

**UNIVERSIDADE DO VALE DO RIO DOS SINOS  
CENTRO DE CIÊNCIAS EXATAS E TECNOLÓGICAS  
CURSO DE CIÊNCIA DA COMPUTAÇÃO**

**Reconhecimento Inteligente  
de Sinais de Trânsito Brasileiros**

**Flávio de Oliveira Alves**

**Prof. Dr. Fernando Santos Osório  
Orientador**

*Monografia submetida como requisito  
parcial para a obtenção do título de  
Bacharel em Ciência da Computação.*

São Leopoldo, junho de 2004

*Dedico este trabalho  
aos meus avós maternos que,  
desde os meus primeiros anos,  
me tomaram por seu filho.  
E também o dedico,  
com muito carinho,  
à minha amada noiva,  
Fabiane Epping.*

## SUMÁRIO

Lista de Figuras .....	5
Lista de Tabelas .....	7
Lista de Abreviaturas .....	8
Agradecimentos .....	10
Resumo .....	11
Abstract .....	12
1 INTRODUÇÃO .....	13
2 OS SISTEMAS DE TRANSPORTE INTELIGENTES (STI) E A SEGURANÇA NO TRÂNSITO .....	15
2.1 O Interesse em Sistemas de Transporte Inteligentes .....	15
2.2 Dispositivos de Segurança Automotiva .....	15
2.3 Pesquisas em Sistemas de Apoio ao Motorista .....	18
2.3.1 Veículos Autônomos .....	18
2.3.2 O Reconhecimento Inteligente de Sinais de Trânsito - RIST .....	22
2.3.3 Trabalhos Anteriores em Reconhecimento de Sinais de Trânsito .....	22
2.3.3.1 Trabalhos Envolvendo apenas a Detecção de Sinais de Trânsito .....	22
2.3.3.1.1 Detecção Utilizando Cores .....	22
2.3.3.1.2 Detecção Utilizando Redes Neurais Artificiais .....	23
2.3.3.2 Trabalhos Envolvendo a Detecção e o Reconhecimento de Sinais de Trânsito .....	23
2.3.3.2.1 Reconhecimento Utilizando Algoritmos Genéticos .....	23
2.3.3.2.2 Reconhecimento Utilizando Análise Relacional de Características .....	24
2.3.3.2.3 Reconhecimento Utilizando Classificação Estatística Não-paramétrica .....	25
2.3.3.2.4 Reconhecimento Utilizando Comparação de Modelos .....	25
2.3.3.2.5 Reconhecimento Utilizando Comparação de Padrões (Pattern Matching) .....	26
2.3.3.2.6 Reconhecimento Utilizando Histograma de Cores .....	27
2.3.3.2.7 Reconhecimento Utilizando Redes Neurais Artificiais .....	27
2.3.3.2.8 Reconhecimento Utilizando Similaridades .....	32
2.3.3.2.9 Reconhecimento Utilizando a Metodologia Padrão do Processamento de Imagens .....	32
2.4 Considerações Finais Sobre Este Capítulo .....	33
3 FUNDAMENTOS DE PROCESSAMENTO DE IMAGENS .....	34
3.1 Conceitos Básicos .....	34
3.1.1 A Representação Digital da Imagem .....	34
3.1.2 Os Passos Fundamentais no Processamento de Imagens .....	35
3.1.3 Equipamentos Envolvidos no Processamento de Imagens .....	36
3.2 Modelos de Cores .....	38
3.2.1 O Modelo de Cores RGB .....	39
3.2.2 O Modelo de Cores HSV .....	41
3.3 A Biblioteca de Processamento de Imagens "Sistema-IMG" .....	42
3.4 Considerações Finais Sobre Este Capítulo .....	43
4 FUNDAMENTOS DE REDES NEURAIS ARTIFICIAIS (RNAs) .....	44
4.1 O Aprendizado de Máquinas .....	44
4.2 Fundamentos das Redes Neurais Artificiais .....	44
4.2.1 Vantagens das RNAs .....	45
4.2.2 Aplicações das RNAs .....	46
4.2.3 Componentes de uma RNA .....	47
4.2.4 O Aprendizado das RNAs .....	51
4.2.5 O Perceptron de Múltiplas Camadas com Aprendizado por Retropropagação de Erro .....	52
4.2.5.1 O Algoritmo de Retropropagação de Erro ( <i>Error Back-Propagation</i> ) .....	52
4.2.5.2 O Algoritmo de Propagação Resiliente ( <i>Resilient Propagation - RPROP</i> ) .....	56
4.2.5.3 O Grau de Generalização de uma RNA .....	58
4.3 A Ferramenta Stuttgart Neural Network Simulator (SNNS) .....	59

4.4	Considerações Finais Sobre Este Capítulo.....	60
5	METODOLOGIA ADOTADA PARA A IMPLEMENTAÇÃO DO RECONHECEDOR INTELIGENTE DE SINAIS DE TRÂNSITO BRASILEIROS (RIST-Br) .....	62
5.1	A Atenção Visual para a Detecção de Objetos de Interesse .....	62
5.1.1	Atenção Visual em Seres Humanos .....	62
5.1.1.1	O Olho Humano.....	63
5.1.1.2	Os Movimentos Oculares em Seres Humanos .....	65
5.1.1.3	O Mecanismo Atencional Humano .....	66
5.1.2	Atenção Visual em Máquinas .....	66
5.2	Experimento Sobre o Reconhecimento Inteligente de Padrões .....	67
5.2.1	Objetivos do Experimento.....	67
5.2.2	Metodologia Adotada no Experimento.....	68
5.2.3	Pré-processamento das Imagens do Experimento .....	69
5.2.4	Aprendizado Neural no Experimento.....	71
5.2.5	Conclusões sobre o Experimento.....	72
5.3	As Fases do Reconhecimento de Sinais de Trânsito.....	72
5.3.1	A Captura das Imagens de Entrada .....	72
5.3.1	O Pré-processamento das Imagens de Entrada .....	72
5.3.2	A Segmentação das Imagens de Entrada .....	72
5.3.3	A Representação e a Descrição das Subimagens Candidatas.....	73
5.3.4	O Reconhecimento dos Sinais de Trânsito nas Subimagens Candidatas.....	73
5.4	Considerações Finais Sobre Este Capítulo.....	73
6	O RECONHECEDOR INTELIGENTE DE SINAIS DE TRÂNSITO BRASILEIROS (RIST-Br) .....	75
6.1	Objetivos e Relevância .....	75
6.2	O Projeto do RIST-Br .....	76
6.2.1	A Captura das Imagens de Entrada .....	76
6.2.1	O Pré-processamento das Imagens de Entrada .....	83
6.2.2	A Segmentação das Imagens de Entrada .....	86
6.2.3	A Representação e a Descrição das Subimagens Candidatas.....	89
6.2.4	O Reconhecimento dos Sinais de Trânsito nas Subimagens Candidatas.....	91
6.2.4.1	As Classes de Sinais de Trânsito Brasileiros Reconhecidas pelo RIST-Br.....	92
6.2.4.2	O Tipo e a Topologia da RNA .....	94
6.2.4.3	O Treinamento da RNA.....	95
6.2.4.4	A Visualização do Reconhecimento do RIST-Br.....	103
6.3	A Implementação do RIST-Br .....	105
6.3.1	O Módulo de Localização de Subimagens Candidatas.....	105
6.3.1	O Módulo de Identificação dos Sinais de Trânsito .....	106
6.3.2	O Módulo de Visualização dos Resultados .....	106
6.3.3	Visão Geral do RIST-Br.....	108
6.4	Resultados Obtidos.....	109
6.5	Perspectivas de Trabalhos Futuros .....	120
6.6	Considerações Finais Sobre Este Capítulo.....	120
7	CONCLUSÃO.....	121
	BIBLIOGRAFIA .....	123

## Lista de Figuras

Figura 2.1 Mini-Baja em Fase de Automatização [Kelber et al. 2003] .....	18
Figura 2.2 Controle Remoto Através de Comunicação por Telefonia Celular [Kelber et al. 2003].....	19
Figura 2.3 O Veículo Autônomo ARGO [Broggi 2000] .....	19
Figura 2.4 Esquema do Interior do Veículo Autônomo ARGO [Broggi 2000] .....	20
Figura 2.5 Reconhecimento de Estradas pelo Veículo Autônomo ARGO [Broggi 2000].....	21
Figura 2.6 Reconhecimento de Outros Veículos.....	21
Figura 2.7 Resultado da Busca com o Algoritmo Genético de Passo [Liu 2002] .....	24
Figura 2.8 Reconhecimento de um Sinal de Velocidade Máxima Permitida [de la Escalera et al. 1994] .....	28
Figura 2.9 Exemplo de Reconhecimento de Sinal de Trânsito [Liu 2001]: (a) imagem original; (b) imagem em tons de cinza; (c) segmentação; (d) reconhecimento; (e) objeto potencial detectado e normalizado; (f) objeto potencial pronto para ser enviado ao módulo de reconhecimento .....	30
Figura 3.1 Exemplo de Imagem Monocromática.....	34
Figura 3.2 Passos Fundamentais no Processamento de Imagens .....	36
Figura 3.3 Os Cinco Elementos Básicos do Processamento de Imagens .....	37
Figura 3.4 Modelo de Cores RGB: .....	39
Figura 3.5 Capturando A Luz - Da Esquerda para a Direita: Filme Fotográfico; Sensores Digitais Convencionais; Sensor Foveon X3 [Foveon 2004] .....	40
Figura 3.6 O Modelo de Cor HSV: (a) Representado como um Hexacone [Foley 1997]; (b1) Representado como um Cone [Pierce 2004]; (b2) Vista em Corte de um Segmento do Cone HSV [Pierce 2004] .....	41
Figura 4.1 A Rede Neural Artificial como uma "Caixa Preta" .....	45
Figura 4.2 Anatomia Simplificada dos Neurônios Biológicos [Costa et al. 2000].....	47
Figura 4.3 Modelo de um Neurônio Artificial $k$ .....	48
Figura 4.4 A Função Logística .....	49
Figura 4.5 Rede Neural Artificial de Múltiplas Camadas .....	50
Figura 4.6 Gráfico de uma Superfície de Erro de uma RNA com Dois Pesos ( <i>Weights</i> ) [Gallagher 2000] .....	55
Figura 4.7 Parada Antecipada, Baseada na Validação Cruzada [Haykin 2002].....	59
Figura 4.8 Exemplos da Interface Gráfica do SNNS .....	60
Figura 5.1 Decomposição da Luz Branca, e Espectro Eletromagnético [Foley 1997] .....	63
Figura 5.2 (a) O Caminho da Luz até a Retina no Olho Humano; .....	64
Figura 5.3 A Retina do Olho Humano [Erickson 2004].....	65
Figura 5.4 Imagens da COIL-100 Database [COIL 2003] .....	68
Figura 5.5 Exemplos de Imagens da COIL-100 .....	68
Figura 5.6 Redução da Resolução e Detecção de Bordas.....	69
Figura 5.7 Trecho do Arquivo de Treinamento Usado no SNNS .....	70
Figura 5.8 Evolução do Erro MSE ( <i>Mean Square Error</i> - Erro Médio Quadrado).....	71
Figura 5.9 Visão Geral do RIST-Br .....	74
Figura 6.1 Câmera Fotográfica Digital Sony DSC-P72 [Sony 2004] .....	76
Figura 6.2 Exemplos de Sinais de Parada Obrigatória: (a) sinal oculto pelos galhos da árvore; (b) sinal parcialmente à sombra da árvore; (c) sinal iluminado pelo Sol; (d) sinal à sombra da árvore; (e) sinal parcialmente oculto; (f) sinal em dia parcialmente nublado .....	78
Figura 6.3 Exemplos de Sinais de Área Escolar: (a) sinal à sombra da árvore; (b) sinal iluminado pelo Sol; (c), (d) e (e) mostram sinais em dia nublado; (f) sinal em dia parcialmente nublado .....	79
Figura 6.4 Exemplos de Sinais de Sentido Obrigatório: (a) sinal iluminado pelo Sol; (b) e (c) mostram sinais em dia nublado; (d) sinal iluminado pelo Sol; (e) sinal sob a sombra; (f) sinal em dia nublado .....	80
Figura 6.5 Exemplos de Sinais de Velocidade Máxima Permitida: de (a) até (d) mostram sinais em dias parcialmente nublados; (e) e (f) mostram sinais em dia nublado .....	81
Figura 6.6 Exemplos de Outros Sinais de Trânsito: (a) e (b) mostram sinal de Saliência ou Lombada iluminado pelo Sol e à sombra, respectivamente; (c) sinal de Proibido Acionar Buzina ou Sinal Sonoro,	

localizado à sombra; (d) sinal de Proibido Virar à Esquerda, iluminado pelo Sol; (e) e (f) mostram sinal de Semáforo à Frente, à sombra e iluminado pelo Sol, respectivamente .....	82
Figura 6.7 Primeiro Passo do Pré-Processamento - Conversão de RGB para HSI (Imagem p370): (a) imagem original; (b) componente de cor (H); (c) componente de saturação (S); componente de intensidade (I) ..	83
Figura 6.8 Imagem Correspondente à Componente de Cor (H) .....	84
Figura 6.9 Resultado da Binarização da Figura 6.8, em Busca do Sinal de Parada Obrigatória (Filtro para a Cor Vermelha) .....	85
Figura 6.10 Exemplos de Subimagens Candidatas: (a) e (c) de 200x200 pixels; e Suas Respectivas Miniaturas de 40x40 pixels (b) e (d).....	87
Figura 6.11 Resultado da Detecção de Subimagens Candidatas pelo Mecanismo Atencional do RIST-Br .....	88
Figura 6.12 Trecho Inicial do Arquivo em Formato Texto da Área, na Imagem Original, Correspondente à Subimagem Candidata - Notar o Cabeçalho (200 200) e os valores para as Componentes R, G e B para os Cinco Primeiros Pixels .....	89
Figura 6.13 Exemplo de Histograma de Cores de Uma Componente (R, G ou B) .....	90
Figura 6.14 Exemplo dos Somatórios dos Pixels de Cada Linha e Coluna de uma Imagem de 4x4 Pixels .....	91
Figura 6.15 Tipos de Sinais de Trânsito Utilizados para Treinamento da RNA .....	94
Figura 6.16 Trecho do Arquivo de Padrões de Treinamento para a RNA 86:172:4 .....	97
Figura 6.17 Curvas (MSE x Épocas) para o Treinamento 1 .....	99
Figura 6.18 Curvas (MSE x Épocas) para o Treinamento 2 .....	99
Figura 6.19 Curvas (MSE x Épocas) para o Treinamento 3 .....	100
Figura 6.20 Comparação da Aprendizagem: (a) Treinamento 2; e (b) Treinamento 3 .....	100
Figura 6.21 Curvas de Aprendizagem do Treinamento 3 .....	101
Figura 6.22 Visualização do Resultado do Reconhecimento de Sinais de Trânsito Pelo RIST-Br (Reconhecimento de um Sinal de Parada Obrigatória) .....	103
Figura 6.23 Ampliação da Região onde foi Reconhecido o Sinal de Parada Obrigatória .....	104
Figura 6.24 Visão Geral do RIST-Br .....	108
Figura 6.25 Exemplos de Detecção e Reconhecimento pelo RIST-Br .....	116
Figura 6.26 Exemplos de Detecção e Reconhecimento pelo RIST-Br .....	117
Figura 6.27 Exemplos de Detecção e Reconhecimento pelo RIST-Br .....	118
Figura 6.28 Exemplos de Detecção e Reconhecimento pelo RIST-Br .....	119

## Lista de Tabelas

Tabela 5.1 Dados do Histograma de Cores .....	69
Tabela 5.2 Matriz de Confusão para a base de Validação .....	71
Tabela 6.1 Tipos de Sinais de Trânsito Fotografados para a Base de Imagens.....	77
Tabela 6.2 Exemplos de Sinais de Cada Classe de Sinais de Trânsito Brasileiros.....	93
Tabela 6.3 Dados dos Arquivos de Treinamento e Validação para a RNA 86:172:4 .....	96
Tabela 6.4 Matriz de Confusão do Conjunto de Validação do Treinamento 3 .....	102
Tabela 6.5 Resultados do RIST-Br: RNA 86:22:4, Treinamento 1 (ver Seção 6.2.5.3), .....	110
Tabela 6.6 Resultados Globais do RIST-Br: RNA 86:22:4, Treinamento 1 (ver Seção 6.2.5.3), limiar para a visualização de 0,10 (ver seções 6.2.5.3 e 6.3.3).....	111
Tabela 6.7 Resultados do RIST-Br: RNA 86:172:4, Treinamento 3 (ver Seção 6.2.5.3), .....	112
Tabela 6.8 Resultados Globais do RIST-Br: RNA 86:172:4, Treinamento 3 (ver Seção 6.2.5.3), limiar para a visualização de 0,900 (ver seções 6.2.5.3 e 6.3.3).....	113
Tabela 6.9 Resultados do RIST-Br: RNA 86:172:4, Treinamento 3 (ver Seção 6.2.5.3), .....	114
Tabela 6.10 Resultados Globais do RIST-Br: RNA 86:172:4, Treinamento 3 (ver Seção 6.2.5.3), limiar para a visualização de 0,600 (ver seções 6.2.5.3 e 6.3.3).....	115
Tabela 7.1 Resultados do RIST-Br: RNA 86:172:4, Treinamento 3 (ver Seção 6.2.5.3, no Capítulo 6), limiar para a visualização de 0,600 (ver seções 6.2.5.3 e 6.3.3, ambas no Capítulo 6) .....	121
Tabela 7.2 Resultados Globais do RIST-Br: RNA 86:172:4, Treinamento 3 (ver Seção 6.2.5.3, Capítulo 6), limiar para a visualização de 0,600 (ver seções 6.2.5.3 e 6.3.3, ambas do Capítulo 6) .....	122

## Lista de Abreviaturas

3-D	Tridimensional, ou Três Dimensões
AHSRA	Advanced Cruise-Assist Highway System Research Association - Associação de Pesquisa para um Sistema Avançado de Assistência ao Tráfego em Auto-Estradas
ASV	Advanced Security Vehicle - Veículo de Segurança Avançada
BMP	Bit Map - Mapa de Bits (Formato de Arquivo Gráfico)
CFRP	Carbon Fiber Reinforced Plastic - Plástico Reforçado com Fibra de Carbono
COIL	Columbia Object Image Library - Biblioteca de Imagens de Objetos de Columbia
DENATRAN	Departamento Nacional de Trânsito
DNIT	Departamento Nacional de Infra-Estrutura de Transportes
EFC	Equações de Formação de Cores
EUA	Estados Unidos da América
FP	Falso Positivo
GCC	GNU C Compiler - Compilador C GNU (GNU : GNU is Not Unix - GNU Não é Unix - é uma abreviatura recursiva)
HSI	Hue, Saturation, Intensity - Cor, Saturação, Intensidade
HSV	Hue, Saturation, Value - Cor, Saturação, Valor
JPEG	Joint Photographic Experts Group - Grupo de Especialistas em Fotografia
MS-DOS	Microsoft Disk Operation System
MSE	Mean Square Error - Erro Médio Quadrado
NAHSC	National Automated Highway System Consortium - Consórcio para um Sistema Nacional de Auto-Estradas Automatizadas
PIXEL	Picture Element - Elemento da Imagem
PROMETHEUS	PROgram for an European Traffic with Highest Efficiency and Unprecedented Safety - Programa para um Tráfego Europeu com a Mais Alta Eficiência e Segurança Sem Precedentes
RGB	Red, Green, Blue - Vermelho, Verde, Azul
RIST	Reconhecimento (ou Reconhecedor) Inteligente de Sinais de Trânsito
RIST-Br	Reconhecedor Inteligente de Sinais de Trânsito
RNA	Rede Neural Artificial
RPROP	Resilient Propagation - Propagação Resiliente
SNNS	Stuttgart Neural Network Simulator - Simulador de Redes Neurais de Stuttgart
STI	Sistemas de Transporte Inteligentes
SVF	Simple Vector Filter - Filtro de Vetor Simples
TGA	Formato de Arquivo Truevision
UNISINOS	Universidade do Vale do Rio dos Sinos



USA	United States of America - Estados Unidos da América
USP	Universidade de São Paulo
VC	Visão Computacional
VP	Verdadeiro Positivo
VSA	Veículo de Segurança Avançada
YIQ	Modelo de cor onde Y significa luminância, enquanto a cromaticidade é descrita pelas componentes I e Q
YUV	Modelo de cor onde a componente Y representa o brilho, e as componentes U e V representam a cor

## **Agradecimentos**

Agradeço especialmente, com muito carinho, à minha amada noiva, Fabiane Epping, que sacrificou incontáveis finais de semana, sempre clamando pela minha presença, para que eu pudesse construir este trabalho. Muito obrigado por tudo, Amor!

