max\{0, \min\{m, f_W(x) + c_1\}\} \quad (x \in W) \quad (2.16)$$

$$f_W^+(x) = \max\{0, \min\{m, f_W(x) + c_2\}\} \quad (x \in W), \quad (2.17)$$

onde c_1 e c_2 ($c_1 \leq c_2$) são duas constantes inteiras. O intervalo $[c_1, c_2]$ define uma folga radiométrica sobre a imagem f_W .

Seja G o subconjunto de Z^2 definido por $G = D \ominus W$, onde \ominus indica a subtração de Minkowski [63], e l um número inteiro entre 0 e m . Os operadores de K_m^D em K_1^G , denotados por ε_i^j e δ_i^{aj} , são definidos por:

$$\varepsilon_i^j(g)(x) = \begin{cases} 1 & \text{se } g(x + w_i) \geq l \\ 0 & \text{caso contrário} \end{cases} \quad (2.18)$$

$$\delta_i^{ai}(g)(x) = \begin{cases} 1 & \text{se } g(x + w_i) \leq l \\ 0 & \text{caso contrário} \end{cases} \quad (2.19)$$

para $g \in K_m^D$ e $x \in G$; para $n = \#W$ e $i = 1, \dots, n$.

O operador ϕ de K_m^D em K_n^G , denominado Operador de Casamento de Padrões, é dado por:

$$\phi = \sum_{i \in n} \lambda^i, \quad (2.20)$$

onde os n operadores λ^i de K_m^D em K_1^G são dados por:

$$\lambda^i = \varepsilon_{f_W^-(w_i)}^i \wedge \delta_{f_W^+(w_i)}^{ai}, \quad (2.21)$$

sendo $n = \#W$ e $i = 1, \dots, n$.

O *pixel* da imagem $\phi(g)$ (a transformação de $g \in K_m^D$ através de ϕ) na posição $x \in G$, tem como valor o número de pixels de $g(x + w_i)$ que se encontram no intervalo $[f_W^-(w_i), f_W^+(w_i)]$.

Em seguida, a imagem g' produzida por ϕ é submetida ao operador $\bar{\psi}$, que encontra seu *pixel* de valor máximo, produzindo a imagem h .

O operador $\bar{\psi}$ de K_n^E em K_1^E que localiza o nível de cinza máximo global de uma imagem pode ser expresso como:

$$\bar{\psi}(f)(x) = \begin{cases} 1 & \text{se } f(y) \leq f(x) \quad (y \in E) \\ 0 & \text{caso contrário} \end{cases} \quad (2.22)$$

para todo $f \in K_n^E$ e $x \in E$.

O Operador Morfológico de Casamento de Padrões consiste na composição do Operador de Casamento de Padrões (expressão 2.20) com o operador de localização de nível de cinza máximo global (expressão 2.22):

$$\bar{\psi} \circ \phi \quad (2.23)$$

O símbolo \circ denota a composição de mapeamentos.

2.4.2.2 Operador de Detecção com Folga

O operador de limiarização definido a seguir localiza padrões em imagens digitais com valores de níveis de cinza acima ou igual a um determinado limiar [7].

Seja $l \in K_n$. O operador Ψ_l^\bullet de K_n^E em K_1 que localiza níveis de cinza maiores ou iguais a l é o operador dado por:

$$\Psi_l^\bullet(f) = \begin{cases} 1 & \text{se } \exists x \in E, f(x) \geq l \\ 0 & \text{caso contrário} \end{cases} \quad (2.24)$$

A composição do operador de limiarização com o operador de casamento de padrões (expressão 2.20) forma o operador de detecção com folga [7]:

$$\Psi_l^\bullet \circ \phi \quad (2.25)$$

A aplicação da composição $\Psi_l^\bullet \circ \phi$ sobre uma imagem g em busca de um padrão de referência f_W , dado um limiar de aceitação l , consiste na limiarização da imagem g' , gerada pelo operador ϕ , para detectar os possíveis pontos de g' com níveis de cinza maiores ou iguais a l .

2.5 OPERAÇÕES MORFOLÓGICAS INVARIANTES A ROTAÇÃO

A morfologia matemática estuda transformações de imagens através de operadores que são invariantes a translações. Alguns autores têm apresentado formas de se utilizar operadores morfológicos para o processamento de imagens com invariância a rotação.

Em [68] é dado um exemplo de como múltiplos elementos estruturantes podem ser aplicados na extração de objetos com diferentes orientações. A Figura 2.4 mostra uma imagem que contém elipses aleatoriamente orientadas. Quatro elementos estruturantes são definidos para representar quatro elipses com diferentes orientações.

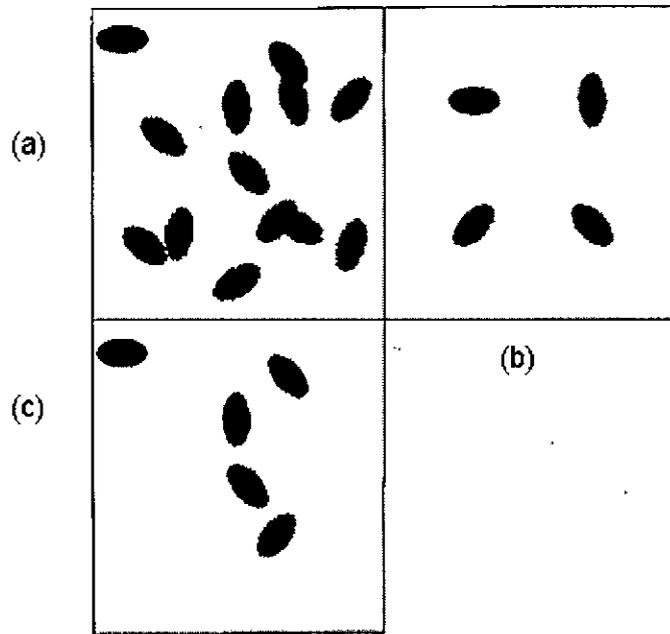


Fig. 2.4 - (a) imagem original; (b) quatro elementos estruturantes; (c) imagem resultante da união de aberturas usando os quatro elementos estruturantes.

FONTE: ([68], p.327).

A operação de detecção neste exemplo é a união de quatro operações de abertura, descrita pela expressão:

$$B = \bigcup_{i=1}^4 (A \circ K_i), \quad (2.26)$$

onde A é a imagem original, B é a imagem resultante, $\{K_i, i = 1, \dots, 4\}$ são os quatro elementos estruturantes mostrados na Figura 2.4(b), e a operação de abertura \circ é definida como

$$A \circ K = (A \ominus K) \oplus K, \quad (2.27)$$

onde \ominus e \oplus são, respectivamente, uma erosão (expressão 2.8) e uma dilatação (expressão 2.5).

Para criar operações morfológicas invariantes a rotação e multiplicação escalar, Heij-

mans e Roerdink [69], [70] desenvolveram a morfologia polar *polar morphology*, observando que erosões e dilatações invariantes a translação dependem somente da adição vetorial, que forma um grupo comutativo em \mathfrak{R}^2 e induz translações de conjuntos por vetores. Eles substituíram o grupo de translações em \mathfrak{R}^2 pelo “grupo polar”, que é o conjunto de pontos planares representados em coordenadas polares (r, θ) e providos da operação de grupo $+_P$ definida por $(r_1, \theta_1) +_P (r_2, \theta_2) = (r_1 \cdot r_2, \theta_1 + \theta_2)$. As operações de dilatação polar e erosão polar são definidas conforme as expressões 2.28 e 2.29, respectivamente [71].

$$X \oplus_P B = \bigcup_{(\rho, \lambda) \in B} X +_P (\rho, \lambda) = \bigcup_{(\rho, \lambda) \in B} \{(r \cdot \rho, \theta + \lambda) : (r, \theta) \in X\} \quad (2.28)$$

$$X \ominus_P B = \bigcap_{(\rho, \lambda) \in B} X +_P (\rho^{-1}, -\lambda) = \bigcap_{(\rho, \lambda) \in B} \{(r/\rho, \theta - \lambda) : (r, \theta) \in X\}. \quad (2.29)$$

Em [71], Maragos define a *morfologia afim*, invariante a rotação, translação e multiplicação escalar através do uso de um grupo $G = \{(M, t) : M \in \mathfrak{R}^{2 \times 2}, \det(M) \neq 0, t \in \mathfrak{R}^2\}$ provido da operação $(L, v) \star (M, t) = (LM, Lt + v)$. A ação do grupo (G, \star) sobre uma imagem $X \subseteq \mathfrak{R}^2$ é denotada por $\Delta_{(M, t)}$ na expressão

$$\Delta_{(M, t)}(X) = \{Mx + t : x \in X\}. \quad (2.30)$$

A matriz M pode ser sempre escrita na seguinte forma:

$$M = \begin{bmatrix} a & b \\ c & d \end{bmatrix} = \begin{bmatrix} r_1 \cos \theta_1 & -r_2 \sin \theta_2 \\ r_1 \sin \theta_1 & r_2 \cos \theta_2 \end{bmatrix}; \quad \begin{array}{l} a, b, c, d \in \mathfrak{R} \\ r_1, r_2 \geq 0, \theta_1, \theta_2 \in [0, 2\pi] \end{array} \quad (2.31)$$

Uma *dilatação afim* de X por $B \subseteq G$ pode ser definida como

$$X \oplus_A B = \bigcup_{(M, t) \in B} \Delta_{(M, t)}(X) = \bigcup_{(M, t) \in B} \{Mx + t : x \in X\}. \quad (2.32)$$

A *erosão afim* de X por B é definida por

$$X \ominus_A B = \bigcap_{(M, t) \in B} \Delta_{(M, t)^{-1}}(X) = \bigcap_{(M, t) \in B} \{M^{-1}x - M^{-1}t : x \in X\}. \quad (2.33)$$

Em *morfologia afim*, o “elemento estruturante” B é uma coleção de pares de parâmetros (M, t) .

Em [72] Roerdink propôs uma extensão da morfologia para incluir invariância a rotação e translação. Alguns pontos da sua teoria são descritos a seguir.

Seja $P(\mathcal{M})$ o conjunto de todos os subconjuntos do grupo \mathcal{M} de movimentações (translações e rotações). Considerando-se o plano \mathbb{R}^2 dos números reais, os elementos de \mathcal{M} são parametrizados como $g_{h,\theta}$, com $h \in \mathbb{R}^2, \theta \in [0, 2\pi]$. Um elemento do grupo atua sobre um ponto $x = (x_1, x_2)^t$ da seguinte maneira:

$$g_{h,\theta} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} + \begin{pmatrix} h_1 \\ h_2 \end{pmatrix}, \quad x = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \in \mathbb{R}^2. \quad (2.34)$$

A imagem de um ponto $x \in \mathbb{R}^2$ sob a movimentação $g \in \mathcal{M}$ é denotada gx .

Define-se τ_h como sendo a única translação que mapeia a origem $\omega = (0, 0)$ do plano para o ponto h . Define-se também r_θ^h como sendo a rotação do plano em torno do ponto $h \in \mathbb{R}^2$ por um ângulo θ . O grupo de translações é denotado por \mathcal{T} , e o grupo de rotações em torno da origem é denotado por \mathcal{R} .

Define-se invariância ao grupo \mathcal{M} no plano \mathbb{R}^2 como sendo o mapeamento $\psi : P(\mathbb{R}^2) \rightarrow P(\mathbb{R}^2)$ que satisfaça

$$\psi(gX) = g\psi(X), \quad \forall X \subseteq \mathbb{R}^2 \quad (2.35)$$

para todo $g \in \mathcal{M}$, onde

$$gX = \{gx : x \in X\} \quad (2.36)$$

é a imagem de X sob a movimentação g .

Para as definições a seguir x, y representam elementos de \mathbb{R}^2 e X, Y representam subconjuntos de \mathbb{R}^2 .

Roerdink define operações de erosão e dilatação invariantes a rotação e translação sobre $P(\mathcal{M})$ ($\varepsilon^{\mathcal{M}}$ e $\delta^{\mathcal{M}}$), e mostra que pode-se construir operações equivalentes a erosões e dilatações invariantes a rotação sobre $P(\mathfrak{R}^2)$ por meio de operações de união e intersecção de erosões e dilatações, como mostrado nas expressões:

$$\varepsilon_Y^{\mathcal{M}}(X) = \bigcup_{r \in \mathcal{R}} X \ominus (rY) \quad (2.37)$$

$$\delta_Y^{\mathcal{M}}(X) = \bigcap_{r \in \mathcal{R}} X \oplus (r\check{Y}), \quad (2.38)$$

onde $X \ominus (rY)$ e $X \oplus (r\check{Y})$ são, respectivamente, a erosão e a dilatação do conjunto X pelos elementos estruturantes rY e $r(\check{Y})$. \mathcal{R} é o grupo de rotações em torno da origem do plano. A reflexão do conjunto Y (expressão 2.2) é denotada aqui por \check{Y} .

CAPÍTULO 3

REDE MORFOLÓGICA NÃO-SUPERVISIONADA

3.1 INTRODUÇÃO

Este capítulo descreve o funcionamento da Rede Morfológica Não-Supervisionada (RMNS), desenvolvida por Silva [7]. Esta rede foi escolhida como mecanismo de aprendizagem do modelo proposto como sugerido em [7] e [73]. Dentre as razões para a escolha da RMNS encontram-se:

- Capacidade de detecção invariante a translação;
- Considerável grau de insensibilidade ao contexto;
- Capacidade de rejeitar entradas que não se aproximem de nenhum dos padrões treinados;
- Capacidade de aprender novos padrões sem alterar os padrões já aprendidos;
- Capacidade de aprendizado limitada apenas pela quantidade de memória disponível;
- Robustez a ruídos na entrada controlada pelo usuário, através do ajuste do parâmetro de limiarização;

A RMNS é uma rede neural com quatro camadas de neurônios que se utilizam dos operadores morfológicos apresentados na seção 2.3 para a detecção de padrões com representação binária ou em níveis de cinza, de maneira invariante a translação.

A forma de aprendizagem é bastante semelhante ao modelo de competição criado por Kohonen [74]. A rede possui também um sistema de *reset* similar ao da rede ART1. No entanto, a utilização de um mecanismo de pré-ativação utilizando o operador de detecção com folga, faz com que a rede tenha suas próprias características.

A RMNS possui capacidade de associar cada padrão apresentado a uma classe já aprendida, ou criar uma nova classe para representar a entrada. Após o treinamento, os pesos de cada neurônio guardam a representação de uma classe aprendida.

3.2 ARQUITETURA

As duas primeiras camadas da rede compõem o mapa morfológico, e são responsáveis pelo treinamento e detecção dos padrões. As duas últimas se encarregam da recuperação do padrão treinado para exibição da saída.

