

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

“Aplicação da Rede Neural Neocognitron  
para Reconhecimento de Atributos Faciais”

**ORIENTADOR:** Prof. Dr. José Hiroki Saito  
**ALUNO:** Marcelo Hiroshi Hirakuri

**São Carlos**  
**Agosto/2003**

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

H668ar

Hirakuri, Marcelo Hiroshi.  
Aplicação da rede neural neocognitron para  
reconhecimento de atributos faciais / Marcelo Hiroshi  
Hirakuri. -- São Carlos : UFSCar, 2003.  
100 p.

Dissertação (Mestrado) -- Universidade Federal de São  
Carlos, 2003.

1. Reconhecimento de padrões. 2. Redes neurais. 3.  
Rede neural neocognitron. I. Título.

CDD: 006.4 (20<sup>a</sup>)

# A GRADECIMENTOS

---

Aos meus pais pela compreensão.

À Clarice Sanches Mariante da Silva pelo amor e carinho.

Ao prof. Dr. José Hiroki Saito, não só pela dedicada orientação, mas pela amizade, paciência e compreensão que esteve sempre presente durante todo o trabalho.

Aos meu amigos que me deram maior apoio: Márcio Toshikazu Dói, Roberto Susa Mihara, Rafael De Lucca Antonio, Vanderson Maldonado Mendes, Claudia Brunetto, Alessandro Rodrigues e Paulo Henrique de Almeida Andrade.

Às pessoas que serviram de modelo para esse projeto.

Aos professores e funcionários do programa de pós-graduação.

# SUMÁRIO

---

1 - Introdução .....	15
1.1. Organização do trabalho.....	15
2 -Visão Cibernética.....	17
2.1.Introdução.....	17
2.2.Visão Natural.....	17
2.2.1.Estrutura do Olho Humano.....	18
2.2.1.1.Sistema Óptico.....	18
2.2.1.2.Sistema Neural.....	20
2.2.2.Córtex Visual Primário.....	22
2.3.Visão Artificial.....	23
2.3.1.Captação de Imagens e Pré-Processamento .....	23
2.3.1.1.Modelo de Imagem.....	24
2.3.2.Processamento Básico .....	25
2.3.3.Processamento Intermediário .....	25
2.3.4.Processamento Avançado.....	25
3 - Reconhecimento Facial .....	27
3.1.Detecção de Faces .....	28
3.1.1.Métodos Top-Down Baseados em Conhecimento .....	28
3.1.2.Métodos Bottom-Up Baseados em Atributos.....	30
3.1.2.1.Atributos Faciais.....	30
3.1.2.2.Textura.....	31
3.1.2.3.Cor de Pele .....	31
3.1.2.4.Múltiplos Atributos .....	31
3.1.3.Template Matching.....	32
3.1.4.Métodos Baseados na Aparência (Similaridade).....	32

3.2.Reconhecimento de Faces .....	32
3.2.1.Redes Neurais .....	33
4 - Redes Neurais.....	34
4.1.Definição .....	34
4.2.Redes Neurais Biológicas.....	34
4.3.Redes Neurais Artificiais.....	35
4.3.1.Arquitetura de Uma Rede Neural Artificial .....	36
4.3.1.1.Redes Neurais Artificiais de Camada Única .....	36
4.3.1.2.Redes Neurais Artificiais Multicamada.....	37
4.3.2.Tipos de Treinamento.....	38
4.4.Exemplos de Redes Neurais .....	39
4.4.1.Perceptron.....	39
4.4.2.Perceptron Multicamadas (Multilayer Perceptron – MLP) .....	41
4.4.3.Rede Neural Recorrente .....	42
4.4.4.Rede Neural Convolutacional .....	44
5 - Neocognitron.....	46
5.1.Introdução.....	46
5.2.Estrutura Hierárquica do Neocognitron.....	46
5.3.A Citologia do Neocognitron .....	49
5.4.Treinamento Não Supervisionado .....	52
5.4.1.Escolha das Células “Vencedoras”.....	52
5.4.2.Reforço das Conexões Intercelulares Variáveis .....	53
5.4.3.Progresso Temporal do Neocognitron.....	54
5.5.Treinamento Supervisionado.....	58
5.5.1.Deteção de Bordas e Linhas .....	58
5.5.2.Deteção de Curvatura e Pontos Finais .....	59
5.6.Reconhecimento de Numerais Escritos a Mão.....	61
5.6.1.Estrutura da Rede .....	61
5.6.2.Treinamento da Rede.....	62
5.6.2.1.Treinamento do nível $U_{S1}$ .....	62
5.6.2.2.Treinamento do nível $U_{S2}$ .....	63

5.6.2.3. Treinamento do nível $U_{S3}$ .....	65
5.6.2.4. Treinamento do nível $U_{S4}$ .....	66
6 - Neocognitron Para o Reconhecimento de Atributos Faciais (NEORAF) .....	68
6.1. Banco de Imagens .....	70
6.2. Extração de Atributos Faciais .....	71
6.3. Estrutura do NEORAF .....	72
6.3.1. Rede Neural para Detecção de Pontos de Controle .....	72
6.3.2. Rede Neural para Reconhecimento de Atributos Faciais .....	77
6.3.2.1. Treinamento não Supervisionado da RNRAF .....	79
6.3.2.2. Plano de Seleção de Semente .....	79
6.3.2.3. Citologia da RNRAF .....	80
6.4. Testes de Desempenho .....	82
6.4.1. Testes com Três Classes .....	83
6.4.2. Testes com Quatro Classes .....	85
6.4.3. Testes com Cinco Classes .....	87
6.4.4. Testes com Seis Classes .....	89
6.4.5. Testes com Sete Classes .....	91
6.5. Esboço Gráfico .....	93
7- Análise de Resultados e Perspectivas Futuras .....	95
Referência bibliográfica .....	97

# I

## NDICE DE FIGURAS

---

Figura 2.1: Olho Humano.....	19
Figura 2.2: Distribuição de bastonetes e cones na retina (adaptado de [Gra95]).....	22
Figura 2.3: Representação de uma imagem em matriz de pixels .....	25
Figura 3.1: Imagem original e imagens de baixa resolução correspondentes: (a) .... n=1, imagem original. (b) n=4. (c) n=8. (d) n=16 .....	29
Figura 3.2: Face típica utilizada em métodos de reconhecimento top-down .....	
baseado em conhecimento [Yan94].....	30
Figura 4.1: Neurônio Biológico.....	35
Figura 4.2: Rede Neural de Camada Única .....	37
Figura 4.3: Rede Neural Multicamada .....	38
Figura 4.4: Perceptron Monocamada.....	39
Figura 4.5: Rede Neural <i>MLP</i> , onde o algoritmo para reajuste dos pesos é..... aplicado em cada camada de forma independente.....	42
Figura 4.6: Rede Recorrente sem unidades ocultas.....	44
Figura 4.7: Exemplo de rede convolucional.....	45
Figura 5.1: Estrutura simplificada da seqüência de estágios do Neocognitron.....	47
Figura 5.2: Distribuição das conexões intercelulares de entrada do Neocognitron...	48
Figura 5.3: Estrutura entre os estágios e camadas de um Neocognitron de três..... estágios .....	49
Figura 5.4: Coluna de células S.....	53
Figura 5.5: Exemplo de treinamento para padrões de linha [Fuk82] .....	55
Figura 5.6: Conexões intercelulares convergindo para células S em uma rede .....	
Neocognitron[Fuk82] .....	56
Figura 5.7: Processo de extração de linhas e bordas .....	59
Figura 5.8: Detecção de pontos extremos de linhas .....	60
Figura 5.9: Estrutura da rede para reconhecimento de numerais .....	61

Figura 5.10: Treinamento utilizado para os 12 planos de célula na camada $U_{S1}$ [Fuk92].....	63
Figura 5.11: Padrões de treinamento usados para treinar os 38 planos de célula da camada $U_{S2}$ [Fuk92].....	64
Figura 5.12: padrões usados para treinar os 35 planos de célula da camada $U_{S3}$ . O campo receptivo da célula está sendo marcado por linhas cruzadas [Fuk92]. .....	66
Figura 5.13: padrões de treinamento usados para treinar os 11 planos celulares da camada $U_{S4}$ [Fuk92] .....	67
Figura 6.1: Etapas da rede neural NEOPC. (a) Pré-Processamento, (b) RNDPC: obtenção dos pontos de controle, (c) RNRF: treinamento e (d) RNRF: reconhecimento facial.....	69
Figura 6.2: Exemplo de imagem do banco principal de imagens.....	71
Figura 6.3: Exemplos de imagens de treinamento e reconhecimento .....	72
Figura 6.4: padrões de bordas a serem detectados pela RNDB.....	74
Figura 6.5: padrões de linhas a serem detectados pela RNDB.....	75
Figura 6.6: padrões de pontos extremos a serem detectados pela RNDPEL.....	75
Figura 6.7: Esquema completo da RNDPC .....	76
Figura 6.8: Execução da RNDPC .....	77
Figura 6.9: Estrutura do RNRAF.....	78
Figura 6.10: Execução da RNRAF .....	80

# ÍNDICE DE TABELAS

---

Tabela 6.1: Dados da RNDPC.....	73
Tabela 6.2: Tabela 6.2 – Dados da RNRAF.....	79
Tabela 6.3: Tabela de limiares para o nariz.....	82
Tabela 6.4: Tabela de limiares para os lábios.....	83
Tabela 6.5: Tabela de limiares para o olho direito.....	83
Tabela 6.6: Taxa de reconhecimento nasal com 3 classes de testes.....	84
Tabela 6.7: Tabela de confusão nasal com 3 classes de testes.....	84
Tabela 6.8: Taxa de reconhecimento labial com 3 classes de testes.....	84
Tabela 6.9: Tabela de confusão labial com 3 classes de testes.....	85
Tabela 6.10: Taxa de reconhecimento de olho direito com 3 classes de testes.....	85
Tabela 6.11: Tabela de confusão de olho direito com 3 classes de testes.....	85
Tabela 6.12: Taxa de reconhecimento nasal com 4 classes de testes.....	86
Tabela 6.13: Tabela de confusão nasal com 4 classes de testes.....	86
Tabela 6.14: Taxa de reconhecimento labial com 4 classes de testes.....	86
Tabela 6.15: Tabela de confusão labial com 4 classes de testes.....	86
Tabela 6.16: Taxa de reconhecimento de olho direito com 4 classes de testes.....	87
Tabela 6.17: Tabela de confusão de olho direito com 4 classes de testes.....	87
Tabela 6.18: Taxa de reconhecimento nasal com 5 classes de testes.....	87
Tabela 6.19: Tabela de confusão nasal com 5 classes de testes.....	88
Tabela 6.20: Taxa de reconhecimento labial com 5 classes de testes.....	88
Tabela 6.21: Tabela de confusão labial com 5 classes de testes.....	88
Tabela 6.22: Taxa de reconhecimento de olho direito com 5 classes de testes.....	89
Tabela 6.23: Tabela de confusão de olho direito com 5 classes de testes.....	89
Tabela 6.24: Taxa de reconhecimento nasal com 6 classes de testes.....	89
Tabela 6.25: Tabela de confusão nasal com 6 classes de testes.....	90
Tabela 6.26: Taxa de reconhecimento labial com 6 classes de testes.....	90

Tabela 6.27: Tabela de confusão labial com 6 classes de testes.....	90
Tabela 6.28: Taxa de reconhecimento de olho direito com 6 classes de testes.....	91
Tabela 6.29: Tabela de confusão de olho direito com 6 classes de testes .....	91
Tabela 6.30: Taxa de reconhecimento nasal com 7 classes de testes .....	91
Tabela 6.31: Tabela de confusão nasal com 7 classes de testes .....	92
Tabela 6.32: Taxa de reconhecimento de olho direito com 7 classes de testes.....	92
Tabela 6.33: Tabela de confusão de olho direito com 7 classes de testes .....	92
Tabela 6.34: Número de classes utilizadas versus número de imagens de..... treinamento e imagens de reconhecimento.....	94

# ÍNDICE DE GRÁFICOS

---

Tabela 6.1: Taxa de acerto de reconhecimento labial .....	93
Tabela 6.2: Taxa de acerto de reconhecimento nasal e de olho direito.....	93

# R ESUMO

---

Este trabalho é relativo à metodologia de reconhecimento de atributos faciais pré-estabelecidos: olho direito, olho esquerdo, nariz e lábio, utilizando uma implementação alternativa da Rede Neural Neocognitron, robusta a distorções e deslocamentos no padrão de entrada. A implementação da rede neural (NEORAF- Neocognitron para o Reconhecimento de Atributos Faciais), consiste de duas sub-redes: Rede Neural para Detecção de Pontos de Controle (RNDPC), que detecta bordas, linhas e extremidades de linhas nos padrões de entrada, durante o treinamento; e Rede Neural para Reconhecimento Facial (RNRF), que é a rede propriamente usada para o reconhecimento, constituído de camadas para manipulação de fatores complexos. A avaliação de desempenho mostrou a viabilidade do sistema para diversas aplicações com excelentes perspectivas futuras. Os testes de desempenho apresentaram taxas de acerto acima de 85%.

# A BSTRACT

---

This work presents a methodology for predetermined facial attributes recognition pre-established. This work realizes the recognition of facial attributes previously established: right eye, left eye, nose and lip, using an alternative implementation of Neocognitron Neural Network, robust for distortions and shifts on input patterns. The implementation of the neural network (NEOFAR - Neocognitron for Facial Attributes Recognition), consists of two sub-networks: Neural Network for Detecting of Control Points (NNDCP), which is used for edge detection, line detection, and line end detection, during the training phase; and Neural Network for Facial Recognition (NNFR), which is used for complex features manipulation, during the recognition phase. The performance evaluation showed the viability of project for many applications with excellent future perspectives. The performance tests showed rightness taxes over 85%.

# 1 INTRODUÇÃO

O interesse em métodos de processamento de imagens digitais tem raízes em duas principais áreas de aplicação: melhoria de informação visual para interpretação humana e o processamento de dados de cenas para percepção automática através de máquinas. Uma das primeiras aplicações de técnicas de processamento de imagens da primeira categoria foi o melhoramento de imagens digitalizadas para jornais, enviadas por meio de cabo submarino de *Londres* para *New York* [Gon92].

Atualmente, novas técnicas de reconhecimento por meio de faces, íris, retina, voz e impressão digital, dentre outras, têm sido abordadas para aplicações em sistemas de reconhecimento automático [Bo100].

Este trabalho é multidisciplinar, relacionando as disciplinas de processamento digital de imagens, redes neurais, reconhecimento de padrões, paradigmas de programação dentre outras.

As redes neurais são apropriadas à detecção e reconhecimento facial. Assim, este projeto está relacionado com a pesquisa de uma arquitetura de rede neural para realização de reconhecimento de atributos faciais. Diversas linhas de pesquisa relacionadas ao reconhecimento facial, tais como detecção, representação e processo, têm despertado grande interesse em decorrência da vasta possibilidade existente de aplicações: sistemas de controle automático de acesso, interfaces homem-máquina, sistemas de vigilância [Fre92], sistemas de cartões inteligentes, reconhecimento de indivíduos de uma base de dados [Phi02], dentre outras.

## 1.1 Organização do Trabalho

A presente dissertação é iniciada por uma introdução, que fornece um breve histórico sobre o processo de reconhecimento de padrões.

O Capítulo 2 realiza o estudo da visão cibernética, com um estudo introdutório sobre visão natural, focado no olho humano estudando as principais estruturas envolvidas no sistema óptico e no sistema neural. Em seguida, é exposta uma breve abordagem sobre visão artificial, que contém os seguintes tópicos: captação de imagens e pré-processamento, processamento básico, processamento intermediário e processamento avançado.

O Capítulo 3 lista os desafios associados com a detecção facial e mostra métodos para a detecção de imagens faciais: métodos *top-down* baseado em conhecimento, métodos *bottom-up* baseado em atributos, *template matching* e métodos baseados na aparência. O Capítulo 3 enfoca também as redes neurais como técnica para reconhecimento facial.

O Capítulo 4 discute as redes neurais, fornecendo definições sobre rede neural biológica e rede neural artificial. São apresentadas as arquiteturas de redes neurais de camada única e multicamada. Em seguida são mostrados alguns exemplos importantes de redes neurais.

O Capítulo 5 expõe a rede neural *Neocognitron*, que é a base para esse projeto de pesquisa. É apresentada a sua estrutura hierárquica, citologia, tipos de treinamento e progresso temporal.

O Capítulo 6 mostra o trabalho de dissertação de mestrado, que consiste em uma adaptação da rede neural *Neocognitron* para reconhecimento de atributos faciais. A rede neural utilizada no projeto é denominada *Neocognitron* para Reconhecimento de Atributos Faciais (*NEORAF*). São abordadas: a base de dados de imagens para treinamento e reconhecimento e a estrutura hierárquica *NEORAF*, com suas sub-redes. Por fim, é feita uma análise dos resultados obtidos da etapa de reconhecimento.

## **2 VISÃO CIBERNÉTICA**

Neste capítulo serão descritos os aspectos de visão cibernética, que está na interface entre visão natural com o seu modelo artificial. Será realizado um estudo do olho humano e algumas de suas funções no processo visual. Serão vistos também tópicos sobre o processamento artificial da visão.

### **2.1 Introdução**

O processo visual desempenhado por homens e animais é o resultado da coordenação do processamento de uma imensa quantidade de dados multidimensionais (posição, formas, cores e outros fatores) utilizando uma enorme quantidade de estruturas biológicas que interagem de uma forma incrivelmente intrincada ainda pouco conhecida.

Embora tenham sido realizados grandes esforços na pesquisa para a construção de sistemas de visão artificial, foram projetados apenas sistemas muito simplificados de visão, geralmente limitados ao tratamento de problemas muito específicos, como, por exemplo, contagem de parafusos em linhas de fabricação.

### **2.2 Visão Natural**

O estudo de visão natural consiste no estudo biológico da visão e possui forte relacionamento com o estudo do cérebro, sendo o córtex visual uma das regiões mais estudadas do cérebro dos primatas. Uma alternativa para o avanço do estudo de visão natural é o princípio da fertilização cruzada, onde os resultados obtidos no estudo da visão natural são empregados para obtenção de novos algoritmos e modelos computacionais e, por outro lado, os resultados de implementação de modelos computacionais podem gerar dados para investigação em visão natural [Cos94]. Assim, as áreas de estudo para a visão, de forma geral, podem ser relacionadas como segue.

Anatomia: objetiva identificar a morfologia das células nervosas, assim como a rede de conexão entre estas células.

Neurofisiologia: consiste em identificar as propriedades funcionais de subsistemas de processamento neural. Por exemplo, em eletro-fisiologia, um micro-eletrodo é posto perto do neurônio a ser analisado. Faz-se a monitoração da resposta elétrica do neurônio em decorrência de diversos padrões de estímulos. Com isso pode-se tentar estabelecer a função do circuito neural do qual tal neurônio faz parte.

Psicofísica: está relacionada com a compreensão do comportamento do sistema visual de um determinado animal a partir de suas manifestações comportamentais.

Engenharia: consiste na implementação de visão artificial utilizando os conhecimentos adquiridos através do estudo da visão natural para a solução de problemas de engenharia.

## **2.2.1 Estrutura do Olho Humano**

O olho é a estrutura biológica encontrada pela luz proveniente da cena que está sendo analisada, consistindo em uma esfera com cerca de 20 mm envolvida por três membranas (a córnea e a cobertura externa da esclerótica, a coróide e a retina) [Gon92].

### **2.2.1.1 Sistema Óptico**

O sistema óptico é responsável pela recepção dos estímulos luminosos para subsequente formação da imagem no interior do olho. A Figura 2.1 ilustra o corte de um olho humano.

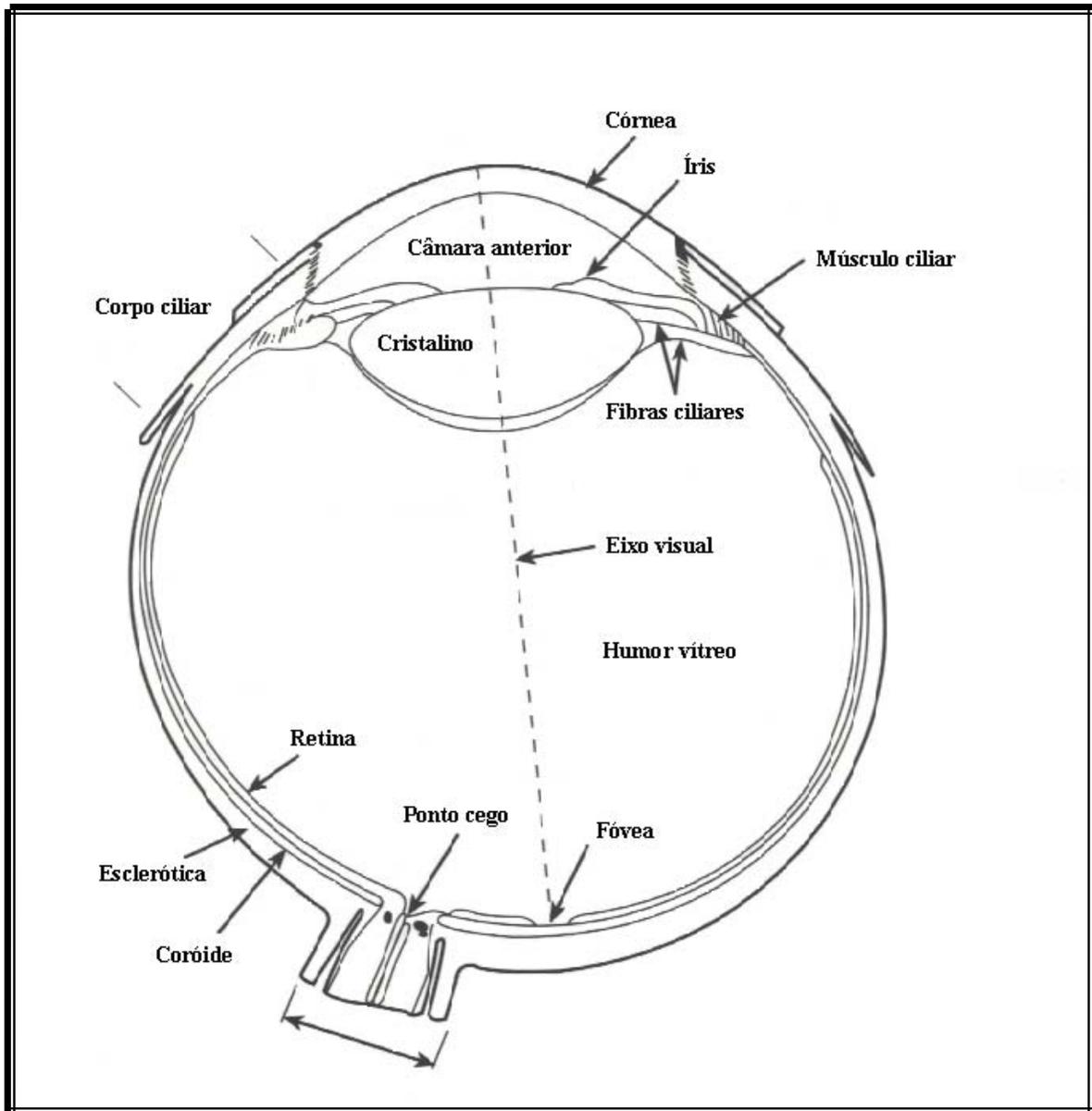


Figura 2.1 – Olho humano.

Dentre os principais elementos do olho humano com suas funções básicas, podem ser citados: córnea, esclerótica, coróide, pupila e cristalino.

A córnea é um tecido resistente e transparente que cobre a superfície anterior do olho. Contígua à córnea, a esclerótica é uma membrana opaca que envolve o restante do globo ocular.

A coróide é a membrana situada diretamente abaixo da esclerótica, possuindo uma rede de vasos sanguíneos que servem como principal fonte de nutrição do olho. Sua cobertura é fortemente pigmentada, o que faz com que a luz não se espalhe pelo globo ocular. Em sua extremidade anterior, a coróide se divide em corpo ciliar e íris. A pupila é a abertura central da íris, com aproximadamente de  $2\text{ mm}$  a  $8\text{ mm}$  de diâmetro, possuindo a função de otimizar a profundidade de campo de visão em função da luminosidade do ambiente.

O cristalino é formado por camadas concêntricas fibrosas e sustentado por fibras (tais como a *Zâmula de Zimm*) que se conectam ao corpo ciliar, sendo colorido por uma pigmentação levemente amarelada, a qual se intensifica com a idade.

### **2.2.1.2 Sistema Neural**

O sistema neural do olho é responsável pela tradução da imagem focalizada pelo cristalino em impulsos eletroquímicos enviados ao cérebro para futuros processamentos [Cos94]. Dentre seus principais elementos envolvidos com o processamento neural, podem ser citadas: a retina, fóvea, ponto cego e fotorreceptores.

A membrana mais interna do olho é a retina, que se estende por toda porção posterior do olho. Quando o olho está apropriadamente focalizado, a luz de um objeto exterior se torna imagem na retina. O processamento neural começa quando os fotorreceptores existentes na retina traduzem o sinal luminoso em estímulos eletroquímicos. Existem duas classes de receptores de luz sobre a superfície da retina: cones e bastonetes.

A retina possui duas regiões especiais: o ponto cego e a fóvea. O ponto cego é o local onde o nervo óptico deixa a retina (observação: o ponto cego não possui fotorreceptores) e a fóvea é uma depressão circular com aproximadamente  $1,5\text{ mm}$  de diâmetro, que se caracteriza possuir um grande número fotorreceptores, principalmente cones.

Os bastonetes são especializados em visão noturna, caracterizando-se por uma baixa resolução espacial e visão monocromática. A sua ampla área de distribuição unida ao fato de estarem conectados a uma única fibra nervosa, reduz a quantidade de detalhes discerníveis por estes receptores de luminosidade. Os bastonetes são, porém, sensíveis a baixos níveis de luminosidade, e sua quantidade é aproximadamente de 75 a 150 milhões [Gon92].