Agradeço aos meus avós maternos, Adail de Oliveira e Clotilde Sena de Oliveira, que me criaram como um filho e sempre me deram todas as oportunidades para crescer intelectualmente.

Agradeço ao meu paciente orientador, Prof. Fernando Osório, pela orientação tranqüila e objetiva, e pelos valiosos conhecimentos e dicas com os quais me instruiu ao longo do meu Curso de Ciência da Computação e ao longo desta Monografia.

E, por último mas não menos importante, deixo meus agradecimentos a todas as pessoas que sempre estiveram, de perto ou de longe, torcendo pelo meu sucesso. Muito obrigado!

## **Resumo**

Atualmente, com o crescimento do tráfego de veículos nas rodovias do Brasil - especialmente nos grandes centros urbanos - têm surgido vários problemas, dentre os quais o mais importante é o crescente número de acidentes com vítimas. A utilização de sistemas computacionais, e outros sistemas de apoio ao motorista, oferece condições para reduzir o número de acidentes nas rodovias do país - e, como consequência disso, contribuir para o bem estar social e econômico da população. Este trabalho tem como objetivo o desenvolvimento de um sistema de reconhecimento inteligente de sinais de trânsito com potencial para apoiar o motorista na tarefa de estar atento às informações apresentadas pelos sinais de trânsito e, assim, contribuir para o aumento da segurança do tráfego de veículos nas rodovias do Brasil.

## **Abstract**

Nowadays, because the increasing traffic of vehicles on the roads of Brazil - specially on the urban centers - many problems are showing up, among which the most important is the growing number of accidents with victims. The use of computational systems, and other systems to give assistance to the driver, offers ways to reduce the number of accidents on the country's roads - and, as a consequence, to contribute to the social and economic welfare of the population. This work intends to develop an intelligent road sign recognition system, which has the potential to give support to the driver in his task of being attentive to the information presented by the road signs and, therefore, to contribute to a growth in the road traffic security on the roads of Brazil.

# 1 INTRODUÇÃO

Atualmente, com o crescimento do tráfego de veículos nas rodovias do Brasil - especialmente nos grandes centros urbanos - têm surgido vários problemas, dentre os quais o mais importante é o crescente número de acidentes com vítimas [DENATRAN 2001].

A utilização de sistemas computacionais, e outros sistemas de apoio ao motorista, tem o potencial necessário para reduzir o número de acidentes [Paromtchik 1997] nas rodovias do país - e, como consequência disso, contribuir para o bem estar social e econômico da população, na medida em que vidas poderão ser poupadas e perdas econômicas poderão ser evitadas.

Um dos sistemas computacionais atualmente estudados para suporte aos motoristas é o reconhecimento de sinais de trânsito. No reconhecimento de sinais de trânsito, imagens do ambiente ao redor do motorista são capturadas e analisadas em busca de algum sinal de trânsito, para que a presença do mesmo seja comunicada para o motorista, através de uma interface visual ou sonora. Com um sistema de apoio dessa natureza, os motoristas poderão dirigir seus veículos com maior segurança pois, sendo o ato de dirigir baseado essencialmente na visão, existem muitas distrações em potencial que podem desviar a atenção do olhar do motorista para longe de um sinal de trânsito.

Sistemas de reconhecimento de sinais de trânsito podem ser projetados e implementados de diversas formas. Este trabalho abordará algumas dessas formas, durante a apresentação do estado da arte do reconhecimento de sinais de trânsito, no Capítulo 2. Nesse capítulo também serão apresentados alguns conceitos e sistemas relacionados à segurança no trânsito.

Para a implementação de um sistema de reconhecimento de sinais de trânsito, devem ser estudados os métodos e técnicas apropriadas para o processamento de imagens. O Capítulo 3 apresentará os fundamentos do processamento de imagens, que envolvem os passos fundamentais pelos quais deve passar uma imagem (e suas informações) para que informações úteis possam ser extraídas da mesma.

O sistema de reconhecimento de sinais de trânsito requer um componente essencial para a sua operação - um módulo capaz de efetuar o reconhecimento de um padrão que esteja presente na imagem do ambiente. Neste trabalho, foi adotada uma técnica de inteligência artificial que utiliza redes neurais artificiais para efetuar esse reconhecimento. Por isso, neste caso, o sistema torna-se um "reconhecedor inteligente de sinais de trânsito". Os fundamentos das redes neurais artificiais são apresentados no Capítulo 4 desta monografia.

Abrangendo os estudos sobre os trabalhos anteriores (Capítulo 2), os fundamentos de processamento de imagens (Capítulo 3), e os fundamentos sobre as redes neurais artificiais (Capítulo 4), foi elaborada uma metodologia para o projeto e desenvolvimento de um reconhecedor inteligente de sinais de trânsito. Essa metodologia é apresentada no Capítulo 5 desta monografia. Finalmente, o projeto e a implementação do reconhecedor inteligente, denominado "Reconhecedor Inteligente de Sinais de Trânsito Brasileiros - RIST-Br", são apresentados no Capítulo 6, juntamente com os resultados obtidos em seus testes.

Assim, neste trabalho, serão apresentados todos os fundamentos, a metodologia e os passos necessários para a implementação computacional do RIST-Br - capaz de detectar e reconhecer um subconjunto dos sinais de trânsito brasileiros - desenvolvida pelo autor.

## **2 OS SISTEMAS DE TRANSPORTE INTELIGENTES (STI) E A SEGURANÇA NO TRÂNSITO**

Neste capítulo serão apresentados alguns Sistemas de Transporte Inteligente (STI) - com destaque para os sistemas de Reconhecimento de Sinais de Trânsito - e algumas de suas influências para a segurança do trânsito nas rodovias de todo o mundo.

### **2.1 O Interesse em Sistemas de Transporte Inteligentes**

O interesse em Sistemas de Transporte Inteligentes (STI) surgiu no mundo há cerca de 20 anos atrás [Bertozzi 2000] [Kelber et al. 2003], quando o problema da mobilidade de pessoas e bens começou a tomar forma, obrigando a busca por soluções para esse problema.

Nos Países desenvolvidos, problemas relacionados à mobilidade através do tráfego de veículos nas rodovias, à segurança, e ao consumo de energia, estão tornando-se cada vez mais sérios [Johansson 2002] [Kelber et al. 2003]. Por conta disso, projetos como o europeu PROMETHEUS (*PROgramM for an European Traffic with Highest Efficiency and Unprecedented Safety* - Programa para um Tráfego Europeu com a Mais Alta Eficiência e Segurança Sem Precedentes), o norte-americano NAHSC (□) e o japonês AHSRA (*Advanced Cruise-Assist Highway System Research Association* - Associação de Pesquisa para um Sistema Avançado de Assistência ao Tráfego em Auto-Estradas) reuniram, ao longo das décadas de 1980 e 1990, instituições de pesquisa, fabricantes de veículos, e governos, para analisar os problemas e procuraram compreender os requisitos e efeitos da implantação de um STI.

### **2.2 Dispositivos de Segurança Automotiva**

Cabe destacar também o programa japonês *Advanced Safety Vehicle* - ASV - (Veículo de Segurança Avançada - VSA) que, coordenado pelo Ministério de Transportes Japonês, e executado de 1991 a 1996, reuniu os fabricantes japoneses de automóveis (Daihatsu, Honda, Isuzu, Mitsubishi, Nissan, Subaru, Suzuki, Toyota) para o desenvolvimento de métodos e dispositivos que aumentassem a segurança do sistema de transportes [Paromtchik 1997] . Do programa ASV, destacam-se os conceitos de segurança da *Mitsubishi Motor Company*, que divide a segurança em ativa (a qual oferece ao motorista sinais de alerta e também afeta o funcionamento do veículo com o objetivo de prevenir acidentes) e passiva (cujo objetivo principal é o de evitar ou reduzir os danos físicos ao motorista e aos passageiros, caso um acidente ocorra).

Os principais dispositivos de segurança ativa mencionados pela Mitsubishi são citados a seguir:

1. Sistema de Alerta ao Motorista Contra Cansaço: o nível de vigilância do motorista é monitorado através do comportamento geral do carro (p.ex.: variações no controle da direção). Para alertar o motorista podem ser utilizadas: a emissão de uma fragrância estimulante; sons de alarme; vibrações no assento; vibrações no volante do veículo.
2. Sistema de Alerta contra Pressão Baixa nos Pneus: sensores monitoram a pressão dos pneus e alertam o motorista caso necessário.
3. Sistemas de Visão Aprimorados: faróis especiais e vidros que repelem a água ajudam o motorista a ter uma melhor visibilidade à noite ou sob chuva.
4. Sistema de Alerta de Ré: monitora veículos que estejam se aproximando por trás e alerta o motorista sobre a aproximação do veículo com sinais sonoros ou visuais.
5. Sistema de Navegação: determina a posição geográfica do veículo, utilizando métodos de localização sofisticados. O motorista é guiado por indicações em mapas e imagens dos locais.
6. Sistema de Alerta de Distância: controladores de velocidade a laser determinam a distância entre o veículo atual e o veículo da frente. Se a distância for muito pequena, um sinal sonoro é acionado para alertar o motorista.
7. Sistema de Alerta Lateral-Ré: câmeras estereo de vídeo monitoram as laterais e a traseira do veículo, para evitar (através de um alerta sonoro) que uma mudança de pista seja efetuada com outro veículo posicionado na lateral para onde o veículo atual tenta se dirigir.
8. Sistema de Alerta de Mudança de Pista: uma câmera monitora as linhas da pista. Se o veículo iniciar uma troca de pista sem o motorista ter acionado o pisca-alerta, um sinal sonoro alerta o motorista.
9. Controle Inteligente de Cruzeiro: um controlador de velocidade a laser e uma câmera de vídeo monitoram os veículos diante do veículo atual. Para prevenir o veículo atual de se aproximar demasiadamente dos que vão à frente, o motor e a transmissão são controlados automaticamente.
10. Sistema Automático de Evasão de Colisões: um sistema de múltiplas câmeras e sensores monitora as condições de tráfego. Se um perigo for detectado, um alerta é transmitido para o motorista e, se uma colisão parece inevitável, o veículo aciona automaticamente as medidas evasivas.
11. Sistema de Regulagem de Velocidade em Curvas: recebe transmissões de dados de curvas que estão adiante a partir de transmissores fixos posicionados próximos a estas curvas. Se o veículo estiver se aproximando da curva com uma velocidade acima da



recomendada, um alerta sonoro é acionado. Se o motorista não reduz a velocidade, o veículo reduz automaticamente a velocidade para o motorista..

E, dentre os dispositivos de segurança passiva para veículos, a Mitsubishi cita:

1. Absorção de Impactos: utilização de materiais leves e capazes de absorver a energia de eventuais colisões, tais como o CFRP (*Carbon Fiber Reinforced Plastic* - Plástico Reforçado com Fibra de Carbono).
2. Sistema de Proteção aos Ocupantes: adicionalmente aos air bags frontais (motorista e passageiro da frente), cada assento dianteiro também é equipado com air bags, para os passageiros de trás. Além disso, o veículo deve ser equipado com air bags laterais, pré-tensionadores dos cintos de segurança e vidros anti-laceração.
3. Sistema de Proteção ao Pedestre: câmeras auxiliam o motorista na visualização de pedestres, e o CFRP da dianteira do veículo ajuda a absorver a energia do impacto em um eventual atropelamento.
4. Sistema de Destravamento das Portas: as portas do veículo são automaticamente destravadas, em caso de colisão - para facilitar a saída do motorista e passageiros, e para facilitar operações de resgate.

Com o desenvolvimento de novas teorias e técnicas, o estudo de STI entra em uma segunda fase, de maior maturidade, em que produtos experimentais começam a ser desenvolvidos, para a resolução de problemas relacionados ao tráfego [Bertozzi 2000] [Paromtchik 1997]. Entre esses produtos, podemos citar, além dos equipamentos já mencionados: (i) sistemas de auxílio à baliza (utilizando sinais de radar, o sistema alerta o motorista sobre a proximidade de obstáculos durante a tarefa de estacionar o [DaimlerChrysler 2003]; (ii) sistemas que vigiam os olhos do motorista em busca de sinais de [Eriksson 1997]; (iii) sistemas baseados em visão para a pilotagem automática de veículos [Kelber et al. 2003] [Broggi 2000] ; (iv) interfaces homem-computador para veículos inteligentes [Cellario 2001]; (v) sistemas para comunicação automática em rede para veículos inteligentes [DaimlerChrysler 2003]; (vi) guias eletrônicos com mapas digitais - que mostram (através de dados e imagens) para o motorista todas as informações relevantes sobre a estrada [DaimlerChrysler 2003]; e, por fim, (viii) os sistemas de reconhecimento de sinais de trânsito [Bénallal 2003] [Fang 2003] [Liu 2002] [Shirvaikar 2004] [Paclík 1998] [Johansson 2002] [Miura et al. 2000] [Lauzière et al. 2001] [Estevez 1996] [Ghica et al. 1995] [de la Escalera et al. 1994] [Kellmeyer 1994] [Nakamura et al. 1999] [Liu 2001] [Vitabile et al. 2001] [Rodrigues 2002] [Ohara et al. 2002] [Jiang et al. 1996] [Gavrila 1999] [Steck 2003], dos quais este trabalho trata.

## 2.3 Pesquisas em Sistemas de Apoio ao Motorista

Nas pesquisas em Sistemas de Apoio ao Motorista, dentre vários projetos destacam-se os trabalhos em: desenvolvimento de sistemas de controle de veículos autônomos; e desenvolvimento de sistemas computacionais para o reconhecimento de sinais de trânsito.

### 2.3.1 Veículos Autônomos

Esta seção apresentará dois exemplos de projetos de sistemas para o controle de veículos autônomos. O primeiro projeto, brasileiro, está sendo desenvolvido na Universidade do Vale do Rio dos Sinos - UNISINOS. O segundo está em desenvolvimento na Universidade de Parma (*Università di Parma*), na Itália.

Em desenvolvimento na Universidade do Vale do Rio dos Sinos - UNISINOS - encontra-se um projeto que trata de tecnologias para automação veicular para prover soluções em mecatrônica e sistemas de apoio ao motorista. Alguns dos objetivos deste projeto, que envolve participantes de diversos cursos da graduação (tais como Engenharia Elétrica e Ciência da Computação) e da pós-graduação (tal como o Mestrado em Computação Aplicada), são [Kelber et al. 2003]:

- Desenvolvimento da tecnologia *Drive-By-Wire*;
- Desenvolvimento de Sistemas de Apoio ao Motorista;
- Desenvolvimento de Sistemas de Supervisão e Comando Remotos;
- Aplicações de Inteligência Artificial em robótica móvel;
- Aumento de segurança nas estradas;
- Exploração de locais de difícil acesso;
- Inspeções em ambientes de risco à saúde humana;
- Automatização de sistemas de transporte;
- Automação rural.

No abrangente projeto de [Kelber et al. 2003], cabe destacar o projeto para um veículo autônomo, cuja plataforma de testes é o veículo do tipo Mini-Baja que pode ser observado nas Figuras 2.1 e 2.2.



**Figura 2.1 Mini-Baja em Fase de Automatização [Kelber et al. 2003]**



**Figura 2.2 Controle Remoto Através de Comunicação por Telefonia Celular [Kelber et al. 2003]**

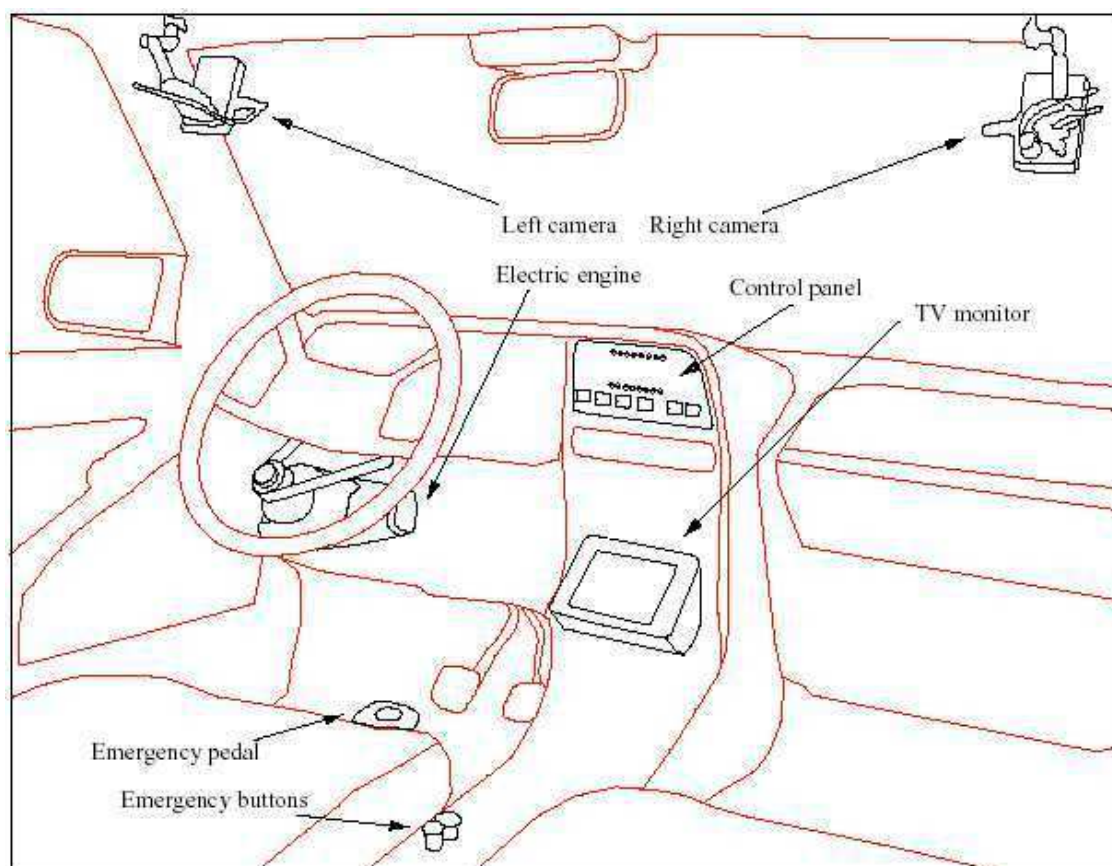
Na Figura 2.2, acima, pode ser observado novamente o Mini-Baja, da UNISINOS, respondendo a um comando de "virar à esquerda" enviado via telefone celular - como resultado parcial do projeto de instrumentação e controle de um veículo autônomo.