Sejam $g \in K_m^D$, $f \in K_m^W$, e $f_W \in K_m^W$, respectivamente o domínio das imagens de entrada, o domínio dos padrões a serem treinados e o domínio dos pesos dos neurônios morfológicos, definidos conforme a seção 2.4.2.1. A camada de entrada da RMNS possui D nós que servem para distribuir o padrão de entrada para os neurônios morfológicos da primeira camada interna (camada morfológica). Cada neurônio da camada de entrada possui conexão com todos os neurônios da camada morfológica, distribuindo cada componente do padrão apresentado a todos os neurônios morfológicos. Todos os pesos possuem valor 1 na camada de entrada, de forma que os valores de entrada não se alterem.

A primeira camada interna possui N neurônios morfológicos que possuem como função de ativação o operador Ψ_l^* o ϕ , de detecção com folga (Expressão 2.25). As saídas destes neurônios são dadas por $z_i \in K_1$, onde $i = 1, \dots, N$. A Figura 3.1 ilustra um neurônio morfológico.

Os pesos w_{ji} são formados por f_{W_j} , $f_{W_j}^+(w_i)$ e $f_{W_j}^-(w_i) \in K_m^W$, com $j = 1, \dots, N$ e $i = 1, \dots, n$. Para as seções seguintes, utiliza-se $w_{ji} = f_{W_j}(w_i)$. Os neurônios desta camada possuem ainda duas sinapses inibidoras, com pesos $w_a^c = 1$ e $w_p^c = 1$, utilizadas para determinar o vencedor durante a etapa de competição. Estes neurônios possuem também um limiar de ativação, $w_{j0} = l$, onde $l \in K_n$, que determina o quão próximo uma entrada precisa estar do padrão representado pelos pesos do neurônio para que ambos sejam considerados da mesma classe.

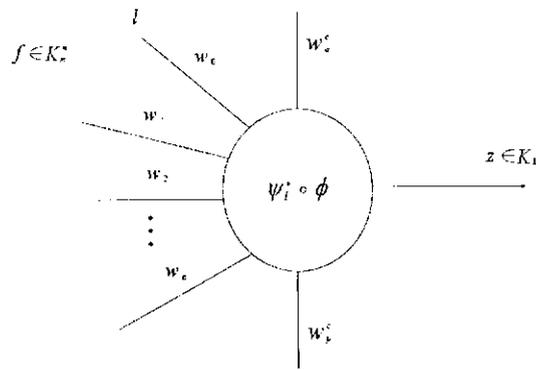


Fig. 3.1 - Representação de um neurônio morfológico da primeira camada interna da RMNS.

FONTE: ([7]).

A segunda camada interna possui N nós responsáveis pela recuperação do padrão encontrado na camada morfológica. Estes neurônios recebem como entrada as saídas z_j da camada morfológica e obtêm como saída um vetor $s_k = (s_1, s_2, \dots, s_m)$. Os pesos são dados por w_{jh} , com $j = 1, \dots, N$, $h = 0, \dots, n$, $w_{j0} = z_j$ e $w_{jh} = w_{ji}$ para $h \neq 0$. A função de ativação é dada pela dilatação especial δ^* da expressão 3.1, que produz como saída o vetor de pesos do neurônio da camada morfológica que tenha produzido saída igual a 1.

$$\delta_{W_j}^*(z_j) = \begin{cases} f_{W_j} & \text{se } z_j = 1 \\ 0_{W_j} & \text{caso contrário} \end{cases} \quad (3.1)$$

A camada de saída possui apenas um neurônio, responsável pela união de todas as saídas dos neurônios da segunda camada interna. Este neurônio possui pesos $w_k = s_j$, onde $s_j = \delta_{W_j}^*(z_j)$, para $j = 1, \dots, N$. A função de ativação é dada por

$$y = \bigcup_{j=1}^N s_j. \quad (3.2)$$

A Figura 3.2 ilustra a arquitetura da RMNS.

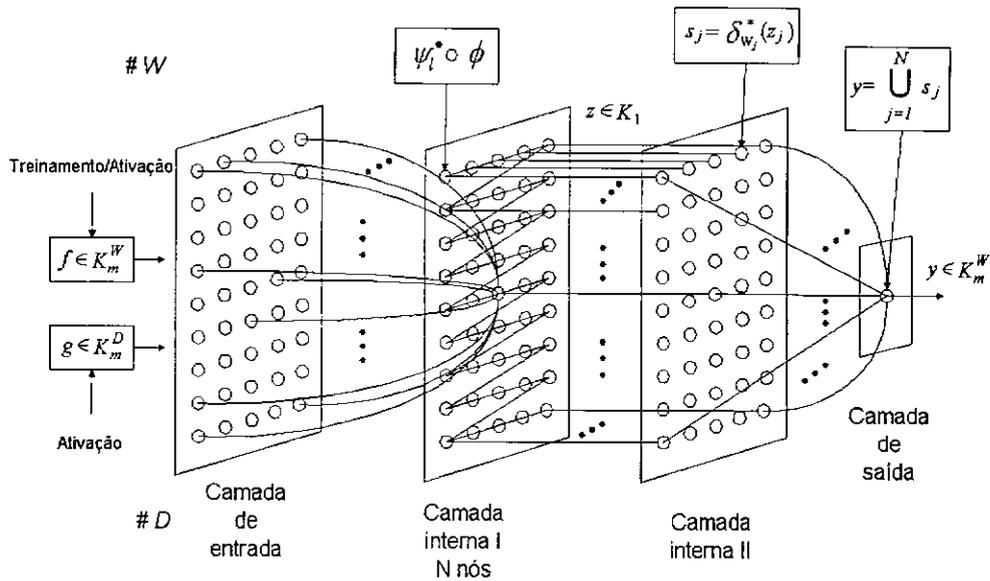


Fig. 3.2 - Exemplo de arquitetura de uma Rede Morfológica Não-Supervisionada.
 FONTE: ([7]).

3.3 OPERAÇÃO DA RMNS

3.3.1 Inicialização

Os pesos da primeira camada são fixados em 1, o que faz com que os sinais de entrada não sejam alterados. Os pesos da camada morfológica são inicializados aleatoriamente, com valores dentro do intervalo utilizado pelos padrões a serem treinados. As demais camadas não participam da etapa de treinamento, e têm seus pesos estabelecidos em tempo de ativação.

3.3.2 Treinamento

Durante o treinamento, cada neurônio da primeira camada recebe um componente do padrão de entrada e o distribui a todos os nós da camada morfológica, onde é verificado se já existe algum nó que represente o padrão de entrada. Os neurônios desta camada realizam uma detecção com folga (seção 2.4.2.2), comparando seu vetor de pesos com a entrada recebida, dada uma folga l . Esta folga indica o grau de similaridade entre os padrões membros de uma mesma classe.

Caso não haja nenhum casamento dentro da folga l , os neurônios morfológicos competem entre si para determinar qual deles representa melhor a entrada.

A competição é baseada na distância Euclidiana entre a entrada x e os pesos W_j :

$$d_j = \sqrt{\sum_i (x_i - w_{ji})^2}. \quad (3.3)$$

O neurônio com a menor distância é o vencedor, denotado por z_j^* e obtido por

$$\|x - W^*\| = \min\{\|x - W_j\|\}, \text{ onde } \|X\| \text{ é o módulo de } X. \quad (3.4)$$

Esta competição pode ser implementada por duas *sinapses* inibidoras especiais, responsáveis por evitar conflitos caso haja mais de um neurônio com a mesma distância.

Estas *sinapses* são definidas por

$$F(d_j) = \begin{cases} 1 & \text{se } \min(d_h) > d_j (\forall h < j), \min(d_k) \geq d_j (\forall k > j), \text{ e } \alpha = 1 \\ 0 & \text{caso contrário} \end{cases} \quad (3.5)$$

onde $j > 1$, d_j é a distância entre os pesos do neurônio e o padrão, d_h é a distância mínima dos neurônios anteriores, d_k é a distância mínima dos neurônios posteriores, $\alpha = 1$ (*reset* = 1) indica que o neurônio pode ser treinado, e $F(d_j)$ é a saída do neurônio j durante a competição. Em caso de conflito, o neurônio de menor índice j é declarado vencedor.

Após a competição, os pesos do neurônio vencedor são atualizados pela regra

$$\begin{aligned} \Delta w_{ji}^{(novo)} &= \alpha_j (x_i - w_{ji}^{(velho)}) z_j^*, \\ w_{ji}^{(novo)} &= (w_{ji}^{(velho)} + \Delta w_{ji}^{(novo)}), \end{aligned} \quad (3.6)$$

onde α é o parâmetro de reset, x_i são as componentes do vetor de entrada $x \in K_m^W$, $w_{ji} \in K_m^W$, e $z_j^* \in K_1$. Apenas o neurônio vencedor, z_j^* , tem seus pesos atualizados.

A seguir é apresentado o algoritmo de treinamento da RMNS, extraído de [7].

- 1) Inicie os pesos com valores aleatórios positivos. Inicializar todos os neurônios da segunda camada (primeira camada interna) com $reset=1$ (variável de controle de disponibilidade de aprendizagem) desligado;
- 2) Apresentar nova entrada x , onde x é o padrão de entrada e é propagado para cada neurônio j ;
- 3) Verificar similaridade entre o padrão apresentado e algum neurônio representante de uma classe já treinada (Equação 2.25);
- 4) Caso exista representante da classe do padrão corrente, executar o passo 2;
- 5) Verificar neurônios disponíveis (neurônios com $reset$ desligado). Caso não exista, termina o treinamento;
- 6) Aplicar competição para determinar menor esforço. Computar a distância d_j entre a entrada e cada neurônio morfológico j , dada pela equação 3.3;
- 7) Selecionar distância mínima entre os neurônios livres (competição). Designar o neurônio de saída com o d_j mínimo para que seja j^* ;
- 8) Atualizar pesos e ligar $reset$ do neurônio vencedor. Atualizar pesos do neurônio j^* (Equação 3.6);
- 9) Retornar ao passo 2, caso ainda exista padrão a ser apresentado;
- 10) Fim.

3.3.3 Ativação

A etapa de ativação da RMNS consiste na apresentação de entradas com padrões a serem reconhecidos pela rede. Os padrões de entrada são comparados com os padrões aprendidos pelos nós da camada morfológica, sendo verificado qual nó é ativado para uma determinada entrada. Caso mais de um neurônio responda a um

dados padrão de entrada, o neurônio de menor índice é escolhido. O resultado é submetido à segunda camada interna, que recupera o padrão aprendido.

O algoritmo de ativação da RMNS, extraído de [7], é mostrado a seguir.

- 1) Inicializar pesos e parâmetros;
- 2) Apresentar imagem à camada de entrada;
- 3) Aplicar ativação com o operador de detecção com folga, sobre a imagem de entrada x com os pesos dos neurônios da camada morfológica (Expressão 2.25);
- 4) Ativar segunda camada interna, responsável pela recuperação de cada padrão detectado pela primeira camada interna, com a função de ativação dada pela Expressão 3.1;
- 5) Apresentar o primeiro padrão detectado na camada de saída (Expressão 3.2);
- 6) Se existe alguma outra imagem, voltar ao passo 1;
- 7) Terminar ativação.

CAPÍTULO 4

O MODELO PROPOSTO

4.1 INTRODUÇÃO

Neste capítulo é descrito o modelo utilizado para reconhecimento de padrões invariante a rotação e translação, bem como a forma como este pode ser incorporado à Rede Morfológica Não-Supervisionada, como sugerido em [73]. Em seguida é feito um comentário comparativo entre a forma de invariância implementada sobre a RMNS e o modelo descrito em [45].

4.2 DESCRIÇÃO DO MODELO

O modelo descrito nesta seção inspira-se no trabalho de [68] e [72] para a definição de operações de casamento de padrões invariantes a rotação e translação. A seguir é apresentada a forma como são realizadas estas operações sobre imagens binárias e em níveis de cinza. A imagem de entrada é chamada de imagem de busca e o padrão procurado é chamado de padrão de referência, conforme a nomenclatura utilizada em [67].

A idéia básica consiste em rotacionar o padrão de referência, para que se possa detectar padrões rotacionados na imagem de busca por meio de um operador morfológico de detecção. É necessário que se possibilite o casamento com figuras rotacionadas de maneira a levar-se em consideração todas as posições possíveis.

Uma forma de se abordar o problema é através da rotação do padrão de referência por diversos ângulos, até que se completem 360 graus, sendo realizado o casamento entre a imagem de busca e o padrão de referência após cada rotação. Uma maneira de aumentar a robustez do sistema seria a definição de um intervalo de rotação tão pequeno quanto possível. Contudo, isto implica em maior número de operações necessárias para se completar um giro de 360 graus. Esta solução pode ser computacionalmente dispendiosa, demandando muito tempo de processamento.

Uma outra maneira de se lidar com o problema é utilizar um conjunto de cópias rotacionadas de um padrão de referência, conforme descrito por [45]. Seja W uma matriz representando o padrão de referência. O conjunto de cópias rotacionadas de W forma uma matriz de padrões de referência chamada de agrupamento invariante a rotação. Este agrupamento compreende todas as rotações de W por ângulos múltiplos de um ângulo α escolhido arbitrariamente.

Uma matriz representando um agrupamento invariante a rotação pode ser construída a partir de W como se segue:

$$\left[\begin{array}{cccc} W & r_{1\alpha}(W) & r_{2\alpha}(W) & \dots & r_{(n-1)\alpha}(W) \end{array} \right] \quad (4.1)$$

onde $r_{i\alpha}(W)$ é a rotação de W por um ângulo $i \cdot \alpha$, com $i = 1, 2, \dots, n - 1$. Dado que α é o intervalo de rotação entre uma cópia e outra, o número de elementos do agrupamento é dado por $n = \lceil 360/\alpha \rceil$, onde $\lceil \cdot \rceil$ representa a função teto, que retorna o maior inteiro menor do que $(\cdot + 1)$.

Um sistema configurado para ser insensível a rotações a cada 30° , possui um conjunto de réplicas do padrão de referência rotacionado por $30^\circ, 60^\circ, \dots, 330^\circ$, sendo necessárias 11 rotações de W no total.

O processo de detecção invariante a rotação envolve a aplicação de W e de cada $r_{i\alpha}(W)$ sobre a imagem de busca. Para cada inclinação de W é produzida uma saída. A união das saídas obtidas produz um resultado invariante a rotação. A forma como isto é feito é detalhada a seguir.

4.2.1 Detecção de Imagens Binárias

Em [68], Song e Delp apresentaram um exemplo de reconhecimento de imagens invariante a rotação utilizando múltiplos elementos estruturantes e a união de operações de abertura morfológica (seção 2.5). Nesta seção é descrita uma forma de se efetuar reconhecimento de imagens binárias invariante a rotação e translação utilizando a união de operações de *Hit* ou *Miss*.