Os cones são especializados em visão diurna, caracterizando-se por uma alta resolução espacial e visão colorida. Eles estão posicionados principalmente na região da fóvea. Sua pequena área de distribuição e o fato de cada cone estar conectado à sua própria fibra lhes dão maior capacidade de discernir pequenos detalhes. Portanto, eles são sensíveis a cores e altos níveis de luminosidade, e sua quantidade total é aproximadamente de 6 a 7 milhões [Gon92].

A densidade de cones se dá principalmente na região da fóvea com aproximadamente 150.000 elementos por milímetro quadrado. A Figura 2.2 mostra a densidade de bastonetes e cones para uma seção do olho direito, passando através da região de emergência do nervo óptico [Gra95].

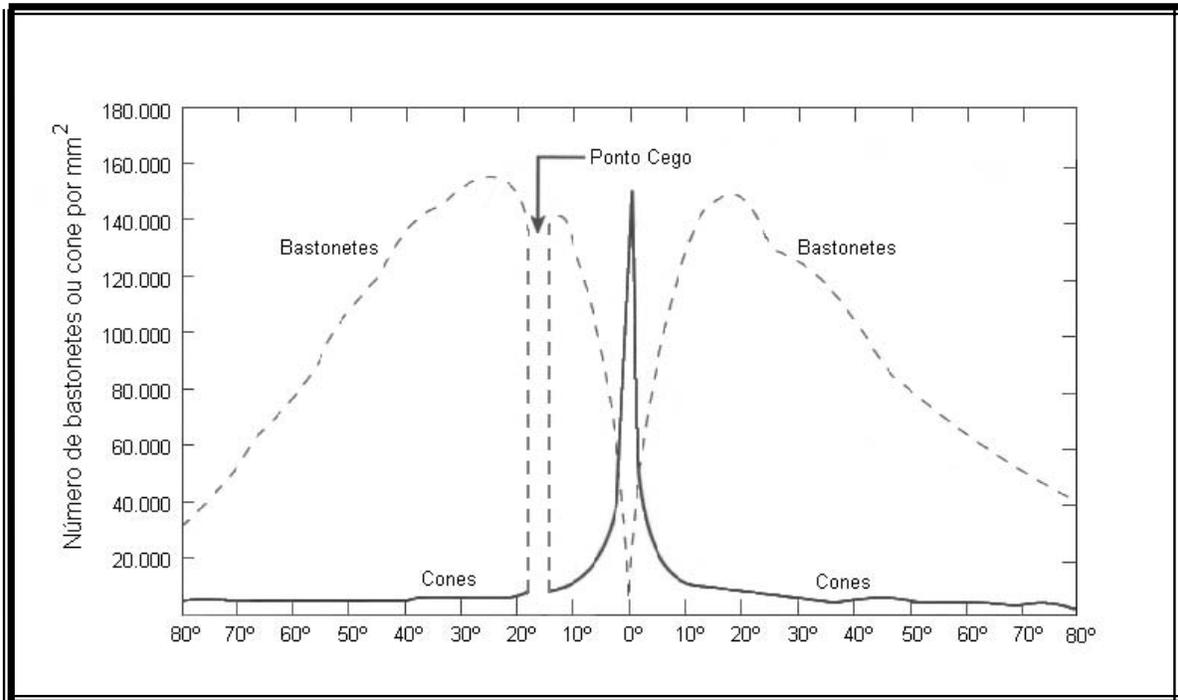


Figura 2.2 – Distribuição de bastonetes e cones na retina (adaptado de [Gra95]).

O nervo óptico consiste dos axônios das células ganglionares, que constituem a última camada neural da retina. Ao deixar o olho, o nervo óptico passa por um conjunto intrincado de estruturas chegando até o núcleo geniculado lateral, ou *LGN* (lateral geniculate nucleus). Porém, este assunto exige maior profundidade e escapa ao escopo desse material [Len01] [Gaz02].

Embora ainda existam muitas discussões a respeito, a principal função da retina é compreendida como a detecção e realce das bordas dos objetos em imagens [Cos94].

### 2.2.2 Córtex Visual Primário

O córtex visual primário, também conhecido como córtex estriado é comumente denotado *V1* ou área 17. O córtex visual primário é a única região do cérebro que recebe sinais aferentes (sinais provindos dos órgãos sensoriais para o cérebro) trazendo informações visuais da retina.

Anatomicamente, o córtex visual primário dos primatas organiza-se em seis camadas, normalmente numeradas de I até VI. Os primeiros resultados importantes sobre o córtex visual primário foram produzidos por *Hubel* e *Wiesel*, durante a década de 60 [Hub67].

Posteriormente ao processamento de informações no córtex visual primário, os resultados obtidos são transferidos para outras regiões visuais de cérebro denotadas V2, V3, V4 e V5. Isso ocorre até que todos os processamentos sejam realizados, ocorrendo, por exemplo, o reconhecimento dentro de uma cena.

## **2.3 Visão Artificial**

Os desenvolvimentos em visão artificial ou computacional têm sido em grande parte empíricos, não havendo um embasamento teórico similar ao existente em processamento digital de sinais. Com isso, existem diversas tendências diferentes sobre o assunto. Serão apresentadas algumas técnicas de visão artificial que tomam como base os processos desempenhados respectivamente pela retina (captação de imagens, pré-processamento, e processamento básico), córtex visual primário (processamento intermediário) e outras regiões visuais de maior hierarquia (processamento avançado).

### **2.3.1 Captação de Imagens e Pré-Processamento**

O processo de captação de imagens pode ocorrer de diversas formas, dependendo do equipamento que é utilizado, tal como: *scanner*, câmera de vídeo, densitômetro, aparelho de raios-x, dentre outros. O princípio comum de operação destes equipamentos consiste em traduzir padrões bidimensionais de grandezas físicas, como a luminosidade, em um conjunto discreto de medidas que é normalmente organizado em forma matricial. Cada elemento da matriz é denominado *pixel* (“*picture element*”). Cada *pixel* é identificado por um par de coordenadas cartesianas  $(x, y)$ .

### 2.3.1.1 Modelo de Imagem

O termo imagem refere-se a uma função de intensidade bidimensional, denotada por  $f(x, y)$ . O valor da amplitude da função  $f$  nas coordenadas espaciais  $(x, y)$  dá a intensidade luminosa da imagem naquele ponto [Gon92]. Como a luz é uma fonte de energia positiva e finita, temos:

$$0 < f(x, y) < \infty, \quad (2.1).$$

A natureza básica de  $f(x, y)$  pode ser caracterizada por dois componentes: a quantidade de luz incidindo na cena sendo observada e a quantidade de luz refletida pelos objetos da cena. Esses componentes são chamados iluminação e refletância, respectivamente, e são representados por  $i(x, y)$  e  $r(x, y)$ . O produto das funções  $i(x, y)$  e  $r(x, y)$  resulta em  $f(x, y)$ :

$$f(x, y) = i(x, y) \cdot r(x, y), \quad (2.2).$$

$$0 < i(x, y) < \infty, \quad (2.3).$$

$$0 < r(x, y) < 1, \quad (2.4).$$

Das equações acima temos que a refletância é limitada entre 0 (absorção total) e 1 (refletância total). A natureza de  $i(x, y)$  é determinada pela fonte de luz e a de  $r(x, y)$  é determinada pelas características dos objetos da cena. Na Figura 2.3 é mostrada a representação de uma imagem em matriz de *pixel* que mapeia  $f(x, y)$  em coordenadas bidimensionais.

$$f(x,y) = \begin{bmatrix} f(0,0) & f(0,1) & \dots & f(0,M-1) \\ f(1,0) & f(1,1) & \dots & f(1,M-1) \\ \vdots & \vdots & \vdots & \vdots \\ f(N-1,0) & f(N-1,1) & \dots & f(N-1,M-1) \end{bmatrix}$$

Figura 2.3 – Representação de uma imagem em matriz de pixels.

Temos três tipos principais de imagens: imagens binárias (cada *pixel* é apenas branco ou preto), imagens em níveis de cinza (cada *pixel* pode assumir um valor em uma escala de diferentes níveis de luminosidade) e imagens coloridas.

### 2.3.2 Processamento Básico

Os algoritmos de processamento básico visam obter a primeira representação da imagem captada, em analogia ao que acontece na retina dos primatas. Uma das principais representações da imagem é em termos de elementos de borda, uma vez que é possível recuperar uma imagem a partir de seus elementos de borda [Cos94].

### 2.3.3 Processamento Intermediário

Geralmente, o processamento intermediário de visão procura realizar a segmentação da imagem em objetos menores e derivar representações de maior abstração da imagem original. Duas formas de segmentação têm sido bastante utilizadas: por regiões e por contornos. Na segmentação por contorno, a representação da imagem se dá em termos de segmentos de reta. Isso se deve ao fato de que uma curva possa ser representada por segmentos digitais de retas [Cos94].

### 2.3.4 Processamento Avançado

O processamento avançado corresponde à combinação de elementos representacionais obtidos nos estágios anteriores em elementos de complexidade crescente até que se atinja a descrição da imagem, analogamente ao que deve acontecer nas regiões

de hierarquia mais elevada do córtex dos primatas. O processamento avançado envolve a manutenção de bases de dados sofisticadas para serem utilizadas em técnicas de reconhecimentos de padrões [Cos94].

### 3 RECONHECIMENTO FACIAL

Neste capítulo serão descritos os seguintes métodos de detecção facial: *top-down* baseados em conhecimento, *bottom-up* baseados em atributos, *template matching* e baseados na similaridade. Neste contexto serão descritas as redes neurais, que podem representar para o campo de reconhecimento facial, uma das possíveis soluções..

No contexto de identificação pessoal, o reconhecimento facial, geralmente se refere ao reconhecimento frontal estático e controlado. Por estático, entende-se que os retratos faciais usados para os sistemas de reconhecimento facial são imagens estáticas. Por controlado, tem-se que características, tais como iluminação e resolução dos dispositivos de aquisição são, geralmente, fixas durante o processo de aquisição de imagem [Hon98].

De modo geral, o reconhecimento facial inicia com a detecção de faces em uma imagem, seguida da normalização da imagem facial somado a mudanças geométricas e de iluminação e, por fim, é identificada a face utilizando uma representação de imagem apropriada, juntamente com um algoritmo de classificação [Sri02]. A localização de face continua a ser um problema quando se utilizam imagens agrupadas ou não controladas. Fortuitamente, no contexto de identificação pessoal, o fundo de plano é controlado ou quase controlado, de modo que a localização de face, geralmente, não é considerada como um grande problema para um sistema biométrico utilizando faces.

Algoritmos que executam ambas tarefas são conhecidos como “algoritmos completamente automáticos” (*fully automatic algorithms*) e aqueles que realizam somente o reconhecimento ou verificação de face são chamados “algoritmos parcialmente automáticos” (*partially automatic algorithms*) [Phi00].

Existem diversas transformações que podem ser aplicadas a uma imagem para sua melhoria antes do processo de reconhecimento facial, tais como translação, orientação, escala e iluminação. Estas variações podem ser restringidas a certos limites, sendo que, para cada tipo de restrição pode existir uma técnica diferente de reconhecimento facial.

### 3.1 Detecção de Faces

O primeiro passo em um sistema de processamento facial é a detecção dos locais onde as faces estão presentes. Segundo [Yan02] a detecção de face pode assim ser definida: para uma dada uma imagem arbitrária, a detecção de faces consiste em determinar se existe ou não qualquer face na imagem e, se existir, retornar o local de cada face presente na imagem e seu conteúdo. Os maiores desafios associados com a detecção de faces são:

- Pose: as imagens de uma face variam com relação à pose de foto (frontal, perfil, olhando pra cima, olhando para baixo, dentre outros);
- Presença ou ausência de componentes estruturais: barba, bigode, óculos e outros atributos podem ou não estar presentes ou não em uma face;
- Expressão facial: a aparência facial é diretamente afetada pela expressão facial da pessoa;
- Oclusão: faces podem ser parcialmente oclusas em uma imagem;
- Orientação de imagem: as imagens faciais variam para diferentes rotações sobre o eixo óptico da câmera;
- Condições de imagem latente: quando a imagem é formada, fatores como a iluminação (espectro e intensidade) e características de câmera (resposta de sensor, lentes) afetam a aparência de uma face.

A seguir serão descritos alguns métodos para detectar faces de uma imagem colorida ou em níveis de cinza.

#### 3.1.1 Métodos *Top-Down* Baseados em Conhecimento

Os métodos *Top-Down* são baseados em regras derivadas do conhecimento humano [Yan02], tais como o relacionamento entre atributos faciais (por exemplo, os olhos). Por exemplo, a posição e distância relativa entre olhos, nariz e boca. Assim, primeiramente,

deve-se utilizar um método para extrair tais atributos e baseado na regra de conhecimento, identificar as faces da imagem.

*Yang e Huang [Yan94]* utilizaram um método baseado em conhecimento hierárquico para detectar faces (*Multiresolution rule-based method*). O sistema consistia de três níveis de regras. No nível mais elevado, todas candidatas à face possíveis são encontradas ao varrer (*scanning*) uma janela sobre a imagem de entrada e aplicar um conjunto de regras em cada local. As regras no mais alto nível são descrições gerais da aparência de uma face enquanto as regras nos níveis mais baixos utilizam os detalhes de atributos faciais. Utiliza-se uma hierarquia de imagens com diferentes resoluções conforme mostra a Figura 3.1. Cada quadrado de células consiste de  $n \times n$  *pixels*, onde a intensidade de cada *pixel* é trocada pela intensidade média de *pixels* naquela região quadrada de células.



(a)

(b)

(c)

(d)

Figura 3.1 – Imagem original e imagens de baixa resolução correspondentes: (a)  $n=1$ , imagem original. (b)  $n=4$ . (c)  $n=8$ . (d)  $n=16$ .

Os exemplos de regras de código usadas para localizar candidatas à face em uma resolução menor incluem: “a parte central da face (a parte sombreada mais escura na Figura 3.2) que tem quatro células com uma intensidade basicamente uniforme”, “a parte arredondada superior de uma face (as partes claras na Figura 3.2) que tem uma intensidade basicamente uniforme” e “a diferença entre a média dos níveis de cinza da parte central e parte arredondada superior”. A imagem de menor resolução (nível 1) é buscada para candidatas à face e serão processadas em resoluções melhores. No nível 2, a equalização local de histograma é realizada nas candidatas seguido da detecção de borda. As regiões das candidatas restantes são examinadas no nível 3 com outro conjunto de regras que

respondem a atributos faciais tais como olhos e boca. Yang et al [Yan94] avaliou um conjunto de teste com 60 imagens, sendo que este sistema localizou faces em 50 destas imagens onde 28 imagens correspondiam a um alarme-falso.

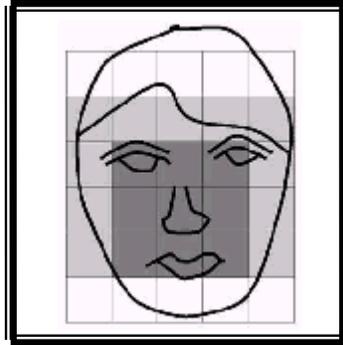


Figura 3.2 - Face típica utilizada em métodos de reconhecimento *top-down* baseado em conhecimento [Yan94].

### 3.1.2 Métodos *Bottom-Up* Baseados em Atributos

Os métodos *Bottom-Up* [Yan02] tentam encontrar atributos invariantes para detecção de faces. A hipótese é baseada na observação de que os humanos podem facilmente detectar faces e objetos em diferentes poses e condições de iluminação e, assim, devem existir propriedades ou atributos que são invariáveis. Geralmente, os atributos são extraídos utilizando-se detectores de bordas. Baseado nos atributos extraídos, um modelo estatístico é construído para descrever os relacionamentos entre tais atributos para verificar a existência de uma face.

#### 3.1.2.1 Atributos Faciais

*Sirohey* [Sir93] propôs um método de localização para segmentar uma face com um plano de fundo não controlado para realizar identificação facial. *Sirohey* utilizou um mapa de borda (*Canny Detector* [Can86]) e heurísticas para remover e agrupar bordas de modo que somente aquelas no contorno da face são preservadas. Uma elipse é ajustada para o limite entre a região da cabeça e o plano de fundo. Este algoritmo teve 80% de exatidão em uma base de dados com 48 imagens com planos de fundo não controlados.

### **3.1.2.2 Textura**

As faces humanas possuem diferentes texturas que podem ser usadas para separá-las de diferentes objetos. *Dai e Nakano* [Dai95] aplicaram um modelo baseado em atributos estatísticos de segunda ordem para detecção de face [Har73]. A informação de cor é incorporada com um modelo de textura facial. Usando o modelo de textura facial, eles projetaram um esquema de varredura para detecção facial em cenas coloridas, no qual, as partes semelhantes à cor laranja, incluindo as áreas da face eram aprimoradas. Uma vantagem desta técnica é que ela pode detectar faces em diferentes angulações e com atributos tais como óculos e barba. A taxa de detecção foi perfeita para um conjunto de 30 imagens com 60 faces.

### **3.1.2.3 Cor de Pele**

*McKenna, Gong e Raja* [Mck98] apresentaram um modelo adaptativo de mistura de cores para rastrear faces através da variação de condições de iluminação. Em vez de contar com um modelo de cor de pele baseado em cores constantes, foi usado um modelo estocástico para estimar a distribuição de cores existentes na imagem em tempo real e adaptar-se às mudanças de acomodação nas condições de visão e iluminação. Resultados preliminares mostram que esse sistema pode rastrear faces dentro de uma escala de luminosidade. Entretanto, este método não pode ser aplicado para detectar faces em uma imagem em nível de cinza.

### **3.1.2.4 Múltiplos Atributos**

Recentemente, numerosos métodos que combinam vários atributos estão sendo propostos para localizar e detectar faces. A maioria deles utiliza atributos globais, tais como cor de pele para encontrar candidatos à face. Depois utilizam atributos locais, tais como o nariz para verificar tais candidatos [Yan02].

Uma técnica típica começa com a detecção de regiões semelhantes à pele. Em seguida, os *pixels* dessas regiões são agrupados usando análise de componentes conectados ou algoritmos de agrupamento. Se o formato de uma região conectada for oval ou elíptico, a região se torna uma candidata à face. Finalmente, atributos locais são usados para a verificação.

### **3.1.3 *Template Matching***

O método *Template Matching* utiliza um padrão de face (geralmente frontal) predefinido ou parametrizado por uma função [Yan02]. Dada uma imagem de entrada, os valores de correlação entre os atributos do padrão de face e os atributos do candidato são computados independentemente para cada atributo. A existência de uma face é baseada nos valores de correlação. Esta técnica é fácil de se implementar, porém inadequada para a detecção de face, pois não é efetiva com problemas como variação de escala, mudança de pose e deformações no formato facial.

### **3.1.4 Métodos Baseados na Aparência (Similaridade)**

Os métodos baseados na aparência baseiam-se em técnicas de análise estatística e aprendizagem para encontrar características relevantes de imagens faciais e não faciais. As características de aprendizagem estão na forma de modelos de distribuição dos padrões de imagem ou funções discriminantes que são conseqüentemente utilizadas para detecção de face.

## **3.2 Reconhecimento de faces**

Um sistema de verificação ou reconhecimento de faces segue os seguintes passos:

1. Captura de uma imagem e uma identidade proposta;
2. Normalização da face na imagem para um padrão de posição, orientação e escala;

3. Verificação da identidade através de alguma técnica; e
4. Aviso de reconhecimento.

### **3.2.1 Redes Neurais**

Dentre as diversas técnicas utilizadas para o reconhecimento facial, destacam-se as que utilizam redes neurais. As redes neurais têm sido aplicadas com sucesso em muitos problemas de reconhecimento de padrões [Fuk92]. A vantagem de seu uso para reconhecimento facial reside na possibilidade da combinação de treinamento supervisionado e não-supervisionado [Bia01]. As redes neurais podem utilizar algoritmos adaptativos para convergir em uma classificação correta. A eficiência de sua classificação está intimamente ligada ao número de padrões de treinamento utilizados e, também quanto à forma de treinamento utilizada (supervisionado e não-supervisionado). As redes neurais e a adaptação de uma rede específica para o reconhecimento facial serão vistas com detalhes nos capítulos subseqüentes.

## **4 REDES NEURAIIS**

No presente capítulo será feita uma analogia entre redes neurais artificiais e redes neurais biológicas. Com relação às redes neurais artificiais, serão estudados: sua arquitetura, tipos de treinamento e exemplos de redes com grande importância na área de reconhecimento de padrões.

### **4.1 Definição**

Uma rede neural artificial é um sistema de processamento de informação que desempenha função análoga à função das redes neurais biológicas. As redes neurais artificiais são desenvolvidas a partir de generalizações de modelos matemáticos de cognição e biologia neural [Fau94].

### **4.2 Redes Neurais Biológicas**

Um neurônio biológico tem três tipos de componentes que são de interesse particular para o estudo dos neurônios artificiais: soma, dendritos e axônio. Os dendritos recebem sinais vindos de outros neurônios. Os sinais são impulsos elétricos que são transmitidos através de uma brecha sináptica por meio de processos químicos. A ação dos transmissores químicos modifica o sinal transmitido [Gon96].

O soma (corpo celular) faz o somatório dos sinais de entrada. Quando o valor de tal somatório ultrapassa um certo limiar, a célula dispara e transmite um sinal através de seu axônio, e este é ligado aos dendritos de outras células através das sinapses [Arb95]. Um neurônio biológico geral é ilustrado na Figura 4.1.

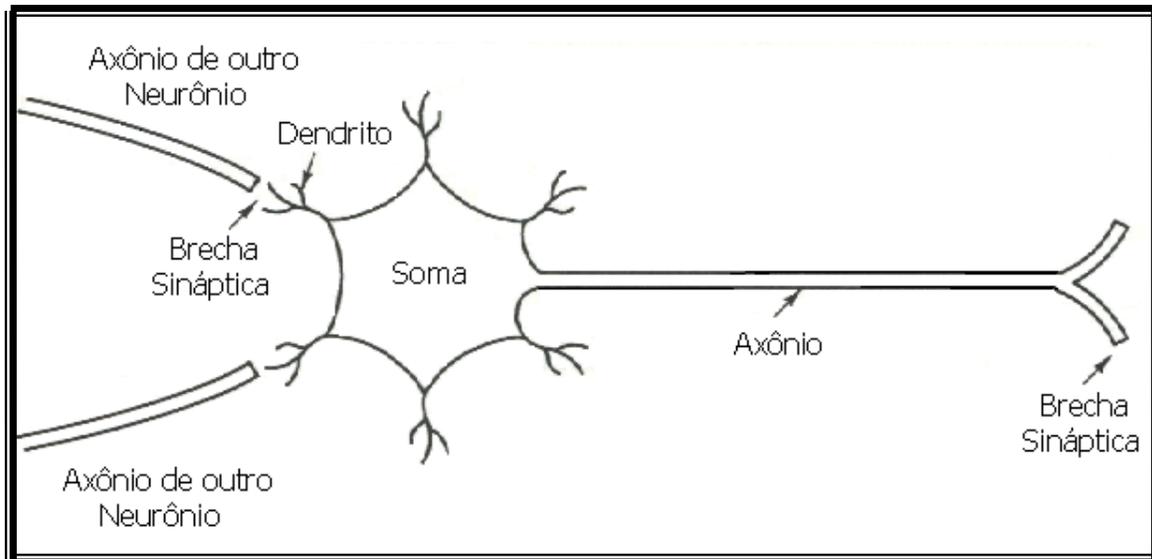


Figura 4.1 – Neurônio biológico.

### 4.3 Redes Neurais Artificiais

A analogia entre rede neural artificial e rede neural biológica é feita assumindo-se que [Fau94]:

- 1- O processamento da informação ocorre nos elementos chamados neurônios;
- 2- Os sinais são passados entre neurônios através de conexões entre neurônios;
- 3- Cada conexão está associada a um peso, que irá multiplicar o sinal transmitido;
- 4- Cada neurônio aplica uma função de ativação na sua entrada global (somatório das suas entradas multiplicadas pelo peso de sua respectiva conexão) para determinar seu sinal de saída.

Uma rede neural consiste de um grande número de elementos processantes chamados neurônios (unidades, células ou nós). Cada neurônio está conectado a outros neurônios através de conexões intercelulares, cada qual com um peso associado. Os pesos representam informações (fatores de multiplicação) que caracterizam a forma de ativação específica de cada neurônio.

Cada neurônio tem um estado interno, chamado ativação (nível de atividade), que é uma função das entradas recebidas. Geralmente, cada neurônio envia sua ativação como um sinal para diversos outros neurônios.

### **4.3.1 Arquitetura de Uma Rede Neural Artificial**

Geralmente, os neurônios são organizados em camadas, sendo que os neurônios em uma mesma camada comportam-se da mesma maneira. Normalmente, os neurônios artificiais têm a mesma ativação e o mesmo padrão de conexão com outros neurônios dentro de uma mesma camada.

As redes neurais artificiais, geralmente, são classificadas como redes de camada única ou redes multicamada. As unidades de entrada não são contadas como uma camada, uma vez que as mesmas não realizam nenhuma computação. O número de camadas em uma rede pode ser definido como o número de camadas de conexões intercelulares existentes na rede.

#### **4.3.1.1 Redes Neurais Artificiais de Camada Única**

Uma rede neural de camada única tem uma camada de conexão intercelular. Podemos dividir uma rede em: unidades de entrada que recebem os sinais externos; unidades de saída, cujos sinais resultantes do processamento podem ser lidos. Na classificação de padrões, cada unidade de saída corresponde a uma categoria particular a qual um padrão de entrada pode pertencer ou não.