Outro projeto para a construção de um veículo autônomo é o Projeto ARGO [Broggi 2000]. Neste projeto, foi implementado um sistema embarcado capaz de reconhecer, a partir de imagens de vídeo, tanto a estrada quanto os outros veículos que nela trafegam - e guiar o veículo autônomo ARGO com sucesso pela estrada. Na Figura 2.3, abaixo, pode ser vista uma foto do veículo autônomo ARGO.



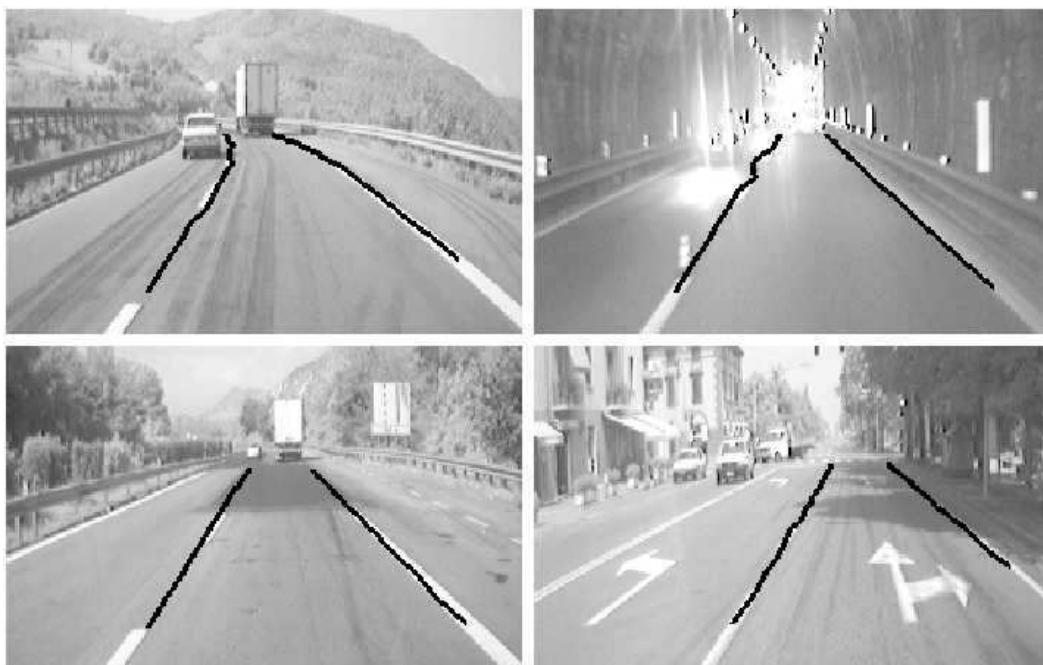
**Figura 2.3 O Veículo Autônomo ARGO [Broggi 2000]**

Na Figura 2.4, a seguir, podem ser observados os itens que compõem o sistema embarcado no Veículo Autônomo ARGO. Notar a presença das duas câmeras de vídeo e do motor elétrico - sob o volante - responsável pelo controle de direção do veículo.



**Figura 2.4 Esquema do Interior do Veículo Autônomo ARGO [Broggi 2000]**

Finalmente, nas imagens das Figuras 2.4 e 2.5, abaixo, podem ser observadas, respectivamente, as representações em vídeo do reconhecimento de estradas e do reconhecimento de outros veículos pelo sistema do ARGO.



**Figura 2.5 Reconhecimento de Estradas pelo Veículo Autônomo ARGO [Broggi 2000]**



**Figura 2.6 Reconhecimento de Outros Veículos  
pelo Veículo Autônomo ARGO [Broggi 2000]**

### 2.3.2 O Reconhecimento Inteligente de Sinais de Trânsito - RIST

O ato de dirigir um veículo baseia-se, quase que inteiramente, em informações visuais provenientes do ambiente. Essas informações visuais são detectadas e interpretadas pelo motorista para que o mesmo possa guiar o veículo em segurança pela rodovia. Parte dessas informações visuais são apresentadas na forma de sinais de trânsito, projetados com a finalidade de contribuir para o bom fluxo do trânsito e para a segurança [DENATRAN 2004] [DNIT 2004].

A presença de um sistema de reconhecimento inteligente de sinais de trânsito (RIST) no veículo deverá ajudar o motorista, através de uma interface homem-computador adequada (p.ex.: alertas sonoros), a: (i) permanecer alerta para os sinais de trânsito; (ii) evitar e resolver situações de risco (p.ex.: presença de obstáculos na pista; estar trafegando a uma velocidade acima do permitido em determinado local; tentar tomar o sentido oposto ao determinado para a via); (iii) manter-se informado sobre rotas a seguir.

Dada a importância de se oferecer apoio ao motorista em sua tarefa, a fim de contribuir para o aumento da segurança do tráfego de veículos nas rodovias, vários estudos sobre sistemas de RIST já foram desenvolvidos, alguns dos quais serão destacados na próxima seção.

### 2.3.3 Trabalhos Anteriores em Reconhecimento de Sinais de Trânsito

Nesta seção serão apresentados alguns trabalhos anteriores sobre o reconhecimento de sinais de trânsito, os quais constituem-se num exemplo do estado-da-arte das pesquisas na área.

Alguns dos trabalhos apresentados nesta seção mencionam a utilização dos modelos de cores RGB (*red, green, blue* - vermelho, verde, azul), HSI (*hue, saturation, intensity* - cor, saturação, intensidade), HSV (*hue, saturation, value* - cor, saturação, valor), e YUV (onde a componente Y representa o brilho, e as componentes U e V representam a cor). Os modelos de cores serão apresentados na seção 3.2 do Capítulo 3, com destaque para os modelos RGB e HSV.

#### 2.3.3.1 Trabalhos Envolvendo apenas a Detecção de Sinais de Trânsito

##### 2.3.3.1.1 Detecção Utilizando Cores

Bénallal (França) & Meunier (**Canadá**) [Bénallal 2003], apresentam um método de segmentação em tempo real de imagens para a detecção de sinais de trânsito utilizando cores. Em seu artigo, os autores apresentam um algoritmo para tal segmentação das cores da imagem - utilizando o modelo de cores RGB. Neste artigo, uma segmentação bem

sucedida de duas imagens é apresentada como resultado da aplicação do método proposto pelos autores.

#### 2.3.3.1.2 Detecção Utilizando Redes Neurais Artificiais

Fang et al. [Fang 2003], da **China (Taiwan)**, desenvolveram um sistema de detecção e acompanhamento de sinais de trânsito em imagens de vídeo. Esse sistema opera em três fases: na primeira, de detecção, duas redes neurais são utilizadas para extrair características de cor (no modelo de cores HSI) e de forma dos sinais de trânsito; a fase de localização (que determina a posição do sinal de trânsito na cena) baseia-se nas características extraídas na fase anterior, utiliza uma abordagem sob a lógica difusa (*fuzzy logic*); finalmente, a fase de acompanhamento do sinal de trânsito (*tracking*) a cada quadro da sequência de vídeo, dado que a velocidade do veículo é conhecida, utiliza um filtro de Kalman. Várias imagens comprovando o funcionamento do sistema são apresentadas, embora os autores não tenham incluído em seu artigo porcentagens de acerto relativas à operação de seu sistema.

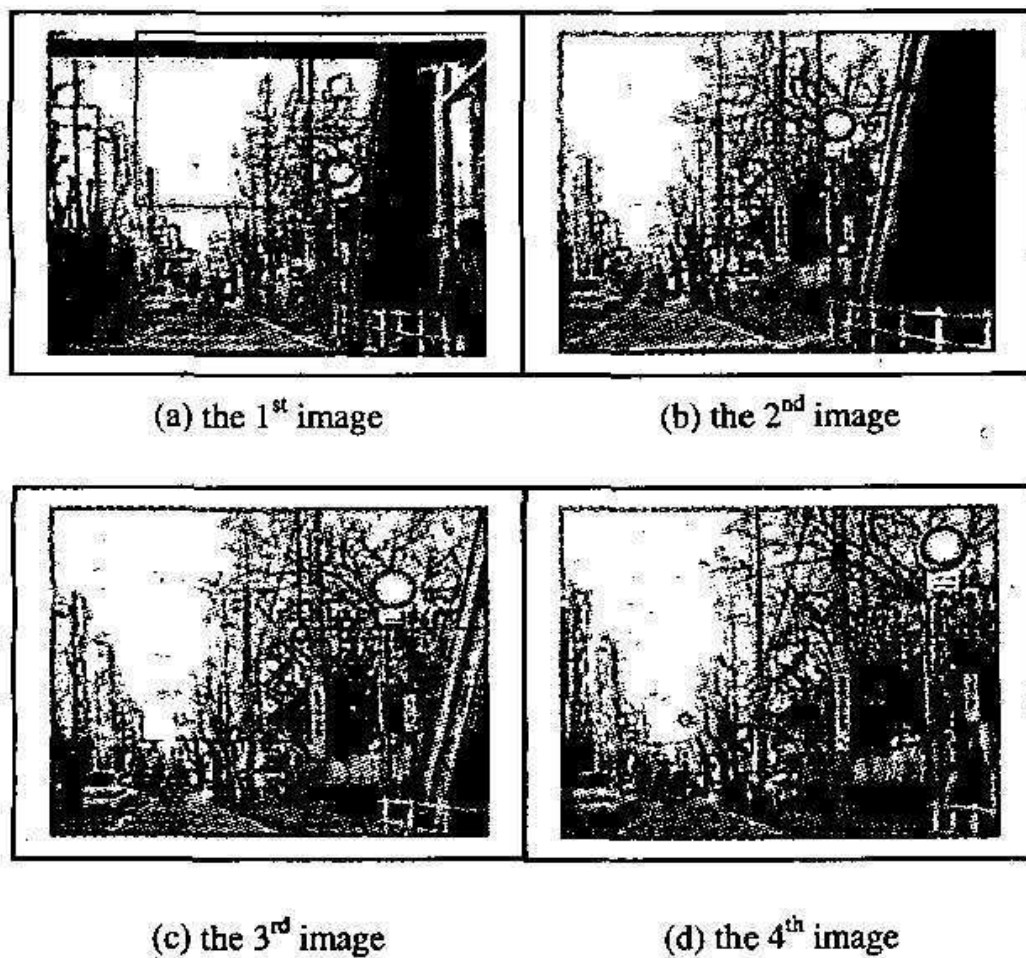
### 2.3.3.2 Trabalhos Envolvendo a Detecção e o Reconhecimento de Sinais de Trânsito

#### 2.3.3.2.1 Reconhecimento Utilizando Algoritmos Genéticos

Liu et al [Liu 2002], da **China**, descreve um sistema de reconhecimento de sinais de trânsito em tempo real para imagens em movimento, baseado em algoritmos genéticos. O sistema captura a imagem a partir de uma câmera de vídeo instalada em um veículo. Para a detecção do sinal de trânsito na imagem, é utilizado um filtro denominado pelos autores de Filtro de Vetor Simples (*Simple Vector Filter - SVF*). O SVF é capaz de, ao mesmo tempo, extrair uma cor específica da imagem (através do modelo de cores HSI) e remover os demais objetos de fundo. O reconhecimento de sinais de trânsito (limitados aos sinais de borda circular) é efetuado através de um algoritmo genético de passo (step genetic algorithm) - modelo elitista - o qual mantém a melhor solução na população. Os autores também empregaram um esquema reprodutivo baseado em níveis, no qual a população de indivíduos (que são as potenciais soluções para o reconhecimento da forma circular do sinal de trânsito) recebe níveis após a ordenação, os quais são baseados em seus valores de adequação à solução do problema.

Como resultados, os autores apresentam algumas imagens de exemplo relativas ao funcionamento do sistema. Algumas delas podem ser observadas na Figura 2.7 abaixo.





**Figura 2.7 Resultado da Busca com o Algoritmo Genético de Passo [Liu 2002]**

#### 2.3.3.2.2 Reconhecimento Utilizando Análise Relacional de Características

Shirvaikar [Shirvaikar 2004], dos **Estados Unidos da América (EUA)**, desenvolveu um sistema para detecção e interpretação automática de sinais de trânsito. O sistema proposto está organizado na seguinte arquitetura: análise espectral de características (onde foi utilizado um modelo de cor RGB do principal componente espectral de cada sinal de trânsito a ser classificado); análise espacial de características (onde são analisadas características como proporção e forma do objeto detectado); análise relacional de características (que verifica os relacionamentos entre objetos detectados na imagem); e, por fim, a localização e interpretação do sinal de trânsito. O autor relata o funcionamento bem sucedido de seu sistema sobre uma base de oito imagens, apresentadas em seu artigo.



#### 2.3.3.2.3 Reconhecimento Utilizando Classificação Estatística Não-paramétrica

Paclík [Paclík 1998], da **República Tcheca**, em sua dissertação de mestrado, apresenta um sistema automático de reconhecimento de sinais de trânsito cujo algoritmo utiliza a informação das bordas dos sinais de trânsito - ao invés da segmentação das cores - para a localização do sinal de trânsito na imagem. A classificação do sinal de trânsito é feita através de uma cascata (abordagem que, segundo o autor: diminui o número de entradas a serem classificadas por ramo da árvore de classificadores; diminui o tempo de classificação; e diminui a possibilidade da ocorrência de erros na classificação de sinais de diferentes classes) de classificadores estatísticos não-paramétricos (classificadores estatísticos paramétricos guardam apenas parâmetros descrevendo as distribuições de probabilidades de amostras de cada classe; já os classificadores estatísticos não-paramétricos, que o autor utiliza, guardam todos os vetores de padrões do conjunto de aprendizado - como chamado pelo autor - e estimam a densidade de probabilidade e seus parâmetros a partir de dados de treinamento). O autor apresenta diversas tabelas de resultados em seu trabalho, para diferentes sinais de trânsito (de um total de 1141 amostras, divididas heterogeneamente entre nove classes, conforme o tipo de sinal de trânsito), onde podem ser observadas médias de erros de reconhecimento na faixa de zero a 26,2%, dependendo do tipo da classe do sinal de trânsito reconhecido.

#### 2.3.3.2.4 Reconhecimento Utilizando Comparação de Modelos

A dissertação de mestrado de Johansson [Johansson 2002], da **Suécia**, trata do reconhecimento de sinais de trânsito de um veículo em movimento. O autor utilizou uma câmera ligada a um notebook para obter as imagens de teste a partir de um veículo - embora os testes com o sistema não tenham sido efetuados em tempo real. A detecção dos sinais de trânsito se dá através da segmentação da imagem pelas cores - no modelo de cores RGB. Uma vez detectados os sinais de trânsito, é efetuada uma nova segmentação no sinal encontrado, para que os dígitos (no caso dos sinais de trânsito de velocidade máxima permitida) sejam extraídos da placa - novamente, a segmentação é feita através das cores, no modelo de cores RGB. Extraídos os dígitos, o reconhecimento dos mesmos é realizado por uma comparação de modelos, em que o dígito extraído é comparado com um modelo matemático de cada um dos dígitos numéricos possíveis. O autor apresenta algumas imagens de seus testes, e alguns resultados correspondentes à velocidade de detecção de uma lista de cerca de 20 imagens. Segundo o autor, as detecções de cada imagem foram efetuadas num tempo de cerca de 200 milissegundos.

#### 2.3.3.2.5 Reconhecimento Utilizando Comparação de Padrões (Pattern Matching)

Miura et al. [Miura et al. 2000], do **Japão**, apresenta um sistema de visão ativa para o reconhecimento em tempo real de sinais de trânsito. Esse sistema, montado em um veículo, é composto por duas câmeras de vídeo (das quais uma está equipada com lentes grande-angulares, e a outra com lentes teleobjetivas) e um microcomputador pessoal equipado com placa de vídeo. O sistema detecta, com a câmera de lentes grande-angulares, a imagem candidata, que será analisada segundo as cores (modelo de cores YUV), a intensidade e a informação de formas. Para cada imagem candidata, a câmera com lentes teleobjetivas é direcionada para o local predito em que o sinal de trânsito estará localizado (assumindo o veículo em movimento) para, dessa forma, capturar uma imagem maior do sinal de trânsito. Em seguida, a identificação do sinal de trânsito é efetuada através de comparação de padrões (*pattern matching*) baseada em correlação normalizada (que, segundo os autores, é robusta às variações de iluminação em exteriores). O reconhecimento de ideogramas, em sinais informativos, é efetuado através da mesma técnica. Os autores apresentam como resultados: (i) para os 17 sinais de trânsito informativos submetidos ao sistema - detecção de 100%, rastreamento de 23,5%, 26 ideogramas detectados e 100% de identificação de ideogramas; (ii) para os 71 sinais de velocidade máxima permitida submetidos ao sistema - detecção de 97,2% e reconhecimento de 46,5%.