Seja \mathcal{R} o conjunto das rotações em torno da origem do plano Z^2 . Os elementos de \mathcal{R} são representados por r_θ , com $\theta \in [0, 2\pi]$, e atuam sobre um ponto $x = (x_1, x_2)^t \in Z^2$ da seguinte maneira:

$$r_\theta \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}, \quad x = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \in Z^2. \quad (4.2)$$

Sejam X e W dois subconjuntos de Z^2 . Pode-se implementar uma operação de casamento de padrões invariante a rotação e translação entre imagens binárias através da união de operações de *Hit* ou *Miss* entre uma entrada X e um conjunto de elementos estruturantes gerados a partir da rotação de um padrão W [75]. Esta operação é denotada por $\otimes^{\mathcal{M}}$ na expressão

$$X \otimes^{\mathcal{M}} W = \bigcup_{r \in \mathcal{R}} X \otimes rW = \bigcup_{r \in \mathcal{R}} [\varepsilon_{rW}(X) \cap \delta_{rW}^a(X)], \quad (4.3)$$

onde

$$rW = \{rw : w \in W\} \quad (4.4)$$

e a anti-dilatação $\delta_{rW}^a(X)$ é dada pela expressão

$$\delta_{rW}^a(X) = [\delta_{rW^c}(X)]^c. \quad (4.5)$$

O processo de detecção invariante a rotação pode ser descrito pelo diagrama de blocos de operadores morfológicos mostrado na Figura 4.1, proposto por [75].

As intersecções das saídas dos blocos ε e δ^a compõem as operações de *Hit* ou *Miss* da entrada X pelos elementos estruturantes $r_{i\alpha}W$, com $i = 0, 1, \dots, n - 1$ para um agrupamento com n elementos. A união das saídas gera um resultado invariante a rotação, uma vez que sempre que o padrão W estiver presente na imagem de entrada, rotacionado ou não, pelo menos uma das operações de *Hit* ou *Miss* produzirá um ponto no local encontrado.

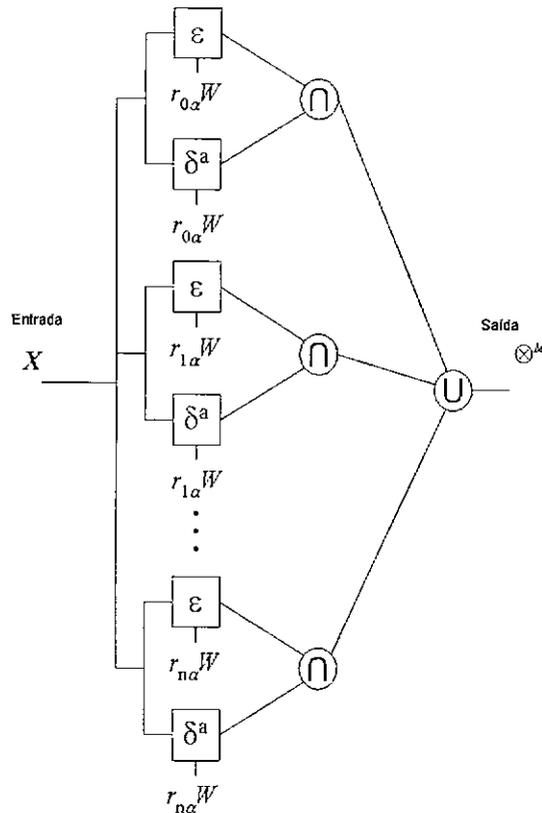


Fig. 4.1 - Diagrama de blocos morfológicos do modelo para detecção de imagens binárias.

4.2.2 Detecção de Imagens em Níveis de Cinza

Nesta seção, o método descrito anteriormente é estendido para imagens em níveis de cinza. O operador de detecção binária é substituído pelo Operador de Detecção com Folga (seção 2.4.2.2).

Sejam D e E dois subconjuntos não-vazios de Z^2 . Seja $g \in K_m^D$ uma imagem de busca. Seja $f \in K_m^E$ uma imagem de referência e $W \subset E$ uma janela. Um elemento de W é denotado por w e indica um ponto (w_1, w_2) . A restrição de f a W , denotada por f_W é uma subimagem de f , delimitada por W . Uma operação de detecção invariante a rotação e translação entre imagens em níveis de cinza pode ser obtida através da união dos resultados da aplicação do operador de detecção com folga

entre a imagem de busca g e um conjunto de padrões de referência $f_{W^i}^i$,

$$(\Psi_i^\bullet \circ \phi)_{f_W}^M(g) = \bigcup_{i=0, \dots, n-1} (\Psi_i^\bullet \circ \phi)_{f_{W^i}^i}(g) \quad i \in Z, \quad (4.6)$$

onde $f_{W^i}^i$ é uma imagem f^i delimitada por uma janela W^i que é equivalente a uma rotação da imagem f_W por um ângulo $\varphi = i \cdot \alpha$, onde α é o intervalo de rotação entre f^i e f^{i+1} . O número de padrões de referência é dado por $n = \lceil 360/\alpha \rceil$.

A versão rotacionada $f_{W^i}^i$ de uma imagem f_W , é dada por:

$$f_{W^i}^i(w_j^i) = f_{W^i}^i(r_{i\alpha}(w_j)) = f_W(w_j) \quad (4.7)$$

onde, para $w_{1j}, w_{2j} \in W$,

$$r_{i\alpha}(w_j) = w_j^i = (w_{1j}(\cos i\alpha) - w_{2j}(\sin i\alpha), w_{1j}(\sin i\alpha) + w_{2j}(\cos i\alpha)) \quad (4.8)$$

para $k = \#W = \#W^i$ e $j = 1, 2, \dots, k$ é a enumeração das posições de W e de W^i . $\#W$ indica a quantidade de elementos de W .

A expressão 4.7 indica que o valor do nível de cinza da imagem f_W na posição w_j é igual ao valor encontrado na posição $r_{i\alpha}(w_j)$ da imagem $f_{W^i}^i$.

A operação dada pela expressão 4.6 produz uma saída positiva sempre que pelo menos um dos padrões $f_{W^i}^i$, com $i = 0, 1, \dots, n$, estiver presente na entrada g . Aqui, uma janela W é um subconjunto não-vazio de Z^2 contendo a origem, e W^i é obtido a partir da rotação das coordenadas de W por um ângulo $i\alpha$.

O diagrama de blocos de operadores morfológicos do modelo para detecção em níveis de cinza é mostrado na Figura 4.2.

Cada bloco composto pelos símbolos ϕ e ψ_i^\bullet representa a aplicação do operador de detecção com folga entre a imagem de entrada g e o padrão de referência f_W rotacionado por um ângulo diferente. Caso um ou mais blocos produzam saída positiva, o resultado da operação de união das saídas será positivo, sendo caracterizada a detecção do padrão na imagem de entrada.

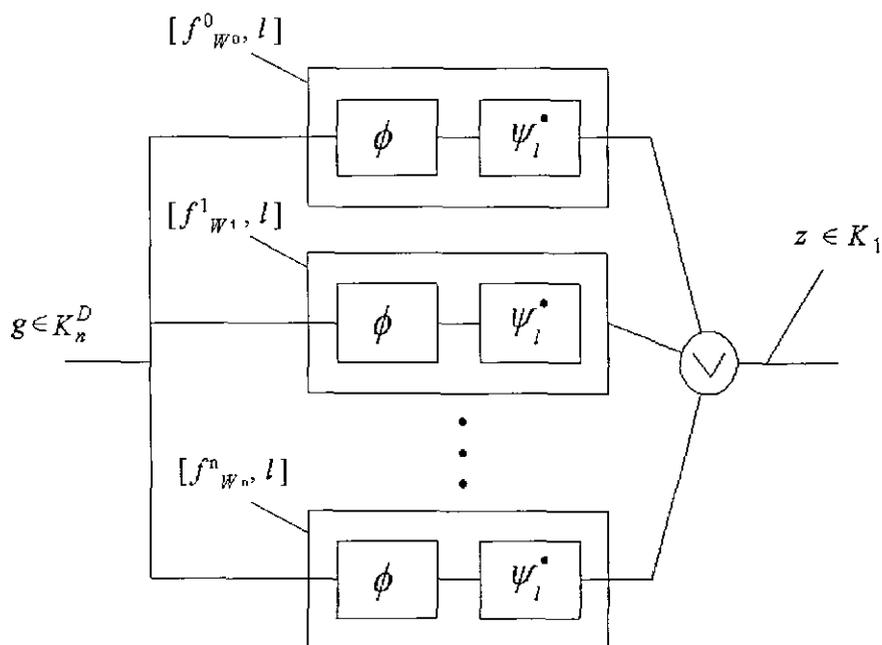


Fig. 4.2 - Diagrama de blocos morfológicos do modelo para detecção em níveis de cinza.

4.3 APLICAÇÃO DO MODELO À RMNS

Esta seção apresenta a aplicação do modelo à RMNS, conforme proposto por [73].

Na RMNS os padrões aprendidos são armazenados nas matrizes de pesos dos neurônios da camada morfológica. A RMNS é treinada com os padrões que representam os elementos estruturantes com orientação 0^0 . Para incluir invariância a rotação na rede, após o treinamento, os pesos dos neurônios morfológicos são rotacionados e replicados em outros neurônios, formando diversos agrupamentos invariantes a rotação, aqui chamados de agrupamentos de neurônios direcionais.

Uma forma de se organizar os agrupamentos é definir uma camada morfológica em tempo de treinamento com apenas uma coluna de neurônios. Todos os padrões treinados com orientação de 0^0 são armazenados em diferentes nós dispostos nesta coluna. Após o treinamento, os pesos de cada nó são rotacionados e copiados para as matrizes de pesos de um conjunto de neurônios posicionados em outras colunas, compondo o mapa morfológico invariante a rotação a ser usado na fase de ativação.

A Figura 4.3 ilustra um mapa morfológico invariante a rotações em intervalos de 90° .

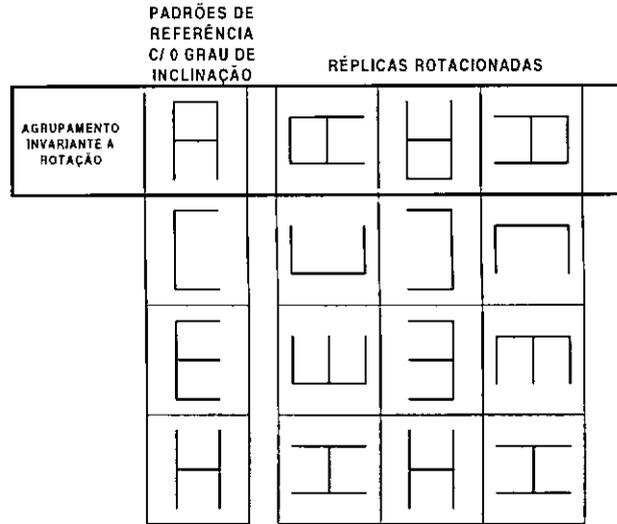


Fig. 4.3 - Exemplo de mapa morfológico invariante a rotação.

A Figura 4.4 mostra a aplicação do mapa morfológico invariante a rotação na arquitetura da RMNS.

Seja z'_i a saída do i -ésimo neurônio da primeira camada interna, com $i = 1, \dots, N$. Seja $q = \lceil (360/\alpha) \rceil$ a quantidade de nós de um agrupamento de neurônios direcionais. A entrada do neurônio j da segunda camada interna, denotada por z_j , é definida conforme a expressão

$$z_j = \bigcup_{i=j \cdot q - (q-1)}^{j \cdot q} z'_i, \quad (4.9)$$

com $j = 1, \dots, N/q$. Para um mapa morfológico invariante a rotação construído conforme a Figura 4.3, onde os padrões com inclinação de 0° se encontram na primeira coluna e as cópias rotacionadas se encontram nas demais colunas, o parâmetro i indexa os neurônios de forma crescente da esquerda para a direita e de cima para baixo no mapa morfológico.

A união de todas as saídas dos nós de um agrupamento garante que seja produzido

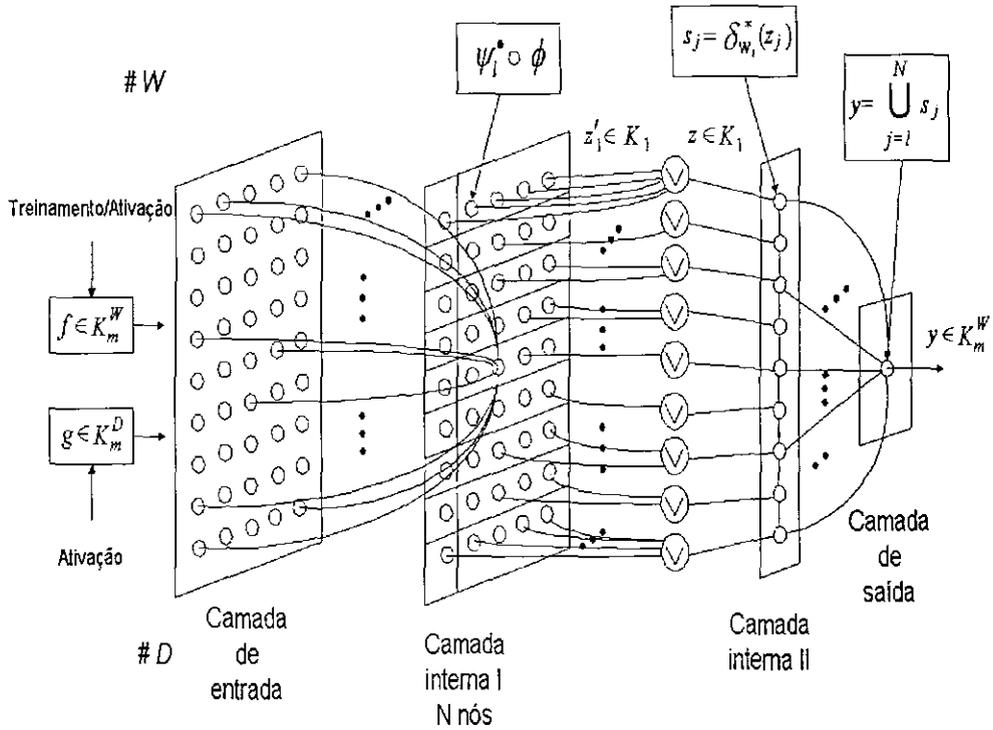


Fig. 4.4 - Modelo de invariância a rotação aplicado à RMNS.

um z_j igual a 1 sempre que um dos neurônios do agrupamento detectar o padrão treinado na imagem de entrada. Esta forma de agrupamento de neurônios é semelhante ao conceito de *slabs* invariantes a rotação e translação apresentado em [42].

O padrão detectado é recuperado na segunda camada interna, dilatando-se z_j pelos pesos do neurônio com orientação 0^0 da camada morfológica. As saídas s_j dos elementos da segunda camada interna são dadas pela expressão

$$s_j = \delta_{W_t}^*(z_j), \quad (4.10)$$

onde W_t é a matriz de pesos do t -ésimo neurônio da camada morfológica, δ^* é a dilatação definida pela expressão 3.1 e t é sempre o índice de um neurônio com orientação 0^0 . O valor de t para s_j pode ser obtido por.

$$t = j \cdot q - (q - 1), \quad (4.11)$$

onde q é a quantidade de nós de um agrupamento de neurônios direcionais.

Na Figura 4.5 é exibido um diagrama de blocos de operadores morfológicos proposto por [73], que ilustra o processo de ativação de uma RMNS com invariância a rotação.