Considerando-se, o neurônio  $Y$ , ilustrado na Figura 4.2, que recebe sinais de entrada vindos dos neurônios  $X_1, X_2, \dots, X_n$ , as ativações desses neurônios são  $x_1, x_2, \dots, x_n$ , respectivamente. Os pesos nas conexões vindas de  $X_1, X_2, \dots, X_n$  para o neurônio  $Y$  são  $w_1, w_2, \dots, w_n$ , respectivamente. A saída do neurônio  $Y$  é calculado pelo somatório dos

sinais dos neurônios  $X_1, X_2, \Lambda, X_n$  multiplicados pelos seus respectivos pesos, que resulta num valor denotado  $Y_{in}$  :

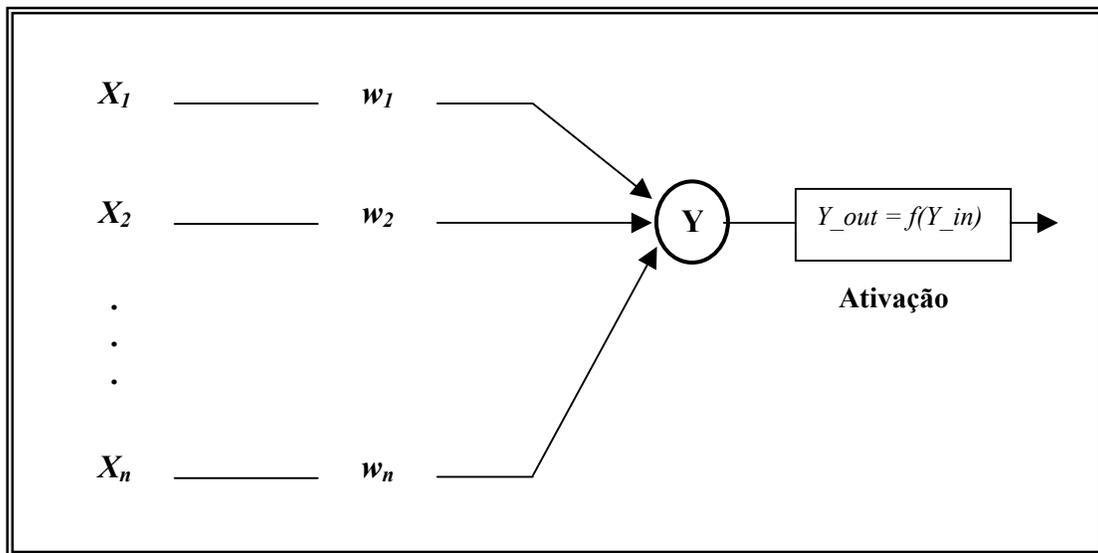


Figura 4.2 – Rede neural de camada única.

$$Y_{in} = w_1x_1 + w_2x_2 + \Lambda + w_nx_n, \quad (4.1).$$

A ativação  $Y_{out}$ , do neurônio  $Y$ , é dada por alguma função a partir de  $Y_{in}$  :

$$Y_{out} = f(Y_{in}), \quad (4.2).$$

#### 4.3.1.2 Redes Neurais Artificiais Multicamada

Uma rede neural multicamada tem um ou mais camadas de unidades ocultas (*Hidden Units*) entre as unidades de entrada e unidades de saída.

Considerando-se, a rede neural artificial, ilustrada na Figura 4.3, semelhante a rede da Figura 4.2, a diferença é que o neurônio  $Y$  está conectado aos neurônios  $Z_1$  e  $Z_2$ , através de conexões com os pesos  $v_1$  e  $v_2$ , respectivamente. O valor da entrada  $Z_1$  é a ativação de  $Y$  multiplicada pelo coeficiente de conexão  $v_1$  e o valor da entrada  $Z_2$  é a ativação de  $Y$  multiplicada pelo coeficiente de conexão  $v_2$ . Tais fórmulas são mostradas a seguir:

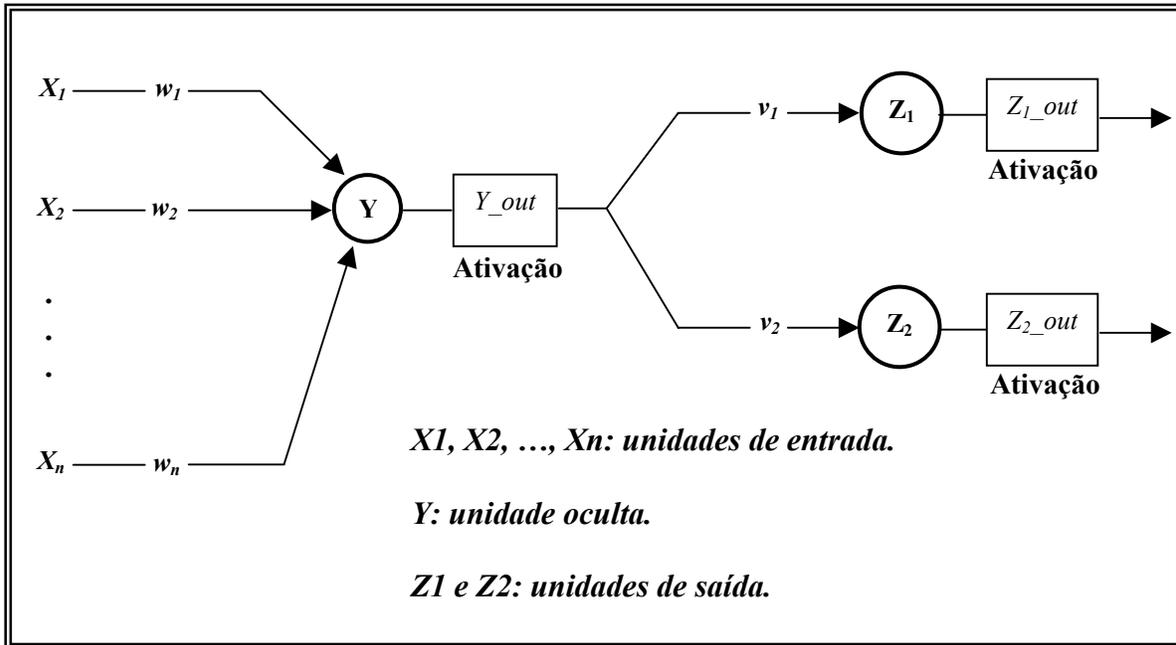


Figura 4.3 – Rede Neural Multicamada.

$$Z_{1\_in} = Y\_out \cdot v_1 = f(Y\_in) \cdot v_1, \quad (4.3).$$

$$Z_{2\_in} = Y\_out \cdot v_2 = f(Y\_in) \cdot v_2, \quad (4.4).$$

As ativações dos neurônios  $Z_1$  e  $Z_2$  serão uma função das entradas das mesmas:

$$Z_{1\_out} = f(Z_{1\_in}), \quad (4.5).$$

$$Z_{2\_out} = f(Z_{2\_in}), \quad (4.6).$$

### 4.3.2 Tipos de Treinamento

Existem dois tipos básicos de treinamento para as redes neurais: supervisionado e não supervisionado, descritos a seguir.

O treinamento supervisionado é realizado apresentando uma seqüência de vetores de treinamento, ou padrões, cada qual associado com um vetor de saída denominado vetor alvo (*target*). Os pesos são ajustados de acordo com o algoritmo de aprendizagem.

O treinamento não-supervisionado é realizado apresentando um grupo de vetores de entrada sem utilizar-se de um vetor alvo. A rede modifica os pesos, de modo que, os vetores de entrada mais similares ativem o conjunto de neurônios. Tal processo é chamado de auto-organização da rede.

#### 4.4 Exemplos de Redes Neurais

Serão mostrados a seguir, alguns exemplos importantes de redes neurais bastante conhecidas.

##### 4.4.1 Perceptron

O *Perceptron* é um modelo de rede neural utilizado para a classificação de padrões linearmente separáveis [Hay94], sendo que existe uma função de decisão para classificar dois conjuntos linearmente separáveis [Gon92] [Gon96]. O *Perceptron* monocamada possui apenas uma camada de conexão intercelular, como mostrado na Figura 4.4, utilizando o método de aprendizado supervisionado.

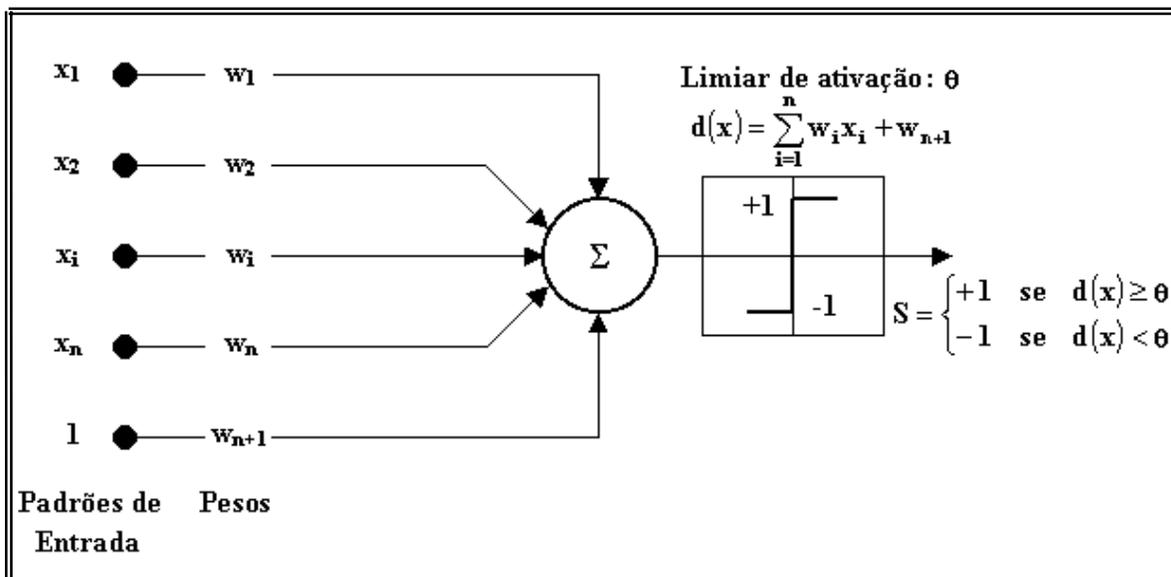


Figura 4.4 – *Perceptron* Monocamada.

O *Perceptron* monocamada possui sinais de entrada  $x_i$ , para  $i=1,2,3,\dots,n$ , coeficientes de conexão  $w_i$  para  $i=1,2,3,\dots,n,n+1$ , um somador  $\Sigma$  que faz a soma ponderada das entradas multiplicadas pelo seu respectivo coeficiente de conexão e um limiar de ativação  $\theta$ . Se a soma ponderada  $d(x)$  for maior ou igual ao limiar  $\theta$ , onde  $d(x)$  é dado pela fórmula:

$$d(x) = \sum_{i=1}^n w_i x_i + w_{n+1}, \quad (4.7).$$

o neurônio dispara, com  $S = +1$ , caso contrário, o neurônio não dispara, indicado com  $S = -1$ . O *Perceptron* monocamada possui treinamento supervisionado com um conjunto de valores alvo para cada padrão de treinamento [Gon96] [Fau94]. O algoritmo é iterativo, reajustando os pesos em um número finito e indefinido de passos. A iteração prossegue até que a saída gerada por cada padrão de treinamento for igual ao seu valor alvo. Assim, a rede irá convergir para valores corretos de pesos em um número finito de iterações. O problema, entanto, acontece quando os padrões não forem linearmente separáveis, pois o *Perceptron* irá entrar em laço infinito tentando convergir em valores corretos. O reajuste dos pesos é dado pela seguinte fórmula:

$$w_i(\text{novo}) = w_i(\text{velho}) + \alpha \cdot t \cdot x_i, \quad (4.8).$$

O parâmetro  $t$  diz respeito ao valor alvo para o padrão de treinamento e o parâmetro  $\alpha$  diz respeito à taxa de aprendizagem, indicando a velocidade de reajuste. Chamando cada padrão de treinamento de  $s$ , temos o seguinte algoritmo de aprendizagem para o *Perceptron*:

- a- Atribuir valores aos pesos  $w_i$ , para  $i = 1,2,3,\dots,n,n+1$ .
- b- Atribuir o valor da taxa de aprendizagem  $\alpha$ .

c- Enquanto a saída gerada por todo padrão de treinamento não for igual ao seu alvo, fazer:

c.1- Para cada par de treinamento:  $s$  e  $t$ , fazer:

c.1.1-  $x_i = s_i$ ;

c.1.2-  $d(x) = w_{n+1} + \sum_{i=1}^n x_i w_i$ ;

c.1.3-  $S = \begin{cases} 1 & \text{se } d(x) \geq 0 \\ -1 & \text{se } d(x) < 0 \end{cases}$

c.1.4- Atualização dos pesos:

Se  $y \neq t$  então

$$w_i(\text{novo}) = w_i(\text{velho}) + \alpha \cdot t \cdot x_i$$

#### 4.4.2 *Perceptron* Multicamadas (*Multilayer Perceptron – MLP*)

O *Perceptron* multicamadas é uma extensão do *Perceptron* monocamada que utiliza o algoritmo iterativo para cada camada da rede. O *MLP*, também possui um limiar  $\theta$  diferente para cada camada da rede. Assim, cada camada da rede irá convergir para um reajuste final de pesos, em um número diferente de passos que outra camada. A Figura 4.5 abaixo mostra a arquitetura do *MLP*:

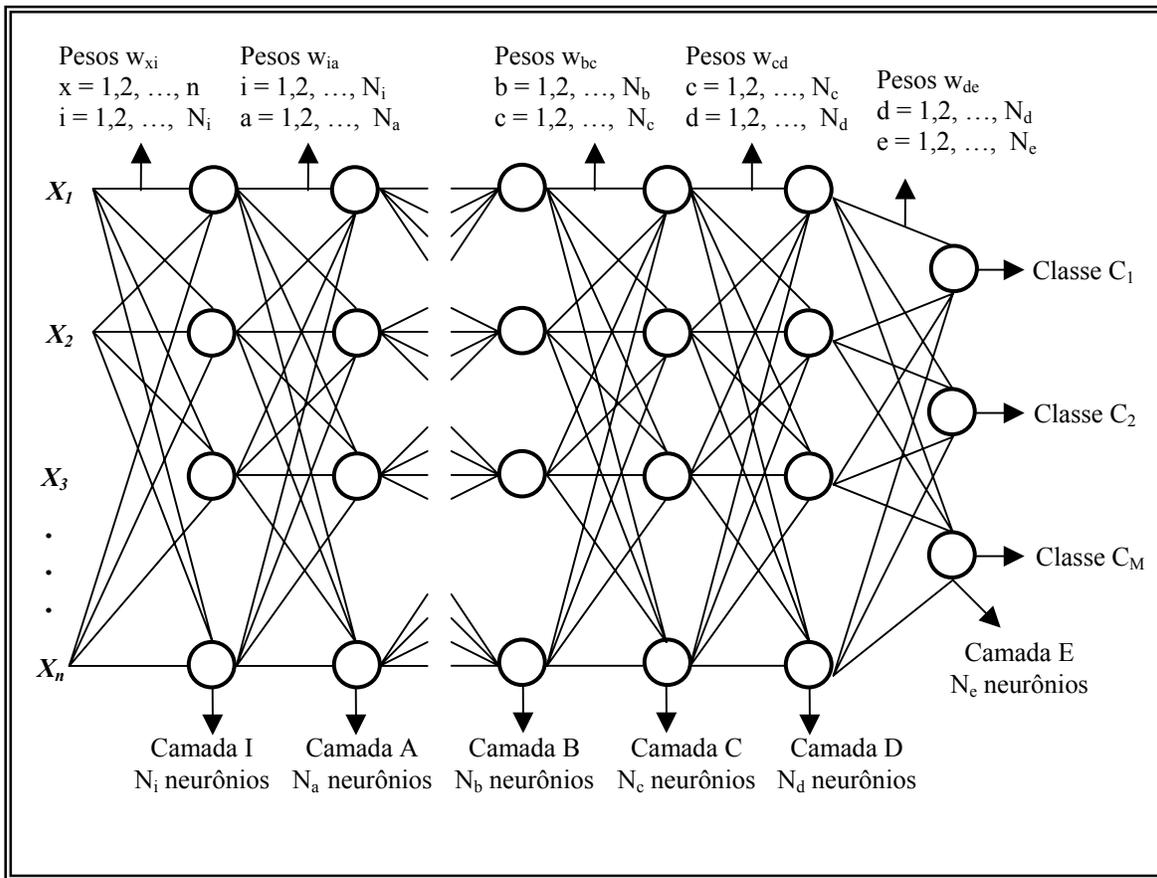


Figura 4.5 – Rede Neural *MLP*, onde o algoritmo para reajuste dos pesos é aplicado em cada camada de forma independente.

#### 4.4.3 Rede Neural Recorrente

A seguir serão descritas as redes neurais recorrentes, que se caracterizam pela existência de interconexões no sentido inverso, além das interconexões no sentido direto, como nas redes neurais Perceptron e *MLP*. Como exemplo temos a rede neural desenvolvida por Hopfield, que é inteiramente interconectada, de modo que, cada unidade está conectada a qualquer outra unidade da rede. A rede tem pesos simétricos sem autoconexões (conexão diretamente para elas próprias), o que pode ser descrito melhor nas fórmulas 4.9 e 4.10.

$$w_{ij} = w_{ji}, \quad (4.9).$$

$$w_{ii} = 0, \quad (4.10).$$

A rede apresentada na Figura 4.6 exemplifica a rede de Hopfield. A atualização é assíncrona e segue uma função conhecida como função *Lyapunov*. Existem duas particularidades com as redes neurais recorrentes:

- Somente uma unidade atualiza sua ativação por vez (baseado no sinal que ele recebe de cada uma das outras unidade);
- Cada unidade continua a receber sinais externos, que serão adicionados aos das outras unidades de entrada vindas da rede.

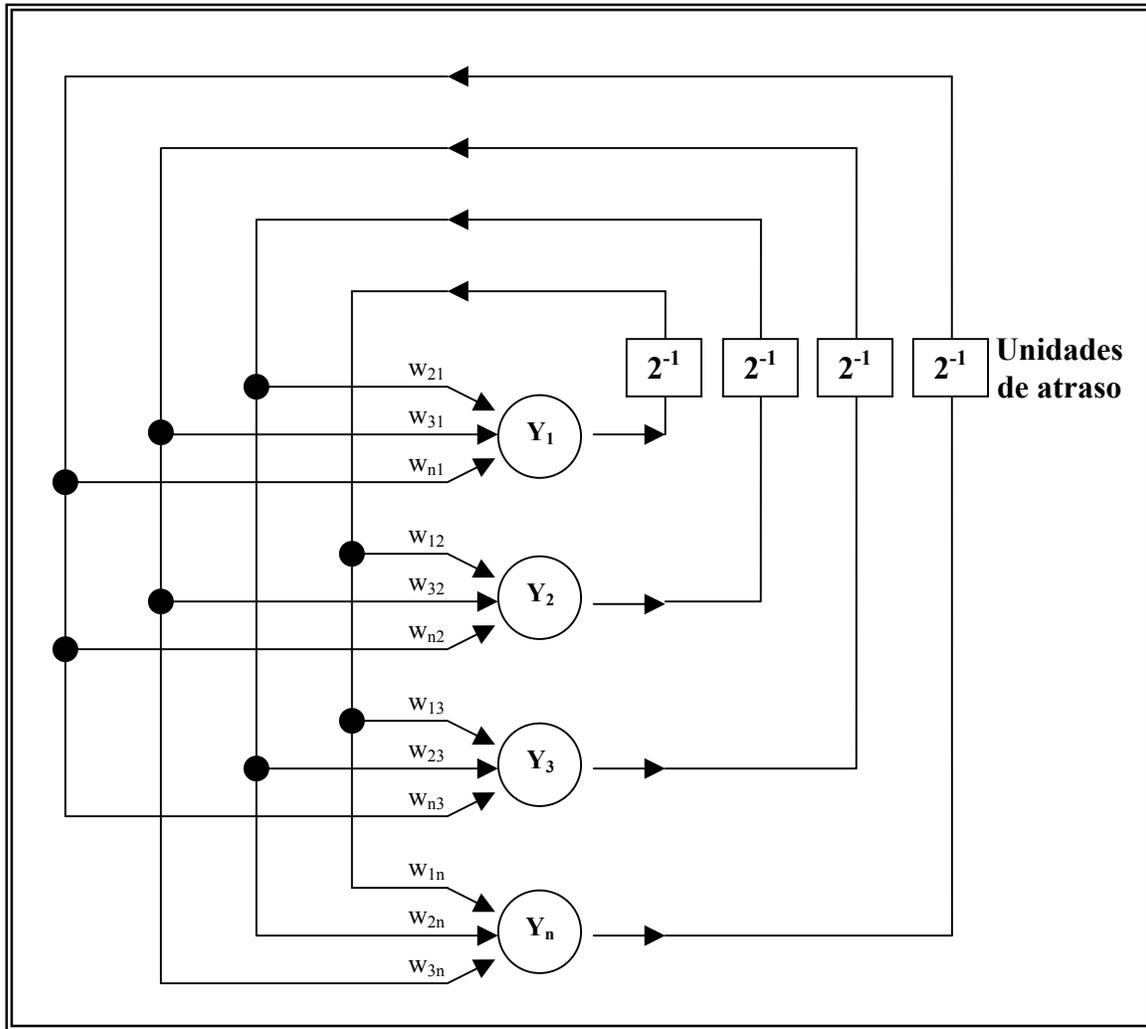


Figura 4.6 – Rede Recorrente sem unidades ocultas.

#### 4.4.4 Rede Neural Convolucional

As redes neurais convolucionais tentam lidar com o problema de deformações e variações nos dados de entrada. Seja, por exemplo, a rede convolucional da Figura 4.7. Ela tem um conjunto de camadas com um ou mais planos. A imagem é apresentada para a camada de entrada. Cada célula em um plano recebe entradas de uma pequena vizinhança de um plano da camada prévia. Os campos receptivos locais em um plano possuem pesos iguais com a finalidade de extrair a mesma característica.

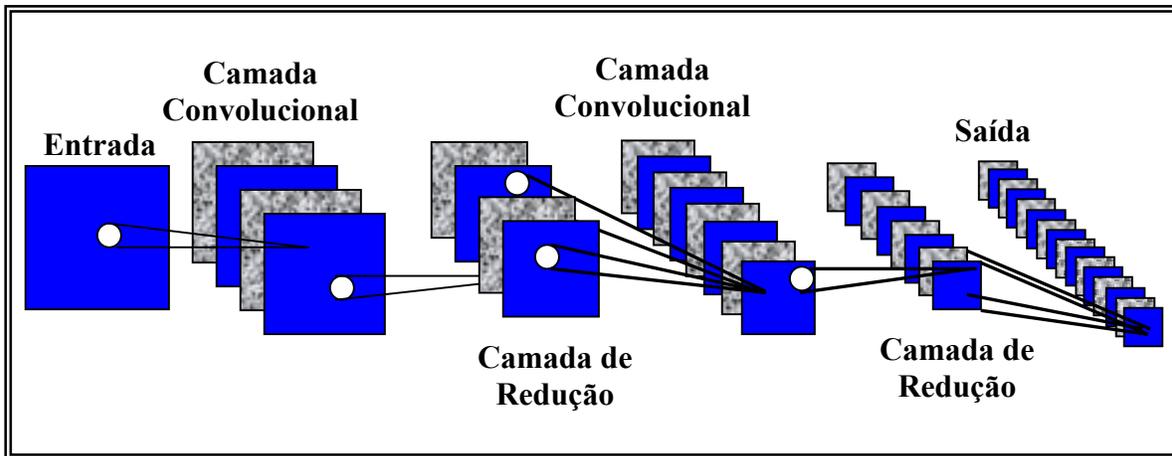


Figura 4.7 – Exemplo de rede convolucional.

Cada plano pode ser considerado como um mapa de características onde são detectadas características pré-estabelecidas. Assim, múltiplos planos são utilizados em cada camada para detecção de múltiplas características. Estas camadas são denominadas camadas convolucionais. As camadas de redução são inseridas entre as camadas convolucionais, sendo responsáveis pelo fator de tolerância a distorções da rede.

Um exemplo de rede neural convolucional é o *Neocognitron*, cuja descrição é feita no capítulo seguinte.

## **5 NEOCOGNITRON**

Neste capítulo é descrita a rede neural convolucional *Neocognitron*, cuja versão alternativa foi implementada no presente trabalho para o reconhecimento de atributos faciais. É enfocada a base teórica da rede, enfatizando sua estrutura hierárquica, citologia, progresso temporal da rede e tipos de treinamento.

O reconhecimento de padrões visuais, tais como o reconhecimento de caracteres, pode facilmente ser feito por humanos, mas é difícil de se projetar uma máquina que possa fazer tal tarefa.

Os métodos de reconhecimento de padrões insensíveis a mudanças de posição e distorções na forma dos padrões de estímulos, necessitam da normalização da posição e forma dos padrões de entrada. Como os métodos de normalização são complexos, é desejável, encontrar um algoritmo para reconhecimento de padrões que possa suportar deslocamentos na posição e distorções na forma do padrão de estímulos [Fuk82].

### **5.1 Introdução**

Proposto por *Kunihiko Fukushima* em 1979, o *Neocognitron* é uma rede neural hierárquica paralela [Fuk80], na qual informação do padrão de estímulo dada na camada de entrada é processada passo a passo em cada nível da rede multicamada. Uma célula em um nível mais avançado, geralmente, responde a atributos compostos de padrões de estímulos e é menos sensível a mudanças na posição do padrão de estímulos. Assim, cada célula no nível mais avançado responde somente a um padrão específico sendo pouco afetada pela posição ou tamanho dos padrões de estímulos [Fuk82] [Fuk88] [Lov97].

### **5.2 Estrutura Hierárquica do *Neocognitron***

O *Neocognitron* é uma rede multicamada hierárquica consistindo de uma cascata de camadas de neurônios. As células são análogicas, ou seja, suas entradas e suas saídas

correspondem a valores não negativos que correspondem à frequência de ativação instantânea de um neurônio biológico [Fuk88].