O trabalho de Lauzière et al. [Lauzière et al. 2001], do **Canadá**, descreve um sistema de identificação de sinais de trânsito baseado em modelos. Esse sistema utiliza, para a detecção dos sinais de trânsito, um método que leva em conta a formação física das cores (numa tentativa de oferecer à detecção a capacidade de superar o problema das variações de iluminação nas imagens coloridas). Esse método oferece "equações de formação de cores (EFCs)", que podem ser utilizadas para "prever" as respostas às cores por parte da câmera de vídeo. No entanto, para que essas previsões sejam calculadas, cada quantidade necessária às EFCs deve ser previamente conhecida. Assim, os autores utilizam dados de medição espectral (através do uso de um espectrofotômetro) dos sinais de trânsito; dados de resposta espectral da câmera de vídeo (obtidos através de um monocromador) e um modelo estatístico das condições de iluminação ao longo do dia. Admitindo as limitações do modelo, os autores relaxam os parâmetros para que ocorra uma adequada detecção das cores. Considerando que no Canadá existem diversos sinais de trânsito semelhantes, cuja diferenciação do significado depende fortemente de suas cores, os autores, além de terem utilizado a detecção através de um modelo físico das cores, utilizam para o reconhecimento uma base de dados de modelos de placas gerada automaticamente. Nessa base de dados, os sinais de trânsito são organizados em classes, sendo que as classes do nível 1 da hierarquia correspondem às oito cores possíveis dos sinais no Canadá. O reconhecedor, de posse da imagem candidata, varre a hierarquia em busca da classe mais adequada, e em seguida percorre essa classe em busca da imagem mais semelhante (através de comparação de padrões, ou *template matching*). Alguns dos resultados apresentados são: de 935 sinais de trânsito, 839 (89,7%) foram detectados em uma classe e 795 (85%) foram reconhecidos.

#### 2.3.3.2.6 Reconhecimento Utilizando Histograma de Cores

Estevez & Kehtarnavaz [Estevez 1996], dos **EUA**, apresentam o desenvolvimento de um sistema para reconhecimento de sinais de trânsito em tempo real. O sistema apresentado foi desenvolvido para reconhecer os seguintes sinais de advertência: Parada, Dê a Preferência e Proibição. A arquitetura do sistema foi organizada em seis módulos: segmentação de cores; localização de bordas; transformada de RGB (para atenuar os efeitos da iluminação da imagem); realce de bordas; extração do histograma; e classificação. O reconhecimento do sinal de trânsito é obtido através do uso de uma máscara semi-retangular, que varre a imagem com a finalidade de extrair as informações de cor para o histograma. Uma vez obtidas as informações do histograma, o mesmo é analisado matematicamente para a definição de qual sinal de trânsito está representado. Os autores utilizam, no módulo de segmentação de cores, o modelo de cores RGB. Os autores apresentam os seguintes resultados: (i) dos sinais de Parada submetidos ao sistema: 50% foram reconhecidos, 41% foram confundidos com os outros dois sinais (Dê a Preferência ou Proibição), e 9% não foram reconhecidos; (ii) dos sinais de Dê a Preferência: 37% foram reconhecidos, e 63% foram confundidos com os outros dois sinais; (iii) dos sinais de Proibição: 94% foram reconhecidos, e 6% não foram reconhecidos.

#### 2.3.3.2.7 Reconhecimento Utilizando Redes Neurais Artificiais

O trabalho desenvolvido por Ghica et al. [Ghica et al. 1995], do **Canadá**, utiliza basicamente redes neurais artificiais para a filtragem e reconhecimento dos sinais de trânsito. Neste sistema, a imagem de entrada é primeiro processada para a extração da cor e da informação geométrica da placa a ser identificada. A cor desejada é extraída através de um limiar, calculado a partir de uma cor de referência no modelo de cores RGB. Em seguida, um filtro morfológico é aplicado para realçar a imagem obtida e também para eliminar objetos pequenos e reunir objetos separados em partes disjuntas devido ao ruído. As coordenadas dos objetos resultantes são determinadas e esses objetos são separados da imagem original conforme essas coordenadas. Em seguida, os objetos são normalizados e enviados como entrada para uma rede neural, a qual efetua o reconhecimento. Sobre a extração das cores, os autores concluem que o método do limiar - utilizado sob o modelo de cores RGB - é inadequado para a detecção de placas em imagens mal iluminadas. Os autores apresentam 100% de sucesso na fase de reconhecimento, para um conjunto "maior do que 100 imagens", uma vez que as mesmas tenham sido detectadas na fase de extração de características.

Arturo de la Escalera et al. [de la Escalera et al. 1994], da **Espanha**, em seu trabalho de reconhecimento de sinais de trânsito para veículos autônomos, apresenta um algoritmo dividido em duas partes principais: a primeira, para a detecção do sinal de trânsito, utiliza a cor e a forma do mesmo para detectá-lo. A segunda parte do algoritmo é responsável pela classificação do sinal de trânsito, e para isso utiliza uma rede neural artificial. A detecção do sinal ocorre em duas etapas: na primeira, a cor é detectada através de um limiar

calculado sobre o modelo de cores RGB; na segunda etapa, a forma do objeto encontrado pelo limiar de cores é analisada. Para a classificação dos sinais de trânsito encontrados, é utilizada uma rede neural artificial cuja entrada, ao invés de corresponder às características resultantes de um pré-processamento, corresponde à própria imagem do sinal a ser classificado. A rede neural foi organizada da seguinte maneira: 900 neurônios na camada de entrada (usados para receber uma imagem de 30x30 pixels); duas camadas ocultas com quinze e cinco neurônios, respectivamente; e uma camada de saída, com dez neurônios - correspondentes aos sinais de trânsito a serem classificados.

Os autores apresentam resultados quantitativos insuficientes para uma avaliação de seu trabalho. Na Figura 2.8, abaixo, pode ser observado o reconhecimento de um sinal de velocidade máxima permitida.



**Figura 2.8 Reconhecimento de um Sinal de Velocidade Máxima Permitida [de la Escalera et al. 1994]**

Kellmeyer & Zwahlen [Kellmeyer 1994], dos EUA, apresentam em seu trabalho um sistema de reconhecimento de sinais de trânsito em imagens de vídeo que utiliza cores e redes neurais artificiais. O sistema apresentado digitaliza a imagem de uma rodovia, pré-processa essa imagem com relação às cores da mesma, e utiliza uma rede neural artificial -

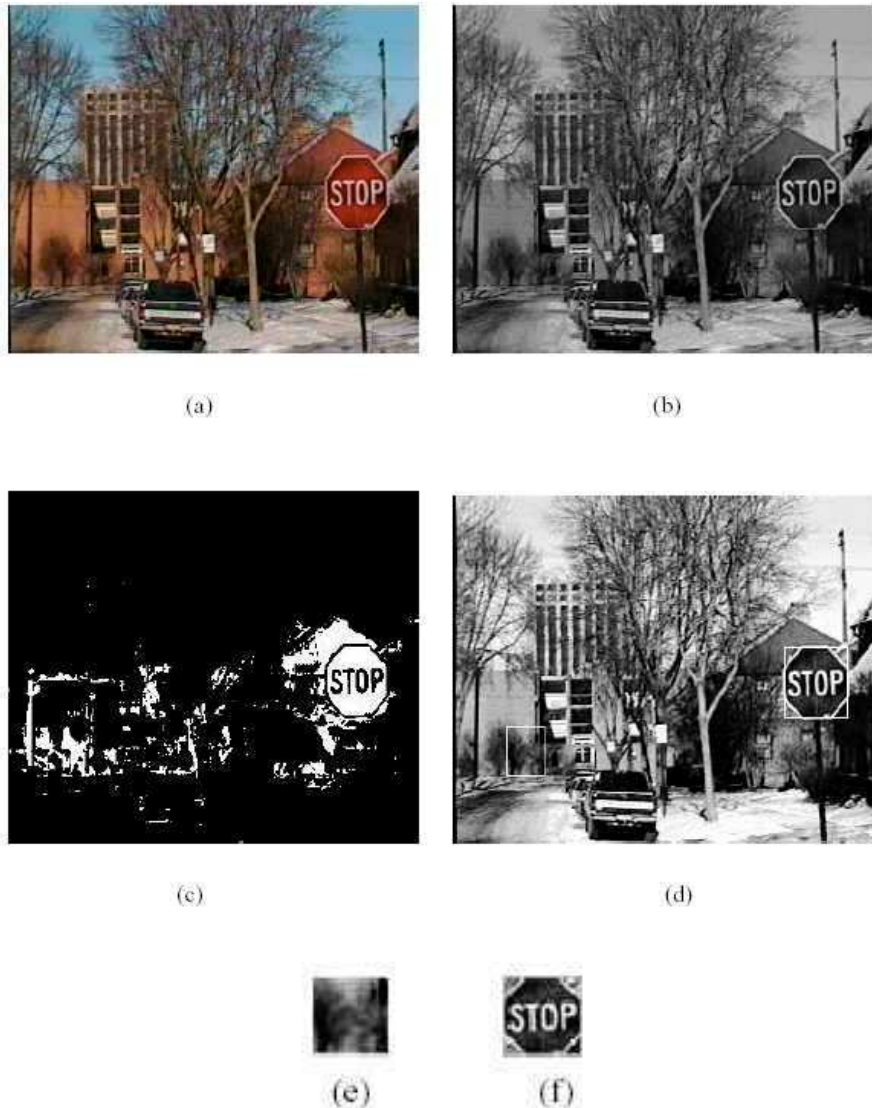
com aprendizado por *backpropagation* - para segmentar a imagem em oito cores importantes para a detecção dos sinais de trânsito. Em seguida, o sistema varre a imagem segmentada em busca de regiões que possivelmente representem os sinais de trânsito procurados. Tendo encontrado essas regiões, o sistema utiliza outra rede neural artificial - também com aprendizado por *backpropagation* - para determinar se a forma da região encontrada corresponde a um sinal de trânsito. Durante o pré-processamento das cores da imagem, as mesmas são convertidas do modelo de cores RGB para o modelo de cores HSI. Entre os resultados que os autores apresentam, estão informações como: (i) em uma gravação de vídeo de sinais desconhecidos para o sistema (efetuada por um operador de dentro de um veículo em movimento, e contendo 40 sinais de trânsito), foram reconhecidos 30 dos 40 sinais de trânsito presentes (resultando num reconhecimento de 75% dos sinais); (ii) dos 40 sinais de trânsito do vídeo, apenas 35 eram suficientemente grandes para o sistema detectá-los - assim, levando em consideração a detecção de 30 em 35, a porcentagem de reconhecimento sobe para 86%.

Nakamura et al. [Nakamura et al. 1999], do **Japão**, apresenta um reconhecedor de sinais de trânsito indicadores de velocidade máxima permitida. O sistema utiliza o modelo de cores RGB para detectar áreas da imagem que possuam a cor vermelha. Em seguida, uma rede neural artificial multi-camadas é utilizada para a detecção do anel circular do sinal de trânsito. Após essa etapa, outra rede neural artificial é utilizada para o reconhecimento dos dígitos indicadores de velocidade (se o dígito da unidade não for zero, o sinal não é de velocidade máxima permitida e é descartado; se for zero, então o sistema passa para o reconhecimento do dígito da dezena - uma vez reconhecido, seu valor é multiplicado por dez, resultando na velocidade máxima permitida expressada pelo sinal de trânsito reconhecido). Os autores apresentam um resultado, para um total de 300 experimentos, de 100% de detecção e de 98,3% de reconhecimento dos sinais de trânsito de velocidade máxima permitida.

Um sistema baseado em visão para o reconhecimento de sinais de trânsito de parada obrigatória é apresentado por Liu & Ran [Liu 2001], dos **EUA**. O sistema dos autores possui dois módulos principais: o módulo de detecção (que utiliza o modelo de cores HSV para segmentar a imagem) e o módulo de reconhecimento (que utiliza uma rede neural artificial para o reconhecimento e outra rede neural artificial para a validação - esta última utilizada com a finalidade de reduzir a ocorrência de falsos positivos, isto é, objetos incorretamente reconhecidos como sinais de parada obrigatória). A detecção leva em consideração, além das cores do objeto candidato: a largura; a proporção; e a simetria do objeto candidato. A rede neural de reconhecimento possui 900 neurônios de entrada, 6 neurônios na camada oculta, e 1 neurônio de saída. A rede neural de validação possui 900 neurônios na camada de entrada, 100 neurônios na camada oculta, e 900 neurônios na camada de saída. Antes de ser enviada para a rede neural artificial, a imagem do objeto candidato tem seu histograma de cores equalizado, para facilitar a análise - segundo os autores. A entrada da rede neural será, então, uma imagem de 30x30 pixels, e a saída será uma resposta se o objeto é ou não um sinal de parada obrigatória. Para uma base de 540 imagens candidatas, os autores apresentam um resultado de aproximadamente 95% de

detecção. Para o reconhecimento, os autores apresentam um erro de 0,025 (sem expressarem porcentagem e sem informarem sobre qual quantidade de imagens).

A Figura 2.9, abaixo, mostra um dos resultados obtidos por [Liu 2001].



**Figura 2.9 Exemplo de Reconhecimento de Sinal de Trânsito [Liu 2001]: (a) imagem original; (b) imagem em tons de cinza; (c) segmentação; (d) reconhecimento; (e) objeto potencial detectado e normalizado; (f) objeto potencial pronto para ser enviado ao módulo de reconhecimento**

Vitabile et al. [Vitabile et al. 2001], da **Itália**, apresenta um sistema de reconhecimento de sinais de trânsito que utiliza uma técnica de agregação dinâmica de pixels no modelo de cores HSV. Esse sistema utiliza uma câmera de vídeo para a captura das imagens, e está organizado nos seguintes blocos principais: (i) aquisição de imagem; (ii) análise de cores; (iii) análise das formas; (iv) classificação neural das regiões candidatas. A etapa de análise de cores é subdividida pelos autores nos seguintes passos: (a) aplicação de filtros HSV padrão; (b) geração de subimagens e determinação dos pixels que serão as sementes (pixels cuja saturação é uma média da saturação dos demais); (c) agregação dos pixels através de um limiar dinâmico, dependendo do valor de saturação das sementes. O objetivo deste método de segmentação de cores é o de realçar os pixels úteis para a detecção do sinal de trânsito, e limitar a presença dos pixels desnecessários. Após a segmentação das cores da imagem, o sistema proposto pelos autores busca pelos formatos (previamente determinados) de sinais de trânsito, através de medidas de similaridade sobre os objetos candidatos. Em seguida, as subimagens candidatas são submetidas à rede neural artificial para a classificação. Alguns dos resultados apresentados são: para 318 sinais de trânsito com anel circular vermelho, foram corretamente classificados 84%; para 246 sinais triangulares com borda vermelha, foram classificados 88%; para 67 sinais de anel circular azul, foram classificados 100%.

Rodrigues [Rodrigues 2002], do **Brasil**, apresenta, em sua dissertação de mestrado, um sistema de localização e reconhecimento de sinais de trânsito utilizando um mecanismo de atenção visual e redes neurais artificiais. Em seu trabalho, o autor organiza seu sistema em dois módulos principais: um módulo de detecção, para localizar os sinais de trânsito em cada quadro de uma sequência de imagens (é utilizada uma câmera de vídeo para a captura das imagens); e um módulo de reconhecimento, para a classificação das imagens detectadas. O módulo de detecção utiliza um mecanismo de atenção baseado em saliência (bottom-up), o qual é constituído a partir de uma Pirâmide Gaussiana, e também a partir de operadores locais de orientação. O módulo de reconhecimento utiliza uma rede neural artificial multicamadas com aprendizado por backpropagation. A entrada dessa rede neural é a própria imagem a ser classificada - há um neurônio na camada de entrada por pixel da imagem. Na camada oculta, após vários testes, o autor chegou a um número de neurônios igual a cinco vezes o número de classes a serem classificadas (sete classes). Na camada de saída, há um neurônio por classe a ser classificada, seguindo uma estratégia "o vencedor leva tudo" (*winner takes all*). O autor apresenta, dentre vários resultados, uma tabela que mostra porcentagens de acerto que variam de 8,23 a 35,29% para as sete classes de sinais de trânsito a serem classificadas pela rede neural artificial.

Ohara & Nishikawa [Ohara et al. 2002], do **Japão**, apresentam um sistema de detecção e reconhecimento de sinais de trânsito usando redes neurais artificiais multicamadas. Após vários passos de pré-processamento para mascarar as áreas irrelevantes da imagem colorida de entrada (áreas sem sinais de trânsito), as áreas candidatas são enviadas para duas redes neurais artificiais. A primeira rede neural é utilizada para identificar a cor da região candidata. A segunda rede neural serve para o reconhecimento da forma do objeto na região candidata. O reconhecimento final do sinal de trânsito é efetuado através de uma

comparação de padrões. Os autores afirmam ter obtido uma porcentagem de reconhecimento de sinais de trânsito acima de 95% "para a maior parte dos dados" (são mostradas taxas de reconhecimento, para 12 imagens, que variam de 87 a 100%).

#### 2.3.3.2.8 Reconhecimento Utilizando Similaridades

O trabalho de Jiang (**China**), Choi (**Coréia do Sul**) & Zheng (**China**) [Jiang et al. 1996] enfoca o reconhecimento morfológico de sinais de trânsito, de maneira a possibilitar o reconhecimento de formas (p.ex.: setas indicadoras de sentido) presentes no interior dos sinais de trânsito sul-coreanos. A técnica de reconhecimento apresentada pelos autores utiliza medidas quantitativas de similaridade baseadas na matemática morfológica (utilizada em processamento não-linear de sinais) e em transformações binárias das imagens. Os autores apresentam 100% de reconhecimento do conteúdo de sinais de trânsito (para 11 sinais de trânsito de indicação de sentido) para quatro diferentes métodos de medida de similaridade. Os autores também afirmam que obtiveram os mesmos 100% de reconhecimento para 22 sinais de trânsito sul-coreanos de advertência.

O trabalho apresentado por Gavrilin (**Alemanha**) & Philomin (**EUA**) [Gavrilin 1999], trata da detecção em tempo real de objetos para veículos inteligentes. Seu sistema utiliza um método de detecção de objetos - baseado em formas - através de transformadas de distâncias. Esse método utiliza uma hierarquia de templates para capturar a variedade de formas dos objetos. Essas hierarquias, por sua vez, podem ser geradas de acordo com o objeto que se deseja detectar - no caso deste trabalho, foram escolhidos os pedestres e os sinais de trânsito. O sistema envolve uma busca de similaridade do mais geral para o mais específico na hierarquia de formas e nos parâmetros de transformação. Os autores apresentam o resultado de 95% de reconhecimento para uma base de 1000 sinais de trânsito (imagens obtidas nas condições: com ou sem chuva; de dia ou à noite), e o resultado de 75-85% de reconhecimento para uma base de 700 imagens de pedestres.

#### 2.3.3.2.9 Reconhecimento Utilizando a Metodologia Padrão do Processamento de Imagens

Em seu sistema de reconhecimento de sinais de trânsito, desenvolvido como trabalho final na Escola de Engenharia e Tecnologia da Informação da Universidade de Berne, da **Alemanha**, Steck & Müller [Steck 2003] dividem as tarefas em cinco módulos: pré-processamento; segmentação; extração de características; localização; e classificação. O pré-processamento consiste em uma normalização das dimensões da imagem de entrada; a segmentação (que utiliza o modelo de cores HSI) é utilizada para detectar os sinais de trânsito na imagem através de suas cores; a extração de características é executada através de uma detecção de bordas e de uma transformação de distância; a localização do sinal de trânsito na imagem é efetuada através de comparação de padrões (template matching); e, finalmente, na fase de classificação, as cores são novamente utilizadas para determinar a



qual classe pertence a imagem candidata localizada. Ainda na fase de classificação, os dígitos do sinal (no caso de um sinal de trânsito de velocidade máxima permitida) são extraídos, normalizados, e classificados através de comparação de padrões.