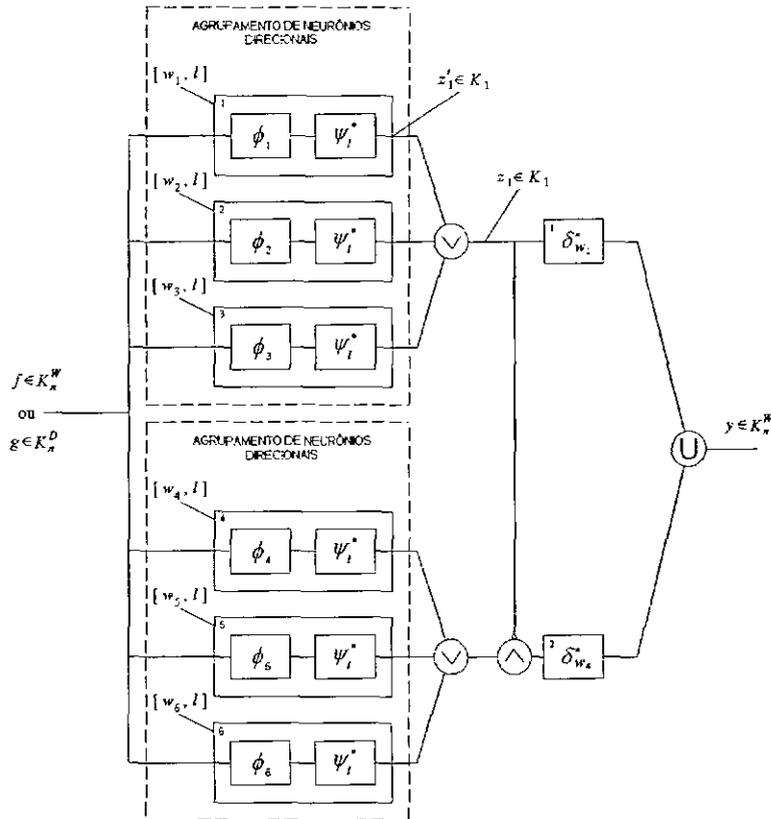


Fig. 4.5 - Modelo morfológico da etapa de ativação da RMNS com invariância a rotação.

Os blocos Ψ_i^* e ϕ representam os neurônios morfológicos, e detectam os padrões. No exemplo mostrado na Figura 4.5, os 2 agrupamentos de neurônios direcionais possuem 3 elementos de processamento. O primeiro elemento de cada agrupamento possui a representação do padrão aprendido com inclinação 0° . Os outros dois são rotacionados por ângulos definidos (45 graus à esquerda e 45 à direita, por exemplo). Uma operação de união produz uma única saída de cada agrupamento. Caso mais de um agrupamento produza saída positiva, a saída z de menor índice é escolhida. Os blocos δ^* realizam a dilatação para recuperação do padrão detectado com orientação 0° .

O algoritmo de ativação da RMNS, refletindo a inclusão de invariância a rotação, é

mostrado a seguir.

- 1) Inicializar pesos e parâmetros;
- 2) Apresentar imagem à camada de entrada;
- 3) Aplicar ativação com o operador de detecção com folga, sobre a imagem de entrada x com os pesos dos neurônios dos agrupamentos invariantes da camada morfológica (Expressão 2.25);
- 4) Produzir uma única saída para cada agrupamento através da expressão 4.9;
- 5) Ativar segunda camada interna, responsável pela recuperação de cada padrão detectado pela primeira camada interna, com a função de ativação dada pela expressão 4.10;
- 6) Apresentar o primeiro padrão detectado na camada de saída (Expressão 3.2);
- 7) Se existe alguma outra imagem, voltar ao passo 1;
- 8) Terminar ativação.

É importante ressaltar que o processo básico de ativação da RMNS não é alterado com a inclusão do mecanismo de invariância a rotação [73]. A única mudança na arquitetura consiste na inserção da operação de união das saídas de cada elemento de um agrupamento responsável pela representação de um padrão e suas rotações. A saída de cada agrupamento é tratada como sendo a saída de um único neurônio da RMNS conforme apresentada no capítulo 3.

4.3.1 Detalhes de Implementação

Para a implementação do método apresentado, alguns cuidados devem ser levados em consideração. Todas as imagens, rotacionadas ou não, foram armazenadas em matrizes retangulares. A rotação de uma matriz representando um padrão de referência pode gerar problemas para a realização do casamento com a imagem de busca. A mudança dos *pixels* para novas coordenadas pode gerar o aparecimento

de “buracos” em certas posições da imagem, e a perda de informações de algumas partes.

Como exemplo, suponha-se que o objetivo seja verificar se o padrão mostrado na Figura 4.6a está presente na imagem de busca mostrada na Figura 4.6b, que encontra-se rotacionada em 15 graus.

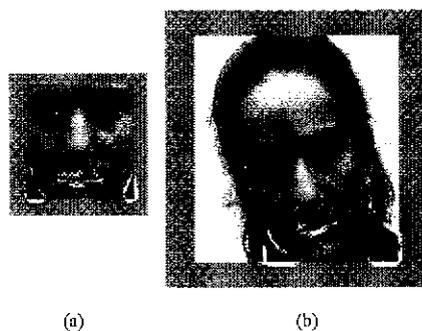


Fig. 4.6 - (a) Padrão de referência (b) Imagem de busca.

Sendo usado o modelo apresentado, o padrão de referência rotacionado por 15 graus, armazenado em uma matriz com as mesmas dimensões do padrão original, resulta na matriz mostrada na Figura 4.7.



Fig. 4.7 - Rotação do padrão de referência.

As partes em preto nos cantos da matriz são *pixels* que possuem valor zero, e não devem ser considerados durante o processo de casamento com a entrada, caso contrário podem provocar erro na busca do padrão de referência, podendo não haver detecção. Além disto, partes do padrão original foram deslocadas para fora dos limites da matriz, havendo perda de informação. Este problema ocorre com todos os ângulos não múltiplos de 90° .

Para contornar este problema, as imagens podem ser representadas por um *array* circular, conforme [47].

A Figura 4.8 mostra a rotação de um padrão circular. Desta forma, após a rotação, todos os *pixels* do padrão de referência permanecem contidos no círculo.



Fig. 4.8 - Rotação de padrão de referência com forma circular.

Neste caso, a imagem de busca deve ser percorrida por uma janela circular. Uma janela circular de raio $r \in \mathfrak{R}$ é um subconjunto de Z^2 que pode ser definido por

$$W = \{(x, y) \mid \sqrt{x^2 + y^2} \leq r\} \quad \forall x, y \in Z. \quad (4.12)$$

No entanto, a representação de padrões em um *array* circular pode não ser adequada em alguns casos. Algumas aplicações podem requerer a busca de padrões com formatos que não se adequem a uma representação circular. A Figura 4.11 da seção 4.3.2 mostra uma situação em que o uso de uma janela circular é inadequado devido a problemas de contexto.

Uma alternativa é representar os padrões em uma matriz quadrada ou retangular, tomando certas precauções para a geração das réplicas rotacionadas de modo a evitar o problema de perda de informação após a rotação da imagem. O resultado da rotação de um padrão de referência f_W é armazenado em uma matriz quadrada M . Após a rotação de f_W , este deve estar totalmente contido na matriz M . A Figura 4.9 ilustra a situação para uma rotação por 45° .

É necessário que M possua um tamanho tal que permita a rotação de f_W por qualquer ângulo, sem que haja perda de informação. Obtém-se as dimensões da menor matriz que satisfaça esta condição com base no comprimento da diagonal que liga o canto superior esquerdo ao canto inferior direito da matriz que armazena f_W . Seja L_W o número de linhas de f_W e seja C_W o número de colunas de f_W , pode-se

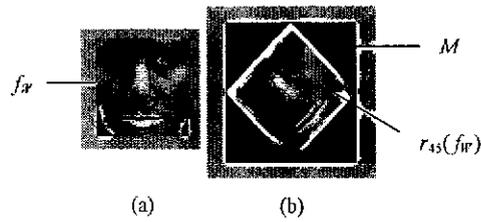


Fig. 4.9 - (a) Padrão de referência com orientação 0° . (b) Rotação de (a) por 45 graus.

obter o número de linhas (L_M) e de colunas (C_M) de M pela expressão

$$L_M = C_M = \left\lceil \sqrt{L_W^2 + C_W^2} \right\rceil. \quad (4.13)$$

4.3.2 Insensibilidade ao Contexto

Em muitas aplicações envolvendo detecção de padrões a dificuldade não está apenas relacionada com o reconhecimento de objetos individuais, mas também com a separação destes do contexto ou fundo da cena. Por exemplo, uma forma de se abordar o problema de reconhecimento de palavras manuscritas é segmentar os caracteres de maneira a separá-los do que os cerca e reconhecê-los isoladamente. Um típico reconhecedor de textos manuscritos usa heurísticas para formar múltiplos caracteres candidatos, cortando a palavra ou juntando traços próximos. O reconhecedor precisa, então, classificar cada candidato como um caracter ou rejeitá-lo como não caracter. Em muitas aplicações é difícil, ou mesmo impossível, conseguir heurísticas robustas de segmentação [76].

Várias técnicas de reconhecimento mencionadas na seção 1.2 requerem que a imagem de entrada seja extraída de seu contexto. É o caso das abordagens que envolvem a extração de características por meio de transformadas de Fourier e Radon, momentos, redes de terceira ordem, casamento estrutural, entre outras. Como exemplo da complexidade desta tarefa, considere-se o caracter chinês mostrado na Figura 4.10. Neste caso, a dificuldade se encontra em compor a figura como um todo, e não como três objetos separados.



Fig. 4.10 - Character chinês

Outro problema ocorre quando os objetos se encontram no interior de outros, como na Figura 4.11. No caso, o fato de o padrão interno tocar a borda da elipse dificulta a sua extração.

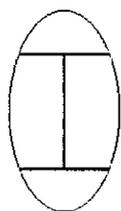


Fig. 4.11 - Padrão procurado inserido em outro objeto

O modelo de detecção proposto neste trabalho contorna estes problemas por meio do deslocamento de uma janela sobre a imagem de entrada, delimitando a região sobre a qual será aplicado o operador de detecção. A janela deve ser de um tamanho tal que contenha o padrão desejado com um mínimo de informações sobre o fundo da cena. Uma vez que o operador de detecção é aplicado diversas vezes sobre a imagem de entrada (de acordo com o número de padrões de referência rotacionados), pode-se utilizar uma janela diferente para cada ângulo de rotação. No exemplo da Figura 4.12 percebe-se que é possível isolar o padrão interno por meio de janelas com inclinações diferentes, representadas por linhas tracejadas.

A insensibilidade do sistema descrito a variações no contexto está restrita à área da janela que delimita o padrão de referência. Isto significa que se houver *pixels* representando o fundo da cena contidos nesta área, variações neste fundo na imagem de busca serão tratadas como ruído.

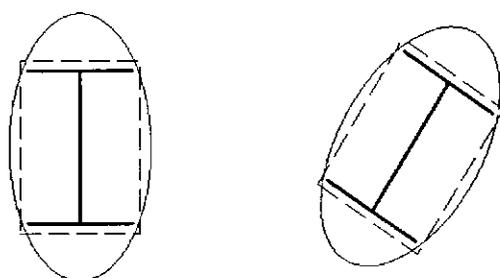


Fig. 4.12 - Extração de objetos por múltiplas janelas

4.3.3 Comentários Sobre Possíveis Ambigüidades

A possibilidade da presença de entradas rotacionadas pode gerar ambigüidades em um sistema de detecção. A Figura 4.13a exibe uma letra que pode ser interpretada como sendo um “Z” ou um “N” rotacionado por 270° . Semelhantemente, na Figura 4.13b a entrada pode representar um “W” ou um “M” rotacionado por 180° .

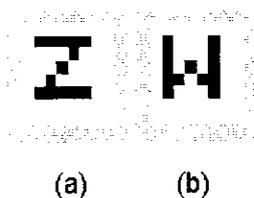


Fig. 4.13 - Exemplos de ambigüidades. (a) A figura pode estar representando a letra “Z” ou a letra “N” rotacionada por 270° . (b) A figura pode representar um “W” ou um “M” rotacionado por 180° .

Nestes casos, a RMNS apresenta como saída o padrão que vier primeiro no mapa morfológico.

Para uma aplicação em reconhecimento de caracteres, sugerimos uma análise contextual para determinar a inclinação do texto analisado. Os caracteres podem ser extraídos do texto em uma etapa de pré-processamento da maneira apresentada em [77]. A sua inclinação pode ser determinada baseado na inclinação dos caracteres à sua volta, caso haja uma situação de ambigüidade. O ângulo de rotação deve ser o mesmo para os diversos caracteres que compõem uma página de texto. Um conjunto

de regras, semelhante ao sistema híbrido proposto por [77], pode ser utilizado para determinar a saída correta. Ativam-se todos os neurônios morfológicos e, se mais de um deles detectar o padrão de entrada, é verificado o ângulo de rotação de outros padrões detectados anteriormente sem ambigüidade. A entrada é, em seguida, submetida somente aos neurônios responsáveis pelo ângulo encontrado. Após esta etapa, somente uma resposta é produzida [78].

4.3.4 Comparação entre RMNS e ART1 com Invariância a Rotação

Alguns comentários são feitos a seguir a respeito da quantidade de memória gasta pela camada morfológica invariante em relação à camada F_2 da rede ART1 modificada de [45].

No algoritmo da rede ART1 modificada, a invariância a translação é atingida também por meio da replicação dos padrões aprendidos, sendo que cada padrão aprendido é replicado transladado por todas as posições possíveis dentro da matriz de pesos de um neurônio. O aumento no número de padrões pode implicar em problemas de limitação de memória. Esta solução pode tornar-se muito dispendiosa no caso de reconhecimento de imagens de alta resolução por um sistema insensível a pequenas rotações, como 1 ou 2 graus.

O Operador de Detecção com Folga, usado na RMNS, já possui a propriedade de invariância a translação. Sendo assim, torna-se desnecessária a replicação dos padrões transladados no modelo proposto neste trabalho. O fato de não ser preciso replicar translações permite uma etapa de treinamento mais rápida.

A invariância a translação na RMNS é atingida percorrendo-se a imagem de entrada com uma janela com as mesmas dimensões dos elementos estruturantes, focalizando a região que será comparada. Embora esta solução, implementada em uma máquina seqüencial, exija mais tempo de processamento durante a ativação do que a forma utilizada pela rede ART1 modificada, ela representa uma grande economia de memória no caso de imagens de alta resolução, além de permitir a detecção independente do fundo da cena ao redor do objeto procurado. No caso da rede ART1,

é necessário que as entradas estejam separadas do contexto. Outra diferença é que a ART1 trabalha apenas com padrões de entrada binários, ao passo que o operador de detecção com folga pode processar padrões em escala de cinza.