A Figura 5.1 é um esboço simplificado da seqüência de níveis do modelo inicial do *Neocognitron* proposto em 1979, consistindo de uma conexão em cascata de estruturas modulares precedidas por uma camada de entrada  $U_0$ , que diz respeito à matriz fotorreceptora. Cada estrutura modular é denominada estágio e possui duas camadas de células:  $S$  e  $C$ . A rede possui também células  $V_S$  na camada  $S$  e células  $V_C$  na camada  $C$ . As conexões de entrada para as células  $C$  têm pesos fixos, enquanto que, as conexões de entrada para as células  $S$  têm pesos variáveis.

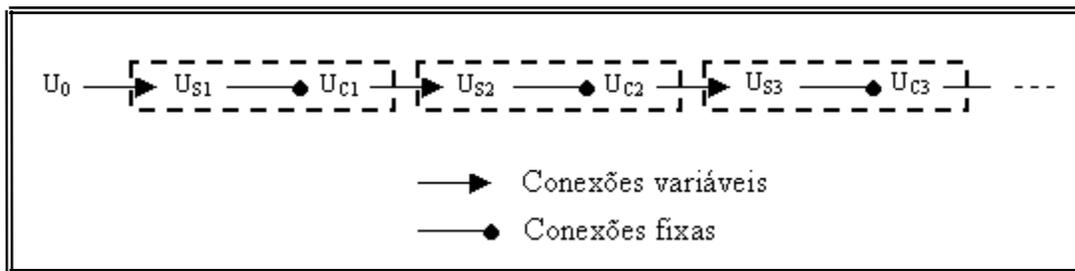


Figura 5.1 – Estrutura simplificada da seqüência de estágios do *Neocognitron*.

Cada camada de células  $S$ , ou  $C$ , é constituída de um certo número de arranjos bidimensionais de células denominados planos celulares. Todas as células em um plano celular têm conexões intercelulares de entrada com a mesma distribuição espacial, sendo que as posições das células na camada precedente de onde suas conexões intercelulares de entrada estão vindo são deslocadas em paralelo. Isso é ilustrado na Figura 5.2.

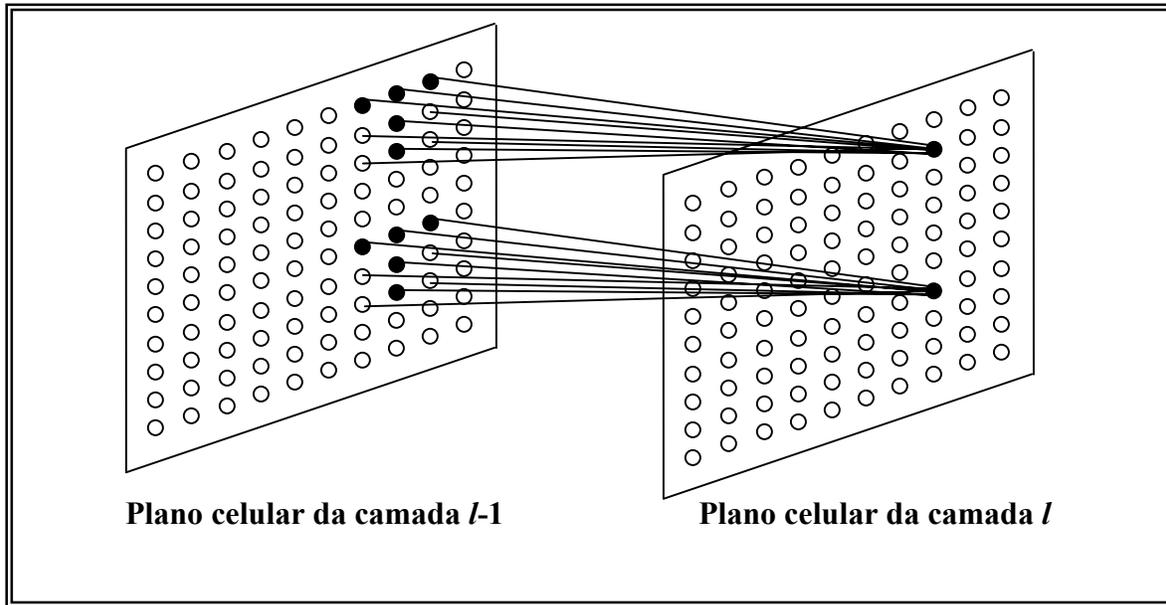


Figura 5.2 – Distribuição das conexões intercelulares de entrada do *Neocognitron*.

A Figura 5.3 é um diagrama esquemático ilustrando as conexões intercelulares entre os estágios e camadas para uma implementação do *Neocognitron* de três estágios. Cada retângulo, dentro dos quais estão os planos celulares (em quadrado), representa uma camada *S* ou *C*, denotados  $U_{Si}$  e  $U_{Ci}$ . Para maior simplicidade, somente uma célula é mostrada por plano, sendo que cada célula recebe conexões intercelulares de entrada de uma área fechada (representada na Figura 5.3 por uma circunferência) da camada precedente. As outras células no mesmo plano têm conexões intercelulares de entrada com a mesma distribuição espacial e as posições das células na camada precedente que estão conectadas aos terminais de entrada são deslocadas em paralelo de célula para célula. Portanto, todas as células em um plano têm campos receptivos com a mesma distribuição, mas em diferentes posições, sendo que os planos celulares da última camada pertencente ao último estágio, denotado por  $U_{C3}$  tem apenas uma célula, a qual corresponde a uma classe na qual pode ser classificado o padrão de entrada.

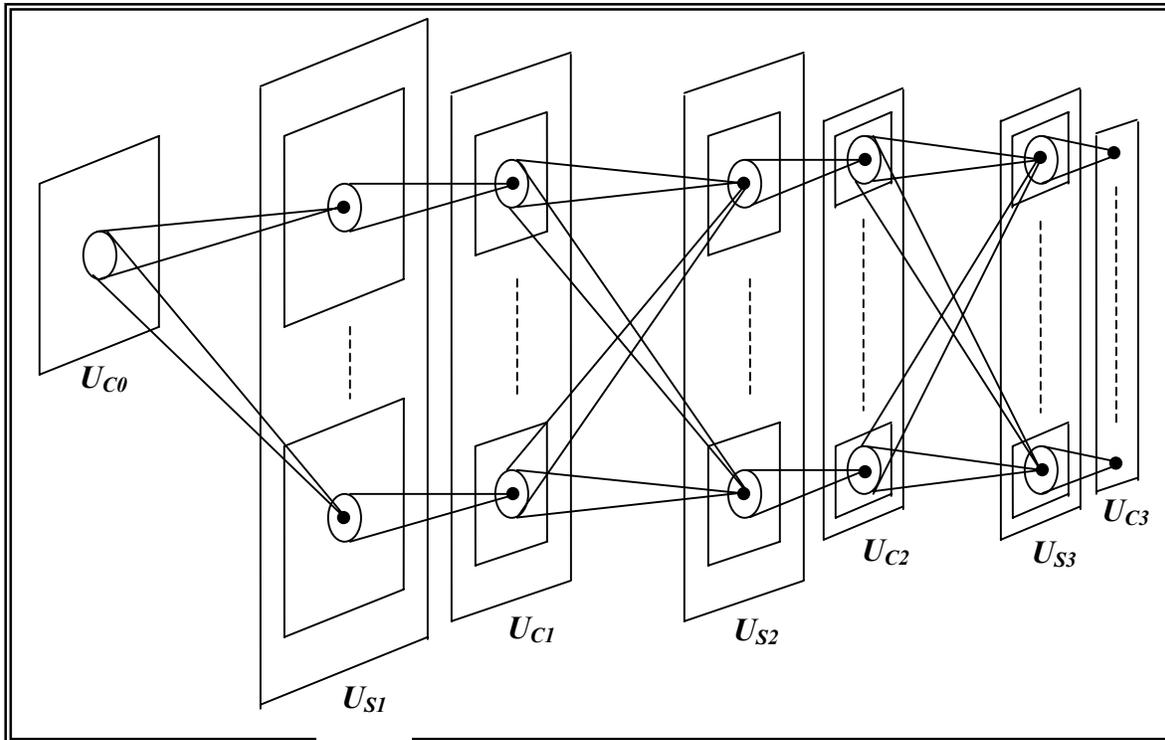


Figura 5.3 – Estrutura entre os estágios e camadas de um *Neocognitron* de três estágios.

### 5.3 A Citologia do *Neocognitron*

As células *S* são extratoras de atributos [Fuk88] e suas conexões intercelulares de entrada são variáveis e modificáveis durante o processo de aprendizagem. As células *S* na menor camada extraem atributos locais tais como linhas ou bordas. As células em camadas mais avançadas extraem atributos globais (compostos) do padrão de treinamento.

As células *S* recebem conexões de entrada de células *C* e  $V_C$  e enviam sinais para as células  $V_S$  e *C*. As conexões entre as células *C* e *S* são excitatórias enquanto que a conexão entre as células  $V_C$  e *S* é inibitória, utilizando um mecanismo “*shunting*” (Dado um conjunto de células que deveriam disparar em um plano, o mecanismo “*shunting*” tem a função de estabelecer uma proporção entre as células do conjunto que realmente dispararam e o total de células do conjunto). Sejam  $u_{C_{l-1}}(k_{l-1}, \mathbf{n} + i)$  as entradas excitatórias e  $v_{C_{l-1}}(\mathbf{n})$  a entrada inibitória, a saída  $u_{S_l}(k_l, \mathbf{n})$  da célula *S* é definida como:

$$u_{sl}(k_l, \mathbf{n}) = r_l \cdot \varphi \left[ \frac{1 + \sum_{k_{l-1}=1}^{K_{l-1}} \sum_{i \in S_l} a_l(k_{l-1}, i, k_l) \cdot u_{cl-1}(k_{l-1}, \mathbf{n} + i)}{1 + \frac{r_l}{1 + r_l} \cdot b_l(k_l) \cdot v_{cl-1}(\mathbf{n})} - 1 \right], \quad (5.1).$$

onde  $l$  representa o estágio da rede;  $k_l$  diz respeito ao plano celular da camada  $l$  que contém a célula sendo analisada;  $k_{l-1}$  indica o plano celular da área de conexão na camada precedente  $l-1$ ;  $K_{l-1}$  denota o índice final do somatório (último plano da camada precedente  $l-1$ );  $\mathbf{n}$  o campo receptivo (diz respeito à posição de uma célula e das células pertencentes a sua área de conexão na camada precedente). Tem-se também,  $a_l(k_{l-1}, i, k_l)$  representando os coeficientes de conexão excitatória,  $b_l(k_l)$ , o coeficiente conexão inibitória e  $S_l$  a área de conexão intercelular de entrada para a célula  $u_{sl}(k_l, \mathbf{n})$ . Assim, para  $l=1$ ,  $u_{cl-1}(k_{l-1}, \mathbf{n})$  será  $u_0(\mathbf{n})$  e temos que  $K_{l-1}=1$ . A função  $\varphi[ ]$  é definida como:

$$\varphi[x] = \begin{cases} x & \text{se } x \geq 0 \\ 0 & \text{se } x < 0 \end{cases}, \quad (5.2).$$

O parâmetro  $r_l$  controla a intensidade de inibição. Quanto maior o valor de  $r_l$ , mais seletiva se torna a resposta da célula a um atributo específico.

A célula  $v_{cl-1}(k_{l-1}, \mathbf{n})$ , que está enviando um sinal inibidor para a célula  $u_{sl}(k_l, \mathbf{n})$ , recebe suas conexões intercelulares de entrada das mesmas células que  $u_{sl}(k_l, \mathbf{n})$  e sua saída pode ser dada por:

$$v_{cl-1}(\mathbf{n}) = \sqrt{\sum_{k_{l-1}=1}^{K_{l-1}} \sum_{i \in S_l} c_{l-1}(i) \cdot u^2_{cl-1}(k_{l-1}, \mathbf{n} + i)}, \quad (5.3).$$

O parâmetro  $c_{l-1}$  diz respeito às conexões intercelulares fixas de entrada para a célula  $v_{Cl-1}(\mathbf{n})$  e  $S_l$  corresponde à área de conexão intercelular de entrada para a célula  $v_{Cl-1}(\mathbf{n})$ .

As células  $C$  são inseridas na rede para permitir erros posicionais [Fuk88], sendo suas conexões intercelulares de entrada fixas durante o processo de aprendizagem. Cada célula  $C$  recebe sinais de um grupo de células  $S$  que extraem os mesmos atributos, mas em posições ligeiramente diferentes, sendo ativada se pelo menos uma célula  $S$  na sua área de conexão estiver ativa. Mesmo se o padrão de estímulo for deslocado em posição, outra célula  $S$  em sua área de conexão tornar-se-á ativa. Sendo assim, a célula  $C$  é menos sensível a deslocamentos na posição do padrão de estímulo.

As células  $C$  recebem conexões de entrada de células  $S$  e  $V_S$  e enviam sinais para as células  $V_C$  e  $S$ . As conexões entre as células  $S$  e  $C$  são excitantes enquanto que a conexão entre as células  $V_S$  e  $C$  é inibidora. Sejam  $u_{Sl}(k_l, \mathbf{n} + i)$  as entradas excitatórias e  $v_{Sl}(\mathbf{n})$  a entrada inibitória, a saída  $u_{Cl}(k_l, \mathbf{n})$  é definida como:

$$u_{Cl}(k_l, \mathbf{n}) = \psi \left[ \frac{1 + \sum_{i \in D_l} d_l(i) \cdot u_{Sl}(k_l, \mathbf{n} + i)}{1 + v_{Sl}(\mathbf{n})} - 1 \right], \quad (5.4).$$

onde  $d_l(i)$  representa os coeficientes de conexão intercelular excitatória para uma célula  $u_{Cl}(k_l, \mathbf{n})$  e  $D_l$  representa a área de conexão intercelular de entrada para uma determinada célula  $u_{Cl}(k_l, \mathbf{n})$ . Sendo  $\alpha$  um parâmetro definido arbitrariamente, a função  $\psi [ \ ]$  é definida como:

$$\psi [x] = \begin{cases} \frac{x}{\alpha + x} & \text{se } x \geq 0 \\ 0 & \text{se } x < 0 \end{cases}, \quad (5.5).$$

A célula inibidora  $v_{sl}(\mathbf{n})$  envia sinais inibitórios para a célula  $C$  e realiza a inibição lateral do sistema, produzindo uma saída proporcional a:

$$v_{sl}(\mathbf{n}) = \frac{1}{K_l} \sum_{k_l=1}^{K_l} \sum_{i \in D_l} d_l(i) \cdot u_{sl}(k_l, \mathbf{n} + i), \quad (5.6).$$

## 5.4 Treinamento Não-Supervisionado

Na proposta inicial de *Kunihiko Fukushima*, a aprendizagem do *Neocognitron* se dava através de treinamento não-supervisionado [Fuk82]. A apresentação repetitiva de um conjunto de padrões de treinamento é suficiente para a organização citológica da rede, não sendo necessária nenhuma informação sobre as categorias nas quais os padrões são classificados. O *Neocognitron* adquire a habilidade de classificar e reconhecer padrões de acordo a similaridade em suas formas.

Toda vez que um padrão for apresentado, muitas células  $S$  serão escolhidas em cada camada  $S$ . Tais células recebem diversas denominações, tais como: “*representative cell*” [Fuk82] ou “*seed cell*” [Fuk88].

As células vencedoras são escolhidas dentre aquelas que produzem as maiores saídas, mas o número de células selecionadas é restrito, de modo que somente uma célula vencedora é escolhida por plano de células  $S$  [Fuk82].

### 5.4.1 Escolha das Células “Vencedoras”

O procedimento para escolha das células vencedoras ocorre da seguinte forma: primeiramente, define-se um grupo de células  $S$ , cujos campos receptivos estão situados dentro de uma área pequena na camada de entrada.

Como exemplo, o arranjo da Figura 5.4 tem um grupo de células  $S$  constituindo uma coluna em alguma camada  $S$ , a qual pode ser chamada coluna  $S$  [Fuk82] ou hiper-coluna  $S$  [Fuk88]. Uma coluna tem células de todos os planos  $S$ , assim, cada coluna  $S$  contém todos tipos de células de extração de atributos. Existem várias colunas em uma camada, sendo que as mesmas poderão ser sobrepostas, possibilitando que uma célula  $S$  esteja contida em duas colunas.

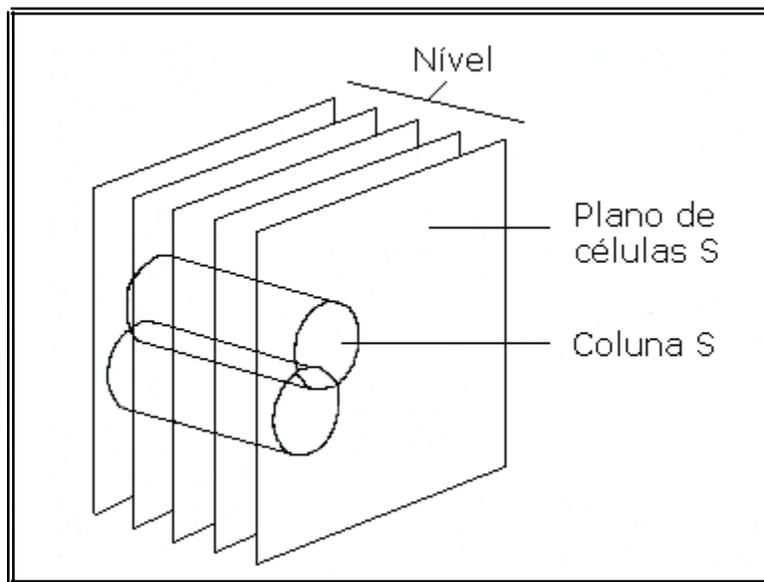


Figura 5.4 – Coluna de células  $S$ .

Em cada coluna, toda vez que um padrão de estímulos for apresentado, a célula  $S$  que produzir a maior saída será escolhida como uma candidata à célula vencedora. Para um dado plano  $S$ , se duas ou mais candidatas aparecerem, somente uma delas será escolhida como célula vencedora. Caso aparecer somente uma, ela será determinada como célula vencedora e, se nenhuma aparecer, nenhuma célula vencedora será escolhida.

#### 5.4.2 Reforço das Conexões Intercelulares Variáveis

O princípio das células vencedoras faz com que cada plano  $S$  se torne responsável por um atributo particular, sendo que não existe nenhuma possibilidade de que dois ou mais planos  $S$  sejam usados para a detecção de um mesmo atributo. De acordo com outro

princípio do *Neocognitron*, não somente as células vencedoras, mas também suas vizinhas têm suas conexões intercelulares reforçadas [Fuk88].

Somente as conexões intercelulares de entrada que apresentam sinais diferentes de zero serão reforçadas. Todas as outras células no mesmo plano  $S$  têm suas conexões intercelulares de entrada reforçadas do mesmo modo que sua célula vencedora. Estas relações podem ser expressas como segue:

Seja a célula  $u_{Sl}(\hat{k}_l, \hat{\mathbf{n}})$  escolhida como vencedora no plano  $\hat{k}_l$  e  $q_l$  uma constante positiva que determina a velocidade do incremento. As conexões intercelulares variáveis  $a_l(k_{l-1}, i, \hat{k}_l)$  e  $b(\hat{k}_l)$  de entrada são reforçadas de acordo com as equações abaixo:

$$\Delta a_l(k_{l-1}, i, \hat{k}_l) = q_l \cdot c_{l-1}(i) \cdot u_{Cl-1}(k_{l-1}, \hat{\mathbf{n}} + i), \quad (5.7).$$

$$\Delta b(\hat{k}_l) = q_l \cdot v_{Cl-1}(\hat{\mathbf{n}}), \quad (5.8).$$

Quando o reforço das conexões intercelulares de entrada da camada  $U_{Sl}$  progride a uma certa dimensão e a resposta da camada  $U_{Sl}$  para cada padrão de treinamento torna-se grande e estável, a saída da camada  $U_{Cl}$ , que é diretamente dirigida pela saída na camada  $U_{Sl}$ , começa a desencadear o reforço das conexões intercelulares variáveis entre  $U_{Cl}$  e  $U_{S2}$ . Durante o estado inicial de reforço, a saída da camada  $U_{Sl}$  mantém-se muito pequena e reforço das conexões intercelulares entre  $U_{Cl}$  e  $U_{S2}$  quase não progride.

### 5.4.3 Progresso Temporal do *Neocognitron*

Será apresentado um exemplo de organização citológica para explicar como funciona o reforço das conexões intercelulares variáveis. Será focado primeiro o princípio, denominado *winner-take-all*, no qual somente as células com saídas máximas têm suas conexões reforçadas.

A rede da Figura 5.5 consiste de doze células: nove fotorreceptoras ( $P1-P9$ ), uma inibidora ( $v$ ) e três  $S$ , chamadas de  $A$ ,  $B$  e  $C$ , sendo representadas por  $u_A$ ,  $u_B$  e  $u_C$ , respectivamente. A estrutura das conexões intercelulares convergindo para cada uma dessas três células  $S$  é idêntica a da rede da Figura 5.6, a qual mostra a estrutura de conexão intercelular entre as camadas  $S$  e  $C$  da rede *Neocognitron*.

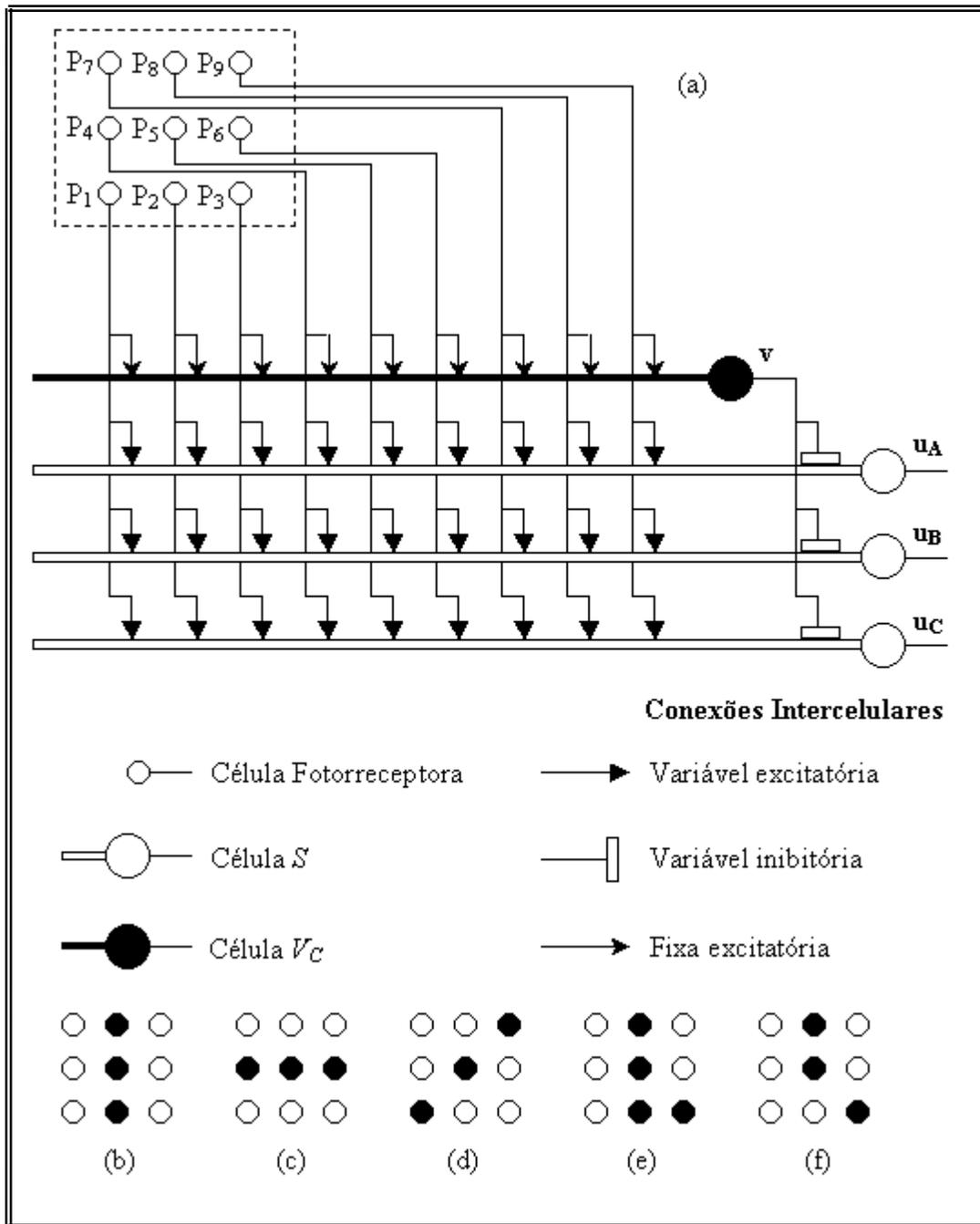


Figura 5.5 – Exemplo de treinamento para padrões de linha [Fuk82].