## **2.4 Considerações Finais Sobre Este Capítulo**

Neste capítulo foram apresentados os sistemas de transporte inteligentes, juntamente com diversas observações sobre a segurança no trânsito - as quais englobam desde a criação de dispositivos de segurança veiculares até o desenvolvimento de sistemas de apoio aos motoristas.

Neste capítulo também foi feita uma breve revisão de uma amostra dos trabalhos que representam o estado da arte do reconhecimento de sinais de trânsito. A partir destes trabalhos, o autor tomou conhecimento de várias técnicas usadas por pesquisadores de diferentes partes do mundo. Essas técnicas são utilizadas para tentar resolver o problema da detecção e do reconhecimento de sinais de trânsito em imagens.

Embora existam trabalhos que não utilizam as cores dos sinais de trânsito para a detecção dos mesmos, essa característica é muito importante - pois destaca o sinal de trânsito em uma paisagem, facilitando a visualização por parte do motorista.

Contudo, o uso das cores para a detecção dos sinais de trânsito traz consigo problemas inerentes à iluminação da cena, o que pode alterar a forma como a cor do sinal de trânsito é percebida. Essas questões serão abordadas com maior detalhe no Capítulo 3, na seção 3.2.

Os trabalhos anteriores, apresentados neste capítulo, demonstram que existe uma determinada maneira de se tratar o problema do reconhecimento de sinais de trânsito. Essa maneira é constituída por passos como: a captura da imagem; detecção da imagem; o pré-processamento dessa imagem; a extração de características de interesse dessa imagem; e, por fim, o reconhecimento dessas características extraídas. Esses passos serão definidos com detalhes no Capítulo 3.

### 3 FUNDAMENTOS DE PROCESSAMENTO DE IMAGENS

Os estudos sobre o processamento de imagens obtêm o seu impulso a partir de duas fontes: o desejo de melhorar a qualidade das informações pictóricas para os seres humanos; e o desejo de construir máquinas com sistemas perceptivos autônomos, capazes de processar dados a partir de uma cena [Gonzalez 1993].

Neste capítulo, serão apresentados alguns conceitos fundamentais relacionados ao processamento de imagens digitais, juntamente com uma breve introdução aos seguintes tópicos: os modelos de cores; e a biblioteca de Processamento de Imagens "Sistema-IMG" [Kim 2004].

#### 3.1 Conceitos Básicos

Nesta seção serão apresentados alguns conceitos básicos sobre o processamento digital de imagens.

##### 3.1.1 A Representação Digital da Imagem

[Gonzalez 1993] define uma imagem monocromática, ou simplesmente uma imagem, como uma função bidimensional de intensidade luminosa  $f(x,y)$ , onde  $x$  e  $y$  denotam coordenadas espaciais e o valor de  $f$  em qualquer ponto  $(x,y)$  é proporcional ao brilho (ou nível de cinza) da imagem naquele ponto.



**Figura 3.1 Exemplo de Imagem Monocromática**

**(Em Tons de Cinza) [Kelber et al. 2003]**

Dessa forma, uma imagem digital é uma imagem  $f(x,y)$  que sofreu uma discretização tanto com relação às coordenadas espaciais quanto com relação ao seu brilho (no caso de tratar-se de uma imagem monocromática). Os elementos constituintes de uma imagem digital são chamados de elementos de imagem, ou simplesmente de *pixels* (abreviação de *picture elements*).

Embora [Gonzalez 1993] trate principalmente de imagens monocromáticas, atualmente existem monitores de vídeo que permitem a visualização de cores correspondentes à combinação de 24 bits de informação por pixel da imagem - o que significa que temos oito bits (256 valores possíveis) disponíveis para cada componente do modelo de cores RGB (apresentado na seção 3.2.1), resultando em mais de 16 milhões de cores ( $256^3$  cores, para 24 bits) que podem ser visualizadas em uma imagem [Foley 1997].

### 3.1.2 Os Passos Fundamentais no Processamento de Imagens

Para o processamento de imagens ocorrer, devem ser seguidos alguns passos fundamentais [Gonzalez 1993].

O primeiro deles é o da aquisição da imagem - para tal, podemos utilizar uma câmera de vídeo, uma câmera fotográfica digital, ou um digitalizador de imagens (*scanner*).

O próximo passo é o de pré-processamento. No pré-processamento, o objetivo é o de processar a imagem de maneira que o resultado desse processamento auxilie na resolução do problema para os processos subsequentes.

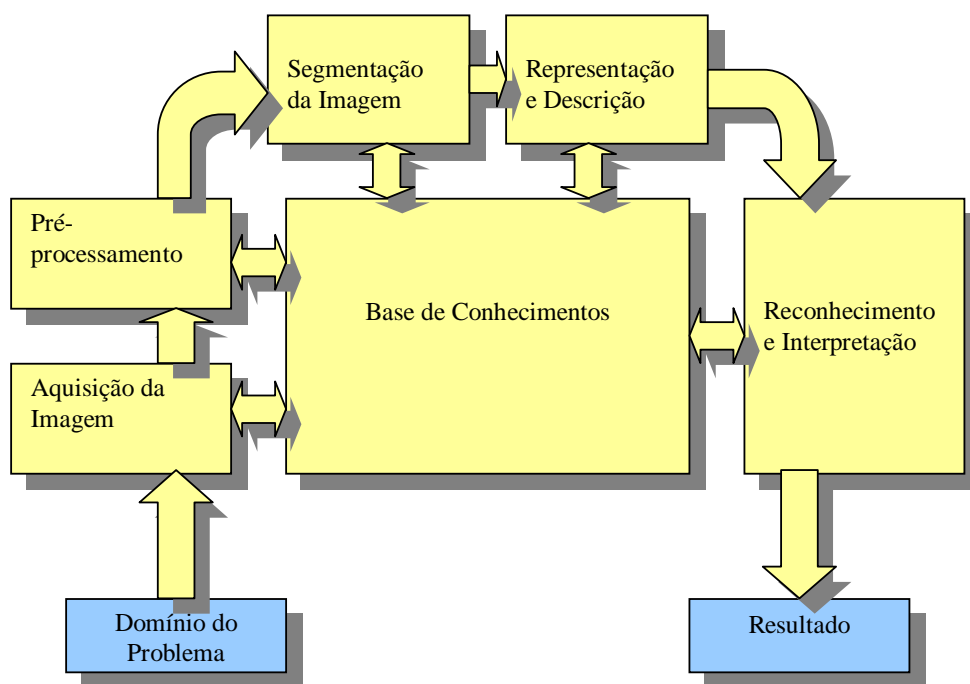
O terceiro passo lida com a segmentação. De maneira geral, a segmentação consiste em se extrair da imagem apenas aquelas áreas que interessam para a resolução do problema.

O quarto passo envolve duas fases: a primeira é a de representação das áreas obtidas a partir da segmentação da imagem. Essa representação poderá ser, por exemplo, através da forma ou do contorno dessas áreas. O importante, neste passo, é que cada uma das áreas extraídas seja representada de maneira adequada para o processamento. O que nos leva para a segunda fase deste quarto passo: a fase de descrição (também chamada de fase de extração de características). Na fase de descrição, as áreas - agora representadas - deverão ser convertidas em um conjunto de informações quantitativas de interesse para a resolução do problema. Essas informações devem ser capazes de fornecer uma diferenciação adequada entre objetos de classes diferentes.

O quinto passo consiste também em duas fases: a do reconhecimento - que fornecerá um rótulo para um objeto, baseado nas informações fornecidas pelos seus descritores; e a da interpretação, que envolve a atribuição de um significado para os objetos reconhecidos.

Em contato com todos esses cinco passos deve existir uma base de conhecimentos. Essa base de conhecimentos é o que dá sentido ao processamento da imagem, uma vez que essa base contém informações relevantes sobre o problema a ser resolvido. Exemplos de informações importantes que podem estar contidas na base de conhecimentos são: conhecimento prévio dos locais na imagem onde a informação desejada poderá estar; ou listas de itens a serem procurados na imagem.

Todos esses passos fundamentais estão ilustrados na Figura 3.2, abaixo.

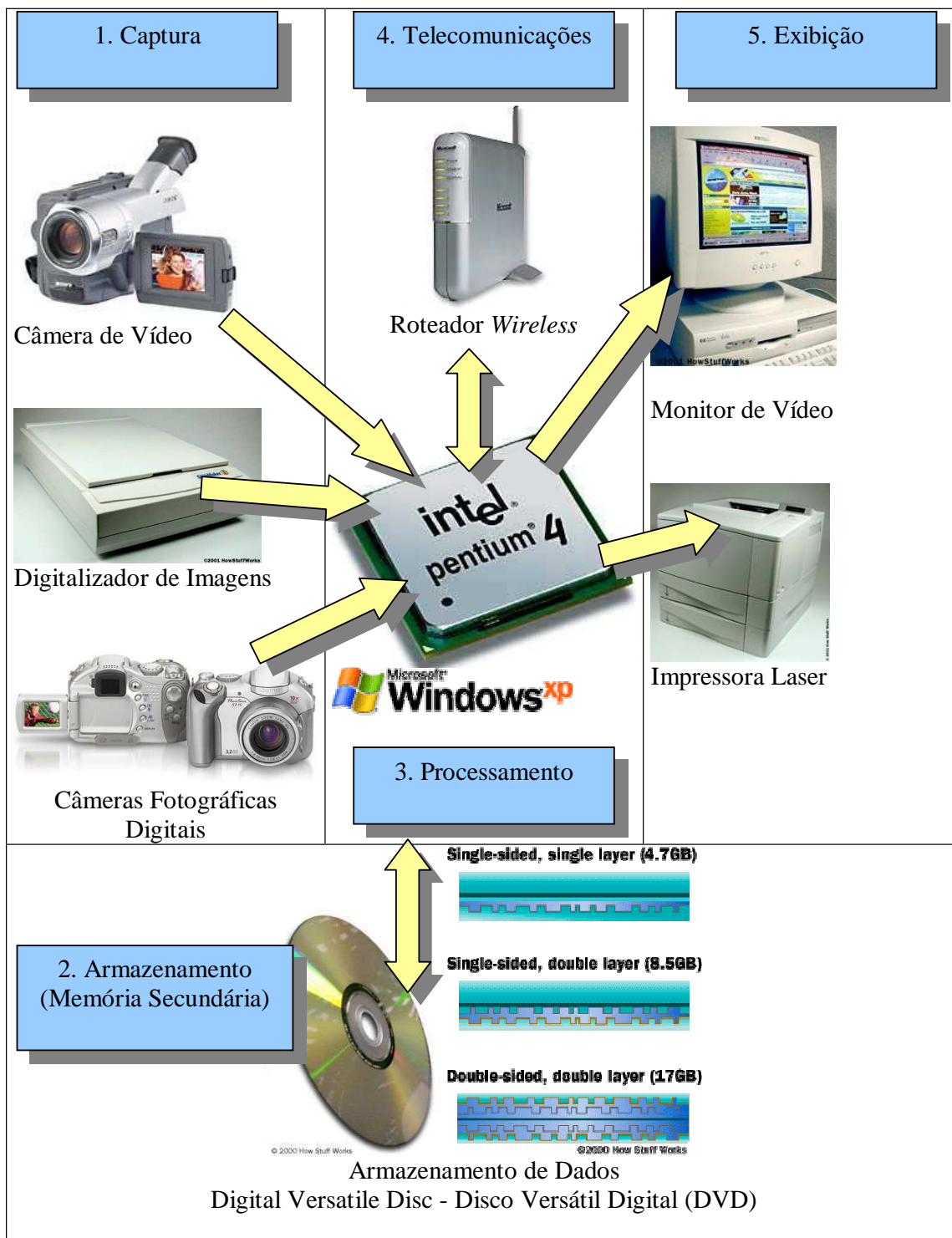


**Figura 3.2 Passos Fundamentais no Processamento de Imagens**

### 3.1.3 Equipamentos Envolvidos no Processamento de Imagens

Os equipamentos envolvidos no processamento de imagens estão organizados em cinco tipos: (i) equipamentos para a captura de imagens; (ii) equipamentos de armazenamento de dados; (iii) equipamentos para o processamento de dados; (iv) equipamentos de telecomunicação; e, por fim, (v) equipamentos de exibição de imagens; [Gonzalez 1993].

A Figura 3.3, a seguir, mostra - através de exemplos - como são os relacionamentos entre os cinco tipos de equipamentos, conforme [Gonzalez 1993].



**Figura 3.3 Os Cinco Elementos Básicos do Processamento de Imagens e Seus Relacionamentos**

### 3.2 Modelos de Cores

O objetivo dos modelos de cores é o de facilitar a especificação das cores conforme algum padrão. A especificação das cores em um padrão, ou modelo, é efetuada através de um sistema de coordenadas de três dimensões (3-D), e através de um subespaço nesse sistema 3-D no qual cada uma das cores possíveis é representada por um único ponto [Gonzalez 1993].

Um modelo de cores é incapaz de representar todas as cromaticidades visíveis, por ser um subconjunto do espaço de cores de todas as cromaticidades possíveis. Esse subconjunto de cores representáveis pelo modelo recebe o nome de *gamut* de cores [Foley 1997]. Cada dispositivo de captura, telecomunicação, ou exibição de cores possui seu próprio *gamut* de cores, de acordo com um modelo de cores específico.

Diferentes modelos de cores foram criados, visando atender aos objetivos de diferentes dispositivos ou finalidades.

Temos atualmente os seguintes modelos mais usados: o modelo RGB (*componentes Red, Green, Blue* - Vermelho, Verde, Azul) foi criado para tubos de raios catódicos coloridos (*color cathode-rays tubes* - CRT), utilizados em monitores de vídeo; o modelo YIQ (Y significa luminância, enquanto a cromaticidade é descrita pelas componentes I e Q) foi criado para a transmissão de sinais de televisão (preto e branco e em cores); o modelo CMY (*componentes Cyan, Magenta, Yellow* - Ciano, Magenta, Amarelo) foi desenvolvido para impressoras coloridas [Gonzalez 1993] [Foley 1997].

Infelizmente, os modelos de cores para equipamentos não oferecem facilidades para o processamento de imagens, porque eles não se relacionam diretamente com as noções intuitivas de cor (*hue*), saturação (*saturation*), e brilho (*brightness*) [Foley 1997].

Assim, foram criados modelos específicos para a manipulação das cores em imagens. Entre os mais utilizados, estão o modelo HSI (*Hue, Saturation, Intensity* (intensidade)) e o HSV (*Hue, Saturation, Value* (valor)) [Gonzalez 1993].

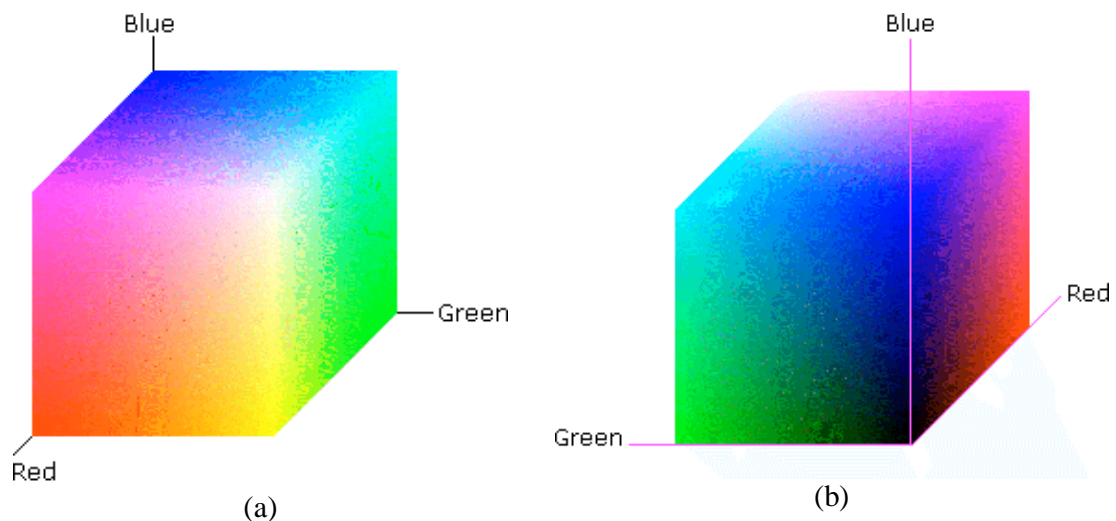
Os modelos de cores frequentemente utilizados no processamento de imagens são o RGB, o YIQ e o HSI [Gonzalez 1993]. Nas seções a seguir, serão brevemente apresentados os modelos RGB e HSV. Embora diferentes em sua nomenclatura e em seus respectivos cálculos de conversão para (ou a partir de) outros modelos de cores, os modelos HSI e HSV possuem os mesmos princípios de interpretação das informações de cromaticidade.

Os algoritmos de conversão do modelo RGB para o modelo HSI, e vice-versa, são apresentados em [Gonzalez 1993]. Algoritmos de conversão do modelo RGB para o modelo HSV, e vice-versa, são apresentados em [Foley 1997].

### 3.2.1 O Modelo de Cores RGB

No modelo de cores RGB (*Red, Green, Blue* - Vermelho, Verde, Azul), cada uma das cores é representada de acordo com os valores de seus respectivos componentes espectrais. Isso equivale a dizer que os componentes espectrais de uma cor correspondem a valores para o vermelho, para o verde, e para o azul [Gonzalez 1993].

Dessa forma, a representação das cores no modelo RGB baseia-se em um sistema de coordenadas Cartesiano tridimensional (3-D). Por conveniência, assume-se que todos os valores de cores foram normalizados, para que as magnitudes das componentes R, G, e B estejam no intervalo  $[0,1]$ . O subespaço de cores de interesse é, assim, o cubo mostrado na Figura 3.4 (a) e (b).



