

Rede Neural por Convolução para  
Reconstrução Estéreo

Júlio Massayuki Otuyama

Florianópolis, fevereiro de 2000



0.327.173-8

UFSC-RU

# Rede Neural por Convolução para Reconstrução Estéreo

Júlio Massayuki Otuyama  
Orientador: Raul Sidnei Wazlawick

Curso de Pós Graduação em Ciência da Computação  
Departamento de Informática e Estatística  
Centro Tecnológico  
Universidade Federal de Santa Catarina

Florianópolis, fevereiro de 2000

Dissertação submetida à Universidade Federal de Santa Catarina para obtenção do Grau de Mestre em Ciência da Computação.

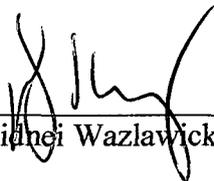
# Rede Neural por Convolução para Reconstrução Estéreo

Júlio Massayuki Otuyama

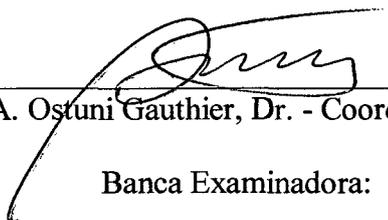
Esta dissertação foi julgada adequada para a obtenção do título de

**Mestre em Ciência da Computação**

na área de concentração Sistemas de Conhecimento e aprovada em sua forma final pelo  
Curso de Pós-Graduação em Ciência da Computação da Universidade Federal de Santa  
Catarina.

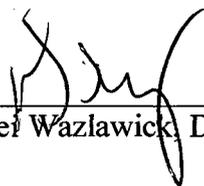


Raul Sidnei Wazlawick, Dr. - Orientador



Fernando A. Ostuni Gauthier, Dr. - Coordenador do Curso

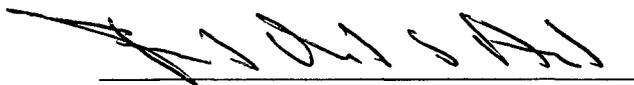
Banca Examinadora:



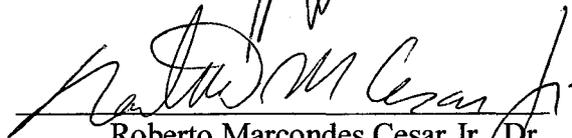
Raul Sidnei Wazlawick, Dr. - Presidente



Rogério Cid Bastos, Dr.



Fernando Mendes de Azevedo, Dr.  
GPEB - EEL - UFSC



Roberto Marcondes Cesar Jr., Dr.  
IME - USP

Florianópolis, 24 de Fevereiro de 2000

# Resumo

Para a modelagem de redes neurais para aplicações de problemas invariantes à translação utilizamos restrições de compartilhamento de pesos. Redes neurais com estas restrições são chamadas de Redes Neurais com Pesos Compartilhados. O compartilhamento de pesos equivale à operação de convolução, permitindo que se utilize diversas técnicas que aceleram o cálculo desta. Esta equivalência é o motivo pelo qual estas redes neurais também são chamadas de Redes Neurais por Convolução. Este trabalho implementa o modelo de Rede Neural por Convolução em uma linguagem apropriada, permitindo a criação de novas topologias de forma flexível, simples e rápida. Supõe-se que nos estágios iniciais do córtex visual ocorra a reconstrução da informação tridimensional a partir de diversos aspectos, entre eles a visão estereoscópica, sombreamento, sobreposição de objetos, textura, perspectiva e movimento. Este trabalho utiliza o modelo de Rede Neural por Convolução para abordar o problema de reconstrução estéreo (visão estereoscópica), utilizando todas as características específicas do problema para obter uma topologia adequada. Em uma segunda linha, este trabalho apresenta uma extensão do modelo de Rede Neural por Convolução utilizando características específicas do cálculo de convolução - a transformada de Fourier. Esta extensão do modelo de Rede Neural por Convolução mantém os pesos no domínio da frequência.

# Abstract

In order to model neural networks for applications of translation invariant problems, weight sharing restrictions are used. Neural networks with such restrictions are called Shared Weight Neural Networks. Weight sharing corresponds to the convolution operation, allowing for the use of several techniques that accelerate its calculation. This equivalence is the reason why such neural networks are also known as Convolution Neural Networks. This project implements the Convolution Neural Network model with an adequate language that allows new topologies to be created in a flexible, simple and fast way. Allegedly, the reconstruction of tridimensional information occurs in the first stages of the visual cortex, based on various aspects, among which are the stereoscopic vision, shape from shading, object overlapping, texture, perspective and movement. This project uses the Convolution Neural Network approach for the purpose of stereo reconstruction (stereoscopic vision), using all specific characteristics of the problem to reach an adequate topology. A second line of this project presents an extension of the Convolution Neural Network model, using specific characteristics of the convolution calculation - the Fourier transform. Such extension of the Convolution Neural Network model represents weights on the frequency domain.

# Sumário

<b>1</b>	<b>Introdução</b>	<b>9</b>
1.1	Apresentação . . . . .	9
1.2	Justificativa e importância . . . . .	10
1.3	Objetivos . . . . .	10
1.3.1	Domínio geral . . . . .	10
1.3.2	Domínio específico . . . . .	11
1.4	Limitações . . . . .	11
1.5	Estrutura . . . . .	12
<b>2</b>	<b>Redes Neurais Biológicas</b>	<b>14</b>
2.1	Neurônios e sinapses . . . . .	14
2.2	Informação através do neurônio . . . . .	15
2.3	Células nervosas sensoriais e motoras . . . . .	16
2.4	Organização entre as células nervosas . . . . .	17
2.4.1	Representação topográfica . . . . .	18
2.4.2	Resposta de células nervosas a estímulos . . . . .	19
2.5	Desenvolvimento das células nervosas . . . . .	21
<b>3</b>	<b>Redes Neurais Artificiais</b>	<b>22</b>
3.1	Modelo de um neurônio artificial . . . . .	22
3.1.1	Função de ativação . . . . .	23
3.2	Arquitetura de redes neurais . . . . .	24
3.2.1	Redes neurais diretas de camada única . . . . .	24
3.2.2	Redes neurais diretas multicamada . . . . .	25
3.3	Algoritmo de aprendizagem . . . . .	27
3.3.1	Back-Propagation . . . . .	27
<b>4</b>	<b>Compartilhamento de Pesos em Redes Neurais</b>	<b>30</b>
4.1	Replicação de redes neurais . . . . .	30
4.2	Modelo de Rede Neural com Pesos Compartilhados . . . . .	31

4.2.1	Conexões locais, reticulado e compartilhamento de pesos . . . . .	33
4.2.2	Canais . . . . .	35
4.3	Algoritmo de aprendizagem . . . . .	36
4.3.1	Adaptação do Back-Propagation . . . . .	37
4.4	Arquiteturas relacionadas . . . . .	38
4.4.1	Rumelhart et al 1985 . . . . .	38
4.4.2	Waibel et al 1989 . . . . .	39
4.4.3	LeCun et al 1990 . . . . .	40
4.4.4	Wilson 1989/1992 . . . . .	40
<b>5</b>	<b>Rede Neural por Convolução</b>	<b>42</b>
5.1	Convolução . . . . .	42
5.1.1	Sinais e sistemas . . . . .	42
5.1.2	Soma de convolução . . . . .	45
5.1.3	Vetores de sinais . . . . .	47
5.2	Modelo de Rede Neural por Convolução . . . . .	49
5.3	Algoritmo de aprendizagem . . . . .	51
5.3.1	Adaptação do Back-Propagation . . . . .	51
5.3.2	RNC e Transformada de Fourier . . . . .	52
5.4	Implementação . . . . .	52
<b>6</b>	<b>Reconstrução Estéreo</b>	<b>56</b>
6.1	Introdução . . . . .	56
6.2	Casamento estéreo . . . . .	57
6.3	Geometria epi-polar . . . . .	58
6.4	Abordagens computacionais . . . . .	59
<b>7</b>	<b>RNC para Reconstrução Estéreo</b>	<b>61</b>
7.1	Detecção de disparidades . . . . .	61
7.1.1	Pré processamento . . . . .	61
7.1.2	Compartilhamento de máscaras de convolução . . . . .	62
7.2	Resultados . . . . .	63
7.2.1	Detecção de disparidades . . . . .	63
7.2.2	Discussão . . . . .	69
<b>8</b>	<b>Conclusão</b>	<b>71</b>
8.1	Recomendações . . . . .	72

# Lista de Figuras

2.1	Neurônio. . . . .	15
2.2	Caminho neural típico. . . . .	16
2.3	Organização topográfica entre a retina e o córtex visual primário. . . . .	18
2.4	Colunas de dominância ocular. . . . .	20
2.5	Colunas de orientação. . . . .	20
3.1	Modelo de neurônio artificial. . . . .	23
3.2	Rede neural direta de camada única. . . . .	25
3.3	Rede neural direta multicamada. . . . .	26
3.4	Rede neural direta multicamada conectada localmente. . . . .	27
4.1	Arquitetura do NETtalk. . . . .	31
4.2	Técnica de janelas para o processamento de imagens. . . . .	32
4.3	Rede neural conectada localmente com camadas organizadas na forma de reticulado unidimensional. . . . .	34
4.4	Rede neural com cada camada composta de canais. . . . .	36
4.5	O Problema T-C. . . . .	39
4.6	Arquitetura do Time-Delay Neural Network (TDNN). . . . .	40
4.7	Rede neural para o reconhecimento de dígitos numéricos manuscritos. . . . .	41
4.8	Arquitetura da Iconic Network. . . . .	41
5.1	Sinal discreto (Seqüência). . . . .	43
5.2	Deslocamento de uma seqüência. . . . .	44
5.3	Sistema. . . . .	44
5.4	Rede Neural por Convolução (RNC). . . . .	50
6.1	Par estéreo. . . . .	57
6.2	Linha epi-polar e disparidade. . . . .	59
7.1	RNC para reconstrução estéreo. . . . .	62
7.2	Padrões de aprendizado. . . . .	64
7.3	Comparativo de saída desejada e saída obtida após treinamento da rede neural. . . . .	65

7.4	Erro na saída da rede neural ao longo das épocas. . . . .	66
7.5	Ativações de canais da camada escondida. . . . .	66
7.6	Máscara de convolução de um neurônio. . . . .	67
7.7	Circunferências sobrepostas em diversas disparidades. . . . .	68
7.8	Círculos sobrepostos em diversas disparidades. . . . .	68
7.9	Detecção de disparidades em uma imagem real. . . . .	69

# Capítulo 1

## Introdução

### 1.1 Apresentação

Os neurônios biológicos são a base de inspiração da teoria de Redes Neurais Artificiais. As Redes Neurais Artificiais são formadas por unidades simples, os chamados neurônios artificiais, que efetuam operações de soma e transferência de sinal. Redes Neurais Artificiais possuem uma ampla gama de aplicações, tal como reconhecimento de padrões e processamento de imagens.

A forma como os neurônios se interligam no córtex e as qualidades decorrentes que podem ser observadas a partir disto é inspiração para criação de arquiteturas de Redes Neurais Artificiais. A partir da teoria de Redes Neurais Artificiais deriva-se o modelo de Rede Neural com Pesos Compartilhados. No córtex visual observa-se a qualidade de extração de características invariante à translação. O modelo de Rede Neural com Pesos Compartilhados possui restrições que produzem qualidade semelhante. Esta qualidade é desejável em determinados problemas, em particular para área de processamento de imagens.

O modelo de Rede Neural com Pesos Compartilhados pode ser rescrito utilizando ferramentas da teoria de processamento de sinais: Convolução e Transformada de Fourier. Disto deriva-se o modelo de Rede Neural por Convolução. O modelo de Rede Neural com Pesos Compartilhados e o modelo de Rede Neural por Convolução são equivalentes, porém a notação utilizada neste último é mais concisa, auxiliando a compreensão. O uso de ferramentas da teoria de processamento de sinais permite um aumento significativo na eficiência do modelo.

O modelo de Rede Neural por Convolução, tal como os modelos de Redes Neurais Artificiais, é adequado ao processamento paralelo. A qualidade de extração de características invariante à translação e a estrutura paralelizável torna o modelo de Rede Neural por Convolução adequado a uma gama de problemas de processamento de imagens, tal como reconhecimento de letras e dígitos ou detecção de bordas.

Nos estágios iniciais do córtex visual, supõe-se que ocorra o processamento da informação visual para detecção de linhas, bordas, agrupamentos ou para reconstrução da informação

tridimensional a partir de visão estereoscópica, sombreamento, sobreposição de objetos, textura, perspectiva ou movimento. Utilizando-se padrões de aprendizado e topologias adequadas, cada um destes problemas pode ser abordado com o modelo de Rede Neural por Convolução.

Reconstrução estéreo (visão estereoscópica) é o problema de reconstruir a informação tridimensional a partir de um par de imagens capturadas simultaneamente com um deslocamento entre as posições de captura. O modelo de Rede Neural por Convolução pode ser aplicado ao problema de reconstrução estéreo, permitindo que se explore características exclusivas deste problema.

## 1.2 Justificativa e importância

O modelo de Rede Neural por Convolução tem importantes aplicações na área de processamento de imagens e sinais (veja [Rumelhart 1986], [Waibel 1989], [LeCun 1990], [Wilson 1989], [Wilson 1992], [Lawrence 1997] e [Spreeuwens 1993]).

Ao utilizar o modelo de Rede Neural por Convolução, é natural que se procure formas de torná-lo mais eficiente. A busca por melhorias na eficiência do modelo geralmente resulta em abordagens de implementação utilizando algoritmos alternativos. Tais implementações alternativas permitem uma visão mais ampla do próprio modelo, o que contribui para enriquecer este.

Abordando diferentes problemas com o modelo de Rede Neural por Convolução, observa-se a necessidade de adaptação ou extensão do modelo para que tais problemas sejam resolvidos de forma mais adequada. Isto possibilita a generalização do modelo de Rede Neural por Convolução, tornando-o mais abrangente.

O problema de reconstrução da informação tridimensional pode se basear em visão estereoscópica, sombreamento, sobreposição de objetos, textura, perspectiva, movimento ou outros. Geralmente, cada um destes problemas é abordado separadamente com algoritmos distintos. A integração de tais algoritmos é um objetivo desejável, caso contrário é possível que haja informação relevante para reconstrução da informação tridimensional deixando de ser utilizada. Tal integração pode vir a gerar resultados não satisfatórios, ou mesmo não ser possível, devido a diferenças intrínsecas entre os algoritmos empregados. O modelo de Rede Neural por Convolução permite a adaptação a cada um destes problemas. O uso de um único modelo para abordar tais problemas torna mais fácil uma possível integração para unificar o processo de reconstrução da informação tridimensional.

## 1.3 Objetivos

### 1.3.1 Domínio geral

O trabalho corrente tem como objetivo geral:

- Melhoria do modelo de Rede Neural por Convolução utilizando elementos da teoria de processamento de sinais.
- Melhoria do modelo de Rede Neural por Convolução observando-se a adaptação deste ao problema de reconstrução estéreo para torná-lo mais abrangente.
- Implementação do modelo de Rede Neural por Convolução.

### 1.3.2 Domínio específico

O trabalho corrente tem como objetivo específico:

- Adaptação do modelo de Rede Neural por Convolução ao problema de reconstrução estéreo.
- Implementação e execução desta adaptação.

## 1.4 Limitações

O modelo de Rede Neural por Convolução tem sido utilizado principalmente para reconhecimento de padrões (veja [Rumelhart 1986], [Waibel 1989], [LeCun 1990] e [Lawrence 1997]). Esta tendência deve-se à forma como se aborda problemas em redes neurais - o treinamento baseado em padrões de aprendizagem, onde dado uma entrada da rede neural espera-se à saída desta uma resposta de acordo com uma categorização previamente definida. Por outro lado, o modelo de Rede Neural por Convolução tem sido utilizado para modelagem de filtros para processamento de imagens (veja [Wilson 1989], [Wilson 1992] e [Spreeuwes 1993]), porém com aplicações restritas.

O trabalho corrente aborda o problema de reconstrução estéreo utilizando o modelo de Rede Neural por Convolução de forma similar a um filtro de processamento de imagem. A enorme gama de possibilidades de entrada com o qual um filtro de processamento de imagens deve lidar dificulta a tarefa de treinar esta rede neural. A escolha de padrões de aprendizagem adequados para esta rede neural é um limitante crítico.

A abordagem do problema de reconstrução da informação tridimensional a partir de visão estereoscópica, sombreamento, sobreposição de objetos, textura, perspectiva ou movimento utilizando o modelo de Rede Neural por Convolução está apenas em seu início. O trabalho corrente aborda apenas o problema de reconstrução da informação tridimensional a partir de visão estereoscópica. A consolidação dos resultados deste trabalho pode vir a ter contribuições provenientes das outras abordagens e vice-versa.

Redes neurais geralmente podem ser implementadas em circuitos digitais, o que aumenta a velocidade de processamento de forma significativa. Porém, na prática, redes neurais são quase sempre implementadas com simuladores em linguagens de programação, devido à versatilidade

e facilidade de acesso a tal recurso. O modelo de Rede Neural por Convolução, mesmo sendo implementado com simuladores em linguagens de programação pode se valer de aceleradores dedicados para processamento de sinais. Porém, no trabalho corrente, estes recursos não estiveram disponíveis.

## 1.5 Estrutura

### Capítulo 1 : Introdução

### Capítulo 2 : Redes Neurais Biológicas

Introdução ao funcionamento de neurônios biológicos e à organização observada entre estes, que serve de inspiração para modelagem deste trabalho. As principais qualidades observadas, decorrentes da estrutura entre os neurônios do córtex visual são: extração de características e invariância ao deslocamento. Entre os tipos de respostas observados entres os neurônios do córtex visual estão: dominância ocular, detecção de orientação e detecção de disparidade estereoscópica.

### Capítulo 3 : Redes Neurais Artificiais

Introdução ao modelo de neurônio artificial, redes neural direta multicamada, rede neural direta conectada localmente e algoritmos de aprendizagem.

### Capítulo 4 : Compartilhamento de Pesos em Redes Neurais

A partir do modelo de rede neural direta conectada localmente, adicionando-se restrições de compartilhamento de pesos, obtém-se o modelo de Rede Neural com Pesos Compartilhados. As qualidades de extração de características com invariância ao deslocamento do modelo de Rede Neural com Pesos Compartilhados são apresentados. O algoritmo de aprendizagem Back-Propagation é adaptado ao modelo de Rede Neural com Pesos Compartilhados.

### Capítulo 5 : Rede Neural por Convolução

O modelo de Rede Neural com Pesos Compartilhados é descrito utilizando a notação de convolução, o que resulta no modelo de Rede Neural por Convolução (RNC). Com uma descrição concisa, utilizando elementos da teoria de Processamento de Sinais, explora-se alternativas para o aumento de eficiência do modelo de Rede Neural por Convolução.

### Capítulo 6 : Reconstrução Estéreo

Introdução à reconstrução estéreo, um problema da área de Visão Computacional. São apresentados: conceitos de casamento estéreo, geometria epi-polar; adaptação à restrição de

invariância à translação; abordagens computacionais usuais a este problema.

#### Capítulo 7 : RNC para Reconstrução Estéreo

O modelo de Rede Neural por Convolução é adaptado ao problema de reconstrução estéreo. Utilizando como inspiração a organização observada em Redes Neurais Biológicas (detecção de disparidade estereoscópica e dominância ocular) e abordagens computacionais de reconstrução estéreo, é apresentada uma topologia de Rede Neural por Convolução para detecção de disparidade. Esta topologia é implementada utilizando-se compartilhamento de máscaras de convolução. Resultados da execução do treinamento da rede neural e da detecção de disparidades são apresentados. Após a execução do treinamento, observa-se na camada escondida da rede neural a presença de detetores de orientação.