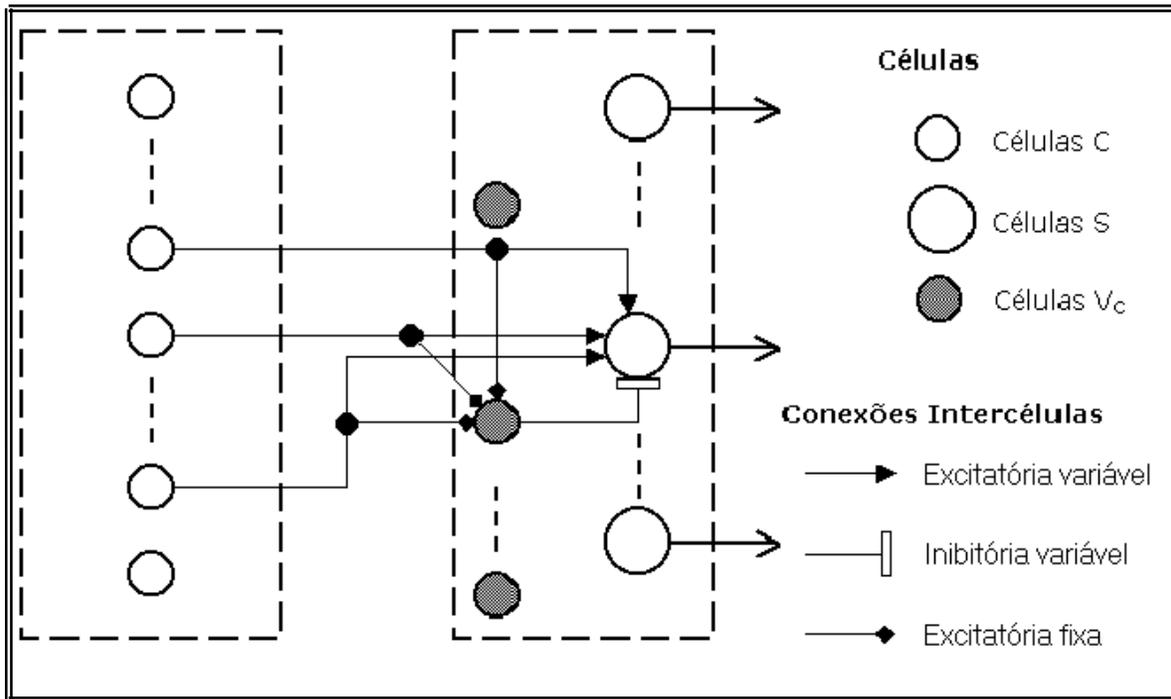


Figura 5.6 – Conexões intercelulares convergindo para células S em uma rede *Neocognitron*[Fuk82].

Assume-se três padrões de treinamento: linha vertical, linha horizontal e linha diagonal, mostrados nas Figuras 5.5 (b-d). As conexões intercelulares excitantes variáveis têm valores iniciais positivos e pequenos, enquanto os valores iniciais das conexões inibitórias são zero.

Assume-se que a linha vertical mostrada na Figura 5.5 (b) é apresentada ao fotorreceptor, todas as células produzem sinais de saída positivos muito pequenos. Caso a célula *A* produza a maior saída, ela terá suas conexões intercelulares de entrada reforçadas de acordo com as equações (5.7) e (5.8). Com isso, as conexões intercelulares das células  $p_2$ ,  $p_5$  e  $p_8$  e  $v$  são reforçadas. Se o valor do parâmetro  $q$  nas equações (5.7) e (5.8) for grande o bastante, comparado com os valores iniciais das interconexões variáveis, a célula *A* adquire a responsabilidade pela linha vertical.

Seja, em seguida, a linha horizontal mostrada na Figura 5.5 (c), apresentada ao fotorreceptor. Uma vez que a célula  $A$  já respondeu seletivamente a uma linha vertical, ela não responderá à linha horizontal, caso o valor de  $r$  não seja muito pequeno. Somente as células  $B$  e  $C$  produzem saídas não-zero. Assume-se que a célula  $C$  responda com maior saída. Neste momento, a célula  $C$  tem suas conexões intercelulares de entrada reforçadas de acordo com as equações (5.8) e (5.9) e adquire a responsabilidade pela linha horizontal.

Seja a linha diagonal, o próximo padrão de treinamento, mostrado na Figura 5.5 (d). Uma vez que as células  $A$  e  $C$  já são responsáveis pelas linhas verticais e horizontais, respectivamente, a célula  $B$  produzirá um sinal de saída não-zero para a linha diagonal. Portanto, a célula  $B$  tem suas conexões intercelulares de entrada reforçadas e torna-se responsável pela linha diagonal.

Caso seja apresentada outra linha horizontal ao fotorreceptor, somente a célula  $C$  já reforçada para este padrão de entrada responderá, tendo suas conexões intercelulares reforçadas novamente para a linha horizontal, o que torna a seletividade da linha horizontal mais estável. Assim, não existe a possibilidade de que duas ou mais células sejam reforçadas para a linha horizontal. O mesmo ocorre com os outros padrões de treinamento.

Caso o número de padrões de treinamento diferentes seja maior que três, nenhuma das células  $S$  produz um sinal de saída positivo para o quarto padrão de treinamento, uma vez que cada uma das células,  $A$ ,  $B$  e  $C$  já foi reforçada e adquiriu responsabilidade para um padrão específico de treinamento. Portanto, no projeto de uma rede, tal como a rede da Figura 5.5 (a), é necessário um número de células  $S$  maior que o número de padrões de treinamento a serem reconhecidos.

Para que alguma célula responda a um padrão de treinamento, o valor do parâmetro  $r$  é escolhido, com o objetivo de que a célula  $A$  responda também ao padrão da Figura 5.5 (d), mas não ao da Figura 5.5 (e). Se nenhuma outra célula não foi ainda treinada para o padrão da Figura 5.5 (d), a célula  $A$ , provavelmente, deve possuir a maior saída e é reforçada para o padrão da Figura 5.5 (d). Caso esse mesmo padrão seja apresentado várias

vezes ao fotorreceptor, a resposta da célula  $A$  para esse padrão aumentará gradualmente e se tornará maior que a resposta ao padrão da Figura 5.5 (b). Quando a resposta ao padrão da Figura 5.5 (d) tornar-se grande o bastante, depois de repetidos reforços, a célula  $A$ , agora, produzirá uma saída diferente de zero para o padrão da Figura 5.5 (e), que tem somente um elemento diferente do padrão da Figura 5.5 (d). Portanto, a célula  $A$  pode, agora, ser reforçada para o padrão da Figura 5.5 (e), para o qual a célula não respondia.

Fazendo uma analogia com o *Neocognitron*, as células fotorreceptoras na Figura 5.5 (a) correspondem à célula da camada  $U_0$  e as células  $A$ ,  $B$  e  $C$  correspondem a células  $S$  da camada  $U_{S1}$ , que têm campos receptivos na mesma posição, mas pertencendo a diferentes planos  $S$ . Geralmente, no *Neocognitron*, cada plano  $S$  tem muitas células  $S$  em paralelo e, cada qual deve ter conexões intercelulares de entrada com a mesma distribuição parcial daquelas células  $A$ ,  $B$  e  $C$ .

## 5.5 Treinamento Supervisionado

Em 1983, *Fukushima* introduziu um esquema de treinamento supervisionado para melhorar o desempenho do *Neocognitron* para o reconhecimento de caracteres manuscritos [Lov97]. Enquanto no treinamento não supervisionado, as células com saída máxima são escolhidas como as células vencedoras, no treinamento supervisionado o “professor” apontará qual célula deve ser a célula vencedora para cada padrão de treinamento. O processo de aprendizagem é o idêntico ao do treinamento não supervisionado, porém não se necessitam dos procedimentos de formação de grupo para a escolha da célula vencedora, devendo apenas, apontar quais atributos devem ser extraídos por quais células [Fuk88]. Como exemplo de treinamento supervisionado serão mostradas: rede para detecção de bordas e linhas e rede para detecção de pontos finais e pontos de curvatura.

### 5.5.1 Detecção de Bordas e Linhas

A Figura 5.7 ilustra o princípio de extração de bordas e linhas. O padrão de treinamento apresentado para cada plano de célula na camada  $U_{S1}$  é uma borda reta com

uma orientação particular. A borda, entretanto, está localizada no centro do campo receptivo [Fuk92].

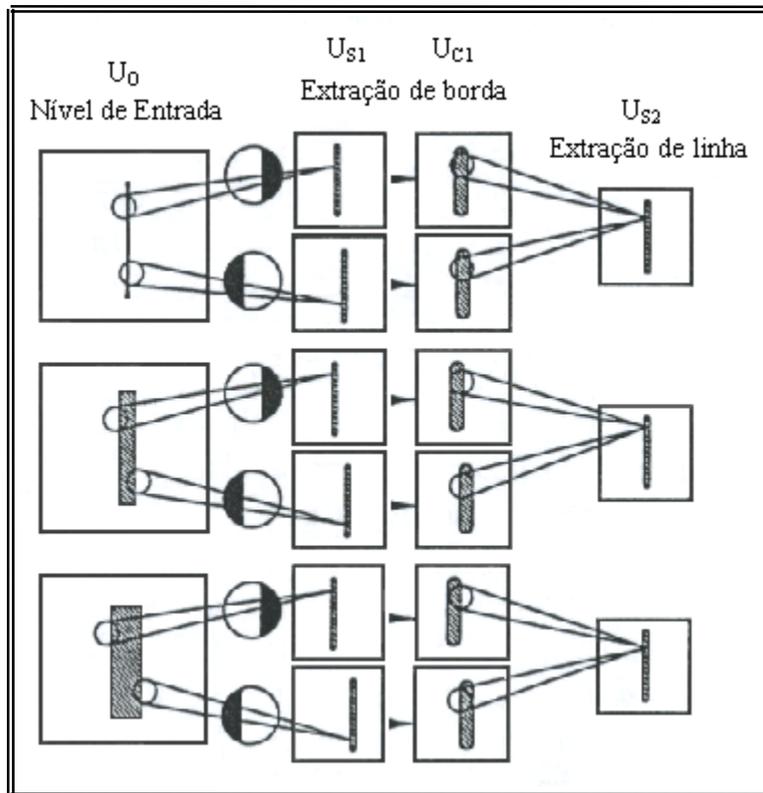


Figura 5.7 – Processo de extração de linhas e bordas.

Uma linha reta com uma orientação particular é usada para treinar um plano de células de  $U_{S2}$ . Somente a apresentação de uma linha de certa largura é necessária para treinar as células para extrair linha de diferentes larguras [Fuk92].

O treinamento utilizado tanto para detecção de bordas quanto para detecção de linhas é supervisionado. Ou seja, as células vencedoras serão selecionadas arbitrariamente por um “professor”.

### 5.5.2 Detecção de Curvatura e Pontos Finais

A camada  $U_{S2}$ , da Figura 5.8, consiste de células de detecção de pontos extremos de segmentos de linha, que correspondem aos pontos de curvatura, quando uma linha é

composta de vários segmentos. Uma célula  $S$  recebe conexões intercelulares de entrada de um plano de célula da camada  $U_{C2}$ . Estas conexões são fixas e tem a distribuição espacial ilustrada na Figura 5.8.

Quando uma linha é apresentada à camada de entrada  $U_0$ , a resposta das células  $S$  aparece em um ponto final de linha. Quando uma linha curvada for apresentada, a resposta das células  $S'$  representa um ponto de curvatura da linha curvada.

A camada  $U_{S3}$  recebe conexões intercelulares de entrada das camadas  $U_{C2}$  e  $U_{C2}'$ , integrando informações sobre a existência de componentes de linha, pontos de curvatura e pontos finais de linhas [Fuk92].

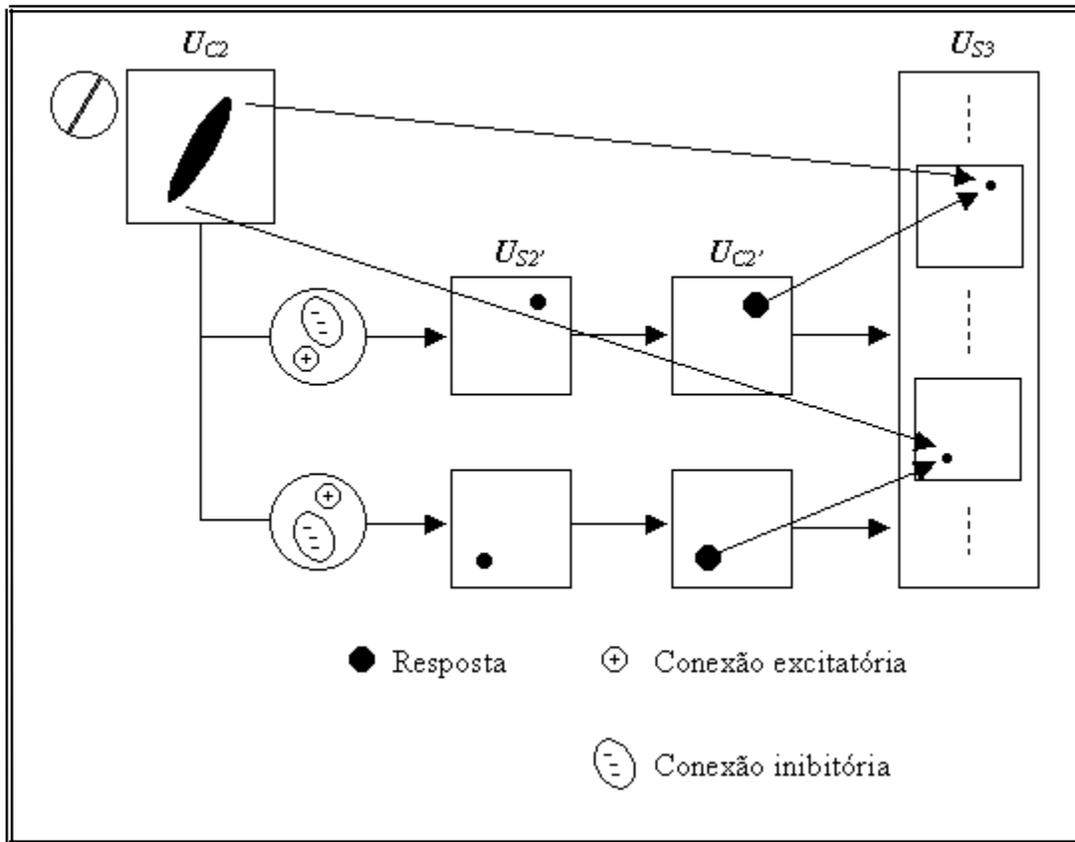


Figura 5.8 – Detecção de pontos extremos de linhas.

## 5.6 Reconhecimento de Numerais Manuscritos

Para demonstrar o funcionamento do *Neocognitron* utilizando um treinamento supervisionado, foi projetado um sistema para reconhecer numerais manuscritos de “0” (zero) a “9” (nove), proposto por Fukushima [Fuk88].

### 5.6.1- Estrutura da Rede

A rede tem quatro estágios de camadas de células *S* e *C*. O número de células *S* e células *C* em cada camada da rede é mostrado na Figura 5.9. Por exemplo, a primeira camada de células *S* contém 12 planos celulares de 19x19 elementos. A camada  $U_{C4}$  no último estágio tem dez planos de células, cada qual somente com uma célula, que corresponde a um dos padrões de “0” a “9”.

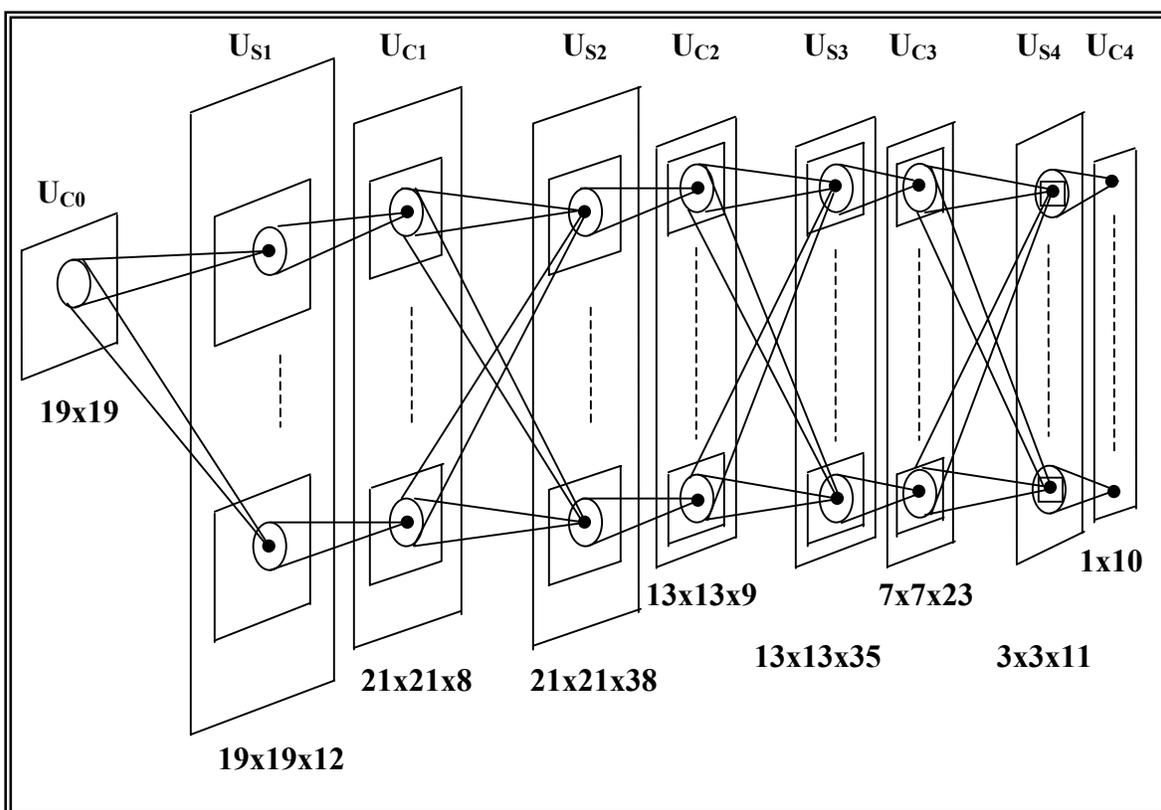


Figura 5.9 – Estrutura da rede para reconhecimento de numerais.

As equações utilizadas no treinamento supervisionado são as mesmas utilizadas no treinamento não supervisionado, porém um “professor” apontará qual célula será a célula vencedora em um determinado plano de célula.

### **5.6.2 Treinamento da Rede**

No treinamento da rede, o “professor” apresentará um conjunto de padrões de treinamentos e indicará qual célula será a célula vencedora para cada padrão de treinamento. Existem quatro etapas de treinamento, correspondente às camadas  $U_{S1}$ ,  $U_{S2}$ ,  $U_{S3}$  e  $U_{S4}$ .

#### **5.6.2.1 Treinamento da Camada $U_{S1}$**

A camada  $U_{S1}$  é treinada para extrair componentes de linha de diferentes orientações. A Figura 5.10 mostra 12 padrões de treinamento usados para treinar os 12 planos celulares da camada  $U_{S1}$ . Essa camada de células tem um campo receptivo 3x3 sendo que a célula central é assumida como a célula vencedora do plano celular. Uma vez que a célula central é a vencedora, somente a área central 3x3 será efetiva para o treinamento conforme é ilustrado na Figura 5.10.

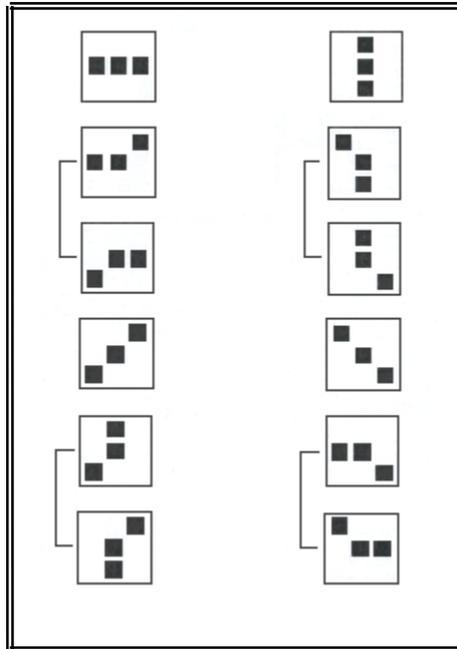


Figura 5.10 – Treinamento utilizado para os 12 planos de célula na camada  $U_{S1}$  [Fuk92].

Um campo receptivo 3x3 é muito pequeno para se extrair todas as partes de uma linha inclinada. Assim, dois planos de células são utilizados para extrair componentes de linha inclinada. As saídas destes dois planos celulares são agrupadas e convergem para um mesmo plano  $C$ . Isso é ilustrado nas linhas verticais de ligação da Figura 5.10.

### 5.6.2.2 Treinamento da Camada $U_{S2}$

A Figura 5.11 mostra os padrões de treinamento usados para treinar os 38 planos de célula da camada  $U_{S2}$ . Uma linha horizontal indica um plano celular. Cada um desses planos celulares pode estar treinando um ou mais padrões que devem ser igualmente classificados. As células desta camada têm um campo receptivo de dimensão 9x9. Novamente, as células centrais do plano de células são escolhidas como células vencedoras. Por isso, somente a área central de dimensão 9x9 dos padrões de treinamento é ilustrada na Figura 5.11. As saídas das células da camada  $U_{S2}$  também são agrupadas, como indicado pelas linhas verticais de ligação.

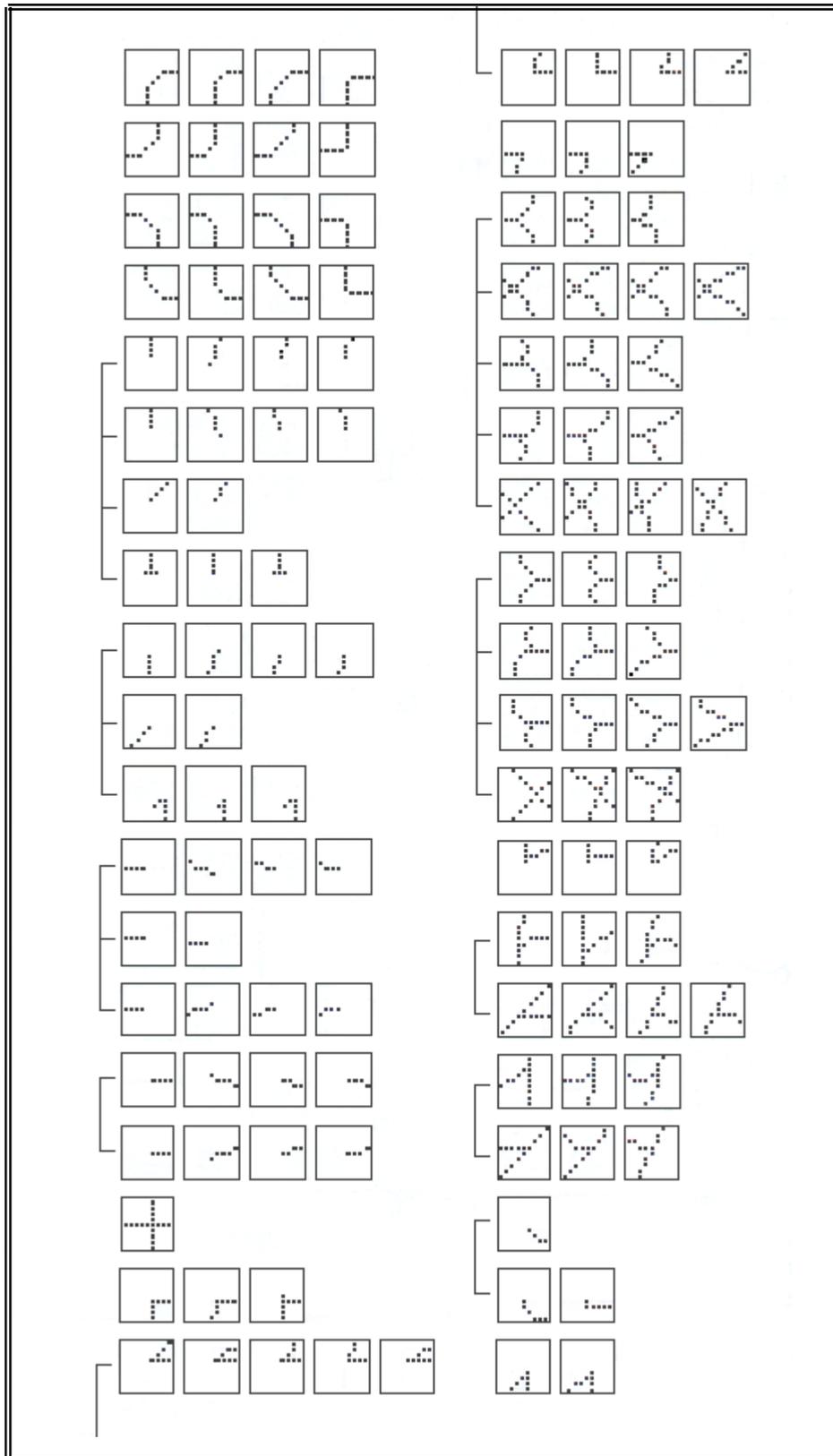


Figura 5.11 – Padrões de treinamento usados para treinar os 38 planos de célula da camada  $U_{S_2}$  [Fuk92].

Como pode ser visto na Figura 5.11, cada padrão de treinamento consiste de uma parte de um número que supostamente aparece durante o processo de reconhecimento. Em outras palavras, são exemplos típicos de padrões deformados que são apresentados à rede como padrões de treinamento. A escolha de padrões de treinamento é mais importante na camada  $U_{S2}$  do que em outras camadas. Caso os padrões sejam bem escolhidos, a rede adquire uma considerável habilidade de reconhecimento de padrões, mesmo que a escolha dos padrões para as outras camadas não seja tão completa.

### **5.6.2.3 Treinamento da camada $U_{S3}$**

A Figura 5.12 mostra os padrões de treinamento utilizados para treinar 35 planos de células da camada  $U_{S3}$ . Desde que os campos receptivos das células  $S$  desta camada sejam bem maiores que os campos receptivos da camada  $U_0$ , as células no centro de um plano de célula não podem sempre ser apontados como células vencedoras. A posição das células vencedoras é marcada por uma cruz em cada padrão de treinamento.