**Figura 3.4 Modelo de Cores RGB:**

**(a) visto a partir do ponto (1,1,1); (b) visto a partir da origem (0,0,0) [Microsoft 2004]**

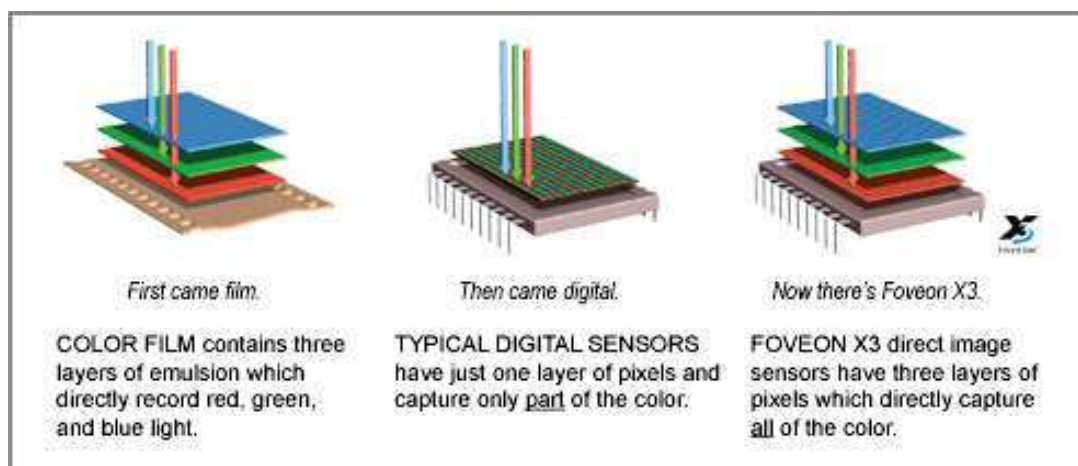
Na Figura 3.4 (a), pode-se perceber que os valores para o vermelho (*Red*), para o verde (*Green*) e para o azul (*Blue*) situam-se em três vértices do bloco, e também que os valores para o ciano, para o magenta, e para o amarelo, situam-se em três outros vértices do bloco. Na origem (observar a Figura 3.4 (b)), localiza-se a cor preta, e na extremidade da diagonal principal (observar a Figura 3.4 (a)) situa-se a cor branca. Cabe notar que os tons de cinza localizam-se sobre a diagonal principal do cubo RGB.

Uma cor no sistema RGB é então definida como sendo um vetor cuja origem coincide com a origem do cubo RGB, e cuja extremidade encontra-se ou no interior do cubo, ou coincidindo com algum ponto de uma de suas faces ou arestas, ou coincidindo com um dos vértices.

As imagens, no sistema RGB, são compostas por três planos independentes de imagem - um para cada componente desse sistema de cor. Assim, quando exibidas em um monitor de vídeo, essas três imagens são combinadas para que o vídeo reproduza a imagem com o resultado da composição das três cores. Equivale a dizer que a cor de cada pixel da imagem, no vídeo, será definida por três valores - um para cada componente do modelo RGB.

As câmeras fotográficas digitais utilizam o sistema RGB para a captura de imagens, o que eleva a importância desse modelo de cores para o processamento de imagens [Gonzalez 1993].

Atualmente, as câmeras fotográficas digitais utilizam sensores eletrônicos que digitalizam o sinal luminoso capturado. Em seguida, as informações obtidas são armazenadas em sua memória secundária. A importância do modelo de cores RGB pode ser melhor percebida através do exemplo apresentado na Figura 3.5 abaixo.



**Figura 3.5 Capturando A Luz - Da Esquerda para a Direita: Filme Fotográfico; Sensores Digitais Convencionais; Sensor Foveon X3 [Foveon 2004]**

A Figura 3.5 demonstra a importância das componentes de cor RGB, pois: no filme fotográfico, existem três camadas de emulsão - cada uma responsável pela captura de uma das componentes RGB; nos sensores digitais convencionais, existe uma matriz de sensores especializados na captura de cada um dos componentes RGB, dispostos de maneira alternada sobre o mesmo plano (o que divide por três a resolução anunciada para as câmeras digitais convencionais, uma vez que um terço dos pontos do sensor capturará a componente R, outro terço capturará a componente G, e o terço final capturará a componente B da imagem); finalmente, temos o recente sistema Foveon X3, cuja

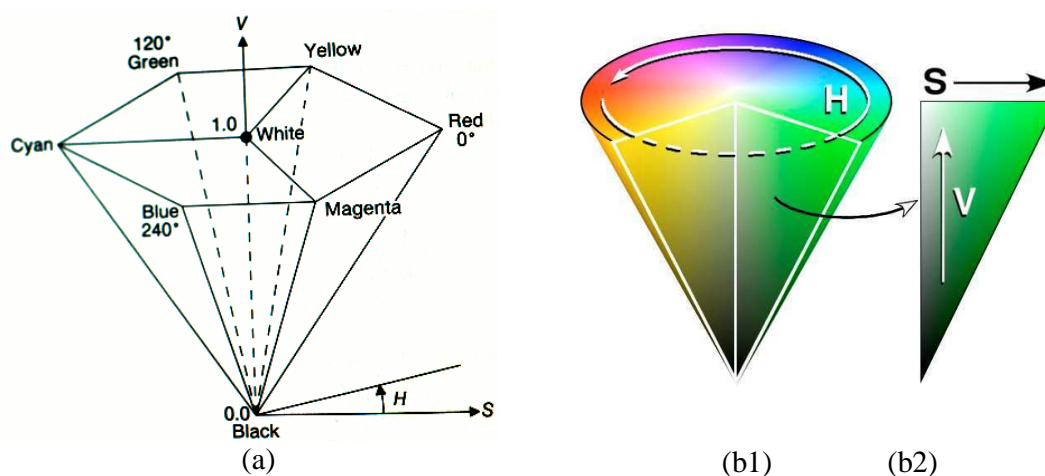


disposição dos sensores é idêntica à dos filmes fotográficos - o que permite que cada coordenada do plano do sensor capture informações sobre as três componentes RGB.

Porém, existe uma desvantagem inerente ao processamento de imagens através da utilização do modelo de cores RGB: a dificuldade em se lidar com a iluminação da imagem [Gonzalez 1993]. Se precisamos processar informações de uma imagem, e parte da mesma encontra-se mal iluminada - ou simplesmente com iluminação diferente da do restante da cena - torna-se muito difícil compensar as diferenças para que a imagem seja processada adequadamente. Existem algumas abordagens que tentam minimizar esse problema [Gonzalez 1993], mas a melhor saída é a utilização de um modelo de cores que, por si, ofereça melhores condições de lidar com a iluminação de uma imagem. Um dos modelos de cores capaz de oferecer tal facilidade é o modelo HSV, que será apresentado na próxima seção.

### 3.2.2 O Modelo de Cores HSV

O modelo de cores HSV (Hue, Saturation, Value - Cor, Saturação, Valor) possui o apelo intuitivo das designações matiz (análoga à cor), sombra (análoga à saturação) e tonalidade (análoga ao valor), provenientes da manipulação de tintas [Foley 1997]. Na Figura 3.6, abaixo, podem ser observadas duas representações tridimensionais do modelo HSV.



**Figura 3.6 O Modelo de Cor HSV: (a) Representado como um Hexacone [Foley 1997]; (b1) Representado como um Cone [Pierce 2004]; (b2) Vista em Corte de um Segmento do Cone HSV [Pierce 2004]**

O sistema de coordenadas usado no modelo de cores HSV é cilíndrico. Utilizando como referência a Figura 3.6 (b1), vemos que a componente H assume valores angulares no sistema de coordenadas cilíndrico (no intervalo  $[0^0, 360^0]$ ), a componente S assume valores

lineares ao longo do raio do cone (no intervalo  $[0,1]$ ), e a componente  $V$  assume valores lineares ao longo do eixo do cone (no intervalo  $[0,1]$ ).

No vértice do cone, o valor ( $V$ ) é igual a zero (o que corresponde à cor preta); Na base do cone, onde estão as cores com maior brilho,  $V=1$  (para  $S=0$ , temos a cor branca). Cabe lembrar que, no plano onde  $V=1$ , as cores não possuem necessariamente o mesmo brilho percebido [Foley 1997]. Ao longo do eixo do cone situam-se os tons de cinza.

A cor ( $H$ ) é definida por um valor angular (Figura 3.6 (a)): um ângulo igual a zero corresponde à cor vermelha; um ângulo de  $120^\circ$  corresponde à cor verde, e assim por diante.

Quando a saturação ( $S$ ) vale zero, o valor da cor ( $H$ ) torna-se irrelevante, sendo tratado como "indefinido". Porém, quando  $S$  é maior do que zero,  $H$  torna-se relevante.

Quando  $V=1$  e  $S=1$ , todos os valores angulares assumidos por  $H$  corresponderão às cores puras (por exemplo, se temos  $V=1$ ,  $S=1$  e  $H=0$ , temos o vermelho puro). Esse fato é análogo ao pigmento puro das tintas, utilizado pelos artistas como ponto inicial para a mistura das cores [Foley 1997]. Assim, um pintor que quisesse decrementar a saturação ( $S$ ), adicionaria um pigmento branco à cor pura (sem alterar o valor ( $V$ )). Sombras são criadas mantendo-se  $S=1$  e decrementando  $V$ . Finalmente, diversos tons podem ser criados pela diminuição dos valores de  $S$  e de  $V$ . Tudo isso levando em consideração que uma mudança em  $H$  equivale à uma mudança no pigmento de cor puro inicial.

Por utilizar uma maneira para representar as cores baseada na percepção humana, e também por fornecer um componente para a cor ( $H$ ) que independe das magnitudes da saturação ( $S$ ) e de valor ( $V$ ) (mesmo quando o valor de  $H$  torna-se "indefinido", ainda assim ele permanece independente), o modelo de cores HSV é um modelo mais adequado do que o RGB quando o processamento de imagens envolve algum passo fundamental (Figura 3.2) que necessite que objetos sejam encontrados na imagem através de suas cores.

Isso significa dizer que a cor ( $H$ ) dos objetos de uma imagem é invariante às mudanças de luz presentes nessa imagem (que, por sua vez, são definidas por  $S$  e por  $V$ ) - o que facilita que uma área de determinada cor seja identificada na imagem, independentemente das condições de iluminação na superfície dessa área [Fang 2003].

### **3.3 A Biblioteca de Processamento de Imagens "Sistema-IMG"**

A Biblioteca de Processamento de Imagens "Sistema-IMG" foi desenvolvida por Hae Yong Kim [Kim 2004], do Departamento de Engenharia de Sistemas Eletrônicos da Escola Politécnica da Universidade de São Paulo (USP), Brasil.

O "Sistema-IMG" consiste em: (i) Biblioteca-IMG - um conjunto de classes, rotinas e "operadores sobrecarregados" para processamento de imagens escrito na linguagem de programação C++ [Stroustrup 2000]; (ii) Programas-IMG - um conjunto de programas para

processamento de imagens escritos com a utilização das rotinas da Biblioteca-IMG [Kim 2004].

Segundo [Kim 2004], a finalidade do "Sistema-IMG" é a de fornecer meios para testar com rapidez novos algoritmos de processamento de imagens. A interface com o usuário se dá através de comandos no *prompt* do MS-DOS [Microsoft 2004].

Dentre as funcionalidades oferecidas pelo "Sistema-IMG" podemos citar: programas para conversão de um modelo de cor para outro (p.ex.: para converter uma imagem no modelo RGB para o modelo HSI); e comandos (para, por exemplo, leitura, gravação e manipulação de arquivos gráficos) que podem ser utilizados no código fonte do programa do usuário, desde que a Biblioteca-IMG seja incluída durante a compilação desse código.

### **3.4 Considerações Finais Sobre Este Capítulo**

Neste capítulo, foram apresentados (na seção 3.2) modelos de cores utilizados com frequência no processamento de imagens. Desses modelos de cores, destaca-se o modelo HSV que, por ser invariante à iluminação da cena, permite que nessa cena sejam facilmente detectados objetos de determinada cor - o que encontra aplicação na detecção de sinais de trânsito através da cor, a qual será apresentada na seção 5.3.2, do Capítulo 5.

Outro destaque deste capítulo são os passos fundamentais no processamento de imagens, apresentados na seção 3.1.2, os quais servem de base para a metodologia (descrita na seção 5.3, do Capítulo 5) adotada pelo autor para o processamento das imagens capturadas.

## **4 FUNDAMENTOS DE REDES NEURAIS ARTIFICIAIS (RNAs)**

As Redes Neurais Artificiais (RNAs) utilizam uma técnica de aprendizado de máquina que modela, matematicamente e de maneira simplificada, o comportamento de redes de neurônios naturais. As RNAs possuem a capacidade de aprender a generalizar uma função a partir de uma base de exemplos (que constituem-se em pares de entradas e saídas, no aprendizado supervisionado). Uma vez completado esse aprendizado, as RNAs tornam-se ferramentas robustas, pois respondem, com o valor adequado de saída, desde entradas com ruídos até entradas totalmente desconhecidas durante seu treinamento. Neste capítulo, serão apresentados alguns conceitos fundamentais relacionados às RNAs.

### **4.1 O Aprendizado de Máquinas**

O campo de estudos sobre o aprendizado de máquinas trata da questão de como construir sistemas computacionais que adquiram experiência na execução de suas tarefas e que passem a funcionar melhor com o auxílio dessa experiência adquirida [Mitchell 1997].

Existem várias técnicas computacionais automáticas de aquisição de conhecimento e de realização de tarefas, das quais podemos citar [Osório 1999]: as árvores de decisão; os sistemas baseados em conhecimentos; os sistemas baseados em regras difusas; os sistemas baseados em casos; os algoritmos genéticos; e as redes neurais artificiais.

Devido às limitações existentes em várias dessas técnicas [Osório 2000], muitos estudos têm sido dedicados às redes neurais artificiais, motivados pela capacidade de processamento do cérebro humano.

### **4.2 Fundamentos das Redes Neurais Artificiais**

O cérebro humano processa as informações de maneira completamente diferente da usada pelos computadores digitais convencionais. Apesar do cérebro ser também similar a um computador, pois é um sistema de processamento de informações, seu modo de funcionamento difere daquele dos computadores digitais em três aspectos principais [Haykin 2002]: o cérebro é altamente complexo; é não-linear; e é paralelo.

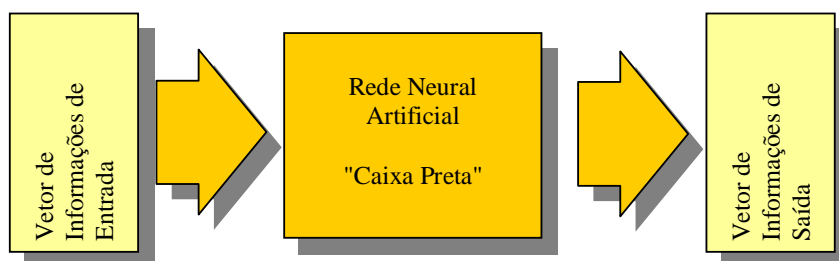
Graças à sua capacidade de organizar os neurônios - seus constituintes estruturais básicos - de forma a processar informações muito mais rapidamente que o computador digital mais rápido existente, o cérebro humano é capaz de realizar com facilidade tarefas complexas (como o reconhecimento de padrões e o controle motor) que demandam recursos computacionais e tecnologia ainda não desenvolvidos. Entre vários exemplos, [Haykin 2002] cita que o cérebro humano é capaz de reconhecer um rosto familiar em meio a uma cena não-familiar em cerca de 200 milissegundos, enquanto que tarefas de menor

complexidade podem levar dias para serem resolvidas por um computador digital convencional.

E é exatamente essa capacidade de processamento do cérebro humano frente aos computadores digitais que motiva os estudos para o desenvolvimento de sistemas computacionais que, embora operem sobre computadores digitais convencionais, sejam capazes de seguir a mesma arquitetura e os mesmos princípios que regem o processamento de informações pelo cérebro humano.

Para que um sistema computacional baseado no funcionamento do cérebro humano - mais especificamente, baseado no funcionamento das redes neurais naturais - possa operar, deve-se primeiro compreender que uma rede neural artificial é uma máquina que é projetada para modelar a maneira como o cérebro realiza uma tarefa particular ou uma função de interesse [Haykin 2002]. A rede neural artificial pode ser implementada fisicamente (através de componentes eletrônicos), ou como uma simulação que será executada por um computador digital.

Segundo [Eberhart 1996], as redes neurais são sistemas de processamento de informações que, em geral, podem ser compreendidas como sendo "caixas pretas", para as quais é fornecida uma entrada e das quais se recebe uma saída. A rede neural artificial, assim, efetua um mapeamento de um vetor de informações de entrada para um vetor de informações de saída (Figura 4.1).



**Figura 4.1 A Rede Neural Artificial como uma "Caixa Preta"**

Cabe salientar duas importantes características das redes neurais artificiais: (i) o seu conhecimento é adquirido através de um processo de aprendizado sobre dados provenientes do ambiente; e (ii) o conhecimento adquirido é armazenado sob a forma de forças de conexão (ou pesos sinápticos) entre os neurônios (as unidades de processamento básicas da rede neural) da rede.

#### 4.2.1 Vantagens das RNAs

Dentre as várias vantagens da aplicação das RNAs à solução de problemas, [Haykin 2002] cita as seguintes:

1. Uma RNA Permite o Mapeamento Entrada-Saída, isto é, em um aprendizado supervisionado, exemplos constituídos por pares de entrada com suas respectivas saídas desejadas são apresentados à rede neural artificial. A cada resposta (saída) da rede neural, os seus pesos são modificados a fim de que seja minimizada a diferença entre a resposta fornecida pela rede e a resposta desejada do exemplo apresentado. O procedimento repete-se para um conjunto de vários exemplos até que não ocorram mais mudanças significativas nos pesos sinápticos da rede neural.
2. Uma RNA tem a capacidade de se adaptar a uma nova aplicação, após um novo treinamento. Adicionalmente, a rede neural artificial pode adaptar-se a um ambiente dinâmico, se for implementada com a capacidade de modificar os próprios pesos sinápticos em tempo real.
3. Uma RNA (especialmente se for implementada fisicamente, através de componentes eletrônicos), devido à sua arquitetura distribuída, é tolerante à falhas (p.ex.: danos em conexões ou em neurônios). Seu desempenho se degrada suavemente em caso de falhas, isto é, a recuperação de um padrão armazenado se dará com uma menor qualidade - mas ainda assim será possível. Seria necessária uma grande quantidade de danos à rede neural para que a sua resposta global fosse seriamente comprometida.