No capítulo 5 é feita uma análise comparativa entre os resultados obtidos pela RMNS e pela rede ART1. Outros detalhes comparativos podem ser encontrados em [7].

CAPÍTULO 5

EXPERIMENTOS

5.1 INTRODUÇÃO

Neste capítulo são mostrados alguns resultados obtidos com o modelo de invariância a rotação aplicado à RMNS sobre imagens binárias e em níveis de cinza. É verificado, através de experimentação, o efeito da alteração do parâmetro de limiarização da rede, bem como a variação da quantidade de neurônios dos agrupamentos invariantes a rotação.

5.2 ESTUDO DE CASO PARA IMAGENS BINÁRIAS

5.2.1 Comparação com a rede ART1

Nesta seção é feita uma comparação entre os resultados obtidos pela rede ART1 invariante de [45] e pela RMNS invariante a rotação em experimentos com imagens binárias. Ambos os modelos foram implementados em Matlab para plataforma PC.

Para as simulações realizadas, utilizou-se o conjunto de treinamento mostrado em [45], com quatro letras representadas em matrizes de 8×8 *pixels*. Os dois sistemas foram configurados para serem invariantes a todas as translações possíveis e a rotações de 90° .

A Figura 5.1 mostra os resultados obtidos com a rede ART1 invariante a rotação após a apresentação das letras “P”, “M”, “K” e “U”. Os quatro padrões de entrada foram codificados em quatro categorias diferentes para uma taxa de vigilância=0.8334. Este parâmetro da rede ART1, estabelecido como sendo um valor entre 0 e 1, controla o grau de similaridade que deve ser atingido para que um padrão de entrada seja classificado como um dos padrões previamente aprendidos. Se o valor estabelecido não for atingido no teste de vigilância, a entrada é considerada uma nova classe e codificada em um nó livre da rede. Quanto mais próximo de 1 for o valor, menor quantidade de ruído será aceita na entrada. Maiores detalhes são descritos no

Apêndice A. O valor 0.8334 foi escolhido por ser o menor valor com o qual a rede é capaz de aprender estes quatro padrões sem cometer erros. As saídas indicadas na figura representam a classe detectada.

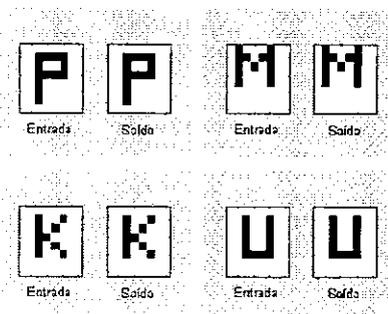


Fig. 5.1 - Treinamento da rede ART1 invariante a rotação e translação com entradas não rotacionadas.

As Figuras 5.2, 5.3 e 5.4 mostram os resultados da aplicação de ruído do tipo “sal” e “pimenta” (do inglês *salt and pepper*) em entradas rotacionadas da rede ART1 invariante treinada anteriormente. O termo “sal” representa a ocorrência de pontos brancos no fundo preto, e o ruído “pimenta” ocorre quando da presença de pontos pretos no fundo branco. Na Figura 5.2 foi utilizada a taxa de vigilância igual a 0.8334. Os padrões foram classificados corretamente, com exceção da letra “P”, que possui uma quantidade maior de ruído. Neste caso, foi criada uma nova classe para representar o padrão corrompido (erro por exclusão). Este efeito indesejável pode ser prevenido se não houver neurônios disponíveis para treinamento, com já comentado em [7]. Desta forma, nenhuma saída é produzida. A Figura 5.3 exhibe os resultados obtidos com a diminuição da taxa de vigilância para 0.7. O teste de vigilância passa a ser menos “rigoroso”, e a rede torna-se menos sensível a ruídos.

A rede ART1 possui problemas para lidar com ruído do tipo “sal”. A Figura 5.4 mostra os resultados de testes realizados com entradas “salgadas”. Por mais alta que seja a taxa de vigilância, a rede aceita entradas com alto nível de ruído “sal”, como ocorreu com a letra “M” na Figura 5.4b. No experimento realizado, foi utilizada uma taxa de vigilância de 0.99. As entradas ruidosas foram detectadas pelos neurônios representantes das letras “P” e “M”. Em ambos os casos a entrada foi aprovada

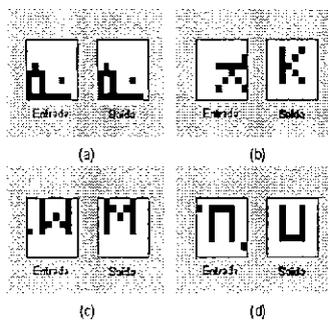


Fig. 5.2 - Ativação da rede ART1 invariante com entradas rotacionadas e ruidosas.

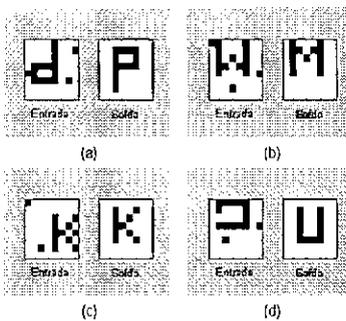


Fig. 5.3 - Efeito da redução da taxa de vigilância na rede ART1.

no teste de vigilância e os pesos foram adaptados de maneira a se aproximarem da entrada.

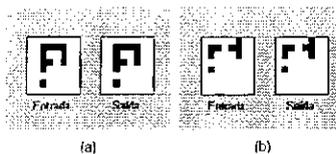


Fig. 5.4 - Problemas com ruído "sal" na rede ART1.

Este problema ocorre porque o teste de vigilância é feito através da verificação da quantidade de pontos 1 da imagem de entrada que se encontram presentes no padrão treinado, o que pode ser observado na expressão

$$\frac{\sum_{i=0}^{N-1} t_{Ji} I_i}{\sum_{i=0}^{N-1} I_i} > \rho, \quad (5.1)$$

onde ρ é a taxa vigilância, I_i é o vetor de entrada e t_{Ji} é o vetor de pesos *top-down* do neurônio vencedor J da fase de competição (ver apêndice A). N representa o

número de nós de entrada, ou o número de *pixels* da imagem de entrada. O vetor t_{J_i} contém a representação do padrão aprendido. Sempre que todos os pontos 1 da imagem de entrada I_i estiverem presentes no padrão treinado t_{J_i} o resultado do teste de vigilância será 1. Os pontos 1 a mais em t_{J_i} não têm influência no resultado. O ajuste dos pesos após a aprovação no teste de vigilância faz com que o ruído “sal” seja “aprendido” pela rede. Isto pode ser explicado pela regra de treinamento da rede:

$$t_{J_i}(t + 1) = t_{J_i}(t)I_i. \quad (5.2)$$

Percebe-se pela expressão 5.2 que apenas o ruído do tipo “sal” é aprendido pela rede. O mesmo não ocorre com ruído do tipo “pimenta”.

A RMNS foi treinada com os padrões mostrados anteriormente. A Figura 5.5 exhibe os pesos de um mapa morfológico composto por quatro neurônios após o treinamento.

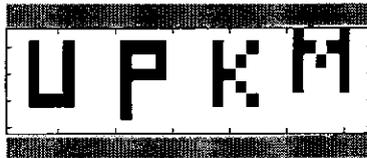


Fig. 5.5 - Pesos do Mapa Morfológico após o treinamento.

Em seguida foram submetidas à rede entradas rotacionadas e ruidosas. A Figura 5.6 mostra os resultados obtidos com o uso do parâmetro de limiarização igual a 2 (aproximadamente 3% da quantidade de *pixels* da imagem). Diferentemente da rede ART1, na RMNS o limiar de ruído é um número inteiro que representa uma contagem dos pontos ruidosos da entrada. As entradas com ruído inferior a 2 pontos produziram saídas corretas. A letra “P” na Figura 5.6d possui 3 pontos de ruído, sendo rejeitada pela rede. Na Figura 5.7 foi utilizado um limiar de 4 pontos, sendo aceitas entradas mais ruidosas.

O ajuste do parâmetro de limiarização tem efeito semelhante à variação da taxa de vigilância na rede ART1. Contudo, na RMNS o limiar atua sobre a quantidade de

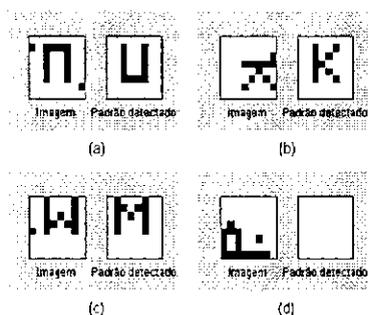


Fig. 5.6 - Saídas da RMNS com invariância a rotação para entradas ruidosas e rotacionadas usando um valor pequeno para o limiar de ruído.

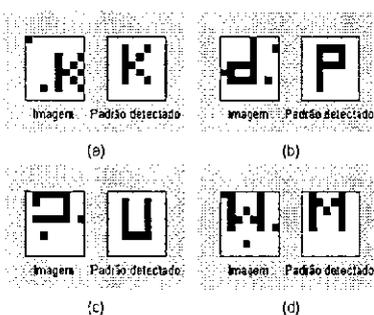


Fig. 5.7 - Aumento do limiar de ruído na RMNS.

pontos 0 e 1 que casam entre a imagem binária de entrada e o padrão de referência, sem que haja problemas com ruído do tipo “sal”, como pode ser visto na Figura 5.8. Novamente foi usado um limiar de ruído de 2 pontos. As entradas com ruído superior a este limite foram corretamente rejeitadas.

Da forma como foi implementada, durante a ativação a RMNS testa a entrada com os nós da camada morfológica e exibe como saída os pesos do primeiro nó que detectar o padrão. Esta ativação difere da ART1, onde ocorre uma competição entre os neurônios para verificar qual deles representa melhor o padrão de entrada.

Em alguns casos, quando a quantidade de ruído permitido na entrada for grande, mais de um neurônio podem responder ao padrão de entrada, e a ausência de uma etapa de competição pode causar resultados indesejados. Na Figura 5.9, a RMNS foi ativada utilizando-se um limiar de 5 pontos. Na Figura 5.9b a entrada foi considerada como sendo a letra “P”, já que a diferença está dentro da margem de erro

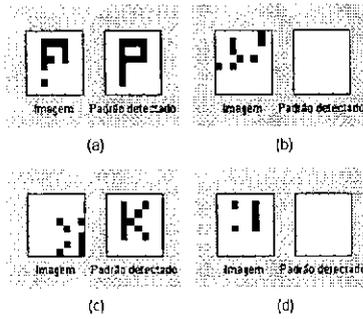


Fig. 5.8 - Comportamento da RMNS para entradas com ruído “sal” usando um valor pequeno para o limiar.

permitido e a letra “P” foi a primeira a satisfazer esta condição. Na Figura 5.10 é mostrado o resultado da ativação da rede ART1 usando uma taxa de vigilância igual a 0.5. Neste caso ocorre uma competição entre os neurônios treinados e a saída exibida representa os pesos do neurônio vencedor. Na Figura 5.10b o neurônio representando o padrão “U” rotacionado por 180° venceu a competição. A mesma etapa de competição pode ser implementada sobre a RMNS se for desejado, conforme mencionado em [7].



Fig. 5.9 - Comportamento da RMNS com limiar muito grande.

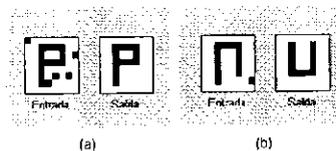


Fig. 5.10 - Comportamento da rede ART1 invariante a rotação com taxa de vigilância muito pequena.

5.2.2 Invariância ao Contexto

Para demonstrar o comportamento da RMNS com respeito a problemas de contexto [7], foram realizadas algumas simulações usando imagens binárias de ferramentas. A Figura 5.11 ilustra os padrões de referência selecionados para o treinamento da rede. Para cada teste a RMNS foi treinada com apenas uma das ferramentas. Os resultados mostrados nas Figuras 5.12, 5.13 e 5.14 foram obtidos usando um limiar de ruído equivalente a 5% do tamanho do padrão. No caso da Figura 5.14, o fato de o alicate estar em contato com outro objeto não dificultou sua detecção.



Fig. 5.11 - Ferramentas usadas para o treinamento da RMNS.

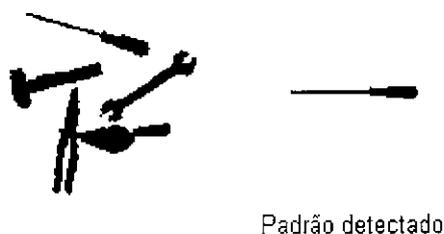


Fig. 5.12 - Saída da RMNS treinada com uma chave de fenda.

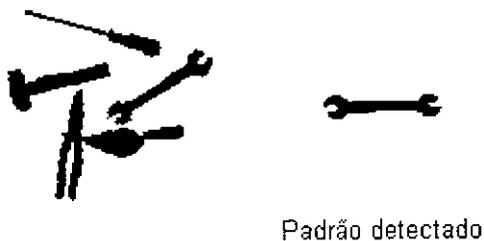


Fig. 5.13 - Saída da RMNS treinada com uma chave de boca.

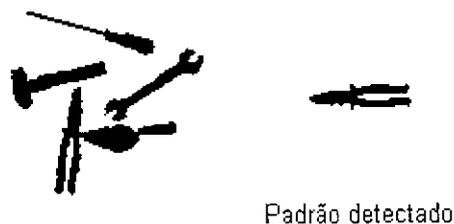


Fig. 5.14 - Saída da RMNS treinada com um alicate.

5.3 ESTUDO DE CASO PARA IMAGENS EM NÍVEIS DE CINZA

Para verificar o comportamento do modelo com entradas em níveis de cinza, foram realizadas simulações com reconhecimento de faces. Um conjunto de quatro faces para treinamento foi selecionado do *MIT Face DataBase*, obtido por *ftp* anônimo¹. Este banco de dados possui 27 imagens para cada pessoa, obtidas em três escalas diferentes, três inclinações da cabeça (centralizada, para a esquerda e para a direita) e três ângulos de iluminação. As imagens possuem resolução de 128x120 *pixels* codificados em 8 *bits* (256 tons de cinza).

Para as simulações realizadas treinou-se a RMNS com as quatro faces selecionadas em sua posição central (0 grau de inclinação) e iluminação frontal. Foi utilizada uma janela de 50x50 *pixels* posicionada manualmente sobre a imagem de maneira a extrair a porção contendo características faciais sem muitos dados desnecessários, como cabelos, fundo da cena etc. Estas regiões extraídas das 4 faces formam o conjunto de padrões de referência mostrado na Figura 5.15, aqui chamadas de *david1*, *joel1*, *thad1* e *rob1*, respectivamente. Deseja-se verificar se um destes padrões de referência encontra-se presente na imagem de entrada. Para a ativação foram usadas as imagens com as três inclinações de cada face com iluminação frontal. O conjunto total de testes é mostrado na Figura 5.16.