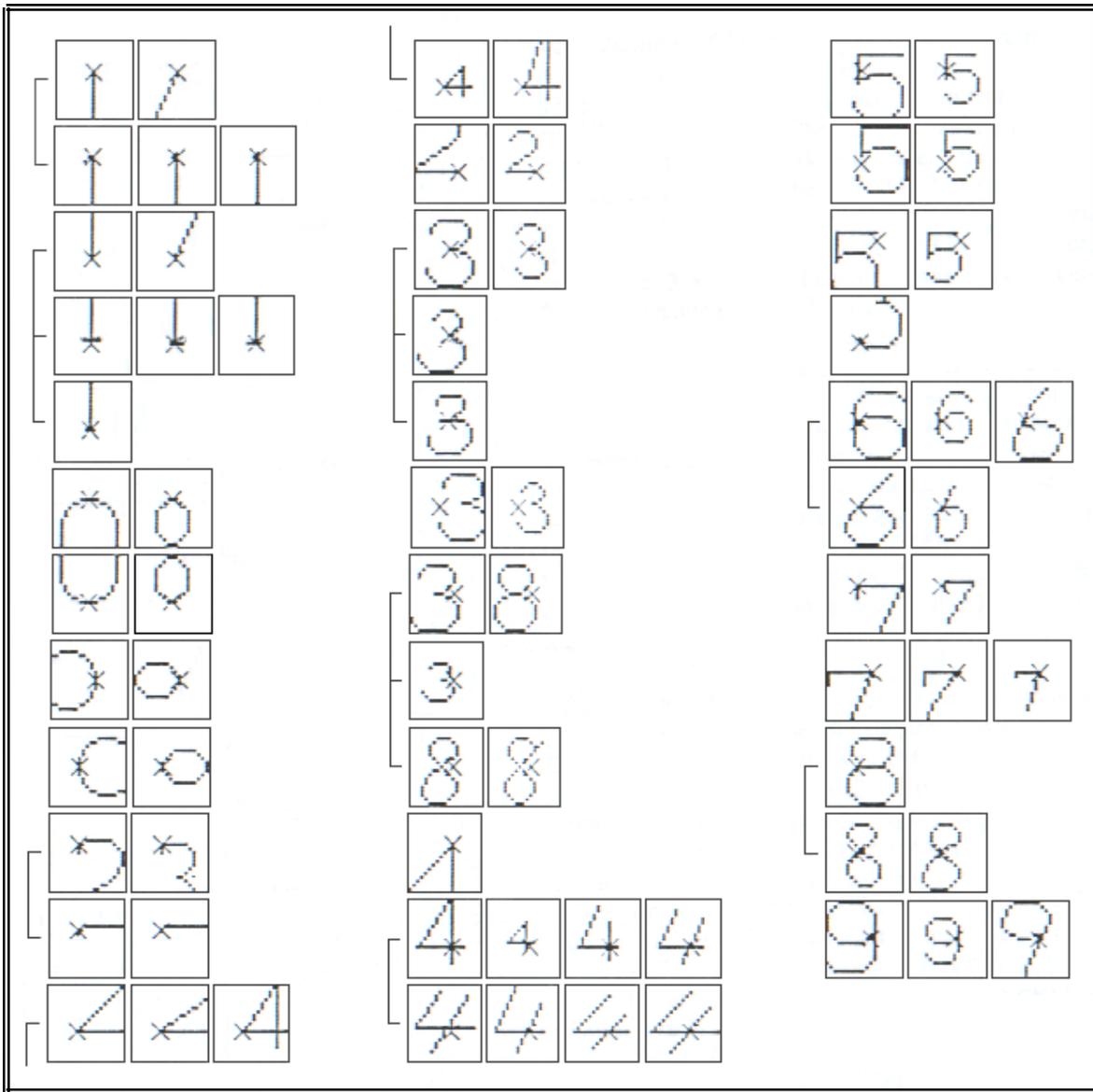


Figura 5.12 – padrões usados para treinar os 35 planos de célula da camada  $U_{S3}$ . O campo receptivo da célula está sendo marcado por linhas cruzadas [Fuk92].

#### 5.6.2.4 Treinamento da camada $U_{S4}$

A Figura 5.13 mostra os padrões de treinamento usados para treinar os 11 planos celulares da camada  $U_{S4}$ . Como em  $U_{S1}$  e  $U_{S2}$ , a célula central do plano será escolhida como célula vencedora. A única observação é que dois planos celulares são utilizados para reconhecer diferentes padrões do número 4, uma vez que algumas distorções possíveis dificultam o seu reconhecimento.

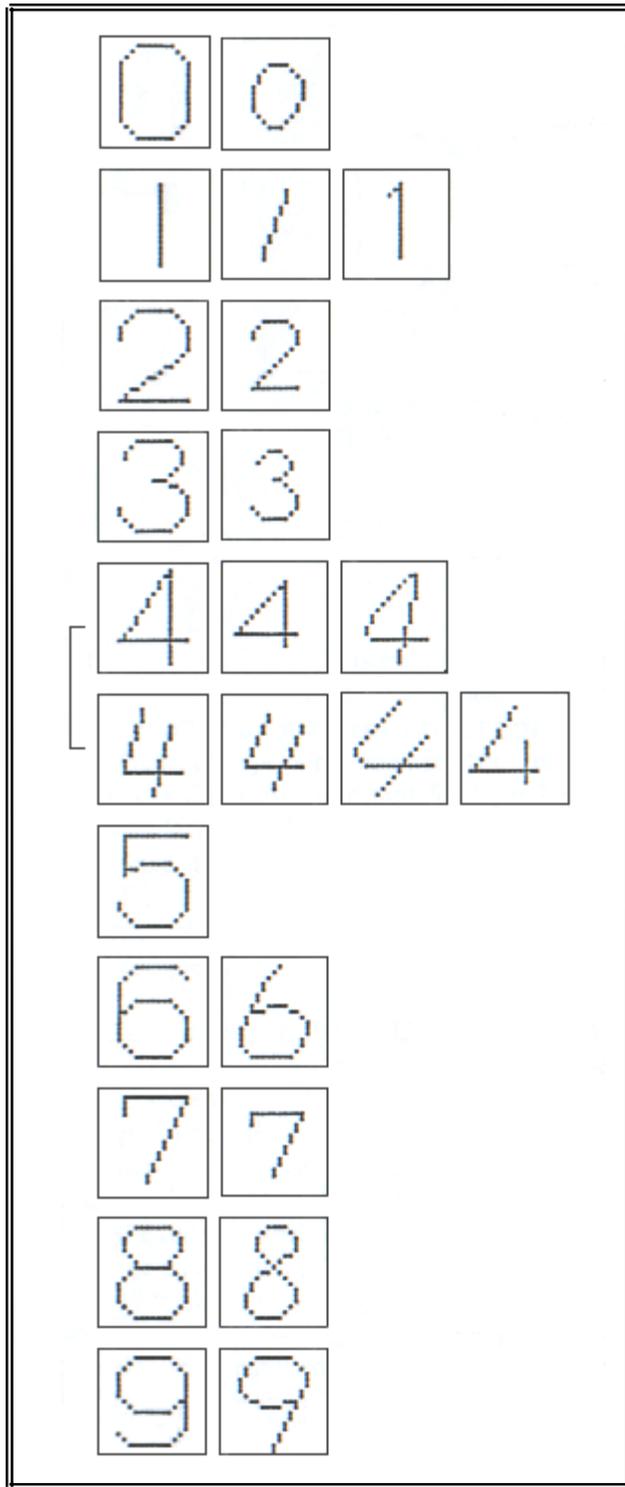


Figura 5.13 – padrões de treinamento usados para treinar os 11 planos celulares da camada

$U_{S4}$  [Fuk92].

## 6 NEOCOGNITRON PARA O RECONHECIMENTO DE ATRIBUTOS FACIAIS (NEORAF)

Neste capítulo é descrito o trabalho de dissertação de mestrado realizado baseado na rede neural *Neocognitron*. Será discutida a estrutura hierárquica do *Neocognitron* para Reconhecimento de Atributos Faciais, enfocando as duas sub-redes neurais que compõem o *NEORAF*: Rede Neural para Detecção de Pontos de Controle (*RNPDC*) e Rede Neural para Reconhecimento Facial (*RNRF*). A seguir é exposta uma análise dos resultados obtidos da etapa de reconhecimento.

O *NEORAF* (*Neocognitron* para Reconhecimento de Atributos Faciais) é um sistema que adapta o *Neocognitron* para realização de reconhecimento de quatro atributos faciais: lábios, nariz, olho direito e olho esquerdo.

A motivação do projeto tem base na capacidade da rede neural *Neocognitron* de tolerância a pequenas distorções no padrão de entrada e o desenvolvimento da rede neural *NEOPC* (*Neocognitron* com Pontos de Controle) [Bia01], um sistema para o reconhecimento facial global.

O *NEOPC* é composto por duas redes neurais: *RNDPC* (Rede Neural para Detecção de Pontos de Controle) e *RNRF* (Rede Neural para Reconhecimento Facial). A primeira etapa do *NEOPC* é o pré-processamento da imagem, seguido da extração de pontos de controle pela rede *RNDPC*, a qual tem três níveis de detecção de pontos de controle: bordas, linhas e pontos extremos. Em seguida, realiza-se a extração de atributos (treinamento) e reconhecimento facial pela rede *RNRF*. A Figura 6.1 ilustra as etapas da rede *NEOPC*.

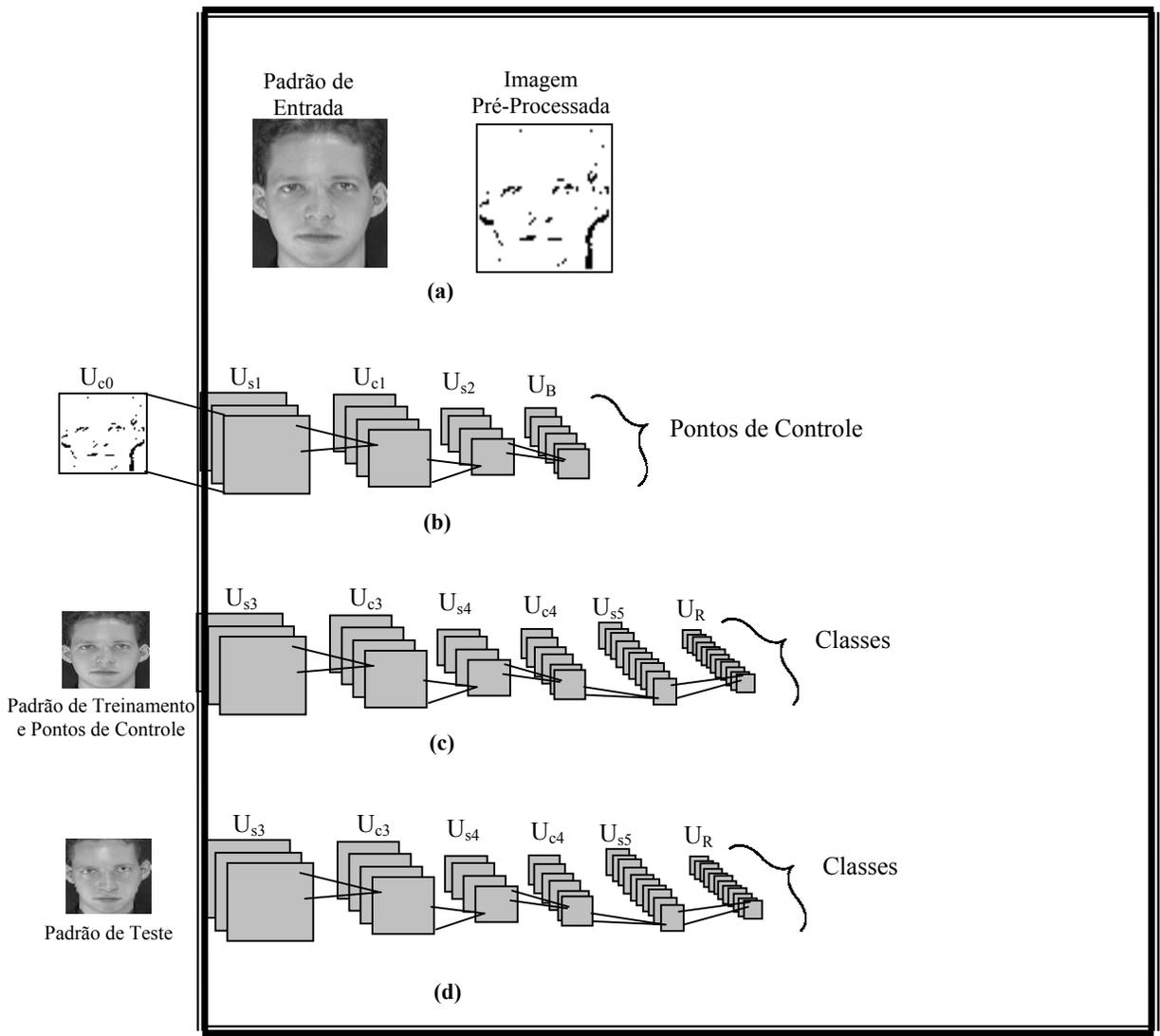


Figura 6.1 - Etapas da rede neural NEOPC. (a) Pré-Processamento, (b) RNDPC: obtenção dos pontos de controle, (c) RNRf: treinamento e (d) RNRf: reconhecimento facial.

O NEOPC analisou imagens de acordo com dois conjuntos distintos:

- Imagens com diferentes poses e expressões faciais. Os testes resultaram em um baixo desempenho, devido à variação causada nos padrões de entrada.
- Imagens com diferentes angulações, com as mesmas poses e expressão facial neutra. Aqui, o desempenho do NEOPC decaiu conforme aumentava o número de classes analisadas.

Um outro aspecto nos testes de desempenho do *NEOPC* foi o número de imagens utilizadas para reconhecimento. Foram utilizadas sempre 5 amostras para o reconhecimento em todos os testes.

O *NEORAF* foi desenvolvido com o objetivo de melhorar os resultados obtidos no *NEOPC*, sendo voltado para o reconhecimento de atributos, ao invés, do reconhecimento global. Nas seções seguintes serão tratados: a base de dados de imagens de atributos faciais utilizada para os testes, as sub-redes neurais envolvidas no processo de reconhecimento de atributos faciais, enfatizando suas estruturas, treinamento e citologia e, por fim serão expostos os testes de desempenho.

## **6.1 Bancos de Imagens**

Para realização de testes e análise de resultados, foi desenvolvido um banco principal com 210 imagens faciais originais de 7 pessoas. As imagens foram obtidas de uma máquina fotográfica digital Kodak 210 A, com resolução 1152x864 pixels. As imagens foram salvas em formato jpeg utilizando compressão através do método de codificação de *Huffman* otimizado. As fotos foram tiradas em três locais diferentes sob diferentes condições de iluminação. A Figura 6.2 ilustra um exemplo de imagem contida no banco principal de imagens.

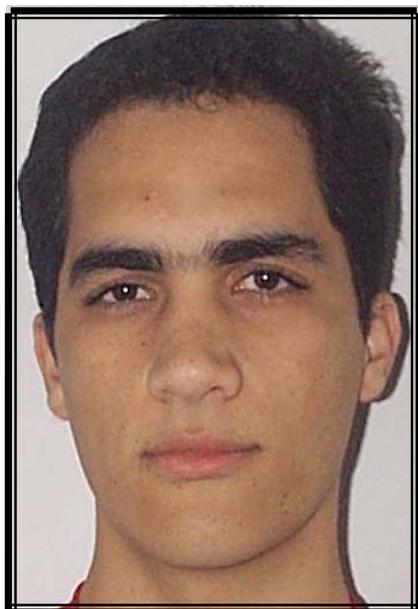


Figura 6.2 – Exemplo de imagem do banco principal de imagens.

## 6.2 Extração de Atributos Faciais

A partir do banco principal de imagens faciais, foi gerado o banco de imagens de atributos faciais. Para a extração dos atributos faciais foi utilizado o editor de imagem *Adobe Photoshop 7.0*, seguindo as seguintes etapas:

- Extração de atributo: são extraídos os atributos faciais (lábios, nariz, olho esquerdo e olho direito) que são utilizados no processo de reconhecimento;
- Redimensionamento da imagem: a imagem do atributo extraído é redimensionada para a resolução 57x57, para que possa ser utilizada pela rede *NEORAF*.

Foram obtidas 210 imagens para cada atributo facial: nariz, olho direito e olho esquerdo. São 30 padrões de atributos para 7 classes de pessoas diferentes. Por fim, foram geradas 180 imagens de lábios, sendo 30 padrões para 6 classes de pessoas. A imagem 6.3 mostra alguns exemplos de imagens prontas para treinamento e reconhecimento.

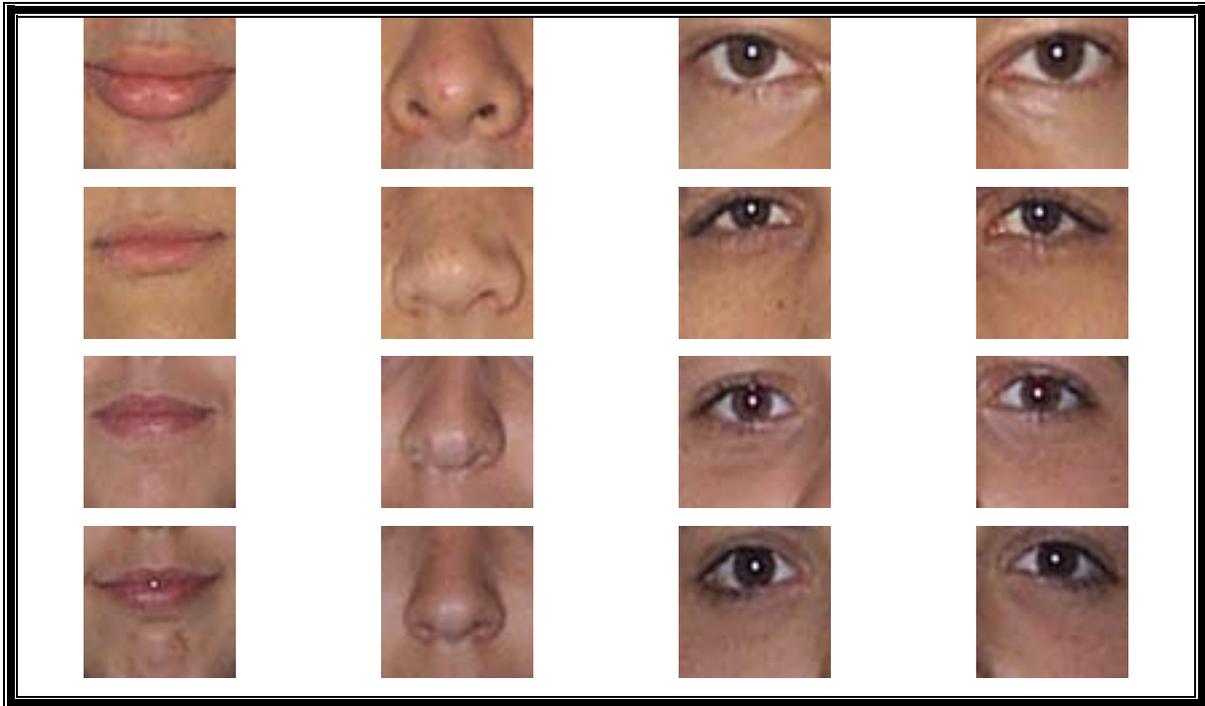


Figura 6.3 – Exemplos de imagens de treinamento e reconhecimento.

### 6.3 Estrutura do *NEORAF*

O *Neocognitron* para Reconhecimento de Atributos Faciais (*NEORAF*) é altamente estruturado e hierárquico, sendo composto por duas sub-redes neurais: Rede Neural para Detecção de Pontos de Controle (*RNDPC*), que realiza a detecção de pontos de controle e Rede Neural para Reconhecimento de Atributos Faciais (*RNRAF*), que realiza o reconhecimento de fatores complexos dos atributos faciais.

#### 6.3.1 Rede Neural para Detecção de Pontos de Controle (*RNDPC*)

A Rede Neural para Detecção de Pontos de Controle pode ser dividida em três sub-redes neurais de acordo com níveis de detecção de pontos de controle: Rede Neural para Detecção de Bordas, Rede Neural para Detecção de Linhas e Rede Neural para Detecção de Pontos Extremos.

A imagem de atributo facial é apresentada à camada de entrada  $U_{C0}$ . A Rede Neural para Detecção de Bordas (*RNDB*) tem a função de detectar 16 tipos possíveis de bordas, como mostrado na Figura 6.4. Assim, existem 16 planos  $U_{S1}$ , cada um responsável por tipo particular de borda. A Rede Neural para Detecção de Linhas (*RNDL*) tenta detectar 8 possíveis tipos de linhas, em diferentes direções, como pode ser visto na Figura 6.5. Por isso, existem 8 planos  $U_{S2}$ , cada um responsável por um padrão de linha. A partir da detecção de bordas e linhas, pode ser realizada a detecção de pontos extremos pela Rede Neural para Detecção de Pontos Extremos de Linhas (*RNDPEL*), que tem a função de detectar 16 possíveis tipos de pontos extremos, como pode ser visto na Figura 6.6. A Figura 6.7 ilustra o esquema completo da *RNDPC*, que possui 16 planos  $U_{S2'}$  e  $U_{C2'}$  responsáveis pela detecção de pontos extremos. A tabela 6.1 ilustra a relação entre níveis, número de planos por camada e área de conexão (conexão intercelular) de células de uma determinada camada.

Tabela 6.1 – Dados da *RNDPC*.

<b>Camada</b>	<b>Dimensão do plano</b>	<b>Área de Conexão</b>	<b>Número de planos celulares</b>
$U_{S1}$	57x57	5x5	16
$U_{C1}$	57x57	5x5	16
$U_{S2}$	57x57	5x5	8
$U_{C2}$	21x21	5x5	8
$U_{S2'}$	21x21	5x5	16
$U_{C2'}$	21x21	5x5	16

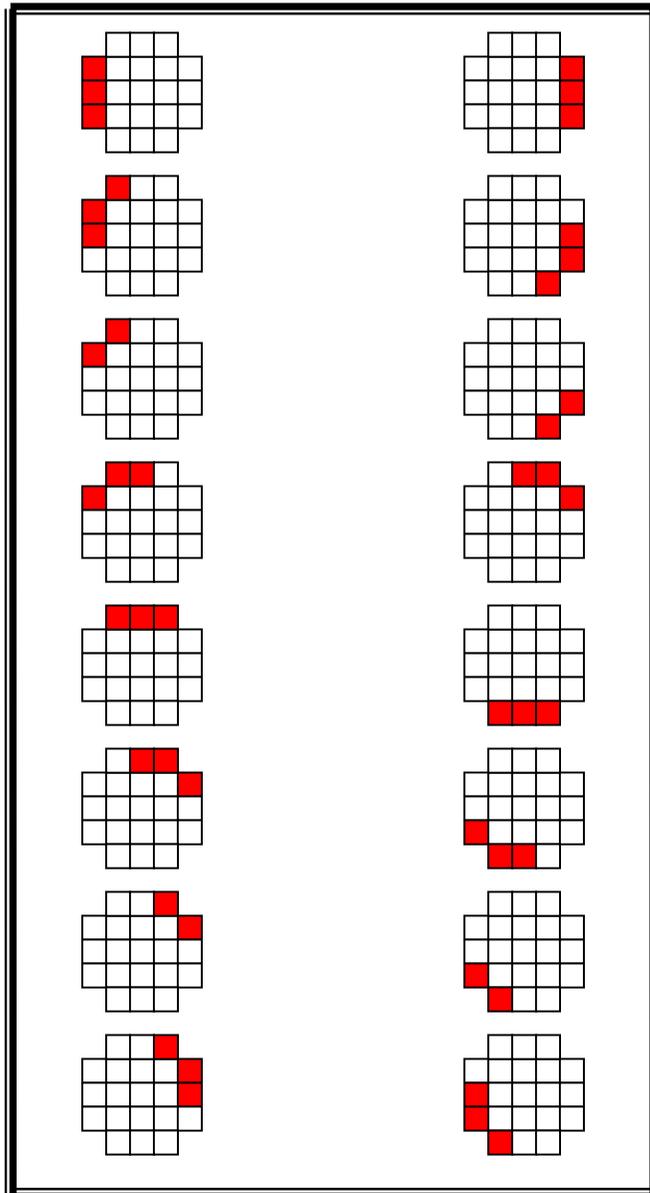


Figura 6.4 – padrões de bordas a serem detectados pela *RNDB*.

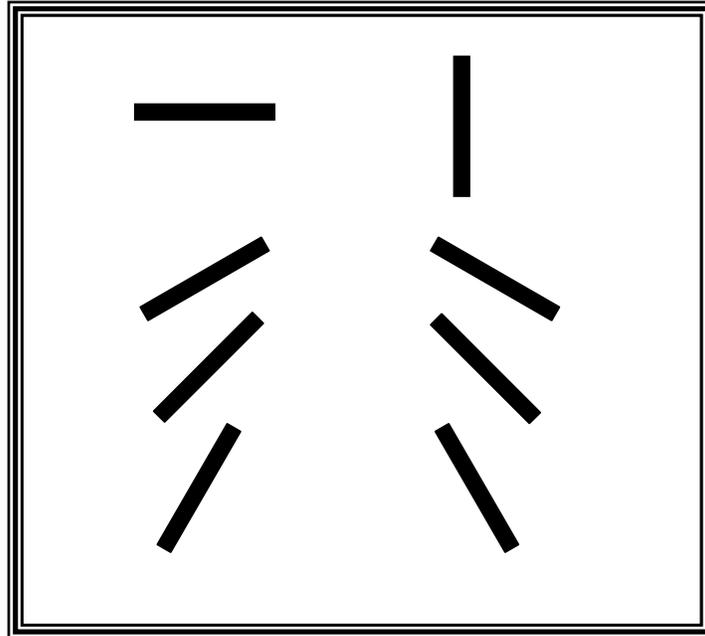


Figura 6.5 – padrões de linhas a serem detectados pela *RNDB*.