#### 4.2.2 Aplicações das RNAs

[Eberhart 1996] cita as seguintes aplicações genéricas para as redes neurais artificiais:

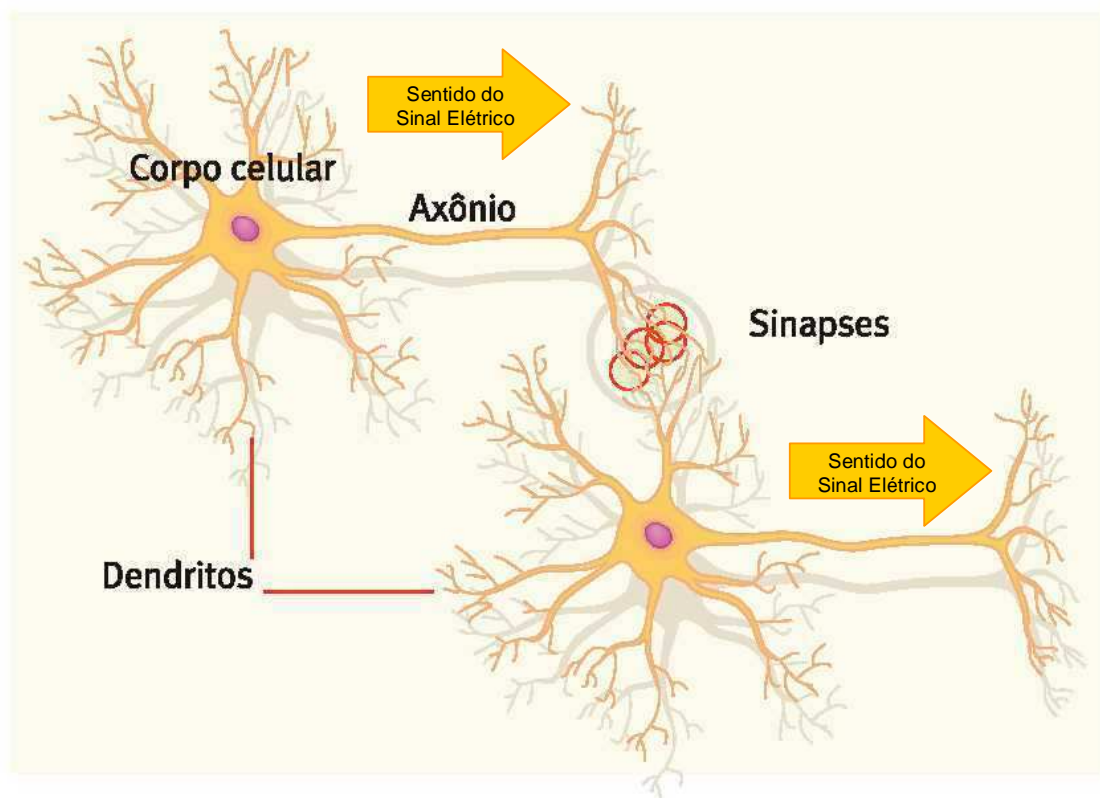
1. Classificação de Padrões: fornecido um padrão de entrada para a rede neural, a mesma responde com uma determinada classe para esse padrão.
2. Associação de Padrões: fornecido um padrão de entrada, a rede neural artificial fornece como saída um padrão correspondente.
3. Recuperação de Padrões: fornecido um padrão de entrada incompleto a rede neural, a mesma responde com o padrão correspondente que apresenta as regiões ausentes no padrão de entrada.
4. Remoção de Ruídos: a partir de um padrão de entrada corrompido por ruídos, fornecido para a rede neural, obtém-se da mesma o padrão correspondente de saída, com o ruído removido ou atenuado.
5. Otimização: dado um padrão de entrada representando um conjunto de valores de um problema a ser otimizado, a rede neural responde com um padrão correspondente a um conjunto de variáveis que representam uma solução otimizada aceitável para o problema.
6. Controle: um padrão de entrada correspondendo à situação atual do sistema e à resposta desejada para o sistema controlado é fornecido para a rede neural, a qual responde com um

padrão representando a sequência de comandos necessária para que a resposta desejada para o sistema controlado seja alcançada.

7. Simulação: dado um padrão representativo de uma situação atual de um sistema simulado, ou de uma série temporal, a rede neural responde com uma sequência estruturada de padrões que simularão o comportamento do sistema simulado ao longo do tempo.

#### 4.2.3 Componentes de uma RNA

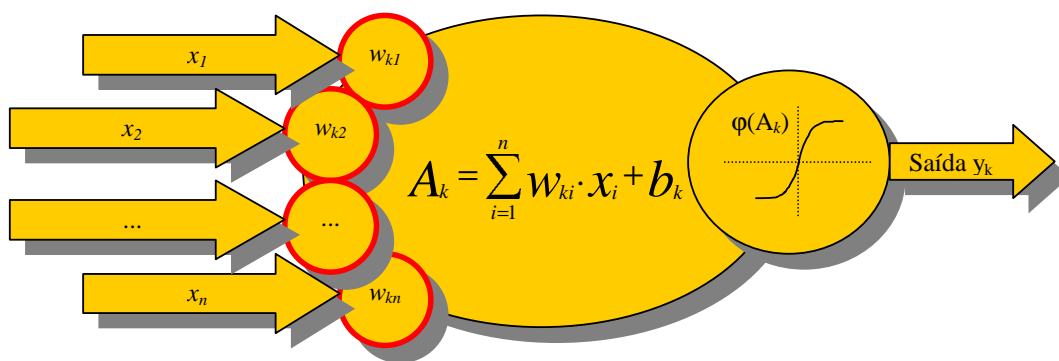
A rede neural artificial modela sua contrapartida biológica através de uma formalização matemática que se inicia em seus elementos de processamento: os neurônios. Um neurônio biológico, conforme pode ser observado na Figura 4.2, é constituído por três partes principais: o corpo celular (o impulso elétrico a ser transmitido é gerado no corpo celular, ou próximo a ele [Haykin 2002]); a árvore dendrítica (onde os sinais elétricos provenientes de outros neurônios são recebidos, através dos dendritos); e o axônio (prolongamento responsável por transmitir o sinal elétrico de saída para os outros neurônios da rede). Ainda na Figura 4.2 podem ser observadas as sinapses, que são as ligações químicas existentes entre o axônio e os dendritos.



**Figura 4.2 Anatomia Simplificada dos Neurônios Biológicos [Costa et al. 2000]**

As sinapses operam através de substâncias transmissoras (liberadas na junção sináptica por um processo pré-sináptico) que agem sobre um processo pós-sináptico (dessa forma, um sinal elétrico pré-sináptico é transformado num sinal químico, que por sua vez volta a tornar-se um sinal elétrico pós-sináptico - o que demonstra que não há uma ligação direta entre o axônio e os dendritos). Em termos elétricos, o elemento no qual se compõe uma sinapse é chamado de dispositivo de dois terminais não-recíproco. Normalmente, assume-se que uma sinapse é uma conexão que pode impor ao neurônio receptor do sinal um estado de excitação ou um estado de inibição, mutuamente exclusivos [Haykin 2002].

Um neurônio artificial é uma unidade de processamento de informações modelada com base no neurônio biológico, e é formada por três elementos básicos [Haykin 2002]: um conjunto de sinapses; um somador; e uma função de ativação. Na Figura 4.3 pode ser observado um modelo de um neurônio artificial.



**Figura 4.3 Modelo de um Neurônio Artificial  $k$**

Cada uma das sinapses do neurônio é caracterizada por um peso  $w_{kn}$ , que multiplicará o valor do respectivo sinal de entrada  $x_n$ . O número de sinapses é dado por  $n$ , enquanto  $k$  representa o número do neurônio em questão, na rede neural.  $A_k$  é o potencial de ativação (também chamado de *campo local induzido*) do neurônio  $k$ , composto pelo somatório dos sinais de entrada ponderados por suas respectivas sinapses e por um *bias*  $b_k$ . O potencial de ativação  $A_k$  é definido por

$$A_k = \sum_{i=1}^n w_{ki} \cdot x_i + b_k \quad (4.1)$$

Se o *bias*  $b_k$  for positivo, ele aumenta a entrada líquida para a função de ativação  $\phi(A_k)$  e, se for negativo, diminui essa entrada líquida. O papel da função de ativação é o de regular o

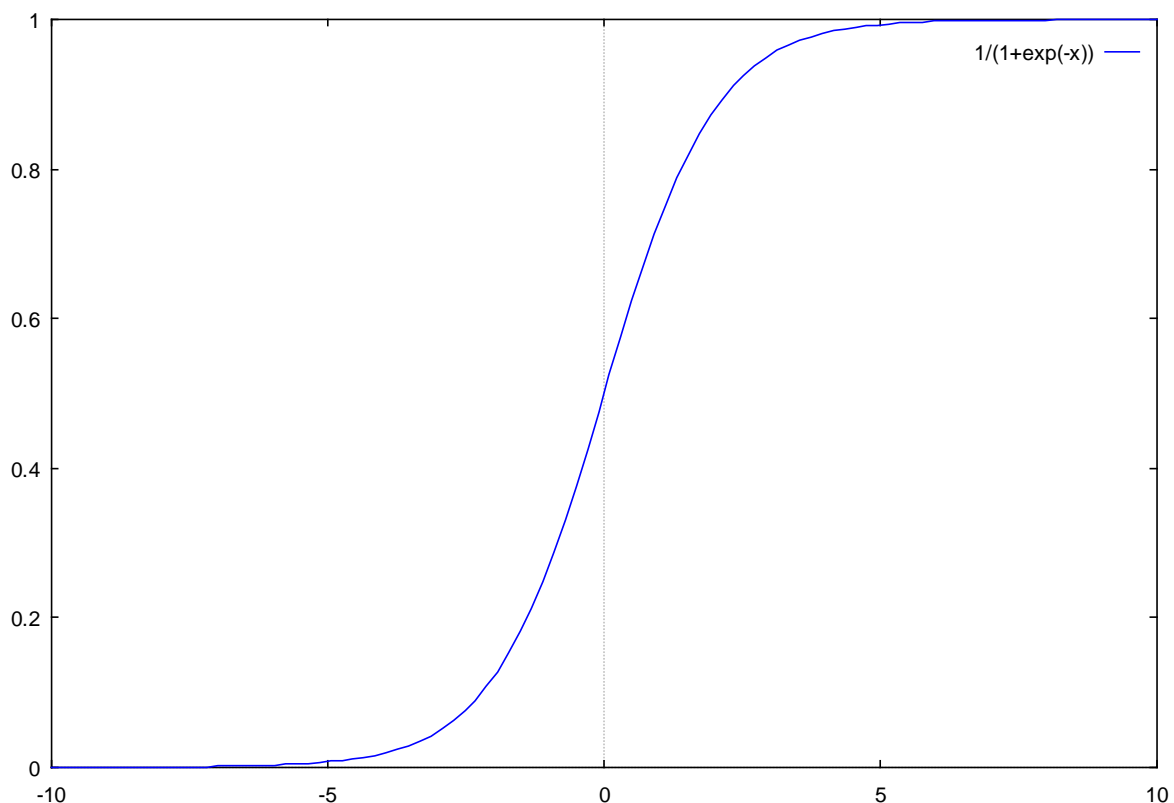


valor final a ser obtido na saída do neurônio. Temos, então, modelados em um neurônio artificial os dendritos (que correspondem às entradas  $x_1$  a  $x_n$ ), as sinapses (que correspondem aos pesos  $w_{k1}$  a  $w_{kn}$ ), o corpo celular (que corresponde ao potencial de ativação  $A_k$  e à função de ativação  $\phi(A_k)$ ) e, finalmente, o axônio (que corresponde à saída  $y_k$ ) do neurônio biológico. A saída  $y_k$  do neurônio é definida por

$$y_k = \phi(A_k) \quad (4.2)$$

Cabe lembrar que o *bias*  $b_k$ , como é um parâmetro externo que afeta o potencial de ativação  $A_k$  do neurônio (Equação 4.1), normalmente é representado no modelo do neurônio artificial como uma sinapse  $w_{k0}$ , de peso  $b_k$ , cujo valor de entrada  $x_0$  é fixado em +1 [Haykin 2002].

A função Sigmóide é a forma mais comum de função de ativação, e é a escolhida com maior frequência para os neurônios de uma rede neural artificial [Haykin 2002]. O gráfico da função logística - um exemplo de função Sigmóide - pode ser observado na Figura 4.4, a seguir.



**Figura 4.4 A Função Logística**

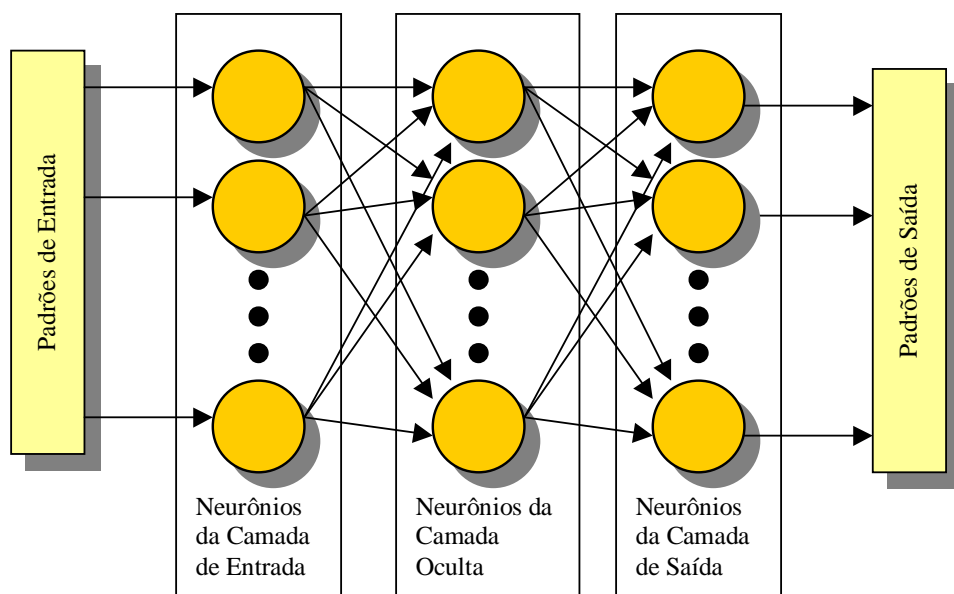
A função logística apresentada na Figura 4.4 é definida por

$$\varphi(A) = \frac{1}{1 + \exp(-a \cdot A)} \quad (4.3)$$

onde  $a$  corresponde ao parâmetro de inclinação da função sigmóide, e  $A$  corresponde ao potencial de ativação do neurônio artificial. Quanto maior o valor de  $a$ , mais a função logística se aproxima da função de limiar, que por sua vez também pode ser usada como função de ativação e é definida por

$$\varphi(A) = \begin{cases} 1 & \text{se } A \geq 0 \\ 0 & \text{se } A < 0 \end{cases} \quad (4.4)$$

A rede neural artificial (RNA), ou rede conexionista, é composta por unidades de processamento (neurônios artificiais) interconectados. Uma RNA é constituída por um grafo orientado e ponderado, onde os nós desse grafo são autômatos simples - os neurônios artificiais. Assim, a rede neural - através das conexões entre os autômatos simples - constitui-se num autômato mais complexo. A Figura 4.5 mostra um exemplo de rede neural artificial de múltiplas camadas.



**Figura 4.5 Rede Neural Artificial de Múltiplas Camadas**

Conforme já foi mencionado, o conhecimento adquirido pelas redes neurais artificiais é armazenado sob a forma de forças de conexão (ou pesos sinápticos) entre os neurônios (as unidades de processamento básicas da rede neural) da rede. A seguir, será apresentada uma visão geral do aprendizado das RNAs.

#### 4.2.4 O Aprendizado das RNAs

A propriedade mais importante das redes neurais artificiais é a sua capacidade de aprender a partir de seu ambiente e de melhorar o seu desempenho através desse aprendizado. Basicamente, o processo de aprendizado das redes neurais artificiais é um processo interativo (informações do ambiente - ou problema - para o qual se quer treinar a RNA devem ser submetidos à mesma, que por sua vez passará por um ajuste de seus pesos sinápticos para a retenção do conhecimento) e iterativo (pois o processo de aprendizado é executado ao longo de vários ciclos) [Haykin 2002].

Existem três estratégias para o treinamento de uma rede neural artificial: o treinamento supervisionado; o treinamento semi-supervisionado; e o treinamento não-supervisionado [Haykin 2002] [Eberhart 1996] [Osório 1999].

No treinamento supervisionado, um professor externo é responsável por fornecer os exemplos - pares de vetores de entrada com seus respectivos vetores de saída para a RNA. O vetor de saída fornecido pelo professor é utilizado para a comparação com o vetor de saída fornecido pela RNA para o vetor de entrada associado, como será explicado com maiores detalhes na seção 4.2.5.

No treinamento semi-supervisionado, também chamado de aprendizado por reforço, a RNA deve aprender através de uma interação contínua com o ambiente. A RNA é projetada para, a partir de uma sequência de estímulos ao longo do tempo, minimizar *uma função de custo para avançar* [Haykin 2002], que é definida como sendo uma expectativa do custo cumulativo de ações tomadas como resposta aos estímulos recebidos.

No treinamento não-supervisionado, também chamado de auto-organização, não há um professor externo para orientar o processo de aprendizado. Dessa forma, o aprendizado não se dá através de exemplos, mas através de uma organização dos dados (clusterização) fornecidos para a RNA

Existem, também, duas maneiras através das quais as RNAs podem ser treinadas: a maneira on-line, através da qual novos padrões podem ser incorporados ao conhecimento armazenado na rede neural; e a maneira off-line, através da qual toda uma base de padrões é apresentada para a RNA - de acordo com o algoritmo de aprendizado escolhido - e, após o treinamento, os pesos finais calculados para a RNA são fixados e a rede é colocada em uso, sem que ocorram mais alterações nesses pesos.

#### 4.2.5 O Perceptron de Múltiplas Camadas com Aprendizado por Retropropagação de Erro

A RNA apresentada na Figura 4.5 - chamada de perceptron de múltiplas camadas - pode ser considerada um exemplo de uma RNA totalmente conectada - o que significa que um neurônio artificial de qualquer camada está conectado a todos os neurônios da camada anterior. Nessa rede neural o sinal progride através da rede para a frente (da esquerda para a direita, na Figura 4.5) de camada em camada.

Um perceptron de múltiplas camadas possui quatro características particulares: (i) o modelo de cada um de seus neurônios possui uma função de ativação não-linear - uma forma de não-linearidade que satisfaz essa característica é a não-linearidade sigmóide definida pela função logística da Equação 4.3 - para que o perceptron de múltiplas camadas apresente uma relação de entrada-saída aprimorada com relação ao perceptron de camada única; (ii) a rede deve possuir uma ou mais camadas ocultas de neurônios - que permitirão que o perceptron de múltiplas camadas resolva problemas complexos através da extração progressiva das características mais significativas dos vetores de entrada; (iii) a rede deve possuir uma alta conectividade entre seus neurônios; (iv) os neurônios da camada de entrada - também chamados de nós de fonte - não realizam quaisquer computações [Haykin 2002].

##### 4.2.5.1 O Algoritmo de Retropropagação de Erro (*Error Back-Propagation*)

Os perceptrons de múltiplas camadas têm sido freqüentemente utilizados para a resolução de diversos problemas, em conjunto com um treinamento supervisionado off-line conhecido por algoritmo de retropropagação de erro (error back-propagation) [Rumelhart 1986].

De maneira geral, o algoritmo de retropropagação do erro consiste em dois passos através das diferentes camadas da RNA.