#### Capítulo 8 : Conclusão

## Capítulo 2

# Redes Neurais Biológicas

O cérebro é o órgão responsável pelo controle de uma gama de atividades complexas. Ele é responsável pela coordenação de nossa interação com o ambiente, através de nossos órgãos sensitivos e motores. A seguir apresentamos os elementos com os quais o cérebro é constituído e a organização observada entre estes elementos, que serve de inspiração para modelagem deste trabalho. Para mais informação consulte: [Hubel 1995] e [Kuffler 1984].

### 2.1 Neurônios e sinapses

O cérebro é composto por células nervosas, também chamadas de *neurônios* (veja ilustração da Figura 2.1 - extraído de [Hubel 1995]) As células nervosas se interligam, recebendo e enviando informação de e para outras células nervosas. A maior parte das células nervosas do cérebro é composta por um corpo celular (*soma*) a partir do qual se ramificam diversas fibras, chamadas *dendritos*, e uma única fibra que se diferencia das outras, chamada *axônio*. Os dendritos se ramificam na forma de árvores, recebendo informação de outras células nervosas. O corpo celular também pode receber informação, tal como os dendritos. O axônio transmite informação de sua célula nervosa para outras células nervosas. O axônio geralmente se ramifica próximo à região onde termina e as diversas terminações aproximam-se mas não tocam os dendritos (e, por vezes, o corpo celular) de outras células. Nestas regiões, chamadas de *sinapse*, a informação é transferida de uma célula nervosa, a célula *pré sináptica*, para a seguinte, a célula *pós sináptica*.

O tipo mais comum de sinapse é a sinapse química. Neste tipo de sinapse a informação é transmitida na forma de substâncias químicas, os chamados *neurotransmissores*. A direção de transmissão de informação nas sinapses é do axônio da célula pré sináptica para o dendrito (ou corpo celular) da célula pós sináptica.

Quase toda célula no sistema nervoso recebe, através de seus dendritos (e do corpo celular), informação de mais de uma outra célula nervosa. A isto chamamos de *convergência sináptica*. Quase todo axônio de uma célula nervosa se ramifica enviando informação para mais de uma

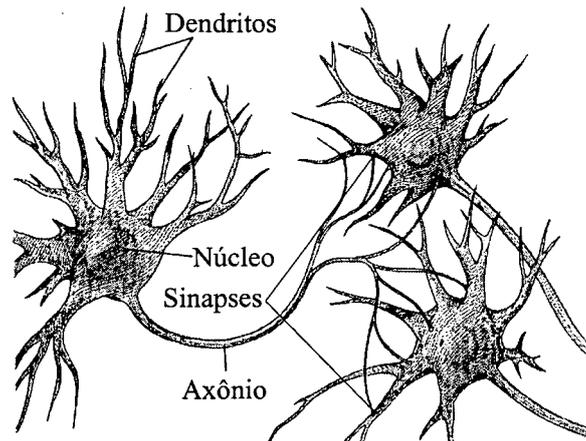


Figura 2.1: Neurônio.

outra célula nervosa. A isto chamamos de *divergência sináptica*.

## 2.2 Informação através do neurônio

Quando neurotransmissores atravessam uma sinapse e atingem o dendrito (ou o corpo celular), a informação química é convertida em uma alteração de potencial elétrico. Há sinapses excitatórias e inibitórias, uma excitando e outra inibindo o estado de ativação da célula nervosa, respectivamente.

A função da célula nervosa é receber toda a informação das sinapses, efetuar a *soma (integração)* das excitações e inibições na base do corpo celular e transferir o resultado ao axônio. A soma na base do corpo celular tem a forma de um potencial elétrico. A informação recebida de cada sinapse geralmente tem um efeito muito pequeno sobre o estado de ativação global do corpo celular. Quanto mais forte a conexão sináptica, maior é a influência da informação recebida.

O axônio geralmente transfere a informação em eventos breves chamados *impulsos nervosos*. Os impulsos nervosos também são conhecidos como *impulsos tudo ou nada (all or none impulses)* devido à ausência de valores intermediários nos potenciais elétricos que assume. À medida que o estímulo na base do corpo celular aumenta, a frequência de disparo dos impulsos nervosos através do axônio aumenta. Existe uma segunda forma menos comum pela qual a informação é transferida através do axônio. Esta outra forma, chamada *sinal gradual (graded signal)*, atua apenas pela dispersão do potencial elétrico ao longo do axônio. Ao contrário dos impulsos tudo ou nada, o sinal gradual pode assumir valores intermediários de potencial elétrico. Por atuar apenas pela dispersão do potencial elétrico, quanto maior comprimento do axônio, maior é a degradação do conteúdo da informação transmitida. Tanto com os impulsos nervosos quanto com o sinal gradual, os terminais do axônio liberam neurotransmis-

sores para sinapse. Quanto maior a frequência de disparo dos impulsos nervosos no axônio, maior a quantidade de neurotransmissor liberado nos terminais do axônio. De forma semelhante, com o sinal gradual, quanto maior o estímulo na base do corpo celular, maior a quantidade de neurotransmissor liberado nos terminais do axônio.

Portanto, ambas formas de transferência de informação através do axônio, por impulsos nervosos ou por sinal gradual, obtêm resultados não muito diferentes, essencialmente analógico [Hubel 1995, pág. 22].

Uma vez que o neurotransmissor é liberado nos terminais do axônio, ele trafega através das sinapses e atinge o dendrito (ou corpo celular) de uma outra célula nervosa, iniciando outro ciclo.

## 2.3 Células nervosas sensoriais e motoras

Além de células nervosas que trocam informações entre si, há células no sistema nervoso que interagem com o ambiente externo, recebendo estímulos sensoriais deste ou exercendo influência neste.

Chamamos de *receptores* às células que recebem informação do ambiente externo. Os estímulos do ambiente externo podem ter a forma de luz para os olhos, deformação mecânica para pele ou tímpanos, ou químicos em nossos sentidos de olfato ou paladar. Em todos estes casos, o efeito do estímulo é a produção nos receptores de um sinal elétrico e conseqüente alteração na taxa de neurotransmissores liberados nos terminais de seus axônios.

*Neurônios motores* são células cujos axônios não terminam em outras células nervosas, mas em células musculares. Toda saída do nosso sistema nervoso tem a forma de contrações musculares, com a exceção de células nervosas que terminam em células glandulares.

Um caminho neural típico inicia-se em um conjunto de receptores, passa por um determinado conjunto de células nervosas e termina em um conjunto de neurônios motores.

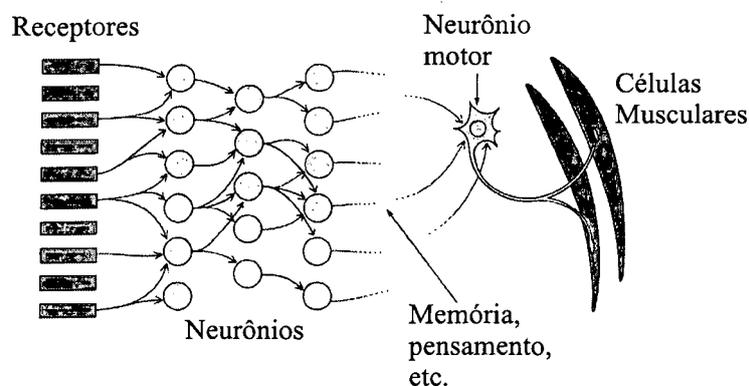


Figura 2.2: Caminho neural típico.

A Figura 2.2 (extraído de [Hubel 1995]) é um exemplo de como ocorre a convergência e

divergência sináptica ao longo de um caminho neural.

Vários sistemas sensoriais fazem uso de pares opostos: quente - frio; preto - branco; rotação da cabeça para direita - rotação da cabeça para esquerda. O motivo para estes pares opostos está provavelmente relacionado com a forma com a qual as células nervosas disparam. A princípio poderia-se imaginar que as células nervosas disparariam a um certo nível mediano, e para um estímulo a célula dispararia a uma taxa inferior e para o estímulo oposto a célula dispararia a uma taxa superior. Porém, como os impulsos nervosos utilizam energia metabólica, provavelmente é mais eficiente que as células não disparem ou disparem a taxas baixas na ausência de um estímulo (e de seu estímulo oposto), e haja dois grupos separados de células nervosas para cada modalidade de estímulo - uma disparando para um estímulo e outra disparando para o estímulo oposto [Hubel 1995, pág. 55].

## 2.4 Organização entre as células nervosas

Em [Hubel 1995] David H. Hubel descreve, de forma brilhante, diversos aspectos do córtex visual em primatas e mamíferos. Ao longo desta obra são apresentados vários tipos de células nervosas e suas respectivas respostas de ativação a determinados estímulos visuais, além da organização observada entre tais células:

“O cérebro contém  $10^{12}$  células, um número astronômico sob qualquer padrão.”

“Uma célula nervosa típica no cérebro recebe informação de centenas ou milhares de outras células e por sua vez transmite informação para centenas ou milhares de outras células. O número total de interconexões no cérebro deve portanto ser algo em torno de  $10^{14}$  a  $10^{15}$ , um grande número, com certeza, mas ainda não um índice de complexidade seguro. Complexidade anatômica é uma questão não apenas de números; mais importante é a complexidade de organização, algo que é difícil de quantificar.”

“Mesmo uma olhada em um livro como *Histologie du Système Nerveux de Cajal* deveria ser suficiente para convencer qualquer um que a enorme complexidade do sistema nervoso é quase sempre acompanhada por um convincente grau de ordenação.” [Hubel 1995, pág. 1,4]

Região receptiva é um conceito importante para compreensão do comportamento das células nervosas. Região receptiva refere-se aos receptores específicos que influenciam a ativação de uma célula nervosa. Para células nervosas no caminho neural da visão, a região receptiva refere-se a uma área na retina na qual a iluminação influencia a ativação da célula nervosa.

A análise sistemática das regiões receptivas de células nervosas demonstram o princípio geral de que células nervosas que processam informação relacionada são agrupadas juntas [Kuffler 1984, pág. 28]. Duas células nervosas ocupando posições próximas no córtex in-

variavelmente possuem regiões receptoras próximas, e estas regiões receptoras geralmente se sobrepõem.

### 2.4.1 Representação topográfica

À agregação de células nervosas chamamos de *estrutura*. Geralmente as células nervosas se organizam na forma de *camadas* dentro de uma estrutura. A retina e o córtex visual primário são exemplos de estruturas. Entre duas estruturas relacionadas há fibras nervosas (axônios) interligando-as. A retina e o córtex visual primário estão relacionadas através de uma estrutura intermediária chamada “Lateral Geniculate Nucleus”.

Observa-se que as células nervosas ao longo de caminhos neurais geralmente apresentam *organização topográfica*. Representação topográfica significa o mapeamento sistemático de uma estrutura e sua seguinte - se dois pontos são vizinhos em uma estrutura, os pontos correspondentes na estrutura seguinte também o serão; por exemplo: à medida que movemos de uma célula nervosa a outra sobre a retina, as células nervosas correspondentes (em termos de ativação) no córtex visual primário traçam um caminho contínuo. Veja ilustração da Figura 2.3 - neste experimento, realizado por Roger Tootell, pode-se observar o mapeamento sistemático que existe entre a retina (estimulada com os círculos) e córtex visual primário (citado em [Hubel 1995]).

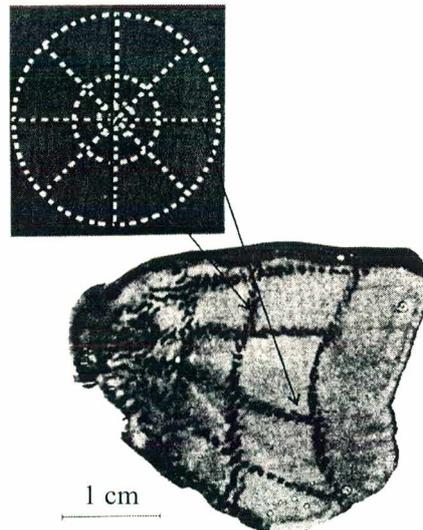


Figura 2.3: Organização topográfica entre a retina e o córtex visual primário.

Devido a convergência sináptica, as células nervosas em estruturas posteriores tendem a possuir regiões receptoras mais largas.

## 2.4.2 Resposta de células nervosas a estímulos

Por volta de 1952, Stephen Kuffler tornou-se o primeiro a registrar as resposta de células retiniais ganglionares a refletores de luz em um mamífero (um gato). Com as respostas obtidas, Kuffler chamou as células em dois tipos: “on-center cell” e “off-center cell”:

“Células on-center respondem melhor a um fecho de luz apontado para parte central de suas regiões receptivas. A iluminação da área ao redor com um fecho ou anel de luz reduz ou inibe os disparos e causa respostas quando a luz é desligada. A iluminação de toda região receptiva causa relativamente disparos fracos pois o efeito do centro e da área ao redor se opõem entre si. Células off-center diminuem ou param seus disparos quando a área central de suas regiões receptivas são iluminadas e aceleram quando a luz é desligada. Luz apontada para a área ao redor de uma região receptiva de uma célula off-center excita a célula correspondente.”  
[Kuffler 1984, pág. 27]

Por volta de 1959, Hubel e Wiesel registraram a resposta de células nervosas do córtex visual de um gato. Tais células respondiam a estímulos de linhas orientadas em uma direção específica e em movimento. Após estes experimentos, seguiram-se outros, cujas observações apresentamos a seguir.

### **Dominância ocular**

Observa-se que células nervosas no córtex visual respondem de forma dominante a estímulos apresentados a apenas um dos olhos (direito ou esquerdo). Estas células nervosas agrupam-se na forma de coluna para cada olho dominante. As colunas de dominância de olho direito e esquerdo se alternam ao longo do córtex visual. Tais agrupamentos de células nervosas são chamados de colunas de dominância ocular [Hubel 1995, pág. 105]. Veja ilustração da Figura 2.4. Esta ilustração é uma reconstrução de colunas de dominância ocular do córtex visual realizada por Le Vay, onde podemos observar a alternância das colunas de dominância de olho direito e esquerdo - citado em [Hubel 1995].

### **Orientação**

Além de se observar colunas de dominância ocular, verificou-se que as células nervosas apresentam também preferência a orientação (a linhas direcionadas). As células nervosas se agrupam para cada orientação preferencial. Ao longo do córtex, a orientação preferencial das células nervosas varia de forma geralmente ordenada, no sentido horário ou anti-horário. Tais agrupamentos de células nervosas são chamados de colunas de orientação [Hubel 1995, pág. 115]. Veja ilustração da Figura 2.5. Nesta ilustração gerada com auxílio de computador, realizado

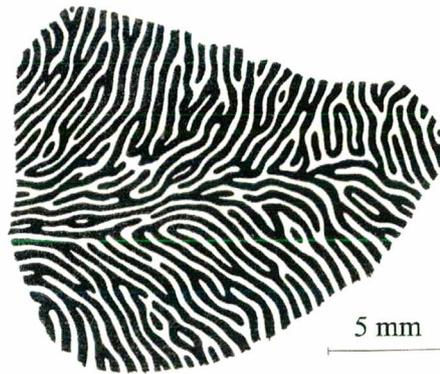


Figura 2.4: Colunas de dominância ocular.

por Gary Blasdel, observamos a sobreposição das respostas no córtex visual primário a estímulos de faixas orientadas, com uma cor diferente representando cada orientação - citado em [Hubel 1995].

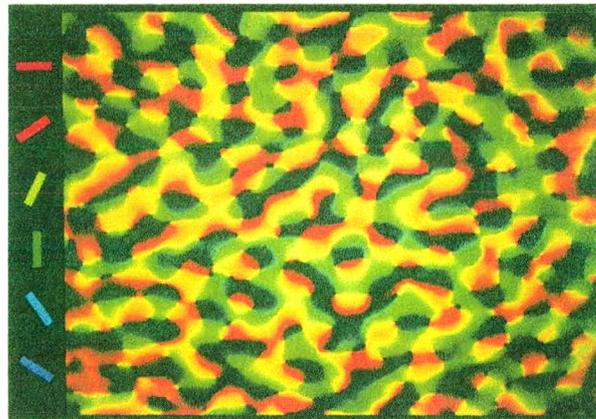


Figura 2.5: Colunas de orientação.

### Disparidade estereoscópica

Mais recentemente, em [Ohzawa 1990] resultados de experimentos demonstram que um tipo específico de neurônio cortical exibe características desejadas de um detetor de disparidade estereoscópica. Tais experimentos foram efetuados observando respostas de neurônios do córtex visual de gatos. Barras luminosas orientadas são utilizadas como estímulo visual. As respostas dos neurônios do córtex visual são comparados com modelos matemáticos para detecção de disparidade.

## 2.5 Desenvolvimento das células nervosas

O desenvolvimento do cérebro humano ocorre principalmente nos primeiros anos de vida, e continua ao longo de todo o resto da vida. Durante este estágio inicial, aproximadamente 1 milhão de sinapses são formadas por segundo.

Treinamento é essencial para que o cérebro efetue todas as tarefas para nossa sobrevivência. *Treinamento* significa a modificação ou moldagem dos circuitos neurais a partir de influências do ambiente. Experimentos observando o sincronismo de células ganglionares da retina de fetos sugerem que células que disparam junto conectam-se juntas, gerando a organização topográfica observada entre a retina e o córtex. Através de experimentos, observa-se que a formação de colunas de dominância ocular em gatos e primatas dependem diretamente da experiência visual durante os primeiros dias de vida [Hubel 1995, pág. 191].

Muitos aspectos do desenvolvimento neural tem recebido estudos intensos, mas ainda existem muitas lacunas para sua compreensão. As pesquisas indicam que há vários mecanismos atuando em conjunto para compor o desenvolvimento neural.

## Capítulo 3

# Redes Neurais Artificiais

Em 1943, em um trabalho pioneiro, McCulloch e Pitts propuseram um modelo de neurônio artificial. A seguir apresentamos elementos básicos da teoria de redes neurais, que utilizamos ao longo deste trabalho. As seções 3.1 e 3.2 são adaptações de [Haykin 1994]. A seção 3.3 é adaptação de [Hertz 1991]. Para mais informação consulte: [Haykin 1994], [Hertz 1991] e [Rumelhart 1986].

### 3.1 Modelo de um neurônio artificial

Um *neurônio artificial* é uma unidade de processamento de informação fundamental para a operação de uma rede neural. A Figura 3.1 mostra um *modelo* para um neurônio. Identificamos três elementos básicos deste modelo de neurônio:

1. Um conjunto de *sinapses*, cada uma das quais é caracterizada por um peso ou força de conexão (*pesos sinápticos*). Dado um sinal de entrada  $x_j$ , na entrada da sinapse  $j$  conectada ao neurônio  $i$ , este sinal é multiplicado pelo peso sináptico  $w_{ij}$ . Na notação utilizada para descrever o peso sináptico, o primeiro subscrito refere-se ao neurônio em questão e o segundo subscrito refere-se à terminação de entrada da sinapse ao qual o peso sináptico se refere. O peso  $w_{ij}$  é positivo se a sinapse associada é excitatória e negativo se a sinapse é inibitória.
2. Um *integrador* para soma dos sinais de entrada, ponderado pelos respectivos pesos sinápticos do neurônio; esta operação constitui-se de um combinador linear.
3. Uma *função de ativação* para limitar a amplitude do sinal de saída do neurônio. A função de ativação também é referida como uma *função limitante* na qual ela limita para um valor finito a faixa de amplitude permitida do sinal de saída. Tipicamente, a faixa de amplitude normalizada da saída de um neurônio está em um intervalo fechado de  $[0, 1]$ .