Após o treinamento, os pesos dos neurônios morfológicos foram replicados em outros

¹<ftp://whitechapel.media.mit.edu/pub/Images/>

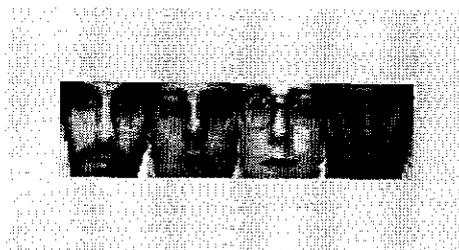


Fig. 5.15 - Conjunto de padrões de referência extraídos do MIT Face DataBase

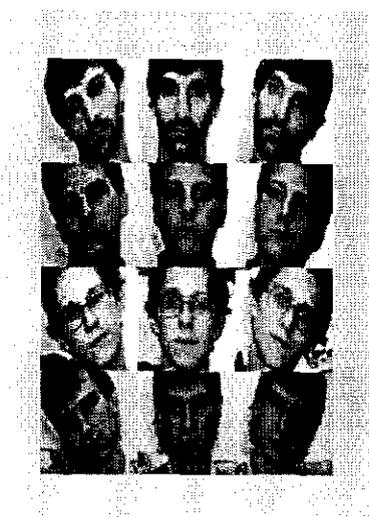


Fig. 5.16 - Conjunto de testes.

nós com inclinações entre 20° e 30° , e entre 340° e 330° com intervalos de 2 graus entre uma cópia e outra. Entradas com inclinações ligeiramente diferentes são tratadas pelo limiar de ruído, que confere à RMNS robustez na detecção de padrões pouco rotacionados [7]. Inicialmente foi utilizado um limiar de ruído de 840 pontos, o que corresponde a 33.6% dos pontos dos padrões de referência. Além disso, foi usada uma folga radiométrica igual a 0.2 (para imagens com 256 níveis de cinza mapeados entre 0 e 1). Esta folga corresponde ao tamanho do intervalo $[c_1, c_2]$ definido na seção 2.4.2.1.

As Figuras 5.17, 5.18, 5.19 e 5.20 mostram alguns dos resultados obtidos. Cada figura exibe um par de entrada e saída obtida.

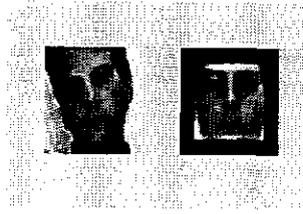


Fig. 5.17 - Ativação com *joel3*.

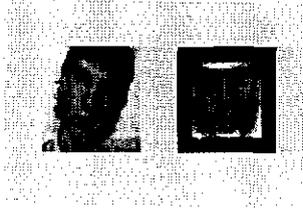


Fig. 5.18 - Ativação com *rob2*.

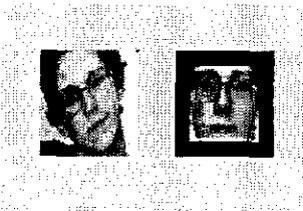


Fig. 5.19 - Ativação com *thad3*.

Quando ativada com um padrão não treinado, a rede não produziu saída, como pode ser visto na Figura 5.21.

Das 12 entradas do conjunto de testes, 9 foram classificadas corretamente e 3 não produziram saída (erro por exclusão), não havendo ocorrido nenhum erro por inclusão (apresentação de uma saída errada).

O aumento do limiar de ruído para 900 pontos (36% dos pontos do padrão) diminuiu a ocorrência de erros de exclusão. Para o conjunto de testes utilizado, obteve-se 11 classificações corretas e 1 entrada não produziu saída. Contudo, o uso de um limiar muito grande aumenta a quantidade de classificações erradas. A apresentação da entrada *wave*, mostrada na Figura 5.21, produziu como saída o padrão *joel*.

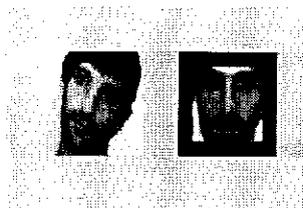


Fig. 5.20 - Ativação com *david2*.



Fig. 5.21 - Ativação com *wave*.

Com o aumento do limiar, mais de um nó morfológico podem ser “disparados” para uma dada entrada. Neste caso, como a rede exibe como saída os pesos do neurônio de menor índice na camada morfológica, o resultado pode não ser correto [7]. Nas simulações realizadas, o limiar foi ajustado de acordo com um pequeno conjunto de treinamento, visando apenas exemplificar a capacidade da rede de discriminar corretamente entradas que sejam diferentes dos padrões treinados. Para aplicações com um maior número de padrões de treinamento, uma solução pode ser a redução da porcentagem do limiar em conjunto com a utilização de imagens de maior resolução. Outra possibilidade é a aplicação de uma etapa de competição sobre as saídas para determinar a que melhor se aproxima da entrada. Esta alternativa é adequada no caso em que se deseja ativar a rede somente com imagens de pessoas que tiveram suas faces treinadas.

5.4 VARIAÇÃO DO LIMIAR DE RUÍDO

Nesta seção são comentados alguns resultados obtidos com a variação do parâmetro de limiarização da RMNS com o intuito de analisar seu efeito sobre entradas rotacionadas. Para as simulações foi utilizado um conjunto de quatro faces selecionadas do

*Manchester Face DataBase*². As imagens originais possuem resolução de 256x256 *pixels* com 256 tons de cinza. Visando redução de tempo de processamento, estas imagens foram redimensionadas para 61x61 *pixels*. Isto resulta em um conjunto de dados com dimensões reduzidas contendo a maior parte das informações relevantes. Procedimento semelhante foi efetuado por [49]. Uma janela de 26x26 *pixels* foi posicionada sobre as faces para extrair as regiões de interesse, mostradas na Figura 5.22, aqui chamadas de *man1*, *man2*, *woman1* e *woman2*, respectivamente. Embora a redução do tamanho das imagens possa resultar em diminuição da taxa de reconhecimento, o tamanho é suficiente para demonstrar os efeitos da variação dos parâmetros da rede.

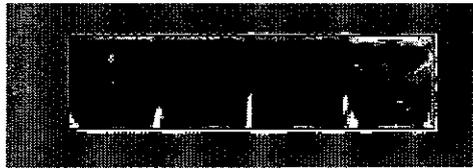


Fig. 5.22 - Conjunto de padrões de referência.

Inicialmente foi treinada uma RMNS com invariância a rotações com intervalos de 10° . Isto significa que os pesos dos neurônios morfológicos serão rotacionados de 10 em 10 graus. No primeiro conjunto de testes deseja-se verificar o efeito da variação do limiar de ruído uma vez fixado este intervalo de rotação arbitrariamente escolhido. As entradas para os testes são obtidas rotacionando as imagens treinadas por 4 ângulos escolhidos entre 320° (-40°) e 40° graus. Estes ângulos são escolhidos de maneira a ficarem entre os intervalos treinados. Desta forma, se os pesos estão rotacionados em 10 e 20 graus, a entrada teste estará rotacionada em 15 graus. A folga radiométrica foi fixada em 0.2 (para imagens com 256 níveis de cinza mapeados entre 0 e 1).

A Figura 5.23 exhibe a saída obtida da ativação com o padrão *woman1* rotacionado por 15° .

²<http://peipa.essex.ac.uk/ipa/pix/faces/manchester/>

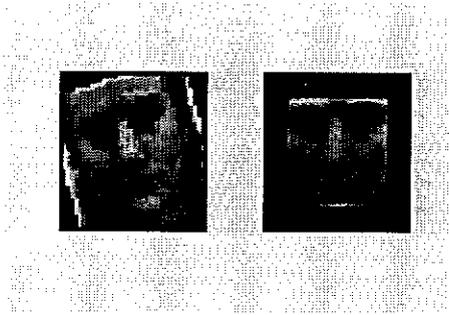


Fig. 5.23 - Ativação com *woman1* rotacionado por 15° .

Foram realizados testes com 7 variações do limiar da rede. Para cada valor do limiar foram feitos 4 testes com cada uma das faces, num total de 112 testes. Os resultados obtidos são mostrados na Tabela 5.1.

TABELA 5.1 - VARIAÇÃO DO LIMIAR

Limiar	101	115	130	150	162	180	210
Erros por Exclusão	7	5	4	4	4	2	0
Erros por Inclusão	0	0	1	2	2	3	3

Na Figura 5.24 são plotadas as quantidades de erros de exclusão e de inclusão em função da variação do limiar. Pelos dados obtidos, percebe-se a redução da quantidade de erros por exclusão com o aumento do limiar, e em certo ponto do gráfico, redução do erro total (inclusão + exclusão). Contudo, este aumento do limiar implica em aumento na quantidade de erros por inclusão, que podem ser menos aceitáveis do que os erros de exclusão, dependendo da aplicação.

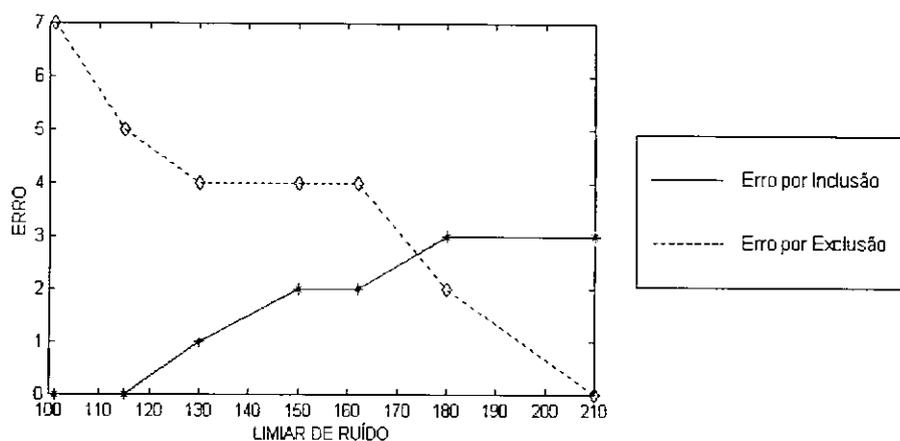


Fig. 5.24 - Quantidade de erros em relação à variação do limiar.

5.5 VARIAÇÃO DO INTERVALO DE ROTAÇÃO

Em seguida foram realizados testes com a variação do intervalo de rotação entre as cópias dos pesos dos nós da camada morfológica. Foi escolhido um valor fixo para o limiar de maneira que não houvesse erros de inclusão. Nos resultados exibidos na Tabela 5.2 e na Figura 5.25 foi utilizado um limiar igual a 115. As entradas consistem nos mesmos padrões do *Manchester Face DataBase* treinados na seção anterior, rotacionados por ângulos entre -24° e 24° que se encontram entre os intervalos treinados, como feito anteriormente. Foram realizados testes com 6 intervalos de rotação diferentes. Para cada intervalo foram feitos 4 testes com diferentes rotações de cada uma das faces, sendo 96 testes no total. Os resultados demonstram, para este conjunto de treinamento, as vantagens de se utilizar intervalos menores de rotação.

TABELA 5.2 - VARIAÇÃO DO INTERVALO DE ROTAÇÃO

Ângulo entre os pesos	2°	4°	6°	8°	10°	12°
Erros por Exclusão	1	2	3	4	6	8

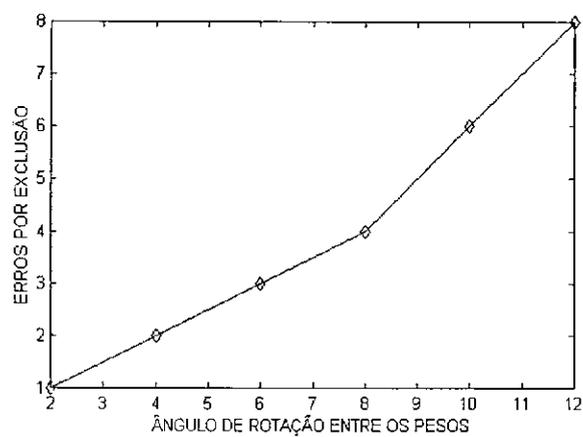


Fig. 5.25 - Quantidade de erros em relação à variação do intervalo de rotação entre os pesos dos nós da RMNS.

CAPÍTULO 6

CONCLUSÕES

Os esforços desempenhados nas últimas décadas na tarefa de reconhecimento de padrões invariante a rotação, bem como a continuidade das pesquisas na área demonstram a dificuldade do problema em questão.

Este trabalho teve por objetivo a modelagem e implementação de um sistema de reconhecimento de padrões invariante a translação e rotação utilizando Morfologia Matemática.

Em comparação com outras técnicas pesquisadas, o uso de operadores morfológicos possui a vantagem de aliar a invariância a translação com a capacidade de detecção de padrões insensível ao contexto, sendo dispensada uma etapa de pré-processamento para a segmentação da imagem de entrada.

O modelo foi implementado sobre a RMNS e comparado com a variação da rede ART1 apresentada em [45] com imagens binárias. Embora tendo apresentado desempenho semelhante com imagens com ruído “pimenta”, a RMNS mostrou-se mais robusta do que a ART1 para lidar com imagens com entradas com ruído “sal”.

O comportamento do sistema sobre imagens em níveis de cinza foi verificado através dos resultados obtidos com reconhecimento de faces rotacionadas. Embora tendo sido feitos testes com poucas imagens, os resultados preliminares sugerem a viabilidade de utilização do sistema em aplicações reais.

Alguns experimentos com a variação do limiar de ruído da rede e do ângulo de rotação entre os pesos dos neurônios dos agrupamentos invariantes a rotação fornecem indicações sobre a influência destes parâmetros na detecção de entradas rotacionadas. Em geral é possível aumentar o ângulo entre os pesos rotacionados com o aumento do limiar da rede, possibilitando a diminuição da quantidade de réplicas dos pesos na camada morfológica. Em futuros trabalhos, pode-se pesquisar métodos

mais precisos para a determinação dos melhores valores para estes parâmetros.

Nos experimentos realizados foi utilizado um valor fixo para a folga radiométrica. Este valor foi estabelecido com base em resultados obtidos em testes iniciais.

Uma série de heurísticas podem ser exploradas como forma de implementação do modelo apresentado, visando economia de memória. Uma alternativa é a replicação dos padrões em agrupamentos com diversas inclinações, com ângulos de rotação múltiplos de um determinado ângulo ϕ . Durante o processo de detecção, uma janela percorre a imagem de entrada em busca de padrões conhecidos. A porção focalizada pela janela é comparada com os padrões armazenados na matriz de elementos estruturantes. Se não ocorrer nenhum casamento similar o bastante, dentro do limite de erro estabelecido, a imagem contida na janela de busca é rotacionada α graus N vezes até completar o intervalo ϕ estabelecido. Por exemplo, para $\phi=90^\circ$, e $\alpha=30^\circ$, no caso de um sistema invariante a rotação a cada 30° , o número de cópias necessárias para representar cada padrão seria 4 (0° , 90° , 180° e 270°), e o número máximo de operações de rotação necessárias para girar a matriz contida na janela de busca dentro do intervalo de 90° , a um passo de 30° , seria 2 [73].