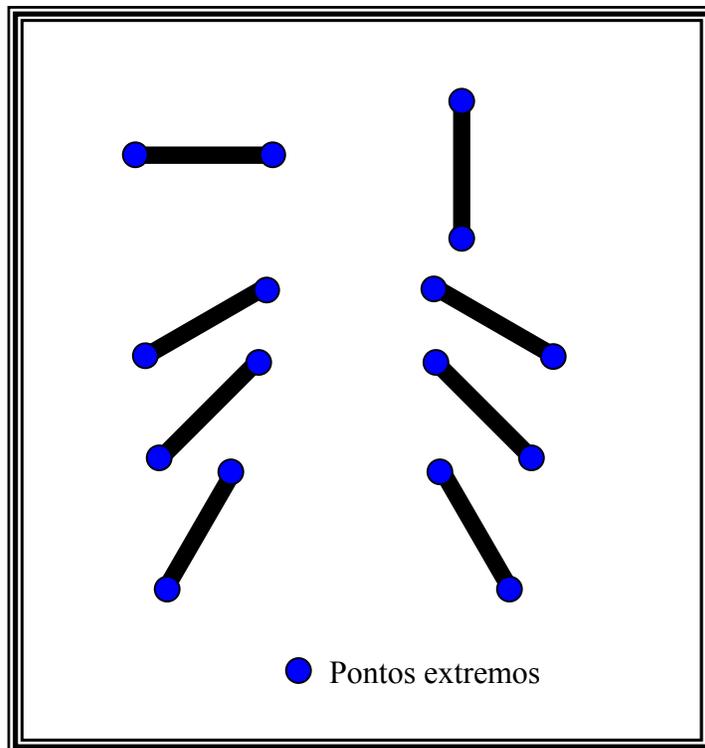


Figura 6.6 – padrões de pontos extremos a serem detectados pela *RNDPEL*.

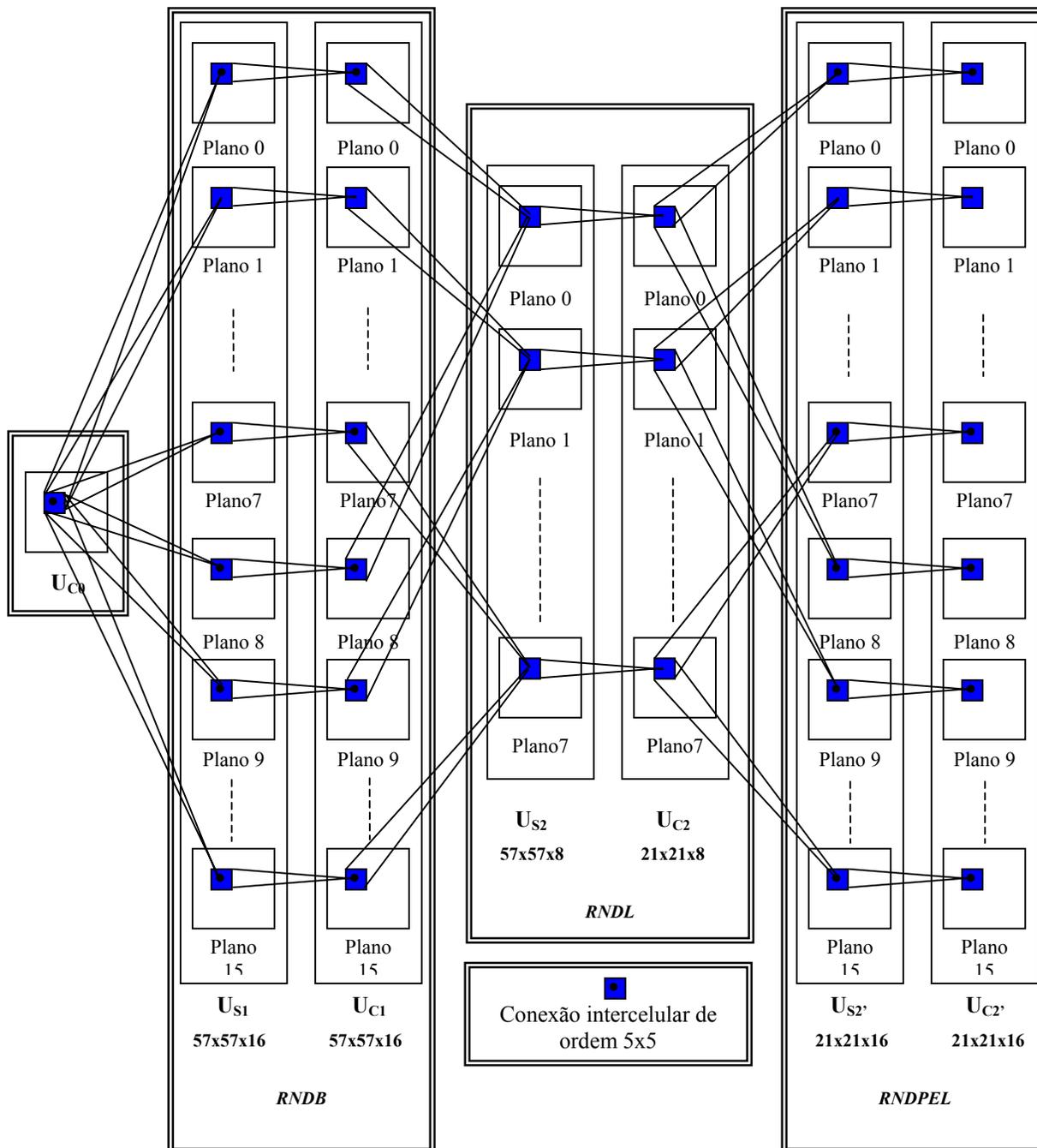


Figura 6.7 – Esquema completo da RNDPC.

A RNDPC recebe um atributo facial como padrão de entrada e realiza a detecção de pontos de controle, passo a passo. Foi utilizado o mesmo tamanho para a área de conexão intercelular entre dois níveis, um quadrado de ordem 5x5, como indicado na Figura 6.7. O

treinamento é supervisionado nas três redes. Na *RNDB* e *RNDL* o “professor” pré-estabelece que plano deve extrair determinado tipo de borda ou determinado tipo de linha, como indicado nas Figuras 6.4 e 6.5. O mesmo acontece na *RNDPEL*, como indicado na Figura 6.6. A Figura 6.8 mostra a interface gráfica do funcionamento da *RNDPC*. Os pontos de controle serão armazenados em um arquivo externo para uso posterior pela *RNRAF*.

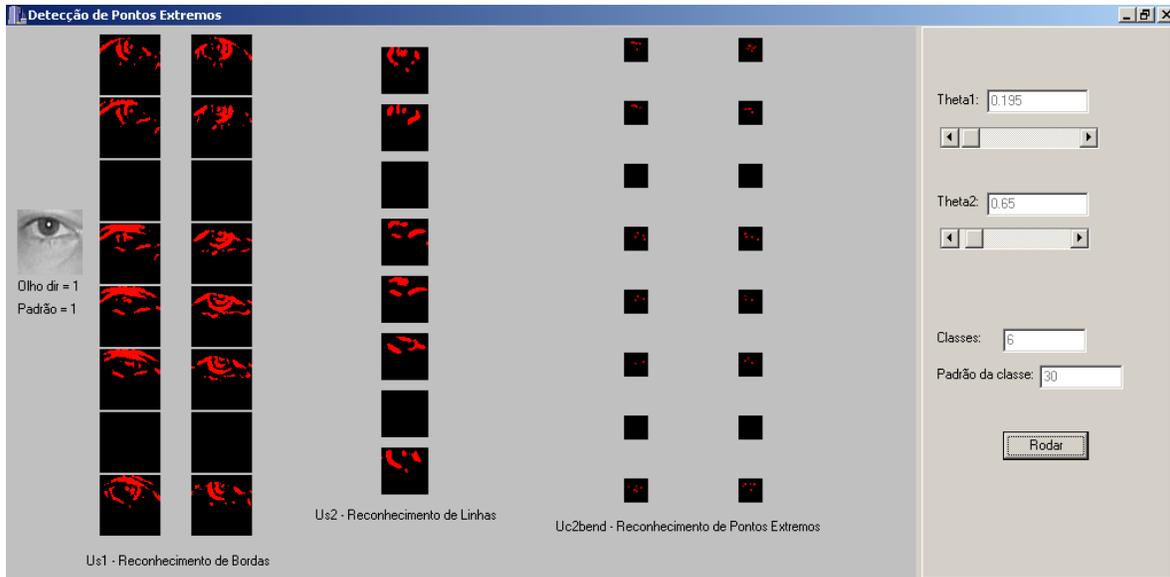


Figura 6.8 – Execução da *RNDPC*.

### 6.3.2 Rede Neural para Reconhecimento de Atributos Faciais (*RNRAF*)

A Rede Neural para Reconhecimento de Atributos Faciais (*RNRAF*) é composta pelos níveis  $U_{S3}$ ,  $U_{C3}$ ,  $U_{S4}$ ,  $U_{C4}$ ,  $U_{S5}$  e  $U_{C5}$ , onde a entrada consiste do padrão de treinamento e dos pontos de controle obtidos pela rede *RNDPC*. A entrada consiste em uma imagem de atributo facial com dimensão 57x57. A Figura 6.9 ilustra a estrutura do *RNRAF*. Embora não seja mostrado, a rede possui também células inibidoras  $V_C$  na camada  $C$ . Porém, diferindo do *Neocognitron* original, não possui células inibidoras  $V_S$  na camada  $S$ .

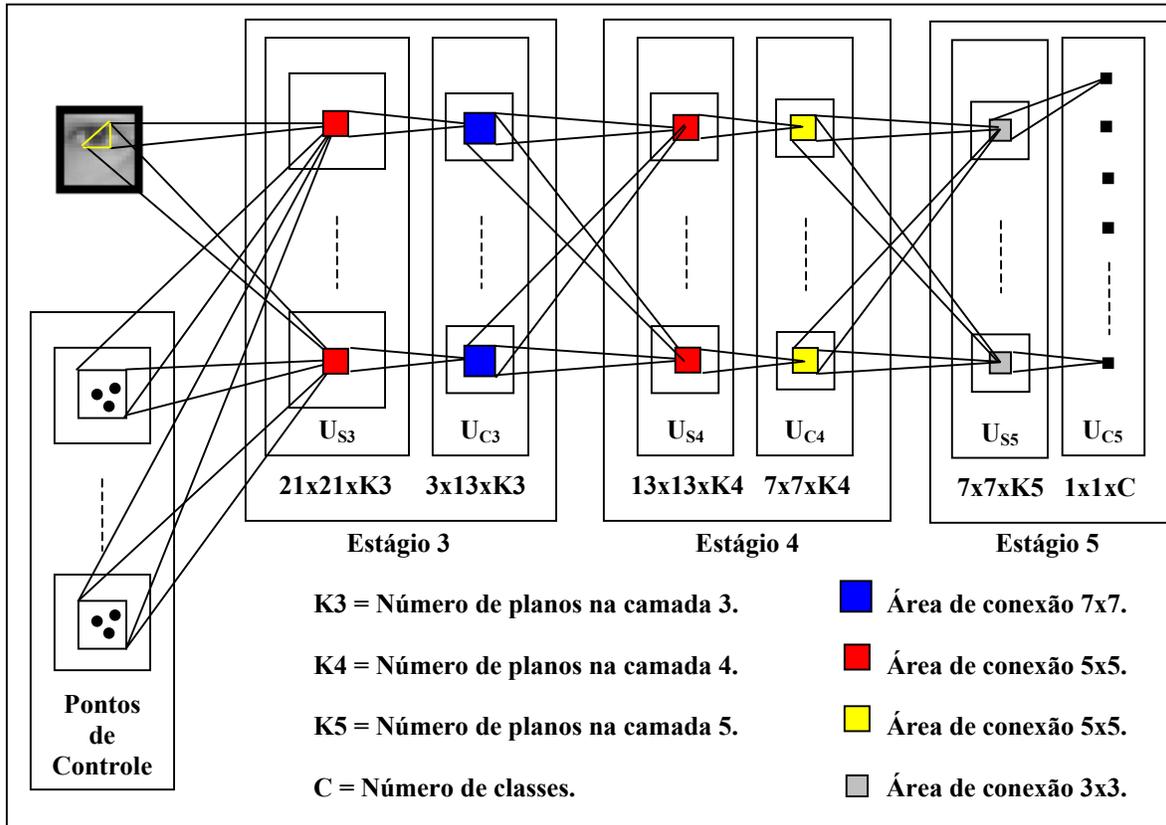


Figura 6.9 – Estrutura do RNRAF.

Considerando que  $L$  represente o estágio da rede,  $a_L$ , a matriz de coeficiente de conexão excitatória de entrada para células na camada  $U_{SL}$ ,  $d_L$ , a matriz de coeficiente de conexão excitatória de entrada para células na camada  $U_{CL}$  e  $K_L$  o total de planos da camada  $U_{SL}$ , pode-se montar a tabela 6.2.

Tabela 6.2 – Dados da *RNRAF*.

Camada ( $U_{XL}$ )	Dimensão do plano	Área de Conexão	Número de planos celulares ( $K_L$ )
$U_{S3}$	21x21	a3: 7x7	$K_3$
$U_{C3}$	13x13	d3: 5x5	$K_3$
$U_{S4}$	13x13	a4: 7x7	$K_4$
$U_{C4}$	7x7	d4: 5x5	$K_4$
$U_{S5}$	7x7	a5: 5x5	$K_5$
$U_{C5}$	1x1	d5: 3x3	C

### 6.3.2.1 Treinamento Não-Supervisionado da *RNRAF*

O treinamento é não-supervisionado, onde imagens de atributos faciais são apresentadas como padrão de entrada para a camada  $U_{S3}$  para realização do processo de auto-organização. Imagens de atributos faciais são apresentadas, junto com informações sobre pontos de controle obtidas no treinamento da *RNDPC*. Com essas informações, a camada  $U_{S3}$  irá extrair características do padrão de entrada. Toda vez que uma característica for extraída, um novo plano deve ser criado. Assim, o número de planos na rede é variável, dependendo da extração de características.

### 6.3.2.2 Plano de Seleção de Semente

Para extrair atributos de imagem de entrada é utilizado um plano de seleção de semente (*seed selection cell*) nos níveis S. Esse plano tem suas conexões intercelulares com valores pequenos e diferentes. Desse modo, se alguma célula do plano de seleção de semente responder a um determinado padrão de entrada e nenhuma outra célula de algum plano S dessa camada responder a esse padrão de entrada, um plano S será criado e terá suas conexões intercelulares de entrada reforçadas para detectar esse novo atributo. A Figura 6.10 exibe o processo de treinamento da *RNRAF*.



Figura 6.10 – Execução da *RNRAF*.

### 6.3.2.3 Citologia da *RNRAF*

Como visto, a *RNRAF* tem quase a mesma estrutura citológica do *Neocognitron*, diferindo apenas na ausência de células inibitórias  $V_S$ . Sejam  $u_{Cl-1}(k_{l-1}, \mathbf{n} + i)$  as entradas excitatórias,  $v_{Cl-1}(\mathbf{n})$ , a entrada inibitória,  $a_l(k_{l-1}, i, k_l)$ , os coeficientes de conexão excitatória e  $b_l(k_l)$  o coeficiente conexão inibitória, a saída  $u_{Sl}(k_l, \mathbf{n})$  da célula  $S$  é definida como:

$$u_{Sl}(k_l, \mathbf{n}) = \left( \frac{\alpha}{1 - \theta} \right) \cdot \Phi \left[ \frac{1 + \sum_{k_{l-1}=1}^{K_{l-1}} \sum_{i \in S_l} a_l(k_{l-1}, i, k_l) \cdot u_{Cl-1}(k_{l-1}, \mathbf{n} + i)}{1 + \theta \cdot b_l(k_l) \cdot v_{Cl-1}(\mathbf{n})} - 1 \right], \quad (6.1).$$

onde a função  $\varphi[x]$  e o parâmetro  $\alpha$  são definidos pelas fórmulas:

$$\varphi[x] = \begin{cases} x & \text{se } x \geq 0 \\ 0 & \text{se } x < 0 \end{cases}, \quad (6.2).$$

$$\alpha = \frac{\theta \cdot b_l(k_l) \cdot v_{Cl-1}(\mathbf{n})}{1 + \theta \cdot b_l(k_l) \cdot v_{Cl-1}(\mathbf{n})}, \quad (6.3)$$

A célula  $v_{Cl-1}(k_{l-1}, \mathbf{n})$ , que está enviando um sinal inibidor para a célula  $u_{Sl}(k_l, \mathbf{n})$ , recebe suas conexões intercelulares de entrada das mesmas células que  $u_{Sl}(k_l, \mathbf{n})$ . Sendo o parâmetro  $c_{l-1}$ , as conexões intercelulares fixas de entrada para a célula  $v_{Cl-1}(\mathbf{n})$  e  $S_l$  a área de conexão intercelular de entrada para a célula  $v_{Cl-1}(\mathbf{n})$ , a sua saída da célula  $v_{Cl-1}(k_{l-1}, \mathbf{n})$  é dada por:

$$v_{Cl-1}(\mathbf{n}) = \sqrt{\sum_{k_{l-1}=1}^{K_{l-1}} \sum_{i \in S_l} c_{l-1}(i) \cdot u^2_{Cl-1}(k_{l-1}, \mathbf{n} + i)}, \quad (6.4)$$

As células  $C$  recebem conexões de entrada de células  $S$  e enviam sinais para as células  $V_C$  e  $S$ . As conexões entre as células  $S$  e  $C$  são excitatórias. Sejam  $u_{Sl}(k_l, \mathbf{n} + i)$  as entradas excitatórias, a saída  $u_{Cl}(k_l, \mathbf{n})$  é definida como:

$$u_{Cl}(k_l, \mathbf{n}) = \frac{\sum_{i \in D_l} d_l(i) \cdot u_{Sl}(k_l, \mathbf{n} + i)}{1 + u_{Sl}(k_l, \mathbf{n} + i)}, \quad (6.5)$$

onde  $d_l(i)$  representa os coeficientes de conexão intercelular excitatória para uma célula  $u_{Cl}(k_l, \mathbf{n})$  e  $D_l$  representa a área de conexão intercelular de entrada para uma determinada célula  $u_{Cl}(k_l, \mathbf{n})$ .

## 6.4 Testes de Desempenho

Conforme foi descrito anteriormente, as imagens utilizadas para teste foram obtidas com uma câmera digital (Kodak 210 A) e sob mesma condição de distância. Foram utilizados três lugares diferentes, sob diferentes condições de iluminação. As fotos foram obtidas com diferenças sutis de angulação e posição do rosto, gerando pequenos deslocamentos e distorção do padrão de entrada. Pequenas diferenças de expressões faciais foram utilizadas, tais como um sorriso leve, expressão de susto e mudança do ponto de vista do fotografado. Os testes foram realizados utilizando imagens diferentes para treinamento e reconhecimento, sendo 15 tanto para treinamento, quanto para reconhecimento. Foram realizados testes com três atributos faciais: nariz, lábios e olho direito. Os limiares utilizados para o reconhecimento de cada atributo estão indicados na tabelas 6.3 – 6.5. Como pode ser visto, foram utilizados valores reais com 4 casas decimais de profundidade para aumentar a precisão da rede.

Tabela 6.3 – Tabela de limiares para o nariz.

<i><b>RNDPC</b></i>					
<b>Bordas</b>			<b>Linhas</b>		
<i>U<sub>S1</sub></i>			<i>U<sub>S2</sub></i>		
<b>0.1950</b>			<b>0.6105</b>		
<i><b>RNRAF</b></i>					
<b>Treinamento</b>					
<i>U<sub>S3</sub></i>	<i>U<sub>S4</sub></i>	<i>U<sub>S5</sub></i>	<i>U<sub>C3</sub></i>	<i>U<sub>C4</sub></i>	<i>U<sub>C5</sub></i>
<b>0.9985</b>	<b>0.8300</b>	<b>0.7400</b>	<b>0.9985</b>	<b>0.8200</b>	<b>0.7400</b>
<b>Reconhecimento</b>					
<i>U<sub>S3</sub></i>	<i>U<sub>S4</sub></i>	<i>U<sub>S5</sub></i>	<i>U<sub>C3</sub></i>	<i>U<sub>C4</sub></i>	<i>U<sub>C5</sub></i>
<b>0.9985</b>	<b>0.7400</b>	<b>0.5000</b>	<b>0.9985</b>	<b>0.7400</b>	<b>0.5000</b>

Tabela 6.4 – Tabela de limiares para os lábios.

<i>RNDPC</i>					
<b>Bordas</b>			<b>Linhas</b>		
$U_{S1}$			$U_{S2}$		
0.1949			0.5900		
<i>RNRAF</i>					
<b>Treinamento</b>					
$U_{S3}$	$U_{S4}$	$U_{S5}$	$U_{C3}$	$U_{C4}$	$U_{C5}$
0.9986	0.8900	0.7800	0.9986	0.8600	0.7800
<b>Reconhecimento</b>					
$U_{S3}$	$U_{S4}$	$U_{S5}$	$U_{C3}$	$U_{C4}$	$U_{C5}$
0.9980	0.8000	0.5000	0.9980	0.8000	0.5000

Tabela 6.5 – Tabela de limiares para o olho direito.

<i>RNDPC</i>					
<b>Bordas</b>			<b>Linhas</b>		
$U_{S1}$			$U_{S2}$		
0.1950			0.6500		
<i>RNRAF</i>					
<b>Treinamento</b>					
$U_{S3}$	$U_{S4}$	$U_{S5}$	$U_{C3}$	$U_{C4}$	$U_{C5}$
0.9955	0.9200	0.8500	0.9955	0.9100	0.8500
<b>Reconhecimento</b>					
$U_{S3}$	$U_{S4}$	$U_{S5}$	$U_{C3}$	$U_{C4}$	$U_{C5}$
0.9955	0.8900	0.5000	0.9955	0.8900	0.5000

#### 6.4.1 Teste com Três Classes

No primeiro conjunto de testes foram analisados os atributos nariz, lábios e olho direito. Como são utilizadas 15 imagens para cada classe, o teste conta com 45 imagens

tanto para treinamento, quanto para reconhecimento. Foram utilizadas duas casas decimais na obtenção dos índices.

A tabela 6.6 mostra a taxa de reconhecimento, que ficou na faixa de 95.56% de acerto. A tabela 6.7, por sua vez, mostra a matriz de confusão, onde as linhas indicam os padrões de entrada e as colunas indicam as classificações feitas pela rede, para esse padrão de entrada. Considere a linha 1, representando a classe 1. A coluna 1 com valor 13, indica 13 classificações corretas. A coluna 2 tem o valor 0, indicando que nenhum padrão da classe 1 foi classificado como pertencente à classe 2. A coluna 3 tem o valor 2, indicando que dois padrões foram classificados com pertencente à classe 3. Por fim, a coluna rejeição, com valor 0, indica que todos padrões da classe 1 foram classificados. Ou seja, nenhum padrão foi rejeitado (não houve abstenção de classificação).

Tabela 6.6 – Taxa de reconhecimento nasal com 3 classes de testes.

<b>Acertos</b>	<b>43</b>	<b>Erros</b>	<b>2</b>	<b>Rejeições</b>	<b>0</b>
<b>Taxa de acerto</b>	<b>95.56%</b>	<b>Taxa de erro</b>	<b>4.44%</b>	<b>Taxa de rejeição</b>	<b>0%</b>

Tabela 6.7 – Tabela de confusão nasal com 3 classes de testes.

<b>Classe/Classificação</b>	<b>1</b>	<b>2</b>	<b>3</b>	<b>Rejeição</b>
<b>1</b>	13	0	2	0
<b>2</b>	0	15	0	0
<b>3</b>	0	0	15	0

Com relação ao reconhecimento labial, obteve-se uma taxa de acerto de 93.93% e uma taxa de rejeição de 6.67%, não ocorrendo erros, indicado na tabela 6.8. A tabela 6.9 indica que as rejeições ocorreram em padrões pertencentes à classe 3.

Tabela 6.8 – Taxa de reconhecimento labial com 3 classes de testes.

<b>Acertos</b>	<b>42</b>	<b>Erros</b>	<b>0</b>	<b>Rejeições</b>	<b>3</b>
<b>Taxa de acerto</b>	<b>93.33%</b>	<b>Taxa de erro</b>	<b>0%</b>	<b>Taxa de rejeição</b>	<b>6.67%</b>

Tabela 6.9 – Tabela de confusão labial com 3 classes de testes.

<b>Classe/Classificação</b>	<b>1</b>	<b>2</b>	<b>3</b>	<b>Rejeição</b>
<b>1</b>	15	0	0	0
<b>2</b>	0	15	0	0
<b>3</b>	0	0	12	3

Com relação ao reconhecimento de olho direito, obteve-se uma taxa de acerto de 86.67% e uma taxa de erro de 13.33%, como indicado na tabela 6.10. Como se pode ver na tabela 6.11, não ocorreram rejeições.

Tabela 6.10 – Taxa de reconhecimento de olho direito com 3 classes de testes.

<b>Acertos</b>	<b>39</b>	<b>Erros</b>	<b>6</b>	<b>Rejeições</b>	<b>0</b>
<b>Taxa de acerto</b>	<b>86.67%</b>	<b>Taxa de erro</b>	<b>13.33%</b>	<b>Taxa de rejeição</b>	<b>0%</b>

Tabela 6.11 – Tabela de confusão de olho direito com 3 classes de testes.

<b>Classe/Classificação</b>	<b>1</b>	<b>2</b>	<b>3</b>	<b>Rejeição</b>
<b>1</b>	13	0	2	0
<b>2</b>	0	15	0	0
<b>3</b>	1	3	11	0

#### 6.4.2 Teste com Quatro Classes

A realização dos testes de desempenho se deram com 4 classes de atributos faciais, totalizando 60 imagens para treinamento e 60 para reconhecimento. Para o reconhecimento nasal, a tabela 6.12 indica uma taxa de acerto de 93.33% contra uma taxa de erro de 5%. Houve ainda uma rejeição, totalizando 1.67%, a qual ocorreu com um padrão da classe 3 como indicado na tabela 6.13.