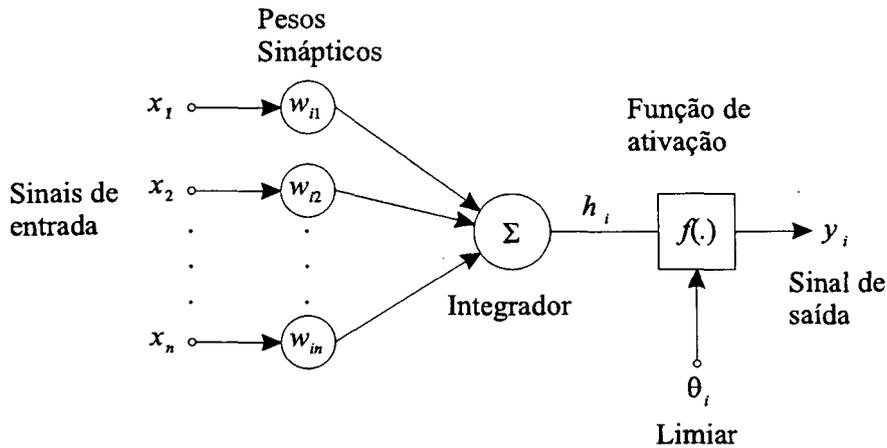


Figura 3.1: Modelo de neurônio artificial.

Podemos descrever um neurônio  $i$  pelo seguinte par de equações

$$h_i = \sum_{j=1}^n w_{ij} x_j \quad (3.1)$$

e

$$y_i = f(h_i) \quad (3.2)$$

onde  $x_1, x_2, \dots, x_n$  são os sinais de entrada;  $w_{i1}, w_{i2}, \dots, w_{in}$  são os pesos sinápticos do neurônio  $i$ ;  $h_i$  é a combinação linear dos sinais de entrada do neurônio  $i$ ;  $f(\cdot)$  é a função de ativação; e  $y_i$  é o sinal de saída do neurônio  $i$ .

O modelo de neurônio mostrado na Figura 3.1 também inclui um *limiar* (*threshold*) externamente aplicado  $\theta_i$  que tem o efeito de reduzir a entrada total da função de ativação. Para ter o efeito contrário, basta utilizar um valor negativo em  $\theta_i$ , o que chamamos de *desvio* (*bias*).

A partir de (3.2) podemos incluir o limiar  $\theta_i$ , o que resulta em

$$y_i = f(h_i - \theta_i) \quad (3.3)$$

Ao longo deste trabalho, utilizamos um limiar de valor constante em todos neurônios. Omitimos o valor do limiar na descrição do neurônio nas formulações a seguir, para facilitar a leitura destas.

### 3.1.1 Função de ativação

A função de ativação, denotada por  $f(\cdot)$ , define a saída do neurônio em termos do nível de ativação em sua entrada. Existem três tipos básicos de função de ativação: função limiar, função linear limitada e função sigmóide.

A função sigmóide é a forma mais comum de função de ativação utilizada na construção de redes neurais artificiais. Ela é definida como uma função estritamente crescente que exhibe propriedade assintótica e de suavidade. Um exemplo de sigmóide é a função logística, definida por

$$f(x) = \frac{1}{1 + \exp(-\alpha x)} \quad (3.4)$$

onde  $\alpha$  é o parâmetro de inclinação da função sigmóide. Ao variar o parâmetro  $\alpha$ , nós obtemos funções sigmóide de diferentes inclinações. A função sigmóide é derivável, uma qualidade importante na teoria de redes neurais artificiais (veja seção 3.3.1).

Utilizamos a função sigmóide como função de ativação ao longo deste trabalho.

## 3.2 Arquitetura de redes neurais

Redes neurais são geralmente organizadas na forma de agrupamentos de nós que chamamos de *camadas*. Há camadas de nós fonte de entrada (nós de entrada) e camadas de nós de processamento (neurônios). *Nós de entrada* não efetuam processamento de sinal, servem apenas para alimentar outros nós com sua informação. *Nós de processamento* são os neurônios, que recebem informação de nós de entrada ou de outros neurônios e efetuam processamento (tal como descrito no modelo de neurônio artificial na seção 3.1).

Redes neurais em que os nós de uma camada enviam sinais apenas para nós de camadas posteriores são chamadas de *redes neurais diretas* (*feed-forward neural networks*). Em contraste com as redes neurais diretas, chamamos de *redes neurais recorrentes* (*recurrent neural networks*) as redes neurais cujas camadas enviem sinais não apenas para camadas posteriores. A existência de ao menos um ciclo de retro alimentação (uma camada lançando sinal para uma camada anterior ou a si mesma) é o que caracteriza as redes neurais recorrentes.

De forma geral, podemos identificar três classes de arquiteturas de redes neurais:

- Redes Neurais Diretas de Camada Única (Single-Layer Feedforward Neural Networks)
- Redes Neurais Diretas Multicamada (Multilayer Feed-Forward Neural Networks)
- Redes Neurais Recorrentes (Recurrent Neural Networks)

Ao longo deste trabalho abordamos apenas redes neurais diretas.

### 3.2.1 Redes neurais diretas de camada única

Na forma mais simples de uma rede neural em camadas, existe apenas uma camada de nós de entrada que lançam seus sinais sobre uma camada de nós de saída (nós de processamento), mas não o contrário. Este é um exemplo de uma rede neural direta. Esta rede é ilustrada

na Figura 3.2. Tal rede é chamada de Rede Neural Direta de Camada Única. A designação “Camada Única” refere-se à camada de saída de nós de processamento. A camada de nós de entrada não é contada pois nenhum processamento é efetuado nesta.

A partir de (3.1) e (3.2) descrevemos a propagação de sinal da camada  $x$  para a camada seguinte  $y$  com o seguinte par de equações

$$h_i = \sum_j w_{ij}x_j \quad (3.5)$$

e

$$y_i = f(h_i) \quad (3.6)$$

onde  $x_j$  é o sinal de ativação do nó  $j$ ;  $w_{ij}$  é o peso sináptico dos neurônio  $i$  referente ao sinal de de entrada do nó  $j$ ; e  $y_i$  é o sinal de ativação do neurônio  $i$ .

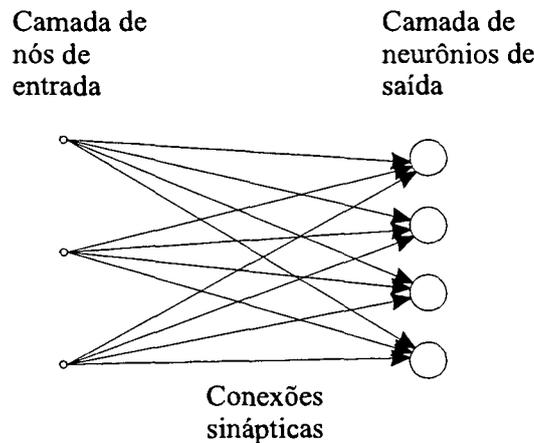


Figura 3.2: Rede neural direta de camada única.

### 3.2.2 Redes neurais diretas multicamada

A segunda classe de rede neural direta distingue-se pela presença de uma ou mais camadas escondidas, cujos nós de processamento são correspondentemente chamados de *neurônios escondidos* ou *unidades escondidas*. A função dos neurônios escondidos é intervir entre a entrada externa e a saída da rede neural. Ao adicionar uma ou mais camadas escondidas, a rede neural está capacitada a extrair *características (features)*. A adição de camadas escondidas é importante na teoria de redes neurais para a resolução de problemas que são não linearmente separáveis (veja “The XOR Problem” em [Rumelhart 1986]).

A partir de (3.5) e (3.6) descrevemos a propagação de sinal da camada anterior  $a^{l-1}$  para a camada  $a^l$  com o seguinte par de equações

$$h_i^l = \sum_j w_{ij}^l a_j^{l-1} \quad (3.7)$$

e

$$a_i^l = f(h_i^l) \quad (3.8)$$

onde  $a_i^l$  é o sinal de ativação do nó  $i$  da camada  $l$ ;  $w_{ij}^l$  é o peso sináptico dos neurônio  $i$  da camada  $l$  referente ao sinal de de entrada do nó  $j$  da camada anterior.

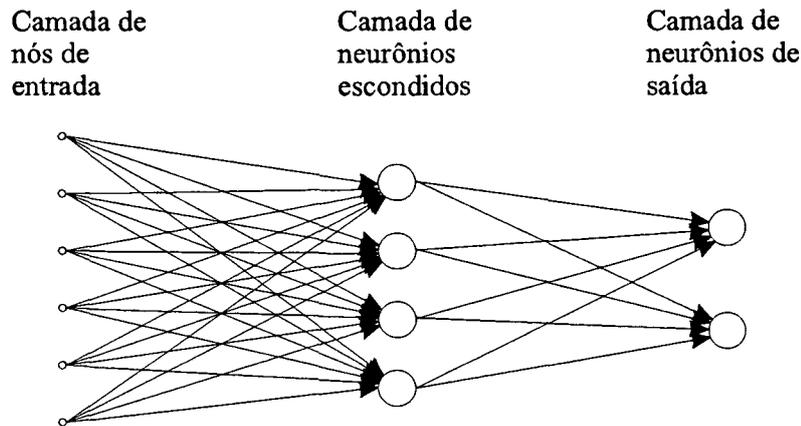


Figura 3.3: Rede neural direta multicamada.

Uma rede neural é chamada *totalmente conectada* (*fully connected*) quando todo nó em cada camada da rede neural é conectado a todos outros nós na camada posterior adjacente. Um exemplo de rede neural direta totalmente conectada é apresentada na Figura 3.3. Se, entretanto, alguma das ligações de comunicação (conexão sináptica) estiver ausente na rede neural, esta é chamada de rede neural *parcialmente conectada* (*partially connected*). Uma forma particularmente interessante deste tipo de rede neural é uma rede neural *conectada localmente* (*locally connected*). Um exemplo de tal rede neural é apresentada na Figura 3.4. Cada neurônio na camada escondida é conectado de forma local (parcial) a um conjunto de nós de entrada que estão na vizinhança próxima. Este conjunto de nós alimentando um neurônio é chamado de região receptiva do neurônio. Da mesma forma, cada neurônio na camada de saída é conectado a um conjunto local de nós da camada escondida. A rede neural da Figura 3.4 possui o mesmo número de nós de entrada, neurônios escondidos e neurônios de saída que os da Figura 3.3. Entretanto, comparando estas duas redes neurais, observamos que a rede neural conectada localmente da Figura 3.4 tem uma estrutura especializada. Na prática, a estrutura especializada construída na modelagem da rede neural conectada localmente reflete uma informação a priori sobre as características do padrão de ativação que está sendo analisado.

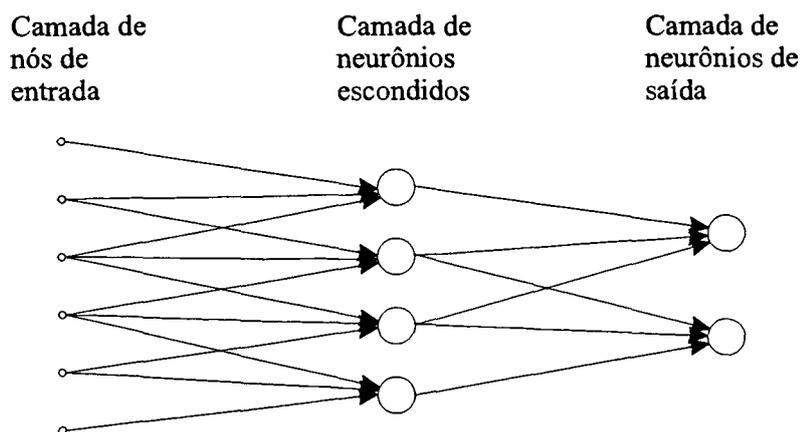


Figura 3.4: Rede neural direta multicamada conectada localmente.

### 3.3 Algoritmo de aprendizagem

Grande parte do trabalho na construção de redes neurais é determinar pesos sinápticos para que a rede neural efetue uma tarefa específica. De forma geral, podemos identificar duas formas de aprendizagem (ou treinamento) utilizados em redes neurais:

- Aprendizagem supervisionada
- Aprendizagem não supervisionada

Aprendizagem *supervisionada* é baseada na comparação direta da saída obtida pela rede neural com as respostas corretas desejadas, para uma dada entrada na rede neural. Esta forma é por vezes chamada de aprendizagem com um professor. Inclui-se o caso especial onde as saídas são classificadas apenas como corretas ou incorretas, sem a informação do valor da resposta correta desejada.

Às vezes o objetivo de aprendizagem não é definido em termos de exemplos corretos específicos. A única informação disponível está na correlação dos dados de entrada. Na aprendizagem *não supervisionada*, espera-se que a rede neural crie categorias a partir destas correlações, e produza uma saída correspondente a cada categoria de entrada.

O algoritmo de aprendizagem empregado em uma rede neural está intimamente ligado à arquitetura utilizada. Ao longo deste trabalho abordamos apenas o algoritmo de aprendizado supervisionado chamado Back-Propagation.

#### 3.3.1 Back-Propagation

O algoritmo de aprendizagem Back-Propagation, também conhecido como Regra Delta Generalizada, baseia-se em algoritmos gradiente descendente. O procedimento executa duas passadas sobre a rede neural: a propagação do sinal e a retro-propagação do erro. O algoritmo Back-Propagation é supervisionado, onde apresentamos um padrão de aprendizagem,

que é composto de uma entrada e uma saída. Denotamos este par entrada e saída por  $e$  e  $s$  respectivamente. A propagação do sinal é efetuada apresentando-se  $e$  à entrada da rede neural, o que produz uma saída. Esta saída é comparada com a saída desejada  $s$ . Se não houver diferença, não há necessidade de treinamento. Caso contrário, efetua-se a retro-propagação do erro para que se calcule fatores de atualização de pesos de modo que se reduza a diferença na saída.

O algoritmo de aprendizagem Back-Propagation requer que a função de ativação seja derivável.

A medição de erro (ou função de custo) usual é

$$Erro = \frac{1}{2} \sum_i (s_i - a_i^L)^2 \quad (3.9)$$

onde  $a^L$  é a saída da rede neural para o respectivo padrão de entrada  $e$  apresentado à entrada da rede neural. O critério de parada do algoritmo geralmente é atingir um erro pequeno o suficiente para todos os elementos do conjunto de padrões. Para o desenvolvimento matemático completo do algoritmo de aprendizagem Back-Propagation consulte [Rumelhart 1986].

Descrição em passos do algoritmo de aprendizagem Back-Propagation:

1. Inicialize os pesos ( $w^l$ ) com valores aleatórios pequenos.
2. Escolha um padrão ( $e, s$ ) dentre os possíveis padrões.
3. Aplique  $e$  à camada de entrada ( $a^0$ ) da rede neural

$$a_i^0 = e_i \quad (3.10)$$

4. Propague o sinal através da rede neural utilizando (3.7) e (3.8)

$$h_i^l = \sum_j w_{ij}^l a_j^{l-1} \quad (3.11)$$

e

$$a_i^l = f(h_i^l) \quad (3.12)$$

para cada  $l$  até que a camada de saída ( $a^L$ ) tenha sido calculada.

5. Calcule o delta para camada de saída ( $\delta^L$ )

$$\delta_i^L = (s_i - a_i^L) f'(h_i^L) \quad (3.13)$$

pela comparação do sinal obtido na camada de saída ( $a^L$ ) com o padrão desejado na camada de saída da rede neural ( $s$ ).

6. Calcule os deltas para as camadas escondidas anteriores pela retro-propagação do erro

$$\delta_i^{l-1} = \left( \sum_j w_{ji}^l \delta_j^l \right) f' (h_i^{l-1}) \quad (3.14)$$

7. Use

$$\Delta w_{ij}^l = \eta a_j^{l-1} \delta_i^l \quad (3.15)$$

onde  $\eta$  é a taxa de aprendizagem, para atualizar os pesos de acordo com

$$wnovo_{ij}^l = w_{ij}^l + \Delta w_{ij}^l \quad (3.16)$$

onde *wnovo* são os pesos da próxima iteração que substituem os pesos da iteração atual  $w$ .

8. Caso o critério de parada não tenha sido atingido, reinicie a partir do passo 2.

Extensões do algoritmo de aprendizagem Back-Propagation utilizam a idéia de *momento*. Para cada conexão podemos associar uma inércia ou momento, de modo que as atualizações de peso tendem a mover em uma direção média que minimiza o erro na saída, evitando oscilações.

## Capítulo 4

# Compartilhamento de Pesos em Redes Neurais

Ao desenvolver a estrutura de uma rede neural podemos nos valer do conhecimento a priori que possuímos do problema abordado. Este conhecimento pode permitir o uso de restrições sobre a estrutura da rede neural que auxiliem na convergência ou na generalização dos resultados. O compartilhamento de pesos em redes neurais é geralmente utilizado para se obter invariância à translação. A seguir apresentamos elementos para descrição do modelo de Rede Neural com Pesos Compartilhados, que é o corpo da arquitetura deste trabalho. Para mais informação sobre compartilhamento de pesos em redes neurais consulte: [Rumelhart 1986] (“T-C problem”), [Waibel 1989] (TDNN - Time Delay Neural Network), [LeCun 1990] e [Wilson 1992] (Iconic Network).

### 4.1 Replicação de redes neurais

Há aplicações de redes neurais que utilizam uma *janela de entrada* e uma *janela de saída* correspondente para analisar apenas uma pequena parcela de cada vez da informação a ser processada. Os dados da janela de entrada são aplicados à entrada da rede neural, resultando geralmente em apenas um nó de saída, que é armazenado na posição designada pela janela de saída. À medida que movemos a janela de entrada sobre os dados de entrada, movemos respectivamente a janela de saída também, sobre os dados de saída. A rede neural utilizada é sempre a mesma, independente da movimentação das janelas. Esta técnica pode ser vista como a *replicação* (ou compartilhamento) de uma rede neural sobre uma série de dados de entrada.

Como os dados da janela de saída dependem apenas dos dados da janela de entrada e a rede neural é sempre a mesma, uma mesma operação é efetuada ao longo de toda entrada e armazenada ao longo de toda saída. Com isto obtemos uma operação invariante à translação das janelas. A técnica de janelas é empregada em [Sejnowski 1987] (NETtalk) para determinar

o fonema a ser pronunciado em uma seqüência de letras (da língua inglesa). Veja exemplo desta rede neural na Figura 4.1.

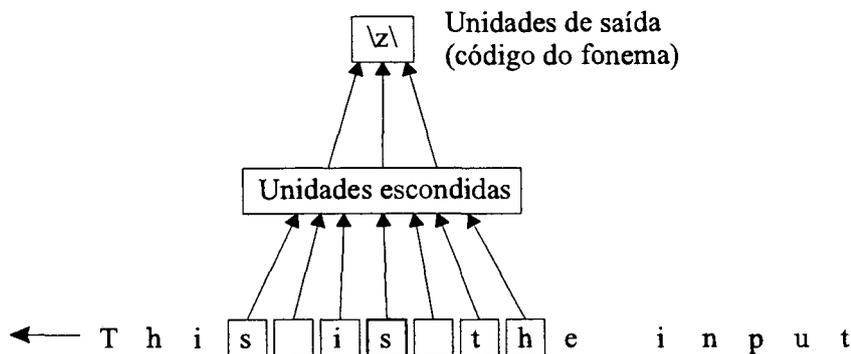


Figura 4.1: Arquitetura do NETtalk.

A técnica de janelas é também empregada no processamento de imagens (veja [Wilson 1992] e [Spreeuwers 1993]). A intensidade luminosa de cada pixel na janela de entrada pode ser associada a um determinado nível de ativação de um nó da rede neural (ou vice-versa). Quanto maior a luminosidade do pixel, maior o estado de ativação do nó da rede neural (ou vice-versa). Veja exemplo na Figura 4.2. O exemplo da Figura 4.1 opera com janelas que se movem ao longo de apenas uma dimensão, enquanto o exemplo da Figura 4.2 opera com janelas que se movem ao longo de duas dimensões.

A replicação de redes neurais com a técnica de janelas torna a ativação dos nós das camadas escondidas completamente independente. Veja ilustração da Figura 4.2. Observe a independência de ativação nos nós das camadas escondidas com a replicação de redes neurais. A convergência e divergência sináptica (veja seção 2.3) sugere que a ativação das células nervosas ao longo de caminhos neurais não seja completamente independente, mas possua um certo grau de correlação. A independência nas ativações dos nós das camadas escondidas pode na verdade significar uma redundância na extração de alguma informação, o que significaria desperdício de processamento.