Dado o fato de que para um grande número de aplicações a orientação dos objetos a serem reconhecidos sofre alterações relativamente pequenas em relação à sua forma padrão, pode-se implementar uma variação deste modelo, em que cada um dos agrupamentos possua cópias rotacionadas dos padrões, com variação de inclinação em um intervalo pequeno, (-45° a 45° , por exemplo). Se, durante a ativação, não ocorrer um casamento correto, significa que a imagem focalizada pode estar rotacionada em mais de 45° em sentido horário ou anti-horário, ou não foi treinada pela rede. É, então, necessário realizar a rotação da matriz contida na janela de busca por um ângulo estabelecido. Após cada rotação, é realizado novamente o casamento com os padrões armazenados. A operação se repete até que se completem os 270° do intervalo entre 45° e 315° (-45°), ou o padrão seja detectado. O padrão contido na janela de busca pode ser rotacionado nos sentidos horário e anti-horário, alternadamente, com um ângulo sucessivamente maior. O tempo de reconhecimento se

torna maior à medida que a inclinação do objeto se aproxima de 180^0 em relação à imagem padrão. Objetos com inclinação inferior a 45^0 podem ser detectados sem a necessidade de rotação da entrada.

Uma possibilidade de implementação sobre a RMNS, mencionada em [7], consiste na criação de uma camada de neurônios direcionais que são carregados por cópias rotacionadas dos pesos de um neurônio selecionado da camada morfológica em tempo de ativação.

Considera-se que os objetivos iniciais do trabalho foram atingidos e espera-se que o modelo proposto possa contribuir como uma ferramenta alternativa para o reconhecimento de padrões invariante a rotação e translação.

REFERÊNCIAS BIBLIOGRÁFICAS

- [1] Pao, Y. **Adaptive pattern recognition and neural networks**, [S.l.] Addison-Wesley, 1989. 309p.
- [2] Jain, A. Advances in statistical pattern recognition. In: Devijver, P. A.; Kittler, J. ed. **Pattern recognition theory and applications**. Berlin: Springer-Verlag, 1987. NATO ASI Series, v. F30, Cap. 1, p. 1-11.
- [3] Deseilligny, M.; Men, H.; Stanon, G. Character string recognition on maps, a rotation-invariant recognition method. **Pattern Recognition Letters**. v. 16, n. 12, p. 1297-1310, Dec. 1995.
- [4] Lisboa, P. G. J.; Perantonis, S. J. Invariant pattern recognition using third-ordered networks and Zernike moments. In: **International Joint Conference on Neural Networks**, Singapore, 1991. **hspacelexProceedings**. Singapore: IEEE, 1991. p. 1421-1425.
- [5] Fukumi, M.; Omatu, S.; Nishikawa, Y. Rotation-invariant neural pattern recognition system estimating a rotation angle. **IEEE Transactions on Neural Networks**, v. 8, n. 3, p. 568-581, May 1997.
- [6] Faria, S. **Uma abordagem morfológica para casamento de padrões**. São José dos Campos. 83 p. (INPE-6346-TDI/597). Dissertação (Mestrado em Computação Aplicada) - Instituto Nacional de Pesquisas Espaciais, 1997.
- [7] Silva, F. A. T. F. **Rede Morfológica Não-Supervisionada-RMNS**. São José dos Campos. Tese (Doutorado em Computação Aplicada) - Instituto Nacional de Pesquisas Espaciais, 1998.

- [8] Goshtasby, A. Template matching in rotated images. **IEEE Transactions on Pattern Analysis and Machine Intelligence**. v. PAMI-7, n. 3, p. 338-344, May 1985.
- [9] Hsu, Y.; Arsenault, H.; April, G. Rotation-invariant digital pattern recognition using circular harmonic expansion. **Applied Optics**, v. 21, n. 22, p. 4012-4016, 15 Nov. 1982.
- [10] Hsu, Y.; Arsenault, H. Optical pattern recognition using circular harmonic expansion. **Applied Optics**. v. 21, n. 22, p. 4016-4019, 15 Nov. 1982.
- [11] Chen, C. H.; Wu, R. Study of a class of invariant recognition for machine vision. **Proceedings of SPIE - Visual Communications and Image Processing**. Cambridge, Massachusetts: SPIE. v. 707, p. 17-22, Sep. 1986.
- [12] Wechsler, H.; Zimmerman, G. 2-D invariant object recognition using distributed associative memory. **IEEE Transactions on Pattern Analysis and Machine Intelligence**. v. 10, n. 6, p. 811-821, Nov. 1988.
- [13] Casasent, D.; Psaltis, D. New optical transforms for pattern recognition. **Proceedings of the IEEE**. v. 65, n. 1, p. 77-84, Jan. 1977.
- [14] Schalkoff, R. **Digital image processing and computer vision**. New York: John Wiley & Sons, 1989. 489 p.
- [15] Altmann, J. On the digital implementation of the rotation-invariant Fourier-Mellin transform. **Journal of Information Processing and Cybernetics**, v. EIK-23, n. 1, p. 13-36, 1987.
- [16] Caelli, T.; Liu, Z. On the minimum number of templates required for shift, rotation and size invariant pattern recognition. **Pattern Recognition**. v. 21, n. 3, p. 205-216, 1988.

- [17] Pintsov, D. Invariant pattern recognition, symmetry, and Radon transforms. **Journal of Optical Society of America A**, v. 6, n. 10, p. 1544-1553, Oct. 1989.
- [18] Simpson, M.; Ervin, P.; Snyder, M. Radon transform applications in optical inspection. **Optical Engineering**, v. 27, n. 2, p. 164-171, Feb. 1998.
- [19] Gindi, G.; Gmitro, A. Optical feature extraction via the Radon transform. **Optical Engineering**, v. 23, n. 5, p. 499-506, Sep./Oct. 1984.
- [20] Lin, C.; Chellappa, R. Classification of partial 2-D shapes using Fourier descriptors. **IEEE Transactions on Pattern Analysis and Machine Intelligence**, v. PAMI-9, n. 5, p. 686-690, Sep. 1987.
- [21] Reeves, A.; Prokop, R.; Andrews, S.; Kuhl, F. Three-dimensional shape analysis using moments and Fourier descriptors. **IEEE Transactions on Pattern Analysis and Machine Intelligence**, v. 10, n. 6, p. 937-943, Nov. 1988.
- [22] Pavlidis, T.; Ali, Farhat. A hierarchical syntactical shape analyzer. **IEEE Transactions on Pattern Analysis and Machine Intelligence**, v. PAMI-1, n. 1, p. 2-9, Jan. 1979.
- [23] Fu, K. **Syntactic pattern recognition and applications**, Englewood Cliffs, NJ: Prentice-Hall, 1982. 596 p.
- [24] Shapiro, L.; Haralick, R. Structural descriptions and inexact matching. **IEEE Transactions on Pattern Analysis and Machine Intelligence**, v. 3, n. 5, p. 504-519, Sep. 1981.
- [25] Bins, L. S. **Uso do casamento estrutural para registro de imagens de satélite**. São José dos Campos. 91 p. (INPE 4734-TDL/345).
Dissertação (Mestrado em Computação Aplicada) - Instituto Nacional de Pesquisas Espaciais, 1988.

- [26] Hu, M. Visual pattern recognition by moment invariants. **IRE Transactions on Information Theory**, v. IT-8, p. 179-187, Feb. 1962.
- [27] Dudani, S.; Breeding, K.; McGhee, R. Aircraft identification by moment invariants. **IEEE Transactions on Computers**, v. C-26, n. 1, p. 39-45, Jan. 1977.
- [28] Khotanzad, A.; Hong, Y. Invariant image recognition by Zernike moments. **IEEE Transactions on Pattern Analysis and Machine Intelligence**, v. 30, n. 5, p. 489-497, May 1990.
- [29] Varella, L.; Passos, E.; Santos, M.; Araujo, R. Reconhecimento de Caracteres com Variações de Escala, Rotação e Translação. In: Simpósio Brasileiro de Computação Gráfica e Processamento de Imagens (SIBGRAPI 92), 5., Águas de Lindóia, 1992. **Anais**. Águas de Lindóia: Sociedade Brasileira de Computação e Instituto Nacional de Pesquisas Espaciais, 1992. p. 301-310.
- [30] Wang, S.; Chen, P.; Lin, W. Invariant pattern recognition by moment Fourier descriptor. **Pattern Recognition - The Journal of the Pattern Recognition Society**. v.27, n. 12, p. 1735-1741, Dec. 1994.
- [31] Wong, W.; Siu, W.; Lam, K. Generation of moment invariants and their uses for character recognition. **Pattern Recognition Letters**, v. 16, n. 2, p. 115-123, Feb. 1995.
- [32] Zhao, D.; Chen, J. Affine curve moment invariants for shape recognition. **Pattern Recognition**, v. 30, n. 6, p. 895-901, 1997.
- [33] Giles, C.; Maxwell, T. Learning, invariance, and generalization in high-order neural networks. **Applied Optics**, v. 26, n. 23, p. 4972-4978, Dec. 1987.

- [34] Reid, M.; Spirkovska, L.; Ochoa, E. Rapid training of higher-order neural networks for invariant pattern recognition. In: IEEE INNS International Joint Conference On Neural Networks, Washington, 1989. **Proceedings**. San Diego: IEEE TAB neural network committee, 1989. p. I-689 - I-692.
- [35] Spirkovska, L.; Reid, M. Connectivity strategies for higher-order neural networks applied to pattern recognition. In: International Joint Conference on Neural Networks, San Diego, 1990. **Proceedings**. San Diego: IEEE Neural Networks Council, 1990. p. I-21 - I-26.
- [36] Perantonis, S. J. Higher-order neural networks for invariant pattern recognition. In: Lisboa, P. G. J. ed. **Neural networks: current applications**. London: Chapman & Hall, 1992. Cap. 11, p. 197-229.
- [37] Kanaoka, T.; Chellappa, R.; Yoshitaka, M.; Tomita, S. A higher order neural network for distortion invariant pattern recognition. **Pattern Recognition Letters**, v. 13, n. 12, p. 837-841, Dec. 1992.
- [38] Kwon, H.; Kim, B.; Cho, D.; Hwang. Scale and rotation invariant pattern recognition using complex-log mapping and augmented second order neural networks. **Electronic Letters**, v. 29, n. 7, p. 620-621, 1st Apr. 1993.
- [39] Fukushima, K. A neural network for visual pattern recognition. **Computer**, v. 21, n. 3, p. 65-75, Mar. 1988.
- [40] Satoh, S.; Kuroiwa, J.; Aso, H.; Miyake, S. **Recognition of rotated patterns using neocognitron**. [on line] <<http://www.aso.ecci.tohoku.ac.jp/shun/research-e.html>>. Dez. 1998.
- [41] Sheng, Y.; Lejeune, C. Invariant pattern recognition using Fourier-Mellin transforms and neural networks. **Journal of Optics**, v. 22, n. 5, p. 223-228, 1991.

- [42] Widrow, B.; Winter, R.; Baxter, R. Layered neural nets for pattern recognition. **IEEE Transactions on Acoustics, Speech and Signal Processing**. v. 36, n. 7, p. 1109-1118, July 1988.
- [43] Srinivasa, N.; Jouaneh, M. A neural network model for invariant pattern recognition. **IEEE Transactions on Signal Processing**. v. 40, n. 6, p. 1595-1598, June 1992.
- [44] Carpenter, G.; Grossberg, S. The ART of adaptive pattern recognition by a self-organizing neural network. **Computer**, v. 21, n. 3, p. 77-88, Mar. 1988.
- [45] Srinivasa, N.; Jouaneh, M. An invariant pattern recognition machine using a modified ART architecture. **IEEE Transactions on System, Man, and Cybernetics**. v. 23, n. 5, p. 1432-1437, Sep./Oct. 1993.
- [46] Hosokawa, M.; Omatu, S.; Fukumi, M. A new approach for pattern recognition by neural networks with scramblers. In: **IEEE International Joint Conference on Neural Networks**, Washington, 1989. **Proceedings**. San Diego: IEEE TAB Neural Network Committee, 1989. p. I183-I188.
- [47] Fukumi, M.; Omatu, S.; Takeda, F.; Kosaka, T. Rotation-invariant neural pattern recognition system with application to coin recognition. In: **International Joint Conference on Neural Networks**, Singapore, 1991. **Proceedings**. Singapore: IEEE, 1991. p. 1027-1032.
- [48] Ranganath, S.; Arun, K. Face recognition using transform features and neural networks. **Pattern Recognition**, v. 30, n. 10, p. 1623-1636, 1997.
- [49] Howell, A.; Buxton, H. **Invariance in radial basis function neural networks in human face classification**. [on line]
<<ftp://ftp.cogs.susx.ac.uk/pub/reports/csrp/csrp365.ps.Z>> . Mar. 1998.

- [50] Loui, A.; Venetsanopoulos, A.; Smith, K. Morphological autocorrelation transform: a new representation and classification scheme for two-dimensional images. **IEEE Transactions on Image Processing**, v. 1, n. 3, p. 337-354, July 1992.
- [51] Beale, R.; Jackson, T. **Neural computing: an introduction**. Bristol: Institute of Physics Publishing Ltd, 1992. 240 p.
- [52] Oliveira, C.; Silva, F.; Oliveira, T. **Processamento de língua natural: uma abordagem simbólica ou conexionista?** São José dos Campos: INPE, 1996. 50 p. (INPE-5971-PRP/194).
- [53] Pham, D.; Liu, X. **Neural networks for identification, prediction and control**, [S.l.], Springer-Verlag, 1995. 238 p.
- [54] Arbib, M. A. **The handbook of brain theory and neural networks**. Cambridge, Massachusetts: The MIT Press, 1995. 1118 p.
- [55] McClelland, J.; Plunkett, K. Cognitive development. In: Arbib, M. A. ed. **The handbook of brain theory and neural networks**. Cambridge, Massachusetts: The MIT Press, 1995. p. 193-197.
- [56] Lisboa, P. G. J. **Neural networks: current applications**. London: Chapman & Hall, 1992. 279 p.
- [57] Carrara, V. **Redes neurais aplicadas ao controle de atitude de satélites com geometria variável**. São José dos Campos. 178 p. (INPE-6384-TDI/603). Dissertação (Doutorado em Engenharia e Tecnologia Espaciais/Mecânica Espacial e Controle) - Instituto Nacional de Pesquisas Espaciais, 1997.
- [58] Wasserman, P. **Neural computing: theory and practice**, New York: Van Nostrand Reinhold, 1989. 230 p.
- [59] Hecht-Nielsen, R. **Neurocomputing**. [S.l.], Addison-Wesley, 1989. 433p.