Tabela 6.12 – Taxa de reconhecimento nasal com 4 classes de testes.

<b>Acertos</b>	<b>56</b>	<b>Erros</b>	<b>3</b>	<b>Rejeições</b>	<b>1</b>
<b>Taxa de acerto</b>	<b>93.33%</b>	<b>Taxa de erro</b>	<b>5.0%</b>	<b>Taxa de rejeição</b>	<b>1.67%</b>

Tabela 6.13 – Tabela de confusão nasal com 4 classes de testes.

<b>Classe/Classificação</b>	<b>1</b>	<b>2</b>	<b>3</b>	<b>4</b>	<b>Rejeição</b>
<b>1</b>	15	0	0	0	0
<b>2</b>	0	15	0	0	0
<b>3</b>	0	0	12	2	1
<b>4</b>	0	0	1	14	0

Com relação ao reconhecimento labial, obteve-se uma taxa de acerto de 95% e uma taxa de erro de 5%, o que pode ser visto na tabela 6.14. Todos os erros ocorreram em padrões da classe 4, como indicado na tabela 6.15.

Tabela 6.14 – Taxa de reconhecimento labial com 4 classes de testes.

<b>Acertos</b>	<b>56</b>	<b>Erros</b>	<b>4</b>	<b>Rejeições</b>	<b>0</b>
<b>Taxa de acerto</b>	<b>93.33%</b>	<b>Taxa de erro</b>	<b>6.67%</b>	<b>Taxa de rejeição</b>	<b>0%</b>

Tabela 6.15 – Tabela de confusão labial com 4 classes de testes.

<b>Classe/Classificação</b>	<b>1</b>	<b>2</b>	<b>3</b>	<b>4</b>	<b>Rejeição</b>
<b>1</b>	15	0	0	0	0
<b>2</b>	0	15	0	0	0
<b>3</b>	0	0	15	0	0
<b>4</b>	0	0	4	11	0

Com relação ao reconhecimento de olho direito, obteve-se uma taxa de acerto de 85% contra uma taxa de erro de 15%, indicado tabela 6.16. A classe 3 foi a mais afetada,

como indicado na tabela 6.17. Isso se deve ao fato da classe 3 possuir uma variedade maior de expressões faciais no seu conjunto de imagens para reconhecimento.

Tabela 6.16 – Taxa de reconhecimento de olho direito com 4 classes de testes.

<b>Acertos</b>	<b>51</b>	<b>Erros</b>	<b>9</b>	<b>Rejeições</b>	<b>0</b>
<b>Taxa de acerto</b>	<b>85.0%</b>	<b>Taxa de erro</b>	<b>15.0%</b>	<b>Taxa de rejeição</b>	<b>0%</b>

Tabela 6.17 – Tabela de confusão de olho direito com 4 classes de testes.

<b>Classe/Classificação</b>	<b>1</b>	<b>2</b>	<b>3</b>	<b>4</b>	<b>Rejeição</b>
<b>1</b>	13	2	0	0	0
<b>2</b>	0	15	0	0	0
<b>3</b>	1	6	8	0	0
<b>4</b>	0	0	0	15	0

### 6.4.3 Teste com Cinco Classes

Os testes de desempenho foram realizados utilizando 5 classes de atributos, totalizando 75 imagens para treinamento e 75 para reconhecimento.

No reconhecimento nasal, obteve-se 89.33% de acerto contra 8% de erro, como pode ser visto na tabela 6.18. Ocorreram duas abstenções. A tabela 6.19 ilustra melhor a distribuição de classificação dos padrões de entrada.

Tabela 6.18 – Taxa de reconhecimento nasal com 5 classes de testes.

<b>Acertos</b>	<b>67</b>	<b>Erros</b>	<b>6</b>	<b>Rejeições</b>	<b>2</b>
<b>Taxa de acerto</b>	<b>89.33%</b>	<b>Taxa de erro</b>	<b>8%</b>	<b>Taxa de rejeição</b>	<b>2.67%</b>

Tabela 6.19 – Tabela de confusão nasal com 5 classes de testes.

<b>Classe/Classificação</b>	<b>1</b>	<b>2</b>	<b>3</b>	<b>4</b>	<b>5</b>	<b>Rejeição</b>
1	15	0	0	0	0	0
2	0	15	0	0	0	0
3	0	0	11	3	0	1
4	0	0	2	13	0	0
5	0	1	0	0	13	1

Com relação ao reconhecimento de lábios, obteve-se uma taxa de acerto de 94.67% e uma taxa de erro de 5.33%, não ocorrendo rejeição, como ilustrado na tabela 6.20. A tabela 6.21 indica que ocorreram erros na classificação de dois padrões pertencentes à classe 4 e na classificação de dois padrões pertencentes à classe 5. Esses padrões foram classificados como pertencentes à classe 3.

Tabela 6.20 – Taxa de reconhecimento labial com 5 classes de testes.

<b>Acertos</b>	<b>71</b>	<b>Erros</b>	<b>4</b>	<b>Rejeições</b>	<b>0</b>
<b>Taxa de acerto</b>	<b>94.67%</b>	<b>Taxa de erro</b>	<b>5.33%</b>	<b>Taxa de rejeição</b>	<b>0%</b>

Tabela 6.21 – Tabela de confusão labial com 5 classes de testes.

<b>Classe/Classificação</b>	<b>1</b>	<b>2</b>	<b>3</b>	<b>4</b>	<b>5</b>	<b>Rejeição</b>
1	15	0	0	0	0	0
2	0	15	0	0	0	0
3	0	0	15	0	0	0
4	0	0	2	13	0	0
5	0	0	2	0	13	0

Com relação ao reconhecimento de olho direito, o sistema obteve uma taxa de acerto de 89.33% contra uma taxa de erro de 4.0%. As rejeições somaram 6.67%, como ilustrado na tabela 6.22. A tabela 6.23 ilustra a distribuição de classificação.

Tabela 6.22 – Taxa de reconhecimento de olho direito com 5 classes de testes.

<b>Acertos</b>	<b>67</b>	<b>Erros</b>	<b>3</b>	<b>Rejeições</b>	<b>5</b>
<b>Taxa de acerto</b>	<b>89.33%</b>	<b>Taxa de erro</b>	<b>4.0%</b>	<b>Taxa de rejeição</b>	<b>6.67%</b>

Tabela 6.23 – Tabela de confusão de olho direito com 5 classes de testes.

<b>Classe/Classificação</b>	<b>1</b>	<b>2</b>	<b>3</b>	<b>4</b>	<b>5</b>	<b>Rejeição</b>
1	14	1	0	0	0	0
2	0	15	0	0	0	0
3	0	2	10	0	0	3
4	0	0	0	15	0	0
5	0	0	0	0	13	2

#### 6.4.4 Teste com Seis Classes

Agora, os testes de desempenho são feitos utilizando 6 classes de atributos, totalizando 90 imagens para treinamento e 90 para reconhecimento. Quanto ao reconhecimento nasal, foi observada uma taxa de 90% de acerto contra 7.78% de erro, indicado na tabela 6.24. Ainda ocorreram duas rejeições, como pode ser vista na tabela 6.25.

Tabela 6.24 – Taxa de reconhecimento nasal com 6 classes de testes.

<b>Acertos</b>	<b>81</b>	<b>Erros</b>	<b>7</b>	<b>Rejeições</b>	<b>2</b>
<b>Taxa de acerto</b>	<b>90.0%</b>	<b>Taxa de erro</b>	<b>7.78%</b>	<b>Taxa de rejeição</b>	<b>2.22%</b>

Tabela 6.25 – Tabela de confusão nasal com 6 classes de testes.

Classes/Classificação	1	2	3	4	5	6	Rejeição
1	15	0	0	0	0	0	0
2	0	15	0	0	0	0	0
3	0	0	9	5	0	0	1
4	0	0	1	14	0	0	0
5	0	0	0	0	14	1	0
6	0	0	0	0	0	14	1

Com relação aos lábios, a rede obteve uma taxa de acerto de 95.56% contra uma taxa de erro de 3.3%, ilustrado na tabela 6.26. A tabela 6.27 permite uma análise mais detalhada do desempenho da rede para o reconhecimento de lábios.

Tabela 6.26 – Taxa de reconhecimento labial com 6 classes de testes.

<b>Acertos</b>	<b>86</b>	<b>Erros</b>	<b>3</b>	<b>Rejeições</b>	<b>1</b>
<b>Taxa de acerto</b>	<b>95.56%</b>	<b>Taxa de erro</b>	<b>3.33%</b>	<b>Taxa de rejeição</b>	<b>1.11%</b>

Tabela 6.27 – Tabela de confusão labial com 6 classes de testes.

Classes/Classificação	1	2	3	4	5	6	Rejeição
1	14	0	0	0	0	0	1
2	0	15	0	0	0	0	0
3	0	0	15	0	0	0	0
4	0	0	1	14	0	0	0
5	0	0	2	0	13	0	0
6	0	0	0	0	0	15	0

Com relação ao olho direito, a rede obteve uma taxa de acerto de 87.78% contra uma taxa de erro de 6.67%, ilustrado na tabela 6.28. A tabela 6.29 permite uma análise mais detalhada do desempenho da rede para o reconhecimento de lábios.

Tabela 6.28 – Taxa de reconhecimento de olho direito com 6 classes de testes.

<b>Acertos</b>	<b>79</b>	<b>Erros</b>	<b>6</b>	<b>Rejeições</b>	<b>5</b>
<b>Taxa de acerto</b>	<b>87.78%</b>	<b>Taxa de erro</b>	<b>6.67%</b>	<b>Taxa de rejeição</b>	<b>5.55%</b>

Tabela 6.29 – Tabela de confusão de olho direito com 6 classes de testes.

<b>Classes/Classificação</b>	<b>1</b>	<b>2</b>	<b>3</b>	<b>4</b>	<b>5</b>	<b>6</b>	<b>Rejeição</b>
<b>1</b>	13	0	0	0	0	2	0
<b>2</b>	0	14	0	0	0	1	0
<b>3</b>	0	2	10	0	0	1	2
<b>4</b>	0	0	0	15	0	0	0
<b>5</b>	0	0	0	0	13	0	2
<b>6</b>	0	0	0	0	0	14	1

#### 6.4.5 Teste com Sete Classes

O quinto conjunto de testes analisou 7 classes de narizes e olhos direitos. Foi um total de 105 imagens para treinamento e 105 imagens para reconhecimento. Foi observada uma taxa de acerto de 89.52% contra 6.67% de erro e 3.81% de rejeição, o que pode ser visto na tabela 6.30. A tabela 6.31 dá uma idéia geral de desempenho.

Tabela 6.30 – Taxa de reconhecimento nasal com 7 classes de testes.

<b>Acertos</b>	<b>94</b>	<b>Erros</b>	<b>7</b>	<b>Rejeições</b>	<b>4</b>
<b>Taxa de acerto</b>	<b>89.52%</b>	<b>Taxa de erro</b>	<b>6.67%</b>	<b>Taxa de rejeição</b>	<b>3.81%</b>

Tabela 6.31 – Tabela de confusão nasal com 7 classes de testes.

Classe/Classificação	1	2	3	4	5	6	7	Rejeição
1	15	0	0	0	0	0	0	0
2	0	15	0	0	0	0	0	0
3	0	0	9	5	0	0	1	0
4	0	0	1	14	0	0	0	0
5	0	0	0	0	14	1	0	0
6	0	0	0	0	0	15	0	0
7	0	0	0	0	0	0	12	3

Observando o reconhecimento de olho direito, foi obtida uma taxa de acerto de 88.57% contra 6.67% de erro, conforme visto na tabela 6.32. A tabela 6.33 mostra a distribuição de classificações dos padrões de entrada.

Tabela 6.32 – Taxa de reconhecimento de olho direito com 7 classes de testes.

<b>Acertos</b>	<b>93</b>	<b>Erros</b>	<b>7</b>	<b>Rejeições</b>	<b>5</b>
<b>Taxa de acerto</b>	<b>88.57%</b>	<b>Taxa de erro</b>	<b>6.67%</b>	<b>Taxa de rejeição</b>	<b>4.76%</b>

Tabela 6.33 – Tabela de confusão de olho direito com 7 classes de testes.

Classe/Classificação	1	2	3	4	5	6	7	Rejeição
1	13	0	0	0	0	2	0	0
2	0	14	0	0	0	1	0	0
3	0	2	10	0	0	1	0	2
4	0	0	0	15	0	0	0	0
5	0	0	0	0	13	0	0	2
6	0	0	0	0	0	15	0	0
7	1	0	0	0	0	0	13	1

## 6.5- Esboço Gráfico

O Gráfico 6.1 ilustra o desempenho das redes para reconhecimento labial, conforme se aumenta o número de classes de treinamento. Já o Gráfico 6.2 ilustra o desempenho do reconhecimento nasal e do olho direito. Enquanto a variação no reconhecimento labial e no reconhecimento de olhos direito é pouca, a variação no reconhecimento nasal é maior.

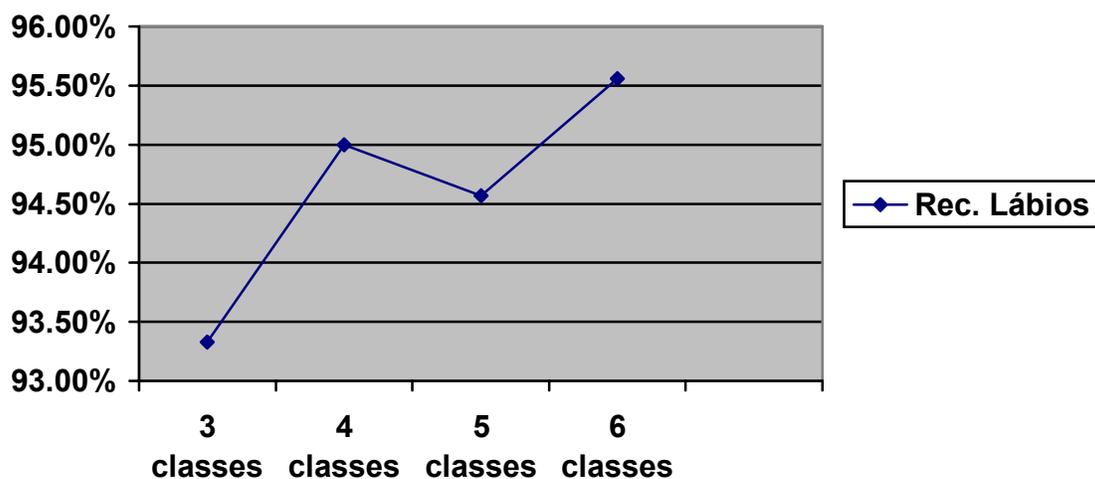


Gráfico 6.1 – Taxa de acerto de reconhecimento labial.

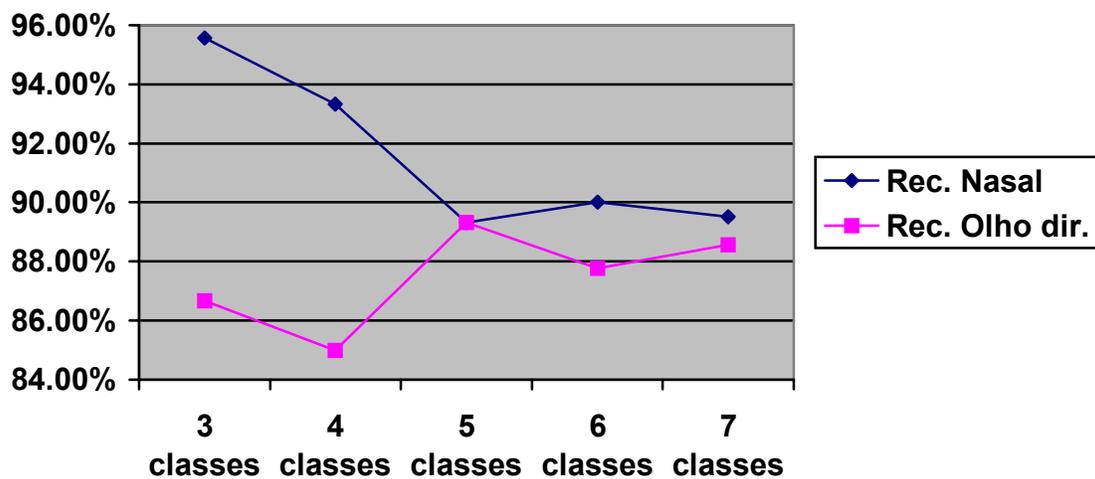


Gráfico 6.2 – Taxa de acerto de reconhecimento nasal e de olho direito.

A tabela 6.34 ilustra os 5 conjuntos de testes utilizados. Como dito anteriormente são utilizados 15 padrões de imagens para cada classe de treinamento e, 15 padrões de imagens para cada classe reconhecimento, distintas umas das outras. No primeiro conjunto de testes utilizam-se 3 classes de treinamento e de reconhecimento, o que totaliza 45 imagens para cada classe. Para o segundo conjunto de testes, com 4 classes de treinamento e de reconhecimento, o sistema conta com mais 15 imagens para treinamento e outras 15 imagens para reconhecimento, totalizando 60 imagens, tanto para treinamento quanto para reconhecimento, e assim por diante. Uma vez que, a adição de uma classe no sistema aumento o número de imagens de reconhecimento, o número total de erros tende a aumentar com a inserção de uma nova classe no sistema. Porém, isso não incorre necessariamente em aumento da taxa de erro. As taxas de erro e acerto geradas pela inserção de uma nova classe vão depender da contribuição e desempenho dessa classe no sistema.

Tabela 6.34 – Número de classes utilizadas versus número de imagens de treinamento e imagens de reconhecimento.

<b>Classes</b>	<b>Treinamento</b>	<b>Reconhecimento</b>
<b>3</b>	<b>45</b>	<b>45</b>
<b>4</b>	<b>60</b>	<b>60</b>
<b>5</b>	<b>75</b>	<b>75</b>
<b>6</b>	<b>90</b>	<b>90</b>
<b>7</b>	<b>105</b>	<b>105</b>

## 7- Análise de Resultados, Conclusões e Perspectivas Futuras

Os resultados possibilitaram validar o *NEORAF*, pois foram obtidas taxas de reconhecimento suficientes para diversas aplicações. Os principais fatores que influenciam o desempenho da rede são:

- A escolha dos padrões de treinamento. Uma boa escolha dos padrões de treinamento resulta num bom desempenho de reconhecimento;
- Maior número de padrões para fase de treinamento aumenta a estabilidade da rede;
- Número de planos criados na etapa de treinamento: se o número de planos criados for muito baixo, a rede terá pouca informação para realizar o reconhecimento do padrão de entrada. Porém, caso esse número seja alto, muita informação contrastante pode estar sendo utilizada pela rede para realização do reconhecimento;
- Os valores ideais de limiares (thresholds) são difíceis de determinar e variam de acordo com diversas variáveis: número de classes, número de padrões de treinamento, tipos de atributos faciais utilizados, etc.
- Expressões faciais. Os diferentes tipos de expressões faciais causam diferentes impactos para o reconhecimento de cada um dos atributos. Por exemplo, uma pessoa fazendo bico não tem tanto impacto no reconhecimento de nariz.

Com base nos testes de desempenho e nos fatores que influenciam o desempenho da *NEORAF*, podem ser feitas várias proposições de estudos futuros, com o intuito de melhorar o desempenho de reconhecimento e, até mesmo, estender a funcionalidade da rede para reconhecimento facial global. Podem ser citados como estudos futuros:

- Estudo para determinação de um limiar ótimo para etapas de treinamento e reconhecimento facial;
- Determinação de fatores relevantes ao treinamento da rede;
- Escolha dos padrões de entrada, objetivando melhorar o desempenho da rede;
- Implementação de um sistema consistindo numa combinação de redes neurais, onde as saídas no *NEORAF* serão utilizadas como entrada para uma outra rede neural que será responsável pelo reconhecimento facial global. Assim, teremos um sistema com dois passos: extração de atributos faciais e reconhecimento facial global.

## REFERÊNCIA BIBLIOGRÁFICA

- [Arb95] Arbib, Michael B., **“Introducing the Neuron”**, The Handbook of Brain Theory and Neural Networks, Massachusetts Institute of Technology, Cambridge, pp. 4-11, 1.995
- [Bia01] Bianchini, Angelo R., **“Arquitetura de Redes Neurais para o Reconhecimento Facial Baseado no Neocognitron”**, UFSCar – São Carlos, julho de 2001.
- [Bol00] Bolle, Ruud M.; Jain, Anil, **“Biometrics: The Future of Identification”**, Computer – Innovate Technology for Computer Professionals, IEEE Computer Society, pp. 46-49, fevereiro de 2000.
- [Can86] Canny, J., **“A Computational Approach to Edge Detection”**, IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol.8, N.6, pp. 679-698, junho de 1986.
- [Cos94] Costa, Luciano F. **Apostila do Curso de Visão Cibernética**, Instituto de Física de São Carlos, JAI 1994.
- [Dai95] Dai, Y.; Nakano, Y., **“Face-Texture Model Based on SGLD and Its Application in Face Detection in a Color Scene”**, Pattern Recognition, Vol. 29, N. 6, pp. 1007-1017, 1996.
- [Fau94] Fausett, Laurene, **Fundamentals of Neural Networks Architectures, Algorithms, and Applications**, Florida Institute of Technology, Prentice Hall Internacional, 1994.

- [Fuk82] Fukushima, Kunihiko, “**Neocognitron: a new algorithm for pattern recognition tolerant of deformations and shifts position**”, Pattern Recognition, Vol. 15, N. 6, pp. 455-469, 1982.
- [Fuk88] Fukushima, Kunihiko, “**Neocognitron: A Hierarchical Neural Network Capable of Visual Pattern Recognition**”, Neural Networks, Vol. 1, pp. 119-130. 1988.
- [Fuk92] Fukushima, Kunihiko; Wake, Nobuaki, “**Improved Neocognitron with Bend-Detecting Cells**”, International Joint Conference on Neural Networks, pp 190-195, Baltimore-Maryland, junho de 1992
- [Gaz02] Gazzaniga, Michael S., Ivry, Richard B., Mangun, George R., “**Cognitive Neuroscience – The Biology of the Mind**”, 22<sup>nd</sup>, Ed. New York, W.W.Norton & Company, 2002.
- [Gon92] Gonzalez, Rafael C.; Woods, Richard E., **Digital Image Processing**, Addison-Wesley Publishing Company, 1994.
- [Gon96] González, Raúl R., **Neural Networks. A Systematic Introduction**, Berlin and Halle, March 1996.
- [Gra65] Graham, CH., “**Vision and Vision Perception**”, John Wiley & Sons, New York, 1965.
- [Gra95] Graf, H.P.; Chen, T.; Petajan, E.; Cosatto, E., “**Locating Faces and Facial Parts**”, Proc. First Int’l Workshop Automatic Face and Gesture Recognition, pp. 41-46, 1995.

- [Har73] Haralick, R.M.; Shanmugam, K.; Dinstein. L., **“Texture Features for Image Classification”**, IEEE Transactions on Systems, Man, and Cybernetics, Vol. 3, N. 6, pp.610-621, 1973.
- [Hay94] Haykin, Simon, **Neural Networks – A Comprehensive Foundation**, Prentice Hall Internacional, 1994.
- [Hon98] Hong, Lin; Jain, Anil, **“Integrating Faces and Fingerprints for Personal Identification”**, IEEE Transaction on Patter Analysis and Machine Intelligence, Vol. 20, N. 12, pp. 1295-1307, dezembro de 1998.
- [Hub67] Hubel, D.H.; Wiesel, T.N., **“Receptive Fields and Functional Architecture of Monkey Striate Cortex”**, From the Department of Physiology Harvard Medical School, Boston, 1967.
- [Len01] Lent, Robert, **“Cem Bilhões de Neurônios – Conceitos Fundamentais da Neurociência”**, São Paulo, Editora Ateneu, 2001.
- [Lov97] Lovell, David R.; Downs, Thomas; Tsoi, Ah C., **“An Evaluation of The Neocognitron”**, IEEE Transactions on Neural Networks, Vol. 8, N. 5, setembro de 1997.
- [Mck98] McKenna, S.; Gong S.; Raja Y., **“Tracking Colour Objects Using Adaptive Mixture Models”**, Image and Vision Computing, Vol.17, N°s. 3 e 4, pp. 223-229, 1998.
- [Phi00] Philips, P. Jonathon; Moon Hyeonjoon; Rizvi, Syed A., **“The FERET Evaluation Methodology for Face-Recognition Algorithms”**, IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 22, N. 10, pp. 1090-1104, outubro de 2000.

- [Sou95] Soulié, Françoise F., “**Applications of Neural Networks**”, The Handbook of Brain Theory and Neural Networks, Michael A. Arbib, Massachusetts Institute of Technology, Cambridge, pp. 94-98, 1995.
- [Sir93] Sirohey, S.A., “**Human Face Segmentation and Identification**”, Technical Report CS-TR-3176, University of Maryland, 1993.
- [Sri02] Srinivas, Gutta; Philomin, Vasanth; Trajkovic, Miroslav, “**An Investigation into to the use of Partial-Faces for Face Recognition**”, Fifth IEEE Internacional Conference on Automatic Face and Gesture Recognition, 20-21 de maio de 2002, Washington DC.
- [Yan94] Yang, G.; Huang, T.S., “**Human Face Detection in Complex Background**”, Pattern Recognition, Vol.27, N. 1, pp. 53-63, 1994.
- [Yan02] Yang, Ming-Hsuan, Kriegman; David J.; Ahuja, Narendra, “**Detecting Faces in Images: A Survey**”, IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 24, N. 1, pp. 34-58, janeiro de 2002.