## 4.2 Modelo de Rede Neural com Pesos Compartilhados

A teoria de redes neurais artificiais tem inspiração em neurônios biológicos. A organização de células nervosas observada no córtex (em particular no córtex visual) serve de inspiração para modelagem de estruturas de redes neurais artificiais com aplicações mais específicas.

No córtex observa-se que células nervosas que processam informação relacionada formam-se em grupos. Mais que agrupamento, observa-se que células nervosas ao longo de caminhos neurais geralmente apresentam organização topográfica, a partir dos receptores (veja seção 2.4). Isto sugere o uso de conexões locais entre neurônios ao longo de um caminho de uma rede neural artificial. No córtex visual observa-se a existência de células nervosas que respondem a

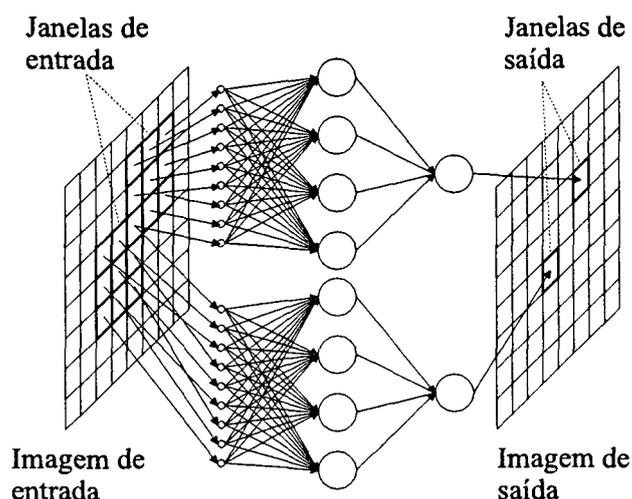


Figura 4.2: Técnica de janelas para o processamento de imagens.

certos estímulos específicos. Como exemplo temos: detetores de linhas orientadas e detetores de disparidade estereoscópica (veja seção 2.4.2). A organização topográfica e a existência de detetores de estímulos específicos sugerem a modelagem de estruturas de redes neurais artificiais que possibilitem a extração de características (features) independente da posição em que o padrão é apresentado à entrada da rede neural artificial.

Estes aspectos do córtex sugerem que problemas como o reconhecimento de padrões mais complexos, tal como letras, possa ser obtido com mecanismos similares, com uma rede neural artificial [Fukushima 1983].

*Invariância à translação* (também conhecido como *invariância ao deslocamento*) é uma propriedade que pode ser aplicada em uma série de problemas relacionados ao processamento de imagens e sinais. Nestes problemas, deseja-se que alguma característica do padrão apresentado seja extraída ou reconhecida, independente da translação ao longo de uma dimensão. Como exemplo, considere um filtro para detecção de bordas em uma imagem. É desejável que o filtro efetue a mesma operação sobre todos pontos da imagem. Desta forma, caso uma parte da imagem seja transladada, a detecção de bordas ocorrerá da mesma forma, apesar do deslocamento.

A invariância à translação pode ser obtida em redes neurais artificiais através de uma técnica chamada *compartilhamento de peso* (*weight sharing*). O compartilhamento de peso é uma restrição que obriga determinados pesos de conexões entre neurônios a possuir exatamente o mesmo valor. Desta forma, os neurônios deste conjunto podem responder de forma igual quando um padrão é apresentado a um ou a outro neurônio. Estas redes neurais são encontradas na literatura com o nome de *Rede Neural com Pesos Compartilhados* (*Shared Weight Neural Network*). A forma de treinamento empregada nestas redes neurais é geralmente uma adaptação do algoritmo de aprendizagem Back-Propagation. A partir de exemplos, do padrão

de entrada a ser processado e do respectivo padrão de saída resultante desejado, este modelo de rede neural pode ser treinado, determinando um filtro (invariante à translação) que efetue a operação requerida.

Em [Rumelhart 1986] a técnica de compartilhamento de pesos é apresentada para resolver o “T-C problem” (reconhecer a letra T ou a letra C em uma matriz de pixels). Em [Waibel 1989] (Time-Delay Neural Network - TDNN) a invariância à translação obtida pelo compartilhamento de pesos é aplicada sobre uma dimensão temporal, para o reconhecimento de características fonético acústicas. Em [LeCun 1990] a técnica de compartilhamento de pesos é utilizada para o reconhecimento de dígitos numéricos manuscritos de imagens. Como veremos adiante (veja capítulo 4), a técnica de compartilhamento de pesos pode ser vista como uma operação de convolução. Por este motivo, estas redes neurais também são chamadas de *Rede Neural por Convolução* (*Convolution Neural Network* ou *Convolutional Neural Network*). Em [Wilson 1992] (Iconic Network) é apresentado uma extensão da álgebra de matrizes para trabalhar com a operação de convolução (matrix convolution), com a qual Wilson define a Iconic Network, uma rede neural que utiliza a técnica de pesos compartilhados. Com esta extensão, a definição usual de redes neurais artificiais pode ser vista como um sub-espço da Iconic Network.

#### 4.2.1 Conexões locais, reticulado e compartilhamento de pesos

Considere uma rede neural direta conectada localmente (veja seção 3.2.2). Podemos organizar as camadas da rede neural na forma de um *reticulado* (*lattice*), onde definimos a posição de cada nó dentro deste reticulado em termos de *coordenadas*. O reticulado pode ser unidimensional, bidimensional ou multidimensional. Utilizamos **negrito** para denotar a notação vetorial - por exemplo  $\mathbf{v} = (v_1, v_2, \dots, v_m)$ . Dado um neurônio  $y(\mathbf{v})$  na posição da coordenada  $\mathbf{v}$  em uma camada, definimos um nó da vizinhança na camada anterior  $x(\mathbf{v} - \mathbf{u})$  como um nó a uma distância  $\mathbf{u}$  da posição da coordenada  $\mathbf{v}$  nesta camada. Os valores que  $\mathbf{u}$  pode assumir determinam o domínio da vizinhança. A região receptiva de  $y(\mathbf{v})$  são os nós de tal vizinhança.

Uma idéia similar à técnica de janelas apresentada na seção 4.1 pode ser utilizada sobre uma camada de neurônios da rede neural para designar um neurônio e os nós de sua região receptiva de modo a impor a restrição de conjuntos de *pesos compartilhados* entre os neurônios. Dado um neurônio de uma camada, representando um único nó da janela de saída, este possui uma região receptiva de nós na camada anterior, representando os nós de sua janela de entrada. Ao analisarmos outro neurônio desta mesma camada, que representa a movimentação da janela de saída, a região receptiva deste outro neurônio deve corresponder à movimentação da respectiva janela de entrada. Com a técnica de janelas podemos descrever o modo com o qual o compartilhamento de pesos é regido dentro de uma camada da rede neural. Um exemplo desta rede neural é apresentado na Figura 4.3. Somente dois conjuntos de pesos compartilhados é

apresentado em cada camada para facilitar visualização..

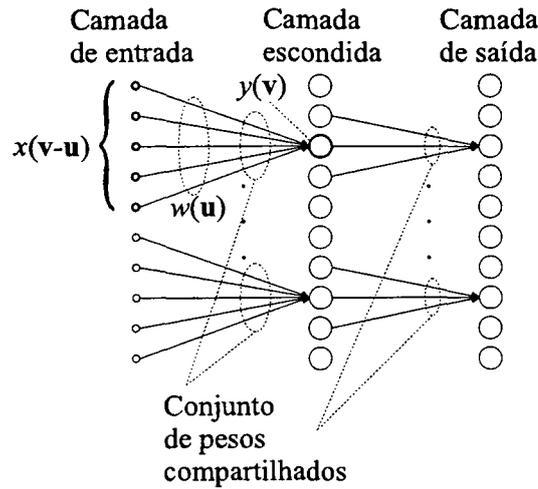


Figura 4.3: Rede neural conectada localmente com camadas organizadas na forma de reticulado unidimensional.

Para descrever a propagação de sinal da camada  $x$  para camada seguinte  $y$  em uma rede neural com camadas organizadas na forma de reticulado e compartilhamento de pesos entre os neurônios em cada camada utilizamos

$$h(\mathbf{v}) = \sum_{\mathbf{u}} w(\mathbf{u}) x(\mathbf{v} - \mathbf{u}) \quad (4.1)$$

e

$$y(\mathbf{v}) = f(h(\mathbf{v})) \quad (4.2)$$

onde  $y(\mathbf{v})$  é o sinal de ativação do neurônio da posição  $(\mathbf{v})$ ;  $f(\cdot)$  é a função de ativação;  $h(\mathbf{v})$  é o a combinação linear referente ao neurônio  $y(\mathbf{v})$ ;  $\mathbf{u}$  percorre a vizinhança da região receptiva de  $y(\mathbf{v})$ ;  $x(\mathbf{v} - \mathbf{u})$  é um nó da região receptiva do neurônio  $y(\mathbf{v})$ ; e  $w(\mathbf{u})$  é o peso sináptico do neurônio  $y(\mathbf{v})$  associado ao nó  $x(\mathbf{v} - \mathbf{u})$ .

A operação em (4.1) é conhecida como operação de convolução (veja seção 5.1.2).

A restrição de conexões locais adicionada ao compartilhamento de pesos em uma camada produz um efeito semelhante à representação topográfica descrito na seção 2.4.1. Em termos biológicos, é pouco plausível que exista algum mecanismo que imponha a restrição de pesos iguais entre as células nervosas do córtex. É muito mais provável que as células nervosas possuam padrões de conexão sináptica semelhantes devido a outros fatores, como o aprendizado - estímulos iguais apresentados em diferentes regiões da retina, por exemplo.

## 4.2.2 Canais

Ao impor a restrição de conjuntos de pesos iguais dentre os neurônios de uma mesma camada, tal camada da rede neural perde a qualidade de distinguir características diversas. Esta camada passa a extrair uma única característica, com invariância à translação. Dado um conjunto de nós alimentando um neurônio, a característica extraída deste conjunto de nós é determinado exclusivamente pelos pesos deste único neurônio. É importante que outras características sejam extraídas deste mesmo conjunto de nós, complementando as características necessárias para que a saída da rede neural aprenda efetivamente os padrões (compare com problemas não linearmente separáveis citado na seção 3.2.2).

Para que a rede neural possa extrair várias características por camada, subdividimos a camada em uma série de agrupamentos de nós que denominamos de *canais*. Cada canal tem a forma de um reticulado de nós. Dado um neurônio  $y_i(\mathbf{v})$  no canal  $i$  na posição da coordenada  $(\mathbf{v})$  em uma camada, definimos um nó da vizinhança na camada anterior  $x_j(\mathbf{v} - \mathbf{u})$  como um nó no canal  $j$  a uma distância  $(\mathbf{u})$  da posição da coordenada  $(\mathbf{v})$  neste canal. Observe que o neurônio  $y_i(\mathbf{v})$  recebe conexões proveniente de todos os canais  $j$  da camada anterior. As conexões locais são definidas ao longo do reticulado do canal, cujo domínio é determinado por  $(\mathbf{u})$ . A região receptiva de  $y_i(\mathbf{v})$  são os nós de tal vizinhança.

Com os canais, a técnica de janelas continua aplicável. Tal como uma camada foi subdividida em uma série de canais, uma janela é subdividida em uma série de sub-janelas, uma sub-janela para cada canal. Desta forma podemos designar uma série de neurônios e os nós de suas regiões receptivas para impor a restrição de compartilhamento de pesos. A série de neurônios representa a janela de saída, e a região receptiva representa a janela de entrada. Note que esta série de neurônios possui exatamente a mesma região receptiva, porém cada neurônio possui seu conjunto próprio de pesos. O compartilhamento de pesos ocorre ao longo do canal, com a movimentação das janelas de entrada e saída. A Figura 4.4 apresenta um exemplo desta rede neural, com canais organizados na forma de reticulado unidimensional.

A partir de (4.1) e (4.2) podemos descrever a propagação de sinais com canais nas camadas da rede neural por

$$h_i(\mathbf{v}) = \sum_j \sum_{\mathbf{u}} w_{ij}(\mathbf{u}) x_j(\mathbf{v} - \mathbf{u}) \quad (4.3)$$

e

$$y_i(\mathbf{v}) = f(h_i(\mathbf{v})) \quad (4.4)$$

onde  $y_i(\mathbf{v})$  é o sinal de ativação do neurônio do canal  $i$  da posição  $(\mathbf{v})$ ;  $h_i(\mathbf{v})$  é o a combinação linear referente ao neurônio  $y_i(\mathbf{v})$ ;  $j$  percorre os canais e  $\mathbf{u}$  percorre a vizinhança da região receptiva de  $y_i(\mathbf{v})$ ;  $x_j(\mathbf{v} - \mathbf{u})$  é um nó da região receptiva do neurônio  $y_i(\mathbf{v})$ ; e  $w_{ij}(\mathbf{u})$  é o peso sináptico do neurônio  $y_i(\mathbf{v})$  associado ao nó  $x_j(\mathbf{v} - \mathbf{u})$ .

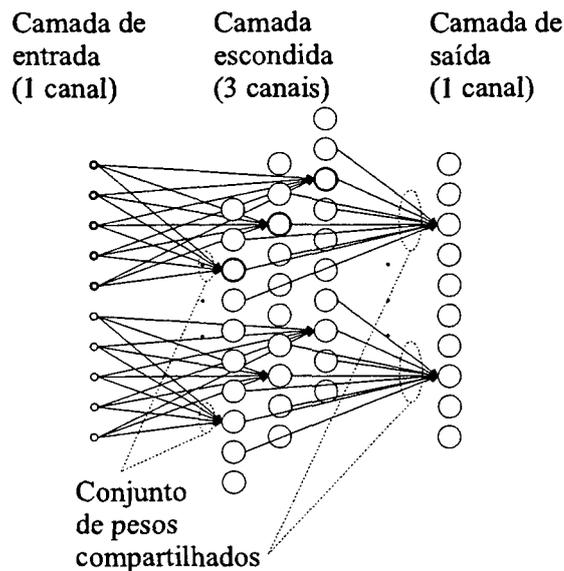


Figura 4.4: Rede neural com cada camada composta de canais.

Ao longo deste trabalho chamamos de *Rede Neural Com Pesos Compartilhados* a uma rede neural direta conectada localmente com camadas agrupadas em canais organizados na forma de reticulado e compartilhamento de pesos, tal como descrito em (4.3) e (4.4).

Com a técnica de janelas, a Rede Neural com Pesos Compartilhados é invariante à translação ao longo dos canais, entre a entrada e a saída de uma camada de neurônios. A invariância à translação é uma propriedade transitiva - se duas camadas consecutivas possuem compartilhamento de pesos, há invariância à translação ao longo dos canais da entrada da primeira camada à saída da segunda camada. Para observar a transitividade, observe a Figura 4.4: ao transladarmos um sinal ao longo dos nós dos canais na camada de entrada ocorrerá uma translação na ativação dos neurônios ao longo dos canais na camada escondida; esta ativação transladada alimenta os neurônios da camada de saída, gerando uma conseqüente translação na ativação dos neurônios ao longo dos canais na camada de saída.

### 4.3 Algoritmo de aprendizagem

A arquitetura da Rede Neural com Pesos Compartilhados permite a adaptação de diversos algoritmos de aprendizagem. A restrição de conexões locais e o compartilhamento de pesos reduz o número de variáveis livres, reduzindo o espaço de soluções com o qual o algoritmo de aprendizagem precisa lidar. Neste trabalho abordamos apenas a adaptação do algoritmo de aprendizagem Back-Propagation, que apresentamos a seguir.

### 4.3.1 Adaptação do Back-Propagation

Uma Rede Neural com Pesos Compartilhados possui uma estrutura de conexões que equivale a de uma rede neural direta conectada localmente. Esta rede pode ser vista como uma rede neural totalmente conectada cujas conexões locais possuem valores não nulos e as conexões restantes (não locais) possuem valores nulos. Portanto, a adaptação do algoritmo de aprendizagem Back-Propagation a uma rede neural direta conectada localmente restringe-se a efetuar a propagação do sinal e a retro-propagação do erro através deste sub-conjunto não nulo de conexões, com o cálculo de fatores de atualização de pesos apenas para ser aplicado a este sub-conjunto.

Além de conexões locais, uma Rede Neural com Pesos Compartilhados possui a restrição de compartilhamento de pesos sobre os canais. A restrição de compartilhamento de pesos não altera a estrutura de conexões locais, apenas obriga certas conexões a possuírem os mesmos valores. Para a adaptação do algoritmo de aprendizagem Back-Propagation isto não altera a propagação do sinal e a retro propagação do erro, pois estes procedimentos dependem apenas da estrutura de conexões. Porém, a atualização dos pesos pelos fatores de atualização deve ocorrer de modo a manter iguais os pesos compartilhados. Para tal, podemos utilizar como fator de atualização a média dos fatores de atualização associados aos pesos compartilhados. Desta forma, este fator de atualização médio mantém todos pesos compartilhados com as mesmas variações e valores (veja [Rumelhart 1986] e [Waibel 1989]).

Descrição dos passos do algoritmo de aprendizagem Back-Propagation adaptado à Rede Neural com Pesos Compartilhados (compare com algoritmo Back-Propagation - seção 3.3.1):

1. Inicialize os pesos ( $w^l$ ) com valores aleatórios pequenos.
2. Escolha um padrão ( $e, s$ ) dentre os possíveis padrões.
3. Aplique  $e$  à camada de entrada ( $a^0$ ) da rede neural

$$a_i^0(\mathbf{v}) = e_i(\mathbf{v}) \quad (4.5)$$

4. Propague o sinal através da rede neural utilizando (4.3) e (4.4)

$$h_i^l(\mathbf{v}) = \sum_j \sum_{\mathbf{u}} w_{ij}^l(\mathbf{u}) a_j^{l-1}(\mathbf{v} - \mathbf{u}) \quad (4.6)$$

e

$$a_i^l(\mathbf{v}) = f(h_i^l(\mathbf{v})) \quad (4.7)$$

para cada  $l$  até que a camada de saída ( $a^L$ ) tenha sido calculada.

5. Calcule o delta para camada de saída ( $\delta^L$ )

$$\delta_i^L(\mathbf{v}) = (s_i(\mathbf{v}) - a_i^L(\mathbf{v})) f'(h_i^L(\mathbf{v})) \quad (4.8)$$

pela comparação do sinal obtido na camada de saída ( $a^L$ ) com o padrão desejado na camada de saída da rede neural ( $s$ ).

6. Calcule os deltas para as camadas escondidas anteriores pela retro-propagação do erro

$$\delta_i^{l-1}(\mathbf{v}) = \left( \sum_j \sum_{\mathbf{u}} w_{ji}^l(\mathbf{u}) \delta_j^l(\mathbf{v} + \mathbf{u}) \right) f'(h_i^{l-1}(\mathbf{v})) \quad (4.9)$$

7. Use

$$\Delta w_{ij}^l(\mathbf{u}) = \eta \frac{1}{n} \sum_{\mathbf{v}} a_j^{l-1}(\mathbf{v}) \delta_i^l(\mathbf{u} + \mathbf{v}) \quad (4.10)$$

onde  $\eta$  é a taxa de aprendizagem e  $n$  o número de elementos somados determinado por  $\mathbf{v}$ , para atualizar os pesos de acordo com

$$wnovo_{ij}^l = w_{ij}^l + \Delta w_{ij}^l \quad (4.11)$$

onde *wnovo* são os pesos da iteração seguinte que substituem os pesos da iteração atual  $w$ .

8. Caso o critério de parada não tenha sido atingido, reinicie a partir do passo 2.

A adaptação de momento pode ser feita associando uma inércia ao conjunto de pesos compartilhados.