- [60] Maren, A.; Harston, C.; Pap, R. **Handbook of neural computing applications**. San Diego: Academic Press, 1990. 450 p.
- [61] Kung, S. **Digital Neural Networks**. Englewood Cliffs, NJ: PTR Prentice-Hall, 1993. 442 p.
- [62] Intrator, N. Competitive Learning. In: Arbib, M. A. ed. **The handbook of brain theory and neural networks**. Cambridge, Massachusetts: The MIT Press, 1995. p. 220-223.
- [63] Banon, G. J. F.; Barrera, J. Bases da morfologia matemática para a análise de imagens binárias. In: Escola de Computação, 9., Recife, 1994. **Anais**. Recife: Departamento de informática da Universidade Federal de Pernambuco, 1994. 230p.
- [64] Banon, G. J. F.; Candeias, A. L. **Restauração de imagens NOAA por morfologia matemática**. São José dos Campos: INPE, 1997. 7 p. (INPE-6132-PRE/2232).
- [65] Serra, J. **Image analysis and mathematical morphology**. London: Academic Press, 1982. 610 p.
- [66] Wilson, S. Training structuring elements in morphological networks. In: Dougherty, E. ed. **Mathematical morphology in image processing**. New York: Marcel Dekker, 1993. p. 1-41.
- [67] Banon, G. J. F.; Faria, S. Morphological approach for template matching. In: Brazilian Symposium of Computer Graphic and Image Processing (SIBGRAPI '97), 10., São Paulo, 1997. **Proceedings**. São Paulo: SBC, 1997. p. 171-178.
- [68] Song, J.; Delp, E. The analysis of morphological filters with multiple structuring elements. **Computer Vision, Graphics, and Image Processing**, v.50, n. 3, p. 308-328, June 1990.

- [69] Roerdink, J.; Heijmans, H. J. Mathematical morphology for structures without translation symmetry. **Signal Processing**, v. 15, n. 3, p. 271-277, Oct. 1988.
- [70] Heijmans, H.; Ronse, C. The algebraic basis of mathematical morphology I. Dilations and erosions. **Computer Vision, Graphics, and Image Processing**, v. 50, n. 3, p. 245-295, June 1990.
- [71] Maragos, P. Affine morphology and affine signal models. **Proceedings of SPIE - Image Algebra and Morphological Image Processing**. San Diego: SPIE. v. 1350, p. 31-43, July 1990.
- [72] Roerdink, J. **On the construction of translation and rotation invariant morphological operators**. Amsterdam: CWI, Centre for Mathematics and Computer Science, Dept of Analysis, Algebra and Geometry, Dec. 1990, 29 p. (Report AM-R9025).
- [73] Silva, F. A. T. F. **Proposta de implementação de invariância a rotação na RMNS**. (Instituto Nacional de Pesquisas Espaciais, São José dos Campos, 1998). Comunicação Pessoal.
- [74] Kohonen, T. The self-organizing map. **Proceedings of the IEEE**, v. 78, n. 9, p. 1464-1480, Sep. 1990.
- [75] Silva, F. A. T. F. **Definição de operadores morfológicos de casamento de padrões invariantes a rotação**. (Instituto Nacional de Pesquisas Espaciais, São José dos Campos, 1998). Comunicação Pessoal.
- [76] LeCun, Y.; Bengio, Y. Pattern recognition. In: Arbib, M. A. ed. **The handbook of brain theory and neural networks**. Cambridge, Massachusetts: The MIT Press, 1995. p. 711-715.

- [77] Ferreira, S. **Um modelo simbólico-conexionista para interpretação de cenas.** São José dos Campos. 97 p. (INPE-6842-TDI/649). Dissertação (Mestrado em Computação Aplicada) - Instituto Nacional de Pesquisas Espaciais, 1998.
- [78] Silva, F. A. T. F. **Sugestão para implementação de OCR invariante a rotação.** (Instituto Nacional de Pesquisas Espaciais, São José dos Campos, 1998). Comunicação Pessoal.

APÊNDICE A

ALGORITMO DA REDE ART1

A ART1 de Carpenter e Grossberg [44] é uma rede não-supervisionada que processa padrões binários por meio de duas camadas de neurônios, F_1 e F_2 , ilustradas na Figura A.1. Cada elemento da camada F_1 é ligado a todos os elementos da camada F_2 por meio de conexões sinápticas chamadas de *bottom – up* e cada elemento da camada F_2 é ligado a todos os nós da camada F_1 por meio de conexões chamadas de *top – down*. O algoritmo da rede ART1, extraído de [51], é apresentado abaixo.

1. Inicialização

$$\begin{aligned} t_{ij}(0) &= 1 \\ b_{ij}(0) &= \frac{1}{1 + N} \end{aligned} \tag{A.1}$$

$$0 \leq i \leq N - 1 \quad 0 \leq j \leq M - 1$$

Inicializa ρ , $0 \leq \rho \leq 1$

onde $t_{ij}(t)$ é a conexão *top – down* e $b_{ij}(t)$ é a conexão *bottom – up* entre o nó i da camada F_1 e o nó j da camada F_2 no instante t . São estes pesos que definem o exemplar especificado pelo nó j de saída. ρ é a taxa de vigilância que determina quão próxima uma entrada precisa estar para casar corretamente com um exemplar armazenado. Há M nós de saída e N nós de entrada.

2. Apresenta nova entrada

3. Calcula o *matching* entre a entrada e os pesos *Bottom – up*

$$\mu_j = \sum_{i=0}^{N-1} b_{ij}(t)x_i \tag{A.2}$$

$$0 \leq j \leq M - 1$$

μ_j é a saída do nó j da camada F_2 e x_i é o i -ésimo elemento do vetor de entrada x , que pode ser 0 ou 1.

4. Seleciona o exemplar de melhor *matching*.

$$\mu_{j^*} = \max_j [\mu_j] \quad (\text{A.3})$$

5. Teste

$$\begin{aligned} \|X\| &= \sum_{i=0}^{N-1} x_i \\ \|T \cdot X\| &= \sum_{i=0}^{N-1} t_{ij^*}(t)x_i \\ \text{se } \frac{\|T \cdot X\|}{\|X\|} &> \rho \end{aligned} \quad (\text{A.4})$$

ENTÃO vai para 7

SENÃO vai para 6

6. Desabilita o melhor *match*, impedindo-o de participar do treinamento.

Faz saída do melhor *match* ser igual a 0. Vai para 3.

7. Adapta o melhor *match*

$$\begin{aligned} t_{ij^*}(t+1) &= t_{ij^*}(t)x_i \\ b_{ij^*}(t+1) &= \frac{t_{ij^*}(t)x_i}{0.5 + \sum_{i=0}^{N-1} t_{ij^*}(t)x_i} \end{aligned} \quad (\text{A.5})$$

8. Repita

Habilita qualquer nó desabilitado, vai para 2.

A Figura A.1 mostra um diagrama da estrutura da rede ART1. São ilustradas apenas algumas das conexões entre os nós das camadas F_1 e F_2 . As conexões *bottom – up* e *top – down* entre estes nós são do tipo “todos para todos”.

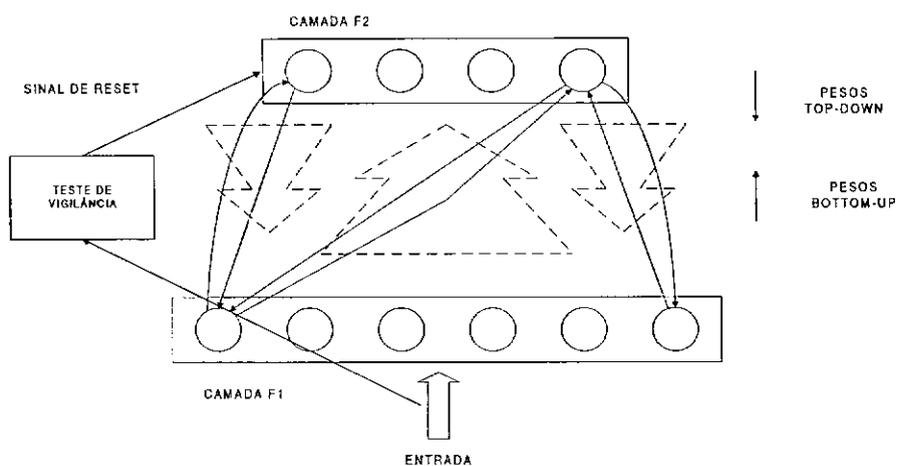


Fig. A.1 - Exemplo de arquitetura de uma rede ART1 com 6 nós de entrada e 4 nós de saída.

FONTE: Adaptada a partir de ([51] p.168).

APÊNDICE B

ART1 INVARIANTE

Srinivasa e Jouaneh [45] acrescentaram à arquitetura da rede ART1 uma nova camada (F_3), formada por um conjunto de agrupamentos de neurônios com base na rede invariante de Widrow et al. [42]. Para ilustrar o funcionamento da rede, considere-se uma imagem binária de 4 x 4 *pixels* como entrada para a camada F_1 da rede ART1. O algoritmo da nova rede segue os seguintes passos:

Passo 1. A imagem binária B de 4 x 4 pontos é convertida para um vetor de entrada I com 16 posições e submetida à camada F_1 da rede ART1.

$$I_i = B_{km} \quad i = 1 \text{ até } 16, \quad (\text{B.1})$$

onde B_{km} são os elementos da imagem B e

$$k = n \left[\frac{i-1}{4} \right] + 1 \quad \text{e } m = i - 4(k-1). \quad (\text{B.2})$$

O operador $n[x]$ retorna a parte inteira de x . Esta entrada é codificada em um nó J na camada F_2 pelo algoritmo da rede ART1 mostrado no apêndice A.

Passo 2. O primeiro agrupamento da camada F_3 é ativado. Este agrupamento contém 16 células correspondendo ao número de *pixels* do vetor I . Cada célula contém 16 unidades conectadas ao J -ésimo nó na camada F_2 . Seja W_1 a matriz de pesos para a célula superior esquerda do agrupamento 1. Os elementos de W_1 são dados por:

$$w_{km} = t_{Ji} \quad i = 1 \text{ até } 16, \quad (\text{B.3})$$

onde o vetor t_J , de 16 componentes é o vetor de pesos *Top-Down*, e w e k são definidos conforme a expressão B.2.

As matrizes de pesos das demais células do agrupamento 1 são construídas a partir de W_1 como se segue:

$$\begin{bmatrix} W_1 & T_{r1}(W_1) & T_{r2}(W_1) & \dots & T_{rm}(W_1) \\ T_{d1}(W_1) & T_{r1}T_{d1}(W_1) & T_{r2}T_{d1}(W_1) & \dots & T_{rm}T_{d1}(W_1) \\ \dots & \dots & \dots & \dots & \dots \\ T_{dk}(W_1) & T_{r1}T_{dk}(W_1) & T_{r2}T_{dk}(W_1) & \dots & T_{rm}T_{dk}(W_1) \end{bmatrix} \quad (\text{B.4})$$

onde o operador T_{di} representa a translação de cada linha da matriz W_1 por i *pixels* para baixo e T_{ri} representa a translação de cada coluna de W_1 por i *pixels* para a direita.

Passo 3. Agrupamentos adicionais são criados para se atingir invariância a diferentes ângulos de rotação. Cada novo agrupamento possui uma cópia da matriz peso W_1 rotacionada por um determinado ângulo. Assim, a matriz W_2 de pesos do canto superior esquerdo do segundo agrupamento será $W_2 = R_\theta(W_1)$, onde R_θ é a rotação de W_1 por θ^0 . Os pesos dos demais neurônios do agrupamento 2 são obtidos substituindo-se W_1 por W_2 em (B.4). O mesmo é feito para os demais agrupamentos, rotacionando-se W_1 por ângulos diferentes.

Passo 4. O conjunto de matrizes de pesos criados é colocado em lugar do nó J na camada F_2 . Os pesos *Top-Down* de um novo nó L são obtidos a partir da matriz de pesos de seu neurônio correspondente na camada F_3 . Por exemplo, para obter os pesos *Top-Down* de um novo nó L de F_2 correspondente à célula superior esquerda do agrupamento 2, tem-se a expressão

$$t_{Li}^{novo} = w_{km}, \quad (\text{B.5})$$

onde w_{km} são os elementos de W_2 e k e m são definidos como na expressão B.2.

Os pesos *Bottom-Up* para o nó L são obtidos por

$$b_{Li}^{novo} = \frac{I_i}{0.5 + \sum_{i=1}^{16} I_i}, \quad (\text{B.6})$$

onde os elementos de I_i são obtidos como se segue:

$$I_i = R_\theta(B) = B_{km} \quad i = 1 \text{ até } 16, \quad (\text{B.7})$$

onde B_{km} são elementos da matriz $R_\theta(B)$, e k e m são definidos como na expressão B.2.

Passo 5. Todos os elementos da camada F_3 são reinicializados com zero e o programa volta ao passo 1 para processar a próxima entrada.

Desta forma, a matriz de pesos para o neurônio superior esquerdo de cada agrupamento de F_3 é baseada nos pesos que codificaram a primeira aparição de uma entrada na camada F_2 . Os agrupamentos da camada F_3 produzem um agrupamento (*cluster*) de nós na camada F_2 com invariância a rotação e translação.

Um diagrama com um exemplo de uma rede invariante a rotações de 90° para entradas com 3×3 *pixels* é mostrado na Figura B.1.

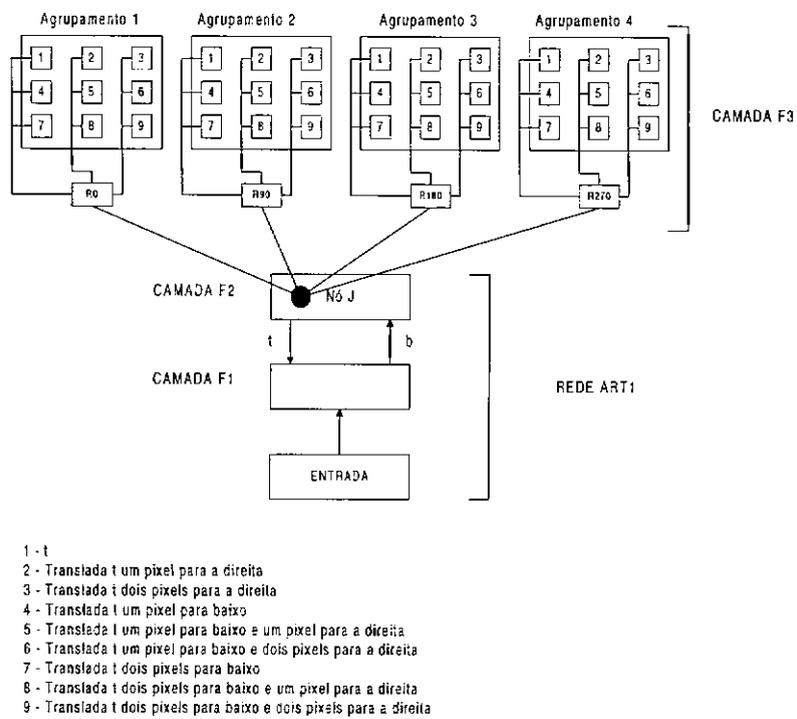


Fig. B.1 - Padrão para conexões de pesos entre o nó da camada F_2 que codificou a entrada binária e as células da camada F_3 em uma rede invariante a todas as translações e rotações por ângulos múltiplos de 90° .

FONTE: Alterada a partir de ([45], p. 1435).