No primeiro passo - de propagação - um vetor de entrada é fornecido aos neurônios da camada de entrada da RNA, e seu sinal se propaga no sentido das conexões sinápticas (para a frente) através de cada uma das camadas da rede. Após esse sinal atingir a camada de saída da RNA, um vetor de saída é produzido pela rede. No passo de propagação, não ocorrem alterações nos pesos sinápticos da rede.

No segundo passo - de retropropagação - ocorre o ajuste dos pesos sinápticos da RNA, de acordo com uma regra de correção de erro, que calcula a diferença entre o vetor de saída

(resposta real) fornecido pela rede neural e o vetor de saída (resposta alvo) associado para o vetor de entrada fornecido. Essa regra, assim, obtém um sinal de erro que é retropropagado pela rede, camada a camada, no sentido oposto ao das conexões sinápticas (para trás), com a finalidade de ajustar os pesos sinápticos da RNA e fazer com que a resposta real da mesma se mova mais para perto da resposta desejada para o respectivo vetor de entrada.

O algoritmo de retropropagação de erro utiliza a técnica de aprendizagem por correção de erro, também referida por regra delta. A regra delta, citada por [Rumelhart 1986], é definida por

$$\Delta_p w_{ji} = \eta \cdot (t_{pj} - o_{pj}) \cdot i_{pi} = \eta \cdot \delta_{pj} \cdot i_{pi} \quad (4.5)$$

onde:

$p$  é o padrão de entrada;

$t_{pj}$  é o valor desejado de resposta (com relação ao padrão  $p$ ) - fornecido pelo professor, pois o aprendizado é supervisionado - para o  $j$ -ésimo neurônio da camada de saída;

$o_{pj}$  é o valor real de resposta (com relação ao padrão  $p$ ) produzido pelo  $j$ -ésimo neurônio da camada de saída;

$i_{pi}$  é o valor do  $i$ -ésimo elemento do padrão de entrada  $p$ ;

$\eta$  é uma constante positiva que determina a taxa de aprendizado quando avançamos em um passo no processo de aprendizagem;

$\delta = t_{pj} - o_{pj}$ ;

$\Delta_p w_{ji}$  é a alteração a ser efetuada no peso da sinapse entre o neurônio  $i$  e o neurônio  $j$ , devido à apresentação do padrão de entrada  $p$ .

Dessa forma, o novo valor do peso para a sinapse  $ji$  será dado por [Haykin 2002]

$$w_{ji}(n+1) = w_{ji}(n) + \Delta_p w_{ji}(n) \quad (4.6)$$

onde  $n$  representa um instante de tempo discreto, que é o passo de tempo de um processo iterativo para o ajuste dos pesos sinápticos do neurônio  $j$ .

A taxa de aprendizado  $\eta$  tem um importante efeito no tempo necessário para que ocorra a convergência para o menor erro [Riedmiller 1993]. Quanto maior o valor dessa constante, maiores as alterações nos pesos [Rumelhart 1986]. Assim, deve ser escolhida uma taxa de aprendizado suficientemente grande que não incorra em oscilações (que impossibilitariam que o erro caísse abaixo de determinado valor). Uma maneira de evitar o surgimento dessas oscilações é a de modificar a regra delta, com a adição de uma constante de *momentum*  $\alpha$  (que pode assumir valores entre zero e um), que é a inércia na descida da superfície de erro

[Osório 1999]. Esse *momentum*  $\alpha$  tem o papel de levar em conta o valor do passo dado anteriormente, isto é, o valor da magnitude anterior de alteração dos pesos. Isso permite uma maior probabilidade de que a busca pelo menor erro não pare em mínimos locais na superfície de busca. O valor do novo peso  $w_{ij}(n+1)$ , com a inclusão do *momentum*  $\alpha$ , é dado por [Eberhart 1996]

$$\Delta w_{ji}(n) = \Delta_p w_{ji}(n) + \alpha \cdot \Delta w_{ji}(n-1) \quad (4.7)$$

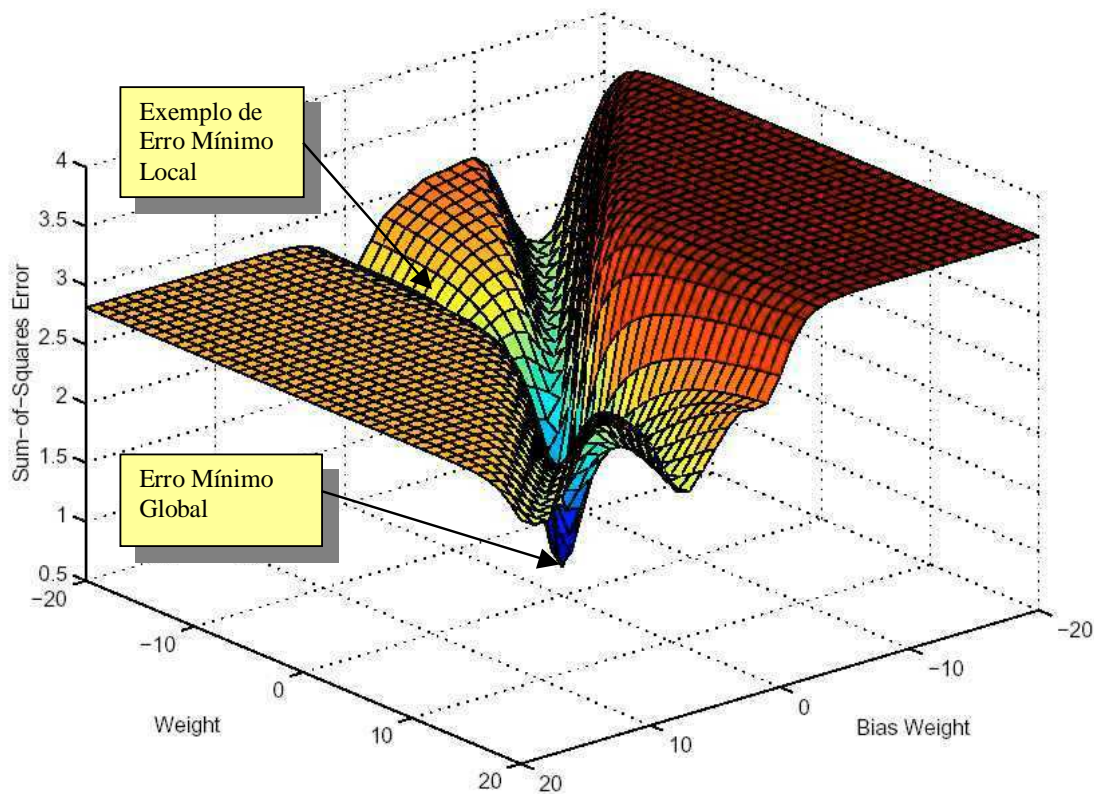
$$w_{ji}(n+1) = w_{ji}(n) + \Delta w_{ji}(n) \quad (4.8)$$

onde  $\Delta w_{ji}(n)$  é a magnitude da alteração anterior nos pesos sinápticos.

O treinamento de uma RNA pode ser visto, então, como uma busca pela minimização do erro (entre a saída desejada e a saída esperada para um determinado padrão de entrada) através do ajuste dos pesos das sinapses.

Cada problema a ser resolvido por uma RNA apresenta uma superfície característica, que representa a relação entre os pesos e o erro na saída da rede. A busca pela minimização do erro da RNA constitui-se, dessa forma, na busca pelo mínimo global de uma superfície (que é o ponto de menor erro possível para a rede - idealmente igual a zero),

Cada valor de erro da RNA corresponde a um conjunto de valores para os pesos de todas as sinapses da RNA. É nessa superfície multidimensional - onde o erro da rede, e cada um dos pesos sinápticos, correspondem às dimensões do espaço em que a superfície se encontra - que a busca pelo erro mínimo é efetuada. Um exemplo de superfície de erro para dois pesos pode ser observado na Figura 4.6.



**Figura 4.6 Gráfico de uma Superfície de Erro de uma RNA com Dois Pesos (*Weights*)**  
[Gallagher 2000]

Uma das formas de efetuar essa busca pelo mínimo global na superfície de erros da rede neural artificial é através da utilização do método da descida do gradiente.

Segundo [Rumelhart 1986], a descida do gradiente minimiza os quadrados das diferenças entre os valores de saída desejado e real da RNA. E uma das maneiras de demonstrar isso é mostrando que a derivada da medida do erro com relação a cada um dos pesos sinápticos é proporcional à alteração nos pesos - ditada pela regra delta - com uma constante de proporcionalidade negativa. Isso corresponde exatamente à busca pelo menor erro (o erro mínimo global), ou minimização da função de erro, através da superfície multidimensional definida pelos pesos sinápticos e pelas magnitudes dos erros para cada conjunto de pesos assumido pela RNA (Figura 4.6).

A descida do gradiente, utilizando o *momentum*, é definida por

$$\Delta w_{ji}(n+1) = -\eta \cdot \frac{\partial E}{\partial w_{ji}}(n+1) + \alpha \cdot \Delta w_{ji}(n) \quad (4.9)$$

Onde  $E$  é a função de erro.

A derivação da regra delta utilizando a descida do gradiente pode ser vista em [Rumelhart 1986].

O algoritmo de retropropagação de erro (*error back-propagation*) apresentado nesta seção utiliza a técnica de aprendizagem por correção de erro (a regra delta) que, por sua vez, utiliza a técnica da descida do gradiente para buscar pelo erro mínimo global (minimização da função de erro).

Assim, como foi colocado no início desta seção, o algoritmo de retropropagação de erro propaga um sinal de erro, camada a camada da rede neural no sentido oposto ao das conexões sinápticas (para trás), com a finalidade de ajustar os pesos sinápticos da RNA para que a resposta real da mesma mova-se mais para perto da resposta desejada para o vetor de entrada. Essa retropropagação do erro garante que a "culpa" pelo erro não seja assumida apenas pela camada de neurônios de saída da RNA, sendo assumida também pelos neurônios das camadas ocultas da rede.

#### **4.2.5.2 O Algoritmo de Propagação Resiliente (*Resilient Propagation - RPROP*)**

Uma melhoria sobre a técnica de aprendizagem por retropropagação de erro é a do algoritmo de propagação resiliente (ou elástica) [Riedmiller 1993]. Neste algoritmo é apresentado um método que efetua uma adaptação direta na alteração dos pesos, baseada em informações de gradiente locais. Para isso, é introduzido, para cada peso, o seu valor individual de atualização  $\Delta_{ji}$ , que determina apenas o tamanho da alteração no respectivo peso. Esse valor individual de atualização  $\Delta_{ji}$  evolui durante o processo de aprendizado, baseado apenas no sinal local da função de erro  $E$ , baseado nas seguintes regras de aprendizado:



$$\Delta_{ji}(n+1) = \begin{cases} \eta^+ \cdot \Delta_{ji}(n), & \text{se } \frac{\partial E}{\partial w_{ji}}(n) \cdot \frac{\partial E}{\partial w_{ji}}(n+1) > 0 \\ \eta^- \cdot \Delta_{ji}(n), & \text{se } \frac{\partial E}{\partial w_{ji}}(n) \cdot \frac{\partial E}{\partial w_{ji}}(n+1) < 0 \\ \Delta_{ji}(n), & \text{caso contrário.} \end{cases} \quad (4.10)$$

Onde  $0 < \eta^- < 1 < \eta^+$ .

A regra de aprendizado faz com que cada vez que a derivada parcial do peso sináptico correspondente troque de sinal - o que indica que a última atualização no peso foi muito grande e o algoritmo "pulou" sobre um erro mínimo local - o valor individual de atualização  $\Delta_{ji}$  seja decrementado por um fator  $\eta^-$ . Se a derivada parcial não trocou de sinal, o  $\Delta_{ji}$  é incrementado por um fator  $\eta^+$ , para que a convergência no sentido de regiões de menor erro seja acelerada.

Uma vez que os valores individuais de atualização  $\Delta_{ji}$  para cada peso sináptico foram determinados, a atualização dos pesos segue a seguinte regra:

$$\Delta w_{ji}(n+1) = \begin{cases} -\Delta_{ji}(n+1), & \text{se } \frac{\partial E}{\partial w_{ji}}(n+1) > 0 \\ +\Delta_{ji}(n+1), & \text{se } \frac{\partial E}{\partial w_{ji}}(n+1) < 0 \\ 0, & \text{caso contrário.} \end{cases} \quad (4.11)$$

$$w_{ji}(n+1) = w_{ji}(n) + \Delta w_{ji}(n+1) \quad (4.12)$$

Porém, existe uma exceção à essa regra, que ocorre quando a derivada parcial troca de sinal - isto é, quando um passo é muito largo e o algoritmo "pula" sobre um mínimo local. A exceção faz com que a última atualização dos pesos seja revertida:

$$\Delta w_{ji}(n+1) = -\Delta w_{ji}(n) \quad , \quad \text{se } \frac{\partial E}{\partial w_{ji}}(n) \cdot \frac{\partial E}{\partial w_{ji}}(n+1) < 0 \quad (4.13)$$

Devido à essa reversão, a derivada parcial deverá trocar novamente de sinal durante o próximo passo do algoritmo. Para evitar outra reversão, não deve ocorrer a adaptação no valor individual de atualização  $\Delta_{ji}$  no próximo passo, o que pode ser obtido atribuindo o valor zero para a derivada parcial do passo  $n$ , na regra da Equação 4.9.

Os valores individuais de atualização  $\Delta_{ji}$  e os pesos sinápticos são atualizados cada vez que a totalidade do conjunto de padrões é apresentado para a rede neural - conforme o aprendizado por épocas.

Um dos aspectos da descida do gradiente é o de que o valor da derivada decresce exponencialmente com a distância entre o peso e a camada de saída da RNA. Assim, pesos muito distantes da camada de saída são menos modificados e, conseqüentemente, aprendem mais lentamente.

Com a utilização do RPROP, o tamanho da atualização dos pesos depende apenas da sequência de troca de sinais das derivadas parciais, não dependendo da magnitude das mesmas. Dessa forma, com os pesos distantes da camada de saída tendo a mesma oportunidade de aprendizado daqueles próximos à camada de saída, o aprendizado é distribuído de maneira homogênea pela RNA - o que garante uma maior eficiência para o processo de aprendizagem da RNA.

#### 4.2.5.3 O Grau de Generalização de uma RNA

Para garantirmos um determinado grau de generalização para o aprendizado de uma RNA, podemos utilizar a técnica chamada de validação cruzada. Essa técnica permite que uma RNA desenvolva uma boa generalização a partir do conjunto de dados de treinamento [Haykin 2002].

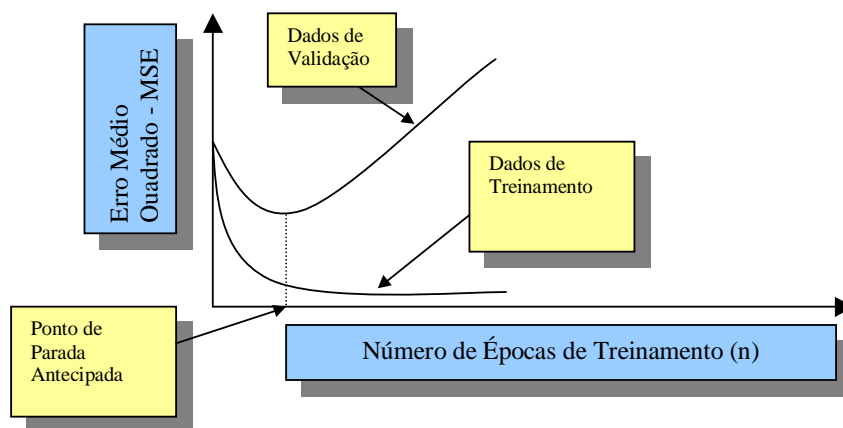
Na validação cruzada, o conjunto de dados para treinamento (pares de vetores de entrada com seus respectivos vetores de saída - no aprendizado supervisionado) é dividido em dois subconjuntos: o conjunto de treinamento; e o conjunto de validação.

Durante o treinamento de uma RNA do tipo perceptron de múltiplas camadas - efetuado por retropropagação de erro - o erro médio quadrado (*mean square error* - MSE) inicia com um valor grande, decresce rapidamente e então continua decrescendo de modo mais lento conforme a RNA segue seu caminho em busca de um mínimo local na superfície de erro. O MSE para um perceptron de múltiplas camadas é definido por

$$MSE = \frac{1}{p} \sum_{j=1}^p \sum_{i=1}^s (t_{ij} - x_{ij})^2 \quad (4.14)$$

onde  $p$  é o número de padrões fornecidos como entrada para a RNA,  $s$  é o número de neurônios na camada de saída da rede,  $t_{ij}$  é o valor objetivo (fornecido pelo professor, no treinamento supervisionado) para o neurônio  $i$  da camada de saída dado um padrão de entrada  $j$ , e  $x_{ij}$  é o valor estimado (fornecido pela RNA) para o neurônio  $i$  da camada de saída dado um padrão de entrada  $j$ .

Como buscamos uma boa generalização para a RNA, torna-se difícil determinar em qual época  $n$  devemos suspender o treinamento, para evitar que a rede neural fique excessivamente ajustada aos dados de treinamento. Assim, efetuamos um processo periódico de treinamento e validação, chamado de treinamento com parada antecipada, que possibilita verificarmos a época mais adequada para suspender o treinamento da RNA.



**Figura 4.7 Parada Antecipada, Baseada na Validação Cruzada [Haykin 2002]**

Conforme a Figura 4.7 acima, observando a curva de aprendizado no treinamento, estaríamos inclinados a permitir que a RNA melhore seu desempenho deixando-a ser treinada para além do ponto mínimo da curva de aprendizado na validação. Entretanto, após esse ponto a RNA sofre um "sobreaprendizado" (*overtraining*), se especializando demasiadamente nos dados de treinamento (a rede decora ao invés de generalizar), o que sugere que o uso do ponto mínimo da curva de aprendizado na validação seja utilizado como uma heurística adequada para a escolha da época de suspensão do treinamento da RNA [Haykin 2002].

### 4.3 A Ferramenta Stuttgart Neural Network Simulator (SNNS)

O SNNS (Stuttgart Neural Network Simulator) é um simulador de redes neurais artificiais criado em 1989, no Instituto de Sistemas Paralelos e Distribuídos de Alto Desempenho (Institute for Parallel and Distributed High Performance Systems / Institut für Parallele und Verteilte Höchstleistungsrechner - IPVR) na Universidade de Stuttgart, na Alemanha

[SNNS 1998]. A partir de 1996 até a versão atual do SNNS (4.2, de 1998), o SNNS tem sido mantido pela Universidade de Tübingen (Wilhelm-Schickard-Institute), da Alemanha.

O SNNS é formado por quatro componentes principais: um *kernel*; uma interface gráfica para o usuário; uma interface para execução em lote - o Batchman; e o compilador de redes neurais artificiais snns2c.

O *kernel* realiza todas as operações sobre as estruturas de dados que representam a rede neural artificial. A interface gráfica para o usuário, construída sobre o *kernel*, fornece uma representação