## 4.4 Arquiteturas relacionadas

A seguir mostramos um resumo dos trabalhos que apresentaram a arquitetura de Rede Neural com Pesos Compartilhados e as principais diferenças destas abordagens.

### 4.4.1 Rumelhart et al 1985

No texto clássico de [Rumelhart 1986] é apresentada a arquitetura de compartilhamento de pesos para resolução do problema de reconhecimento da letra "T" ou "C" ("The T-C Problem").

A rede neural apresentada possui uma camada de entrada e uma camada escondida que são organizados na forma de reticulado bidimensional e uma camada de saída que possui apenas

um único neurônio. A camada escondida possui pesos compartilhados. O único neurônio na camada de saída é totalmente conectado à camada anterior. O neurônio de saída é treinado para responder 1 se a entrada for um “T” e 0 se a entrada for um “C”. Veja a estrutura desta rede neural na Figura 4.5.

O conceito de canais não é apresentado, mas é ressaltado que a camada escondida possui um único detetor de características replicado em diferentes posições. O algoritmo de aprendizagem utilizado é uma adaptação do algoritmo Back-Propagation.

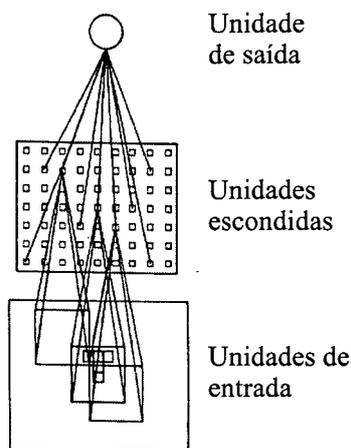


Figura 4.5: O Problema T-C.

#### 4.4.2 Waibel et al 1989

Em [Waibel 1989] é apresentado a arquitetura Time-Delay Neural Network (TDNN) para o reconhecimento de características fonético acústicas.

A rede neural apresentada possui camadas com canais organizados na forma de reticulado unidimensional, que são chamados de “units”. Cada canal da camada de entrada é modelado para operar com uma dimensão temporal com a introdução de retardos (delays) entre os nós, ao longo do reticulado. Desta forma a invariância à translação ocorre sobre uma dimensão temporal. Cada canal da camada de entrada representa a energia de uma frequência (no espectro de frequências) ao longo do tempo. A rede neural apresentada possui uma camada de entrada com 16 canais, a primeira camada escondida com 8 canais, a segunda camada escondida com 3 canais e uma camada de saída com 3 neurônios. As duas camadas escondidas possuem compartilhamento de pesos nos canais. Cada um dos 3 neurônios da camada de saída recebe a soma de um canal da segunda camada escondida. Cada neurônio da camada de saída é treinado para reconhecer a um dos fonemas: “B”, “D” ou “G”. Veja a estrutura desta rede neural na Figura 4.6.

O algoritmo de aprendizagem utilizado é uma adaptação do algoritmo Back-Propagation.

Generalizações da arquitetura TDNN específicas para o processamento de sinais são encontradas com o nome de FIR neural networks (veja [Back 1994]).

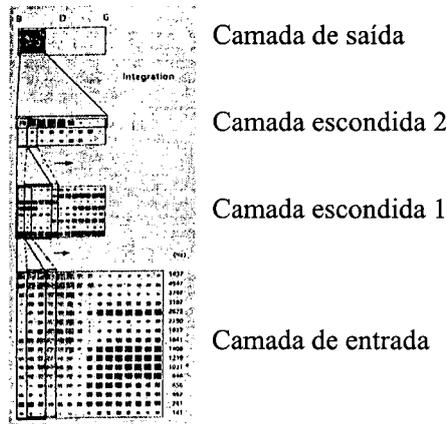


Figura 4.6: Arquitetura do Time-Delay Neural Network (TDNN).

#### 4.4.3 LeCun et al 1990

Em [LeCun 1990] é apresentado uma abordagem para o reconhecimento de dígitos numéricos manuscritos com uma rede neural que possui pesos compartilhados.

A rede neural apresentada possui camadas com canais organizados na forma de reticulado bidimensional, que são chamados de “feature maps”. Após cada “feature map” é efetuado uma redução de resolução, através de cálculo de média (averaging) e sub-amostragem (subsampling). A rede neural apresentada possui uma camada de entrada com 1 canal, a primeira camada escondida com 4 canais com compartilhamento de pesos, a segunda camada escondida com as respectivas reduções de resolução, a terceira camada escondida com 12 canais com compartilhamento de pesos, a quarta camada escondida com as respectivas reduções de resolução, e a camada de saída com 10 neurônios totalmente conectados à camada anterior. Cada neurônio da camada de saída é treinado para reconhecer um dígito numérico distinto. Veja a estrutura desta rede neural na Figura 4.7.

O algoritmo de aprendizagem utilizado é uma adaptação do algoritmo Back-Propagation.

Para o processamento simultâneo de mais de um dígito, esta rede é também chamada de Space Displacement Neural Network (SDNN) (veja [Matan 1992]).

#### 4.4.4 Wilson 1989/1992

Em [Wilson 1989] e [Wilson 1992] é apresentado a arquitetura Iconic Network. A arquitetura é utilizada para o processamento de imagens, em especial da área de Morfologia Matemática.

A arquitetura apresentada possui descrição igual a da definição de Rede Neural com Pesos Compartilhados, porém utiliza restrições sobre os valores de pesos e ativação de neurônio (a

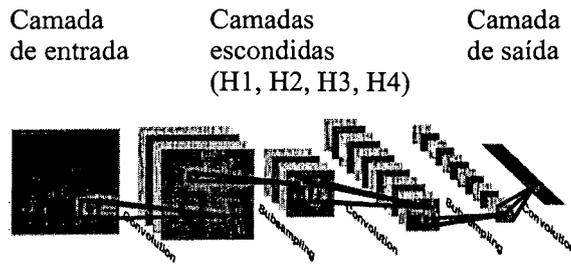


Figura 4.7: Rede neural para o reconhecimento de dígitos numéricos manuscritos.

função de ativação não é derivável). Tais restrições estão relacionadas ao uso de Morfologia Matemática. Devido a estas restrições, não é possível a adaptação do algoritmo de aprendizagem Back-Propagation. A rede neural apresentada possui camadas com canais organizados na forma de reticulado bidimensional, que são chamados de “component sublayer”.

Para o treinamento das camadas escondidas é apresentado um método de treinamento competitivo não supervisionado. Para o treinamento da camada de saída é utilizado aprendizado supervisionado baseado em Simulated Annealing, aprendizado Hebbiano e regras de aprendizado Morfológico.

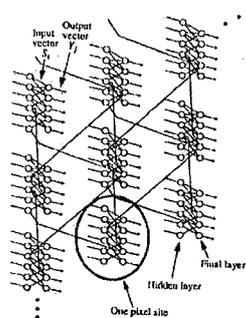


Figura 4.8: Arquitetura da Iconic Network.

## Capítulo 5

# Rede Neural por Convolução

Uma outra forma de descrever o modelo de Rede Neural com Pesos Compartilhados é utilizando a notação de convolução. A seguir apresentamos o modelo de Rede Neural por Convolução e aspectos de implementação deste modelo, efetuado neste trabalho. Esta implementação dá suporte aos requisitos necessários para a abordagem de reconstrução estéreo, o objetivo final deste trabalho (seção 7.1).

### 5.1 Convolução

Convolução é uma operação amplamente utilizada para o processamento de imagens e sinais. A partir do modelo de Rede Neural com Pesos Compartilhados (capítulo 4), esta operação é a base para a descrição da Rede Neural por Convolução (seção 5.2). As seções 5.1.1 e 5.1.2 são adaptações de [Dudgeon 1984]. Para mais informação sobre convolução consulte: [Dudgeon 1984], [Oppenheim 1989] e [Lim 1990].

#### 5.1.1 Sinais e sistemas

##### Sinais

Um *sinal discreto*, também conhecido como uma *seqüência*, é uma função definida sobre um conjunto de inteiros. Desta maneira

$$x = \{x(v), v \in Z\} \quad (5.1)$$

Um único elemento da seqüência é chamada de *amostra*. Então  $x(v)$  representa uma amostra da seqüência  $x$  no ponto  $(v)$ . Valores de amostras podem ser reais ou complexos. Se  $v$  é interpretado como variável, a referência a  $x(v)$  é interpretada como referência a toda seqüência. Um exemplo de sinal discreto é apresentado na Figura 5.1.

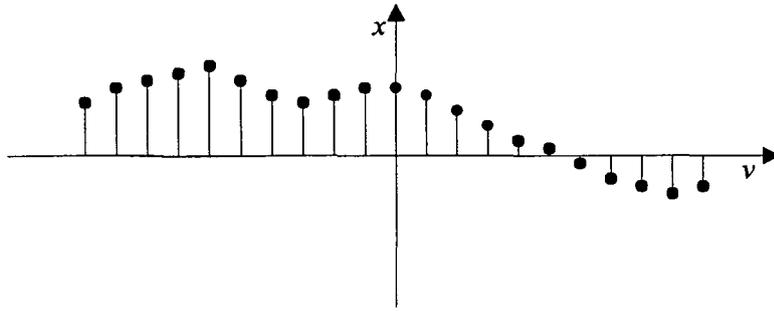


Figura 5.1: Sinal discreto (Seqüência).

Seqüências estendem-se ao infinito desde que  $v$  possa ter qualquer valor inteiro. Na prática, entretanto, a maioria das seqüências tem valores de amostragem que são conhecidos somente em uma região finita. Ao invés de restringir o domínio da definição da seqüência, podemos supor que os valores de amostragem fora da região finita possuem valores nulos.

**Operações com sinais** Sinais podem ser combinados ou alterados por uma variedade de operações.

Sejam  $x$  e  $y$  sinais discretos. Tais sinais podem ser adicionados para produzirem um terceiro sinal,  $z$ . A adição é efetuada amostra a amostra de modo que o valor particular de uma amostra  $z(v)$  é obtida pela adição do valor das duas amostras correspondentes  $x(v)$  e  $y(v)$ .

$$z(v) = x(v) + y(v) \quad (5.2)$$

Seqüências podem também ser multiplicados por uma constante para formar uma nova seqüência. Seja  $c$  uma constante. Podemos formar uma seqüência  $y$  a partir do escalar  $c$  e da seqüência  $x$ , pela multiplicação de cada valor de amostra de  $x$  por  $c$ .

$$y(v) = cx(v) \quad (5.3)$$

Uma seqüência  $x$  pode também ser linearmente deslocada para formar uma nova seqüência  $y$ . A operação de deslocamento simplesmente desliza toda a seqüência  $x$  para uma nova posição na reta ( $v$ ). Os valores das amostras de  $y$  são relacionados com os valores das amostras de  $x$  com

$$y(v) = x(v - u) \quad (5.4)$$

onde ( $u$ ) é a quantia de deslocamento. Um exemplo de deslocamento de uma seqüência é apresentada na Figura 5.2.

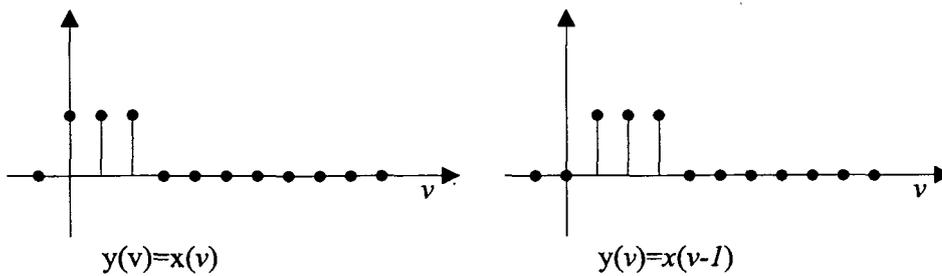


Figura 5.2: Deslocamento de uma seqüência.

### Sistemas

*Sistemas* transformam sinais. Formalmente, um sistema é um operador que mapea um sinal (a entrada) em outro sinal (a saída). A Figura 5.3 ilustra isto apresentando um sistema que mapea  $x$  em  $y$ . O operador incorporado neste sistema é representado por  $T[\cdot]$ , daí podemos escrever

$$y = T[x] \quad (5.5)$$

O operador  $T[\cdot]$  pode representar uma regra ou conjunto de regras para o mapeamento de um sinal de entrada em um sinal de saída, ou mesmo uma lista de sinais de saída que correspondem a vários sinais de entrada.

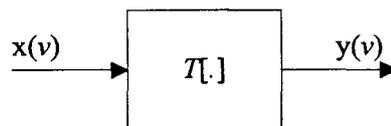


Figura 5.3: Sistema.

**Sistemas Lineares** Um sistema é dito *linear* se e somente se ele satisfaz duas condições:

- se o sinal de entrada é a soma de duas seqüências, o sinal de saída é a soma das duas seqüências correspondentes;
- ao multiplicar o sinal de entrada por um escalar, a saída produzida é também multiplicada pelo escalar.

Portanto, se  $L[\cdot]$  representa um sistema linear, e

$$y_1 = L[x_1] \quad (5.6)$$

$$y_2 = L[x_2] \quad (5.7)$$

então

$$ay_1 + by_2 = L[ax_1 + bx_2] \quad (5.8)$$

para todos sinais de entrada  $x_1$  e  $x_2$  e todas constantes  $a$  e  $b$ .

**Sistemas Invariantes ao Deslocamento** Um sistema *invariante ao deslocamento* é um para o qual um deslocamento na seqüência de entrada implica um deslocamento correspondente na seqüência de saída. Para

$$y(v) = T[x(v)] \quad (5.9)$$

o sistema  $T[\cdot]$  é invariante ao deslocamento se e somente se

$$T[x(v-u)] = y(v-u) \quad (5.10)$$

para todas as seqüências  $x$  e para todos deslocamentos inteiros ( $u$ ).

### 5.1.2 Soma de convolução

A soma de *convolução* é definida por

$$y(v) = \sum_{u=-\infty}^{\infty} x(u)h(v-u) \quad (5.11)$$

A definição da soma de convolução pode ser derivada a partir da definição de sistemas lineares e de sistemas invariantes ao deslocamento, possuindo por isso tais propriedades; dado um sistema linear e invariante ao deslocamento, este sistema pode ser completamente caracterizado por um  $h$ , que é chamado de *resposta impulsiva* - veja [Dudgeon 1984, capítulo 1]. A resposta impulsiva é por vezes chamada de *máscara de convolução* (*convolution mask*) ou *núcleo de convolução* (*convolution kernel*).

Se fizermos a substituição das variáveis  $v-u = w$ , a equação (5.11) pode ser escrita em uma forma alternativa

$$y(v) = \sum_{w=-\infty}^{\infty} h(w)x(v-w) \quad (5.12)$$

com a qual observamos que a convolução é uma operação comutativa.

Utilizamos um asterisco (\*) como notação para denotar a convolução. Desta forma, as equações (5.11) e (5.12) podem ser escritas como

$$y = x * h = h * x \quad (5.13)$$

## Soma de convolução multidimensional

A soma de convolução definida em (5.11) pode ser generalizada para sinais discretos multidimensionais. Podemos definir a soma de convolução multidimensional como

$$y(\mathbf{v}) = \sum_{\mathbf{u}} x(\mathbf{u}) h(\mathbf{v} - \mathbf{u}) \quad (5.14)$$

onde  $\mathbf{v} = (v_1, v_2, \dots, v_m)$  e  $\mathbf{u} = (u_1, u_2, \dots, u_m)$ . Utilizamos negrito para denotar a notação vetorial. Para denotar a operação de convolução multidimensional, utilizamos o mesmo asterisco (\*). Desta forma (5.14) pode ser escrito como

$$y = x * h \quad (5.15)$$

## Transformada de Fourier

É possível demonstrar-se que uma seqüência pode ser escrita como uma soma ponderada de exponenciais complexas usando a transformada multidimensional de Fourier.

Uma propriedade extremamente importante da transformada de Fourier é o *teorema de convolução*: a transformada da convolução de duas seqüências é igual ao produto de suas transformadas (veja [Dudgeon 1984, capítulo 1]). Para denotar o par da transformada de Fourier utilizamos a notação

$$x \leftrightarrow X \quad (5.16)$$

onde  $x$  (letra minúscula) é a resposta impulsiva e  $X$  (letra maiúscula) é a freqüência de resposta. Com esta notação, o teorema de convolução pode ser escrito como

$$y = h * x \leftrightarrow Y = HX \quad (5.17)$$

Outra propriedade da transformada de Fourier é a linearidade. Se

$$x_1 \leftrightarrow X_1 \quad (5.18)$$

e

$$x_2 \leftrightarrow X_2 \quad (5.19)$$

então

$$ax_1 + bx_2 \leftrightarrow aX_1 + bX_2 \quad (5.20)$$

para quaisquer números complexos  $a$  e  $b$ .

**Transformada discreta de Fourier** A transformada discreta de Fourier (DFT) é uma representação exata da transformada de Fourier para seqüências de extensão finita. A DFT possui grande importância pois pode ser calculada em um número finito de operações.

O inverso da DFT do produto da DFT de duas seqüências é chamada de convolução circular. A soma de convolução (também chamada de convolução linear) pode ser calculada utilizando-se a convolução circular (veja [Dudgeon 1984, capítulo 2]). A DFT pode ser calculada de forma eficiente através dos algoritmos de transformada rápida de Fourier (FFT) (veja [Dudgeon 1984, capítulo 2]).

### 5.1.3 Vetores de sinais

Sinais podem ser agrupados na forma de vetores, o que pode ser útil em diversas situações. Podemos definir operações básicas com vetores de sinais, como soma e produto pontual.

Seja  $x$  e  $y$  vetores de seqüências

$$x = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} \quad (5.21)$$

$$y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} \quad (5.22)$$

onde cada  $x_i$  e  $y_i$  é uma seqüência.

A soma de vetores de seqüências é definida como a soma ponto a ponto

$$z_i = x_i + y_i \quad (5.23)$$

onde as duas seqüências são somadas amostra a amostra.

De forma similar, o produto pontual de vetores de seqüências é definido como o produto ponto a ponto

$$z_i = x_i y_i \quad (5.24)$$

onde as duas seqüências são multiplicadas amostra a amostra.

Sistemas podem também operar com vetores de sinais. Um caso particular destes sistemas pode ser visto como o produto de um vetor de sinais de entrada por uma matriz de impulsos de resposta, resultando em um vetor de sinais de saída.

Seja  $a$  um vetor de impulsos de resposta

$$a = \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & & \\ \vdots & & \ddots & \\ a_{m1} & & & a_{mn} \end{bmatrix} \quad (5.25)$$

onde cada  $a_{ij}$  é uma resposta impulsiva.

Podemos definir o produto matricial de vetores de seqüências como

$$y_i = \sum_j a_{ij} * x_j \quad (5.26)$$

onde  $y$  é o vetor de seqüências resultante.

Para denotar o produto matricial de vetores de seqüências utilizamos um asterisco (\*), de modo que 5.26 é escrito como

$$y = a * x \quad (5.27)$$

Esta definição pode ser vista como uma extensão do produto de vetores e matrizes usual, onde a operação de convolução substitui a multiplicação de escalares (veja: [Dudgeon 1984, capítulo 1] - Vector Input-Output Systems; [Wilson 1992] - Matrix Convolution).

### Transformada de Fourier

Para calcular o produto matricial de vetores de seqüências podemos utilizar o teorema de convolução da transformada de Fourier. A princípio, para o cálculo do produto matricial de vetores de seqüências seria necessário duas transformadas de Fourier e uma inversa de transformada de Fourier para cada operação de convolução. Porém, podemos utilizar um número menor de transformadas de Fourier observando que a transformada de Fourier de uma seqüência do vetor de seqüências pode ser reutilizado entre os somatórios. Podemos ainda reduzir o número de inversas de transformadas de Fourier utilizando a propriedade de linearidade da transformada de Fourier, aplicando a inversa da transformada de Fourier somente após o somatório. Desta forma, podemos descrever o produto matricial de vetores de

seqüência em:

1. Cálculo das transformadas de Fourier

$$\begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} \leftrightarrow \begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_n \end{bmatrix} \quad (5.28)$$

$$\begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & & \\ \vdots & & \ddots & \\ a_{m1} & & & a_{mn} \end{bmatrix} \leftrightarrow \begin{bmatrix} A_{11} & A_{12} & \dots & A_{1n} \\ A_{21} & A_{22} & & \\ \vdots & & \ddots & \\ A_{m1} & & & A_{mn} \end{bmatrix} \quad (5.29)$$

2. Somatório dos produtos das transformadas de Fourier

$$\begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_m \end{bmatrix} = \begin{bmatrix} A_{11}X_1 + A_{12}X_2 + \dots + A_{1n}X_n \\ A_{21}X_1 + A_{22}X_2 + \dots + A_{2n}X_n \\ \vdots \\ A_{m1}X_1 + A_{m2}X_2 + \dots + A_{mn}X_n \end{bmatrix} \quad (5.30)$$

3. Cálculo das inversas das transformadas de Fourier

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \end{bmatrix} \leftrightarrow \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_m \end{bmatrix} \quad (5.31)$$

## 5.2 Modelo de Rede Neural por Convolação

O modelo de Rede Neural com Pesos Compartilhados pode ser rescrito utilizando a notação de convolação. A partir de (4.3) e (4.4), utilizando a notação de convolação de (5.15), temos

$$h_i = \sum_j w_{ij} * x_j \quad (5.32)$$

e

$$y_i = f(h_i) \quad (5.33)$$

onde  $y_i$  são os sinais de ativação dos neurônios do canal  $i$ ;  $h_i$  são as combinações lineares dos neurônios do canal  $i$ ;  $x_j$  são os sinais de ativação dos nós do canal  $j$ ; e  $w_{ij}$  são os pesos compartilhados (máscara de convolução) dos neurônios do canal  $i$  referente aos sinais de entrada dos nós do canal  $j$ .

A Figura 5.4 apresenta a propagação de sinal de (5.32) e (5.33). Cada canal de neurônios recebe informação dos canais da camada anterior ponderado pela máscara de convolução respectiva. O modelo de Rede Neural por Convolução é definido com neurônios com conexões locais, porém do ponto de vista de canais, cada canal de neurônios recebe conexões de todos os canais da camada anterior (é completamente conectado). Quando o reticulado possui dimensão zero, ou seja é apenas um número escalar, a descrição (5.32) e (5.33) recai para definição usual de redes neurais como a apresentada em (3.5) e (3.6) (veja [Wilson 1992, pág. 132]). Neste caso, um canal equivale a um nó da rede neural.

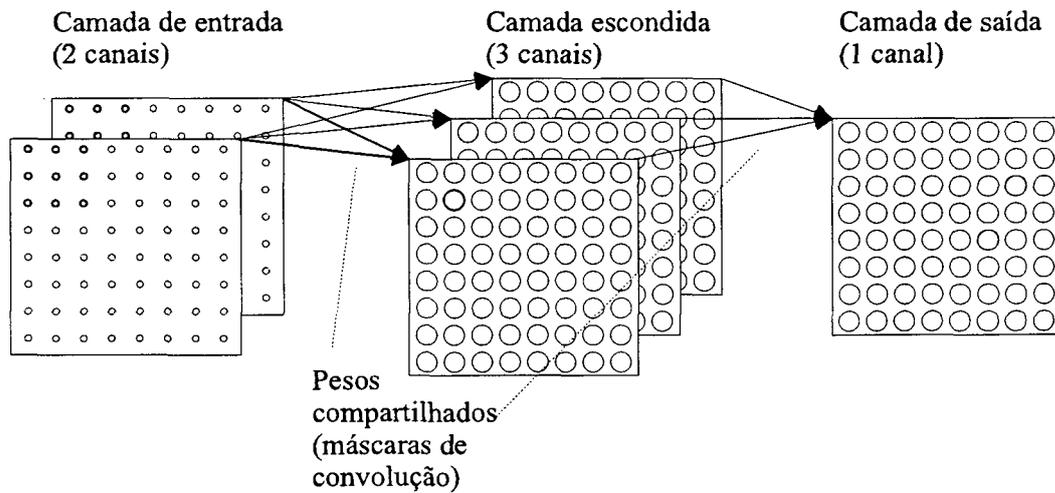


Figura 5.4: Rede Neural por Convolução (RNC).

Utilizando a notação do produto matricial de vetores de seqüência em (5.27) a propagação de sinal descrita em (5.32) e (5.33) pode ser escrita com

$$h = w * x \quad (5.34)$$

e

$$y = f(h) \quad (5.35)$$

(veja [Wilson 1992, pág. 132]).

Ao longo deste trabalho chamamos de *Rede Neural por Convolução* (RNC) a uma rede neural tal como a descrita nesta seção. O modelo de Rede Neural por Convolução equivale ao modelo de Rede Neural com Pesos Compartilhados, porém com uma notação mais concisa.

## 5.3 Algoritmo de aprendizagem

### 5.3.1 Adaptação do Back-Propagation

O algoritmo de aprendizagem Back-Propagation adaptado ao modelo de Rede Neural com Pesos Compartilhados pode ser rescrito utilizando a notação de convolução. A partir dos passos descritos na seção 4.3.1, utilizando a notação de convolução de (5.15), temos

1. Inicialize os pesos ( $w^l$ ) com valores aleatórios pequenos.
2. Escolha um padrão ( $e, s$ ) dentre os possíveis padrões.
3. Aplique  $e$  à camada de entrada ( $a^0$ ) da rede neural

$$a_i^0 = e_i \quad (5.36)$$

4. Propague o sinal através da rede neural com

$$h_i^l = \sum_j w_{ij}^l * a_j^{l-1} \quad (5.37)$$

e

$$a_i^l = f(h_i^l) \quad (5.38)$$

para cada  $l$  até que a camada de saída ( $a^L$ ) tenha sido calculada.

5. Calcule o delta para camada de saída ( $\delta^L$ )

$$\delta_i^L = (s_i - a_i^L) f'(h_i^L) \quad (5.39)$$

pela comparação do sinal obtido na camada de saída ( $a^L$ ) com o padrão desejado na camada de saída da rede neural ( $s$ ).

6. Calcule os deltas para as camadas escondidas anteriores pela retro-propagação do erro

$$\delta_i^{l-1} = \left( \sum_j w_{ji}^l * \delta_j^l \right) f'(h_i^{l-1}) \quad (5.40)$$

onde

$$\delta_i^l(\mathbf{v}) = \delta_j^l(-\mathbf{v})$$

7. Use

$$\Delta w_{ij}^l = \eta \frac{1}{n} (a_j^{l-1} * \delta i_i^l) \quad (5.41)$$

onde

$$\delta i_j^l(\mathbf{v}) = \delta_j^l(-\mathbf{v})$$

e  $\eta$  é a taxa de aprendizagem, para atualizar os pesos de acordo com

$$w_{novo}^l_{ij} = w_{ij}^l + \Delta w_{ij}^l \quad (5.42)$$

onde *w<sub>novo</sub>* são os pesos da próxima iteração que substituem os pesos da iteração atual *w*.

8. Caso o critério de parada não tenha sido atingido, reinicie a partir do passo 2.

Nos passos 6 e 7 temos uma operação conhecida por correlação a qual, através de substituição de variáveis, é escrita na forma de operação de convolução.

### 5.3.2 RNC e Transformada de Fourier

Na adaptação do algoritmo de aprendizagem Back-Propagation, a operação de maior custo computacional é a operação de convolução. A operação de convolução ocorre nos passos 4, 6 e 7, operando com os pesos ou resultando na taxa de atualização destes. Estas operações podem ser implementadas utilizando o produto matricial de vetores de seqüências, tal como descrito na seção 5.1.3. Dada a dualidade resposta impulsiva e impulso de frequência da transformada de Fourier, é possível manter os pesos (*w*) exclusivamente no domínio da frequência (*W*). A vantagem de manter os pesos no domínio da frequência é a redução do número de transformadas de Fourier utilizadas.

## 5.4 Implementação

A implementação do modelo de Rede Neural por Convolução foi efetuada utilizando o ambiente de desenvolvimento Matlab versão 5.1 da Mathworks sobre sistema operacional Windows 95.

Esta implementação do modelo de Rede Neural por Convolução permite:

- modelagem de topologias com um número qualquer de camadas escondidas;
- modelagem de máscaras de convolução com um número qualquer de dimensão.

A implementação utiliza funções do toolbox de processamento de sinais. A função de convolução do toolbox de processamento de sinais implementa a definição de soma de convolução, sendo adequado para máscaras de convolução de tamanho reduzido. Alternativamente a função de convolução é implementada utilizando os algoritmos da transformada rápida de Fourier, que é mais eficiente quando as máscaras de convolução possuem tamanho maior.

O módulo principal da implementação possui aproximadamente 30 funções com 1300 linhas de código.

Para melhor organização, ao abordar um problema, dividimos a estruturação deste em quatro scripts:

- netdef.m - definição da topologia da rede neural;
- initw.m - inicialização dos pesos da rede neural;
- initpat.m - inicialização dos padrões de entrada e saída da rede neural;
- exec.m - execução da rede neural, com parâmetros de taxa de aprendizagem, momento, etc.

Para exemplificar a criação de novas topologias, apresentamos o arquivo de definição de topologia (netdef.m) de um problema de detecção de bordas:

```

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% Julio M. Otuyama - 10/1999
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

function [net]=netdef()

disp('inicializando topologia da rede');

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% Valores utilizados no dimensionamento da rede neural
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
net.value={
    'x',      32
    'y',      32
    'feature', 4
    'pat',    1
};

net.nlayer=0;

```

```
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
```

```
% camada
```

```
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
```

```
% act - ativacao
```

```
net.nlayer=net.nlayer+1;
```

```
net.act{net.nlayer}={'x','y','pat'};
```

```
% w - pesos
```

```
net.w{net.nlayer}.pat={'pat'};
```

```
net.w{net.nlayer}.pre={};
```

```
net.w{net.nlayer}.pos={'feature'};
```

```
net.w{net.nlayer}.conv={
```

```
    'x',          9
```

```
    'y',          9
```

```
};
```

```
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
```

```
% camada
```

```
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
```

```
% act - ativacao
```

```
net.nlayer=net.nlayer+1;
```

```
net.act{net.nlayer}={'x','y','feature','pat'};
```

```
% w - pesos
```

```
net.w{net.nlayer}.pat={'pat'};
```

```
net.w{net.nlayer}.pre={'feature'};
```

```
net.w{net.nlayer}.pos={};
```

```
net.w{net.nlayer}.conv={
```

```
    'x',          5
```

```
    'y',          5
```

```
};
```

```
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
```

```
% camada
```

```
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
```

```
% act - ativacao
```

```
net.nlayer=net.nlayer+1;
```

```
net.act{net.nlayer}={'x','y','pat'};
```

No início do script há rótulos que são associados a números. Estes rótulos são utilizados para manter a consistência do tamanho do reticulado ao longo das camadas da rede neural. O tamanho de cada dimensão de uma camada é dado em termos dos rótulos.

É obrigatório que exista no mínimo duas camadas (camada de entrada e camada de saída). Cada camada é associada aos pesos que a conectam à camada anterior. Para adicionar ou remover camadas, basta adicionar ou remover um conjunto consecutivo de definição de “pesos” e de “ativação”.

Para este exemplo, a camada de entrada possui as dimensões  $32 \times 32 \times 1$  (dimensão  $x$ , dimensão  $y$ , e dimensão de padrões); a camada escondida possui dimensões  $32 \times 32 \times 4 \times 1$  (dimensão  $x$ , dimensão  $y$ , dimensão de canais de características, e dimensão de padrões); e a camada de saída possui dimensões  $32 \times 32 \times 1$  (dimensão  $x$ , dimensão  $y$ , e dimensão de padrões). As máscaras de convolução associadas aos pesos de cada canal da camada escondida possuem dimensões  $9 \times 9$  e as máscaras de convolução associadas aos pesos do canal da camada de saída possuem dimensões  $5 \times 5$ .

Uma vez definida a topologia da rede neural, a inicialização dos padrões de entrada e saída e dos pesos utiliza estas definições.

## Capítulo 6

# Reconstrução Estéreo

Reconstrução estéreo, também chamado de visão estéreo, é o ramo da visão computacional que analisa o problema da reconstrução da informação tridimensional de objetos a partir de um par de imagens capturadas com um deslocamento lateral, simultaneamente. A reconstrução estéreo tem aplicações que vão da robótica à análise de imagens aéreas. A seguir apresentamos os elementos utilizados na modelagem de problemas de reconstrução estéreo, que servem de base para modelagem de uma Rede Neural por Convolução para reconstrução estéreo. Para mais informação sobre reconstrução estéreo consulte: [Weinshall 1994] e [Haralick 1992].

### 6.1 Introdução

Os seres humanos, entre outros animais<sup>1</sup>, possuem um par de olhos localizados na parte frontal da cabeça, o que resulta na aquisição de duas imagens do mundo externo com um pequeno deslocamento lateral. Tal deslocamento gera pequenas diferenças entre as imagens, quase imperceptíveis quando observadas separadamente. Mas são justamente estas pequenas diferenças que nos permitem ter uma viva percepção tridimensional do mundo externo. A estas diferenças chamamos de *disparidade*. Chamamos de *par estéreo* o par de imagens capturadas simultaneamente com um deslocamento de posição da câmera - veja exemplo na Figura 6.1 (foto extraída da Missão Mars Pathfinder). Para ter a percepção tridimensional do par estéreo, seu olho direito deve mirar a imagem direita e seu olho esquerdo deve mirar a imagem esquerda. A sensação resultante é a de se observar apenas uma imagem, com uma nítida sensação de volume dos objetos. Outras formas de se visualizar pares estéreos baseiam-se em equipamentos como o stereoscope e óculos com polarizadores.

---

<sup>1</sup>Na natureza, há muitos animais que possuem um par de olhos localizados na parte frontal da cabeça. Tal característica é típica de caçadores, tal como leões, gatos, gaviões, corujas. Isto é importante para sobrevivência de sua espécie pois tais animais precisam de uma correta percepção da distância da presa, o que representa o sucesso ou fracasso na captura da caça. Por outro lado, animais que tipicamente são presas de caçadores, tais como coelhos, gazelas ou roedores, têm olhos posicionados lateralmente, o que não lhes dá uma boa percepção de distância, mas lhes permite observar continuamente praticamente tudo que há em sua volta. Ao menor sinal de perigo batem em retirada, o que garante sua sobrevivência.

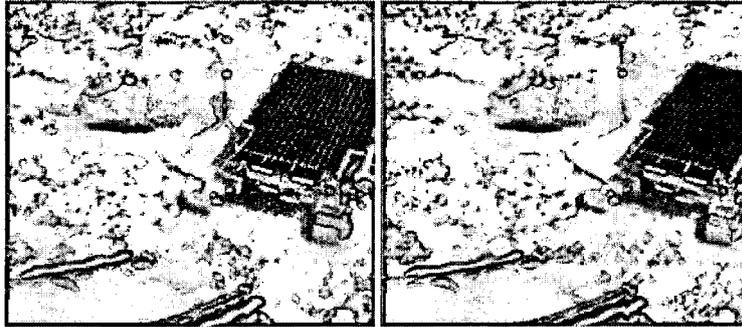


Figura 6.1: Par estéreo.

A partir de duas câmeras posicionadas de forma adequada, com suas localizações e direcionamentos conhecidos, é possível determinar a localização de qualquer ponto neste espaço, desde que este ponto possa ser reconhecido dentro de cada uma das imagens do par estéreo, com as respectivas posições nestas imagens. Por isso, é necessário que haja uma região comum, de proporção razoável, presente em ambas imagens. A determinação da localização de um ponto dentro deste espaço pode ser obtido por interpretação geométrica (triangulação). Chamamos de *stereopsis*<sup>2</sup> a capacidade de determinação da distância de profundidade baseado na informação de um par estéreo.

## 6.2 Casamento estéreo

O maior problema para reconstrução da informação tridimensional a partir de um par estéreo está no fato de capturarmos imagens de uma infinidade de *elementos* distintos, que devem ser pareados para determinação de suas localizações neste espaço tridimensional. Neste contexto, elemento é alguma propriedade que relacione a informação contida em uma imagem do par estéreo com um ponto no espaço tridimensional. A escolha desta propriedade deve permitir que, dado um ponto no espaço, este ponto se reflita em um elemento correspondente em cada imagem do par estéreo. Ao problema de correspondência de elementos chamamos de *casamento estéreo*.

Para efetuar o casamento estéreo, geralmente supõe-se que os elementos a serem casados presentes no par estéreo obedeçam a algumas regras:

<sup>2</sup>Os princípios de disparidade e stereopsis foram enunciados pela primeira vez em 1838 por Sir Charles Wheatstone, o mesmo inventor da ponte de Wheatstone. Em seu artigo, Wheatstone descreve como Leonardo da Vinci quase descobriu tais princípios. Leonardo atribuía a sensação de profundidade que resultava do uso de dois olhos ao fato de vermos um pouco além ao redor dos objetos, pela esquerda com o olho esquerdo e pela direita com o olho direito. Como exemplo de um objeto sólido, ele escolheu uma esfera - ironicamente o único objeto cuja forma permanece a mesma quando vista de direções distintas. Wheatstone observa que se Leonardo tivesse escolhido um cubo ao invés de uma esfera ele teria percebido com certeza que as duas projeções das retinas são diferentes, e que estas diferenças envolviam deslocamentos horizontais (citado em [Hubel 1995]).

- Similaridade - Dentre os diversos elementos presentes no par estéreo, esta regra dá preferência ao casamento de elementos similares (ou idênticos) presente em ambas imagens.
- Unicidade - Uma vez que os elementos presentes no par estéreo são obtidos do mundo real, pode-se supor que para um elemento presente em uma imagem, exista no máximo um único casamento correspondente na outra imagem.
- Continuidade - Supondo que os objetos capturados no par estéreo possuem superfícies contínuas, dado um elemento em uma imagem cujo par foi determinado na outra imagem, para um outro elemento próximo na primeira imagem é razoável supor que seu par na outra imagem esteja na vizinhança do primeiro par.
- Ordem seqüencial - Quando dois elementos estão posicionados um ao lado do outro em uma imagem, e um dos elementos, digamos o da esquerda, é casado com seu par na outra imagem, pode-se supor que o elemento da direita é casado com um par que está posicionado também à direita, em relação ao primeiro par. Este casamento preserva a ordem dos elementos em ambas imagens.

As duas últimas suposições geralmente são verdadeiras, porém podem falhar com contra exemplos. Experimentos mostram que a segunda suposição não é verdadeira para reconstrução estéreo ocorrida no córtex visual (veja Braddick - citado em [Marr 1982]).

Entre as dificuldades de se efetuar o casamento estéreo, temos:

- Regiões com pouca ou nenhuma variação de intensidade luminosa, o que dificulta a extração de qualquer informação que distinga dois elementos distintos.
- Regiões parcialmente ocultas. São regiões que aparecem em uma imagem, mas estão ocultas atrás de algum objeto na outra imagem. Desta forma alguns elementos em uma imagem podem ou não ser pareados com um elemento da outra imagem.
- Efeitos de reflexão e refração sobre ou através dos objetos. Como as imagens são capturadas com um pequeno deslocamento, as características de reflexão ou refração dos objetos podem gerar intensidades luminosas distintas, para um mesmo ponto, em cada uma das imagens.

### 6.3 Geometria epi-polar

A geometria básica envolvida no processo de reconstrução estéreo adiciona uma restrição extremamente útil para o processo de busca de casamento estéreo. Pode-se mostrar através de geometria básica que um ponto localizado no espaço tridimensional encontra-se sobre uma *linha epi-polar* em cada imagem. A linha epi-polar é determinada a partir de dois planos: o plano que corta o próprio ponto no espaço tridimensional e o par de pontos de captura da

imagem; o segundo plano é o plano imaginário onde encontra-se o par estéreo. A região de interseção destes dois planos define a linha epi-polar. A restrição epi-polar permite que a busca do casamento estéreo seja efetuada sobre uma linha, reduzindo o casamento estéreo a uma busca unidimensional sobre o par estéreo.

A disparidade pode ser observada sobre a linha epi-polar como a distância entre cada elemento pertencente ao par estéreo - veja Figura 6.2. Supondo que a distância do par de pontos de captura do par estéreo até os pontos localizados no espaço tridimensional seja muito maior que as distâncias entre estes pontos no espaço tridimensional, dado um ponto em uma determinada profundidade, sua disparidade corresponde a um valor que é invariante ao deslocamento lateral do ponto. Caso haja uma movimentação lateral do ponto no espaço tridimensional, sem alteração de sua profundidade, haverá uma correspondente movimentação lateral da posição da disparidade, porém o valor da disparidade se mantém inalterado. A partir disto, podemos modelar o problema de reconstrução estéreo como um problema invariante à translação lateral dos pontos no espaço tridimensional. Uma vez determinado a disparidade dos pontos, é possível construir um mapa de profundidades correspondente.

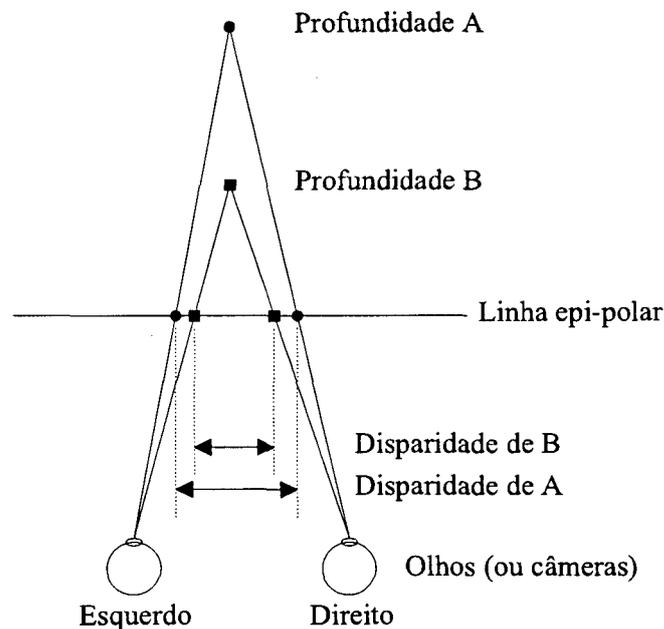


Figura 6.2: Linha epi-polar e disparidade.

## 6.4 Abordagens computacionais

Há várias abordagens computacionais para o problema da reconstrução estéreo. Podemos dividi-las basicamente em três categorias:

- Baseado em área;

- Baseado em características;
- Banco de Filtros.

Baseado em área - dado duas visões de uma pequena parte de uma superfície de um objeto no espaço tridimensional, é razoável supor que estas visões possuam certa similaridade. Os algoritmos baseado em área comparam áreas da imagem de um par estéreo com possíveis áreas correspondentes da outra imagem do par estéreo. A medida de similaridade se baseia na intensidade luminosa presente dentro de cada área.

Baseado em características - dado um objeto, há características destes objetos às quais podemos dar um significado físico, tal como bordas ou quinas. A partir das imagens do par estéreo, podemos efetuar um pré processamento para extrair estas características, às quais chamamos de primitivas, associando-as a atributos como posição, direção, informações sobre sua vizinhança, etc. O casamento estéreo busca primitivas com similaridade baseada em seus atributos.

Banco de Filtros - esta abordagem baseia-se na idéia de aplicar a operação de convolução sobre as imagens direita e esquerda do par estéreo com banco de filtros lineares ajustados a um número de diferentes orientações e escalas (veja [Weinshall 1994]). Os filtros são modelados de acordo com observações das células nervosas do córtex visual primário. Para cada posição na imagem, este banco de filtros gera um vetor que caracteriza os elementos. O casamento estéreo busca vetores de similaridade máxima.

## Capítulo 7

# RNC para Reconstrução Estéreo

### 7.1 Detecção de disparidades

O uso de Redes Neurais por Convolução concentra-se principalmente em problemas de reconhecimento de padrões - veja seção 4.4.

O problema de reconstrução estereo pode ser abordado em termos do reconhecimento de elementos a uma determinada disparidade (veja detecção de disparidades, seção 2.4.2). Podemos descrever o padrão de aprendizagem da rede neural como um par de canais de entrada, representando o par estereo, e um canal de saída, representando o mapa de disparidade de uma profundidade. Para obter outras disparidades, basta incluir mais canais à saída, um canal para cada disparidade. Cada imagem do par estereo é associada a um canal na entrada da rede neural, de forma que a intensidade luminosa de cada pixel é associada a um determinado nível de ativação no nó correspondente na rede neural. Quanto maior a luminosidade do pixel, maior o estado de ativação no nó correspondente. Na saída da rede neural, os nós do canal representando o mapa de disparidade devem ter ativações quando ocorre a detecção de disparidade de dois elementos correspondentes no par estereo. Veja exemplo na seção 7.2.1.

#### 7.1.1 Pré processamento

Uma topologia como a apresentada na Figura 5.4 pode ser utilizada para modelar o problema de reconstrução estereo. Porém, podemos refinar a topologia da rede neural com informações específicas do problema abordado. A observação do caminho neural da visão, da retina ao córtex visual, indica que haja um processamento prévio de informação antes da detecção de disparidades (veja dominância ocular - seção 2.4.2). Algoritmos de reconstrução estereo efetuam um pré processamento em cada imagem do par estereo para extrair primitivas, para então efetuar o casamento estereo destas primitivas (veja seção 6.4). Isto sugere que seja importante um processamento prévio das informações de cada imagem do par estereo, e este processamento prévio deve ser aplicada de forma igual a ambas imagens.

Podemos modelar uma topologia na rede neural que reflita este processamento prévio para cada imagem do par estéreo utilizando o conceito de compartilhamento - veja Figura 7.1. Cada canal de entrada (imagem direita e esquerda) lança sinais para camada escondida, de modo que as máscaras de convolução são compartilhadas. Os sinais dos canais direito e esquerdo se mantêm independentes e processados da mesma forma entre a camada de entrada e a camada escondida.

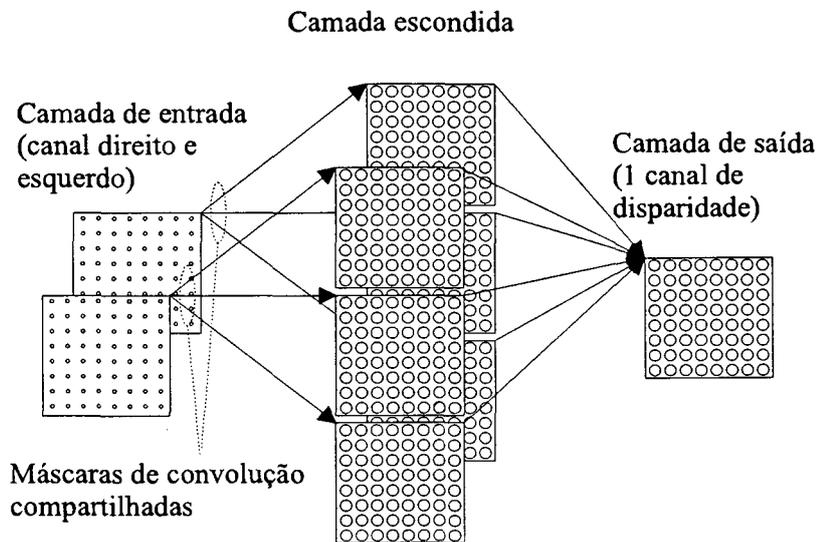


Figura 7.1: RNC para reconstrução estéreo.

Esta abordagem é similar à abordagem computacional de Banco de Filtros (veja seção 6.4). A abordagem computacional de Banco de Filtros modela máscaras de convolução para obter respostas a diferentes orientações e escalas espaciais, similar às respostas observadas no córtex visual primário. A abordagem com Rede Neural por Convolução não toma suposições a respeito das ativações ou máscaras de convolução; porém é desejável que as camadas escondidas, após treinamento, respondam de forma similar às respostas do córtex visual primário.

### 7.1.2 Compartilhamento de máscaras de convolução

O compartilhamento de máscaras apresentado na Figura 7.1 pode ser modelado com a mesma estrutura de Rede Neural por Convolução, utilizando um artifício para obtenção deste compartilhamento. O par de canais da camada de entrada pode ser visto como um reticulado tridimensional - a primeira e a segunda dimensão percorrem a imagem vertical e horizontalmente, e a terceira dimensão percorre um par ou outro. De forma similar, um par de reticulados bidimensionais da camada escondida pode ser visto como um reticulado tridimensional, que representa um canal tridimensional da camada escondida. Utilizando uma máscara de convolução cuja terceira dimensão seja unitária tem o efeito de uma máscara de convolução bidimensional que é compartilhado ao longo da terceira dimensão.

## 7.2 Resultados

### 7.2.1 Detecção de disparidades

Para o aprendizado de disparidade foi utilizado um conjunto de padrões de entrada e saída apresentado na Figura 7.2. Cada imagem possui tamanho 32x32 pixels. As máscaras de convolução da camada escondida possuem tamanho 9x9 e as da camada de saída possuem tamanho 5x5. Há 6 canais para extração de características de cada imagem do par estéreo na camada escondida. A primeira e segunda coluna são os padrões apresentados à camada de entrada (canais esquerdo e direito). A terceira coluna é o padrão desejado na camada de saída (canal de disparidade). Em (a), (b), (c), (d), (e) e (f) temos seis padrões para aprendizagem.

O estado de ativação dos neurônios é associado a intensidades luminosas, nas figuras que apresentam canais de camadas de neurônios. Quanto maior o estado de ativação no nó, maior a luminosidade do pixel correspondente na imagem.

Estes padrões foram escolhidos para se aprender a detectar disparidades a partir de linhas (padrões (a), (b) e (c)) e bordas (padrões (d), (e) e (f)). Os padrões (a) e (d) possuem disparidade correta (com resposta de disparidade). Os padrões (b) e (e) possuem disparidade próxima (sem resposta de disparidade). Os padrões (c) e (f) possuem disparidade distante (sem resposta de disparidade).

Os pares estéreo de disparidade próxima e disparidade distante podem ser gerados a partir de um par estéreo original através de deslocamento lateral das imagens. Para obter um par estéreo que possui disparidades mais próximas que o de um original, deslocamos o par esquerdo para direita e o par direito para esquerda. Quanto maior o deslocamento, maior é a variação de disparidade resultante. De forma similar, porém com direções contrárias, para obter um par estéreo que possui disparidades mais distantes que o de um original, deslocamos o par esquerdo para esquerda e o par direito para direita. Da mesma forma, quanto maior o deslocamento, maior é a variação de disparidade resultante.

Utilizando uma taxa de aprendizagem de 0,0004 e um momento de 0,5, após 1000 iterações obteve-se os resultados apresentados na Figura 7.3. A Figura 7.4 apresenta o erro na saída da rede neural ao longo das épocas. Em um computador IBM PC Pentium II 300 MHz com 32 Mb de memória RAM a execução deste treinamento dura 7320 segundos (aproximadamente 2 horas). Após o treinamento os pesos da rede neural estão determinados. Em um computador IBM PC Pentium II 300 MHz com 32 Mb de memória RAM a execução da propagação do sinal da entrada (par estéreo) à saída (mapa de disparidade) da rede neural para dura 2.1 segundos.

Na Figura 7.5 são apresentados ativações dos canais da camada escondida. Como a extração de características ocorre de forma igual entre a camada de entrada e a camada escondida para a imagem direita e esquerda do par estéreo, apresentamos os canais referentes a apenas um dos lados. A primeira coluna são ativações de um canal da camada de entrada. As outras

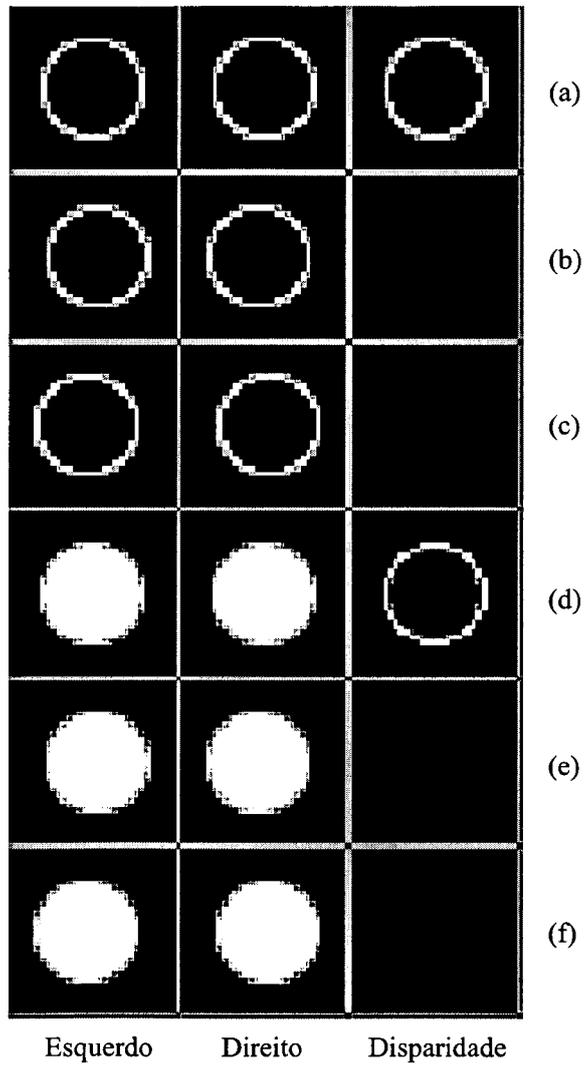


Figura 7.2: Padrões de aprendizado.

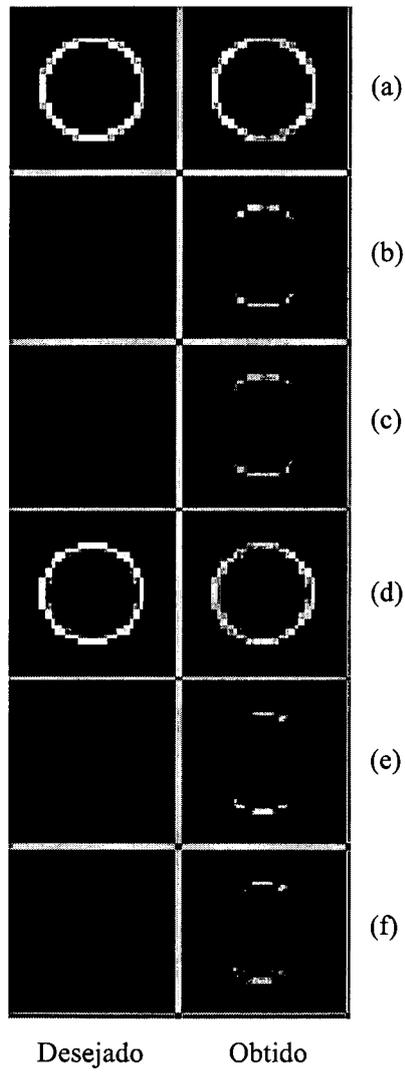


Figura 7.3: Comparativo de saída desejada e saída obtida após treinamento da rede neural.

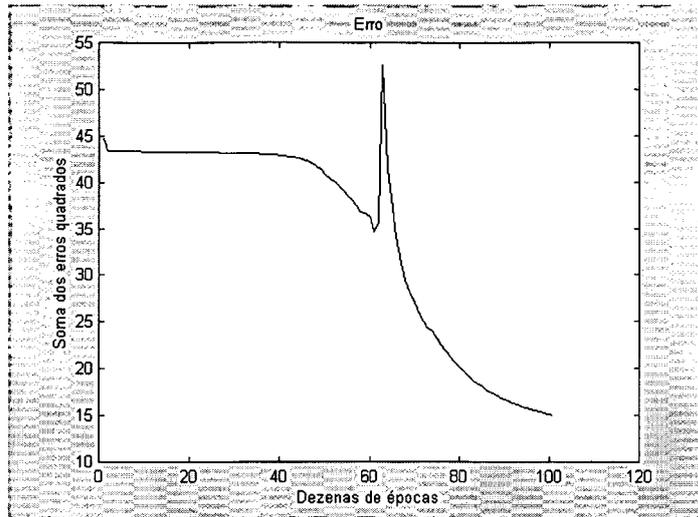


Figura 7.4: Erro na saída da rede neural ao longo das épocas.

colunas são ativações da camada escondida respectivas a tal entrada (6 canais da camada escondida). Observe que alguns canais da camada escondida respondem a linhas ou bordas verticais (canais 2 e 6) - compare com colunas de orientação, seção 2.4.2. Note que não existe ordem quanto ao tipo de característica que é extraída. Para outras execuções de aprendizado, nada garante a extração dos mesmos tipos de características.

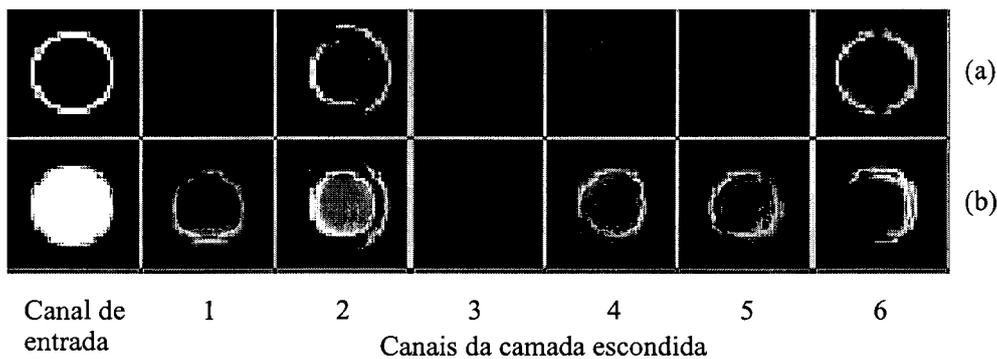


Figura 7.5: Ativações de canais da camada escondida.

A máscara de convolução referente ao neurônio do canal 2 da camada escondida é apresentado na Figura 7.6. Observe que a máscara de convolução possui uma certa continuidade, o que permite extrair qualitativamente um significado para sua função. Podemos supor que tal máscara de convolução possua a função de extrair a característica de diferenciais de luminosidade em uma determinada direção, devido a sua forma de vale e monte ao longo da direção horizontal.

Na Figura 7.7 e Figura 7.8 apresentamos os resultados de padrões diferente dos apresen-

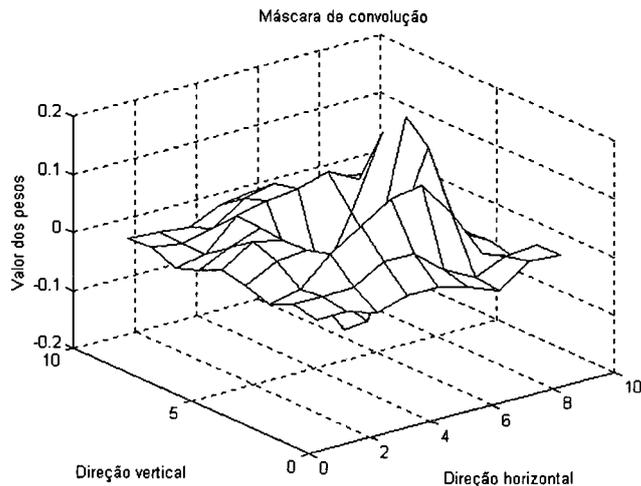


Figura 7.6: Máscara de convolução de um neurônio.

tados à rede neural durante treinamento. Os padrões consistem de sobreposições de circunferências e de círculos. O padrão em (a) consiste de circunferências (ou círculos) à mesma disparidade. O padrão em (b) consiste de circunferências (ou círculos) a disparidades distintas - a acima à esquerda possui disparidade mais distante, a abaixo à direita possui disparidade mais próxima. O padrão em (c) é semelhante ao padrão (b), porém as diferenças de disparidades são maiores.

Observe que no padrão (b), a inibição de disparidade incorreta é mais intensa que no padrão (c). É necessário incluir padrões de inibição de disparidade incorreta para uma grande gama de disparidades, para que a inibição de disparidades possua um efeito abrangente.

Observe na Figura 7.7 que as regiões onde duas circunferências se cruzam geram pouca resposta de disparidade. Para o padrão (a), onde todas circunferências possuem a mesma disparidade isto é um efeito não desejável.

Na Figura 7.8 cada círculo possui uma intensidade luminosa distinta. No padrão (a), onde todos círculos possuem a mesma disparidade, a resposta de ativação para os círculos de intensidade luminosa menor é mais fraca, o que não é desejável. A resposta de disparidade deve ser independente da intensidade luminosa.

A rede neural foi treinada para detectar apenas uma disparidade. Para obter artificialmente a extração de outras disparidades com a mesma estrutura e mesmos pesos de rede neural, deslocamos cada imagem do par estéreo para direita ou esquerda. Para obter respostas de disparidades mais distantes, deslocamos a imagem esquerda para direita e a imagem direita para esquerda. Para obter respostas de disparidades mais próximas, os deslocamentos são contrários. Um exemplo de disparidades em imagem obtida de ambientes reais é apresentado na Figura 7.9. Em relação a (b), (a) responde a disparidades mais distantes, e (c) responde

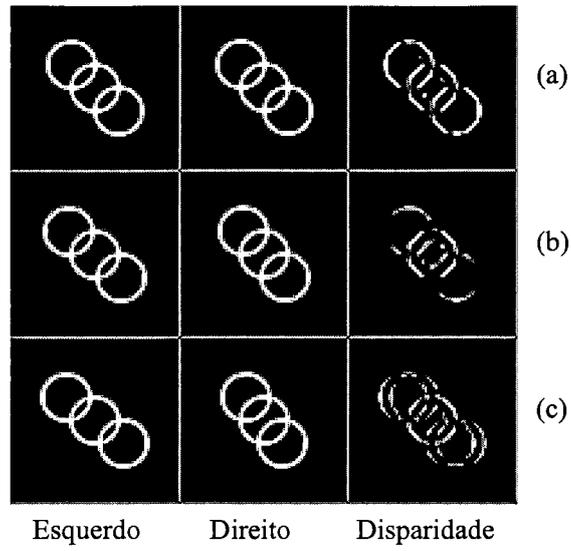


Figura 7.7: Circunferências sobrepostas em diversas disparidades.

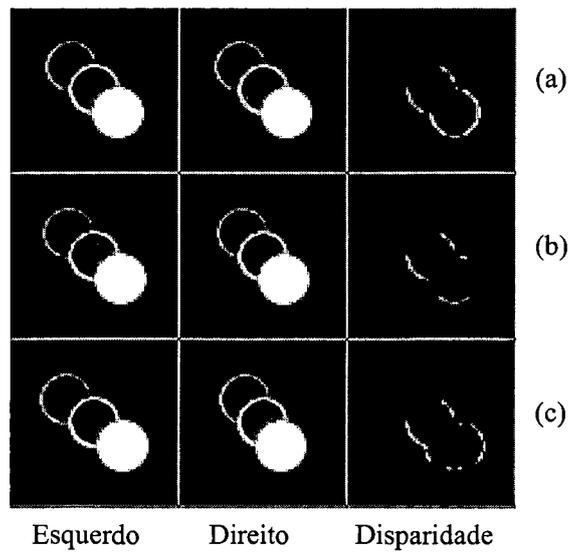


Figura 7.8: Círculos sobrepostos em diversas disparidades.

a disparidades mais próximas. As imagens de resposta de disparidade foram auto equalizadas para melhor visualização. Pode-se observar uma fraca resposta de disparidade ao redor de alguns contornos de objetos, porém tais resultados ainda são insuficientes para se efetuar a reconstrução estéreo da imagem, bem como uma comparação dos resultados deste trabalho com outros algoritmos de reconstrução estéreo.

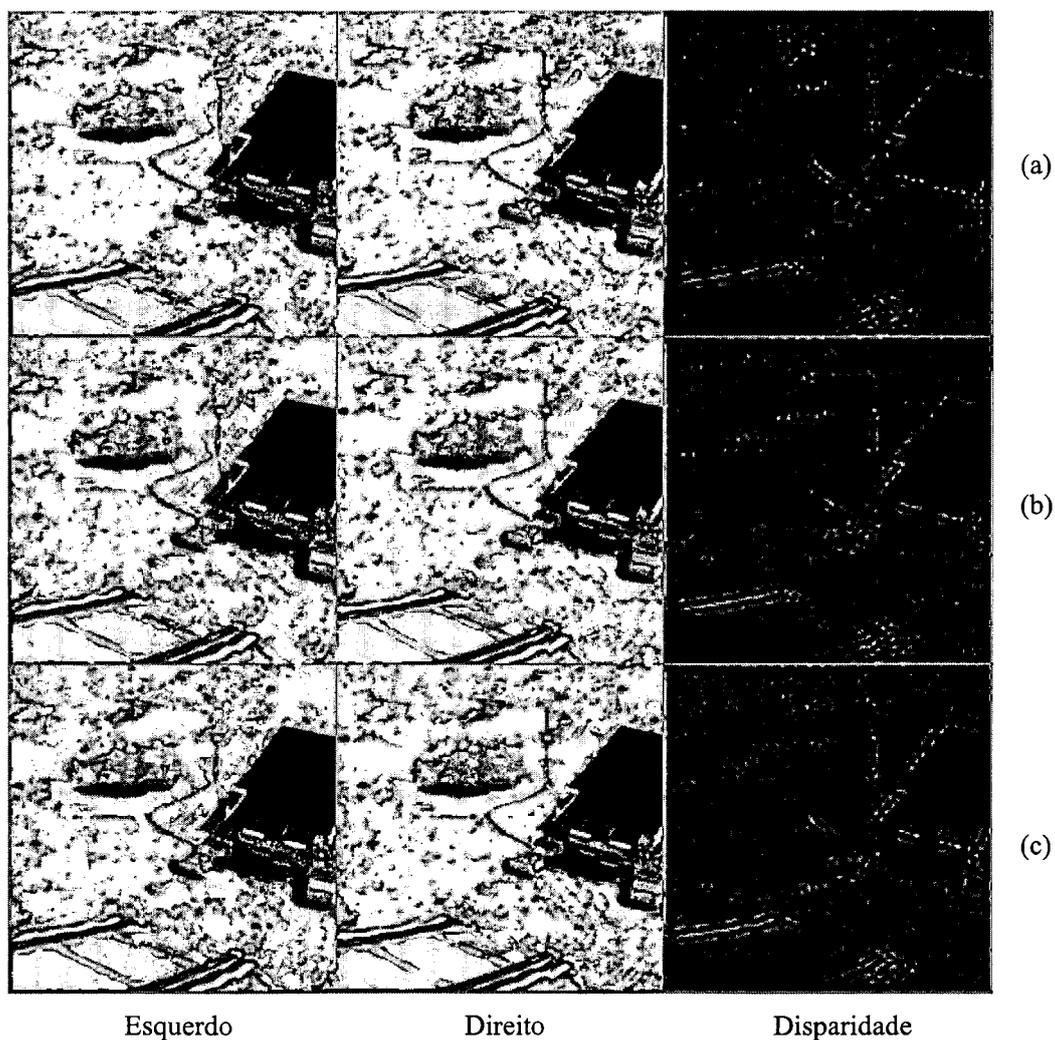


Figura 7.9: Detecção de disparidades em uma imagem real.

### 7.2.2 Discussão

Os resultados apresentados não são o suficiente para obter a reconstrução estéreo que se equipare aos algoritmos de reconstrução estéreo atualmente existentes. Porém, a observação dos resultados indica direções com as quais este modelo pode ser melhorado. Os pontos mais relevantes para melhoria dos resultados são:

- Incluir novos exemplos ao conjunto de aprendizado - a rede neural não generalizou de forma correta situações que não constavam do conjunto de aprendizado (cruzamento de linhas, intensidades luminosas). Neste modelo de rede neural, a correta generalização depende da inclusão de novos padrões de aprendizado. A inclusão de muitos exemplos de aprendizado pode tornar o tempo de processamento pouco viável.
- Determinar padrões mais adequados para o uso do aprendizado supervisionado. Os padrões de aprendizado de entrada e saída da rede neural são apenas suposições do que pode ocorrer nos estágios iniciais do córtex visual. A escolha de padrões mais adequados pode permitir melhores resultados. De fato, a extração de características na camada escondida é um reflexo dos padrões apresentados à rede neural.

Outro fator a influenciar os resultados do aprendizado é o tamanho das máscaras de convolução adotadas. De forma geral, observamos que para:

- máscara de convolução pequena - há problemas para convergência durante o aprendizado. Para o caso da disparidade, há um tamanho mínimo de máscara, sem o qual nunca ocorrerá detecção de disparidade.
- máscara de convolução grande - há problemas de generalização, principalmente para as máscaras de convolução ligadas a entrada e saída da rede neural. A máscara de convolução passa a extrair características pouco genéricas, tal como a extração de circunferências de um determinado diâmetro.

O tamanho adotado para máscara de convolução influencia diretamente o tempo de processamento do modelo. Quanto maior a máscara de convolução, maior o tempo de processamento. Para máscaras de convolução pequenas, é mais eficiente utilizar o cálculo de convolução baseado na definição de soma de convolução. Para máscaras de convolução grandes, é mais eficiente utilizar o cálculo de convolução baseado no teorema de convolução da transformada de Fourier.

Geralmente é possível criar um protótipo de padrão de aprendizado de tamanho reduzido, com máscaras de convolução de tamanho reduzido, para testes iniciais (escalabilidade). Após obter testes satisfatórios, é razoável utilizar padrões de aprendizado equivalentes de tamanho maior, com máscaras de convolução maiores, para o resultado final.

## Capítulo 8

# Conclusão

- Melhoria do modelo de Rede Neural por Convolução utilizando elementos da teoria de processamento de sinais.
- Melhoria do modelo de Rede Neural por Convolução baseado em exemplares de aplicação.
- Implementação do modelo de Rede Neural por Convolução .
- Adaptação do modelo de Rede Neural por Convolução ao problema de reconstrução estéreo.
- Implementação e execução desta adaptação.

Apresentamos uma abordagem para resolução do problema de reconstrução estéreo utilizando o modelo de Rede Neural por Convolução, com os respectivos resultados de execução. O modelo de Rede Neural por Convolução é orientado ao tratamento de problemas invariantes à translação, característica que se obtém com restrição de compartilhamento de pesos. Adequamos o problema de reconstrução estéreo à invariância à translação, de modo que podemos adaptá-lo diretamente ao modelo de Rede Neural por Convolução. O problema de reconstrução estéreo foi também explorado no aspecto do pré processamento que se observa tanto em caminhos neurais da retina ao córtex quanto em abordagens de algoritmos computacionais - o pré processamento igual para imagem direita e esquerda do par estéreo. Este pré processamento é obtido com o próprio modelo de Rede Neural por Convolução, utilizando-se artifícios originados da própria necessidade de adaptação. Observações para tornar mais eficiente o cálculo do produto matricial de vetores de seqüências foram apresentados, refletindo na eficiência do modelo de Rede Neural por Convolução. Sugestões de manter os pesos do modelo de Rede Neural por Convolução no domínio da frequência foram apresentadas, permitindo redução no número de operações executadas.

Como principais contribuições, este trabalho apresenta:

- Adaptação do Modelo de Rede Neural por Convolução ao problema de reconstrução estéreo, explorando o máximo de informações específicas do problema abordado;
- Observações para tornar o Modelo de Rede Neural por Convolução mais eficiente em termos computacionais;
- Implementação do Modelo de Rede Neural por Convolução e da adaptação deste ao problema de reconstrução estéreo.

Os resultados obtidos com o treinamento da rede neural ainda não se equiparam aos resultados que podem ser obtidos com algoritmos de reconstrução estéreo atualmente existentes, porém indicam direções com as quais pode-se melhorar a abordagem de reconstrução estéreo e o modelo de Rede Neural por Convolução.

## 8.1 Recomendações

O modelo de Rede Neural por Convolução pode ser explorado em diversas direções:

- Extensões do modelo de Rede Neural por Convolução;
- Aplicações do modelo de Rede Neural por Convolução.

A teoria existente para o cálculo de convolução e processamento de sinais pode ser explorada para tornar o modelo de Rede Neural por Convolução mais eficiente. Esta direção é importante no aspecto de complexidade computacional do modelo.

A adaptação de algoritmos de aprendizagem para o modelo de Rede Neural por Convolução, em especial algoritmos não supervisionados, pode ser de grande valia para obtenção de melhores resultados de extração de características nas camadas escondidas. O uso de excitação e inibição lateral entre canais das camadas escondidas também pode auxiliar na organização das características nas camadas escondidas.

Em termos de aplicações, o modelo de Rede Neural por Convolução tem sido utilizado na literatura principalmente para o reconhecimento de padrões (veja seção 4.4). Recentemente o modelo de Rede Neural por Convolução tem sido utilizado para o reconhecimento de faces (veja [Lawrence 1997]). A presente abordagem possui o intuito de mostrar como o modelo de Rede Neural por Convolução pode ser utilizado para o processamento de imagens inspirado nos estágios iniciais do córtex visual. Supõe-se que outros problemas têm lugar nos estágios iniciais do córtex visual, entre eles o problema de reconstrução da informação tridimensional a partir de:

- sombreamento;
- sobreposição de objetos;

- perspectiva;
- textura;
- movimento.

O modelo de Rede Neural por Convolução pode ser utilizado para tratar destes problemas, o que permite uma possível integração destas redes neurais, no qual a extração da informação tridimensional possa ocorrer com a soma de diversos fatores, o que tornaria a rede mais robusta contra casos particulares de reconstrução incorreta. Uma grande dificuldade para modelar tais aplicações é determinar como deve ser o padrão de saída da rede neural, no caso de aprendizado supervisionado. Uma possibilidade é o uso de diferenciais de profundidade, onde a informação de profundidade é dada em termos relativos a pontos vizinhos (valores não absolutos).

A abordagem destes outros problemas pode expor novas observações relevantes para adaptação de algoritmos de aprendizagem e extensões do modelo de Rede Neural por Convolução.

# Referências Bibliográficas

- [Back 1994] Back, Andrew; Wan, Eric A.; Lawrence, Steve; Tsoi, Ah C.; 1994. A Unifying view of some training algorithms for multilayer perceptrons with FIR synapses. In Vlontzos, J.; Hwang J.; Wilson E. *Neural Networks for Signal Processing 4*. IEEE press, p. 146-154.
- [Dudgeon 1984] Dudgeon, Dan E.; Mersereau, Russel M.; 1984. *Multidimensional Digital Signal Processing*. Englewood Cliffs, NJ : Prentice-Hall.
- [Fukushima 1983] Fukushima, Kunihiko; Miyake, Sei; Ito, Takayuki; 1983. Neocognitron: A Neural Network Model for a Mechanism of Visual Pattern Recognition. *IEEE Transactions on Systems, Man and Cybernetics*, v. SMC-13, n. 5 (september/october), p. 826-834.
- [Haralick 1992] Haralick, Robert M.; Shapiro, Linda G.; 1992. *Computer and Robot Vision*. v. 2. Reading, Mass. : Addison-Wesley.
- [Haykin 1994] Haykin, Simon S.; 1994. *Neural Networks: a Comprehensive Foundation*. Saddle River, NJ : Prentice Hall.
- [Hertz 1991] Hertz, John A.; Palmer, Richard G.; Krogh, Anders S.; 1991. *Introduction to the Theory of Neural Computation*. Redwood City, CA : Addison-Wesley.
- [Hubel 1995] Hubel, David H.; 1995. *Eye Brain and Vision*. New York : Scientific American Library.
- [Kuffler 1984] Kuffler, Stephen W.; Nicholls, John G.; Martin, A. R.; 1984. *From Neuron to Brain: a Cellular Approach to the Function of the Nervous System*. 2. ed. Sunderland, Mass. : Sinauer Associates.
- [Lawrence 1997] Lawrence, Steve; Lee Giles, C.; Tsoi, Ah C.; Back, Andrew D.; 1997. Face Recognition: A Convolutional Neural Network Approach. *IEEE Transactions on Neural Networks*, v. 8, n. 1, p. 98-113.
- [LeCun 1990] LeCun, Y.; Boser, B.; Denker, J. S. et al.; 1990. Handwritten Digit Recognition with a Back-Propagation Network. In David Touretzky. *Advances*

*in Neural Information Processing Systems 2 (NIPS \*89)*. Denver : Morgan Kaufman.

- [Lim 1990] Lim, Jae S.; 1990. *Two-Dimensional Signal and Image Processing*. Englewood Cliffs, NJ : Prentice Hall.
- [Marr 1982] Marr; D.; 1982. *Vision*. San Francisco : W. H. Freeman.
- [Matan 1992] Matan, O.; Burges, C.; LeCun, Y.; Denker, J.; 1992. Multi-Digit Recognition Using a Space Displacement Neural Network. In Moody, J.; Hanson, S.; Lipmann, R. *Advances in Neural Information Processing Systems 4*, p. 488-495, San Mateo, CA : Morgan Kaufmann.
- [Ohzawa 1990] Ohzawa, Izumi; DeAngelis, Gregory C.; Freeman, Ralph D.; 1990. Stereoscopic Depth Discrimination in the Visual Cortex: Neurons Ideally Suited as Disparity Detectors. *Science*, v. 249 (August), p. 1037-1041.
- [Oppenheim 1989] Oppenheim, Alan V.; Schaffer, Ronald W.; 1989. *Discrete-Time Signal Processing*. Englewood Cliffs, NJ : Prentice-Hall.
- [Rumelhart 1986] Rumelhart, David E.; Hinton, G. E.; Williams, R. J.; 1986. Learning Internal Representations by Error Propagation. In *Parallel Distributed Processing*, Cambridge : M. I. T. Press, v. 1, p. 318-362.
- [Sejnowski 1987] Sejnowski, T. J.; Rosenberg, C. R.; 1987. Parallel Networks that Learn to Pronounce English Text. *Complex Systems*, v. 1, p. 145-168.
- [Spreeuwers 1993] Spreeuwers, Luuk; 1993. Image Filtering with Neural Networks. In: *SMBT Meeting on Measurement and Artificial Neural Networks* (3. : Nov. 1993 : Utrecht, Netherlands) Proceedings. p. 37-52.
- [Waibel 1989] Waibel, Alexander; Hanazawa, Toshiyuki; Hinton, Geoffrey et al.; 1989. Phoneme Recognition Using Time-Delay Neural Networks. *IEEE Transactions on Acoustics, Speech and Signal Processing*, v. ASSP-37, n. 3 (March), p. 328-339.
- [Weinshall 1994] Weinshall, Daphna; Malik, Jitendra; 1995. Review of Computational Models of Stereopsis. In Papathomas, Thomas V.; Chubb, Charles; Gorea Andrei et al. *Early Vision and Beyond*. Cambridge : A Bradford Book, p. 33-41.
- [Wilson 1989] Wilson, Stephen S.; 1989. Vector Morphology and Iconic Neural Networks. *IEEE Transactions on Systems, Man, and Cybernetics*, v.19, n. 6 (November/December), p. 1636-1644.

[Wilson 1992] Wilson, Stephen S.; 1992. Translation Invariant Neural Networks. In Souécek, Branko. *Fast Learning and Invariant Object Recognition : the Sixth-Generation Breakthrough*. New York : John Wiley & Sons, p. 125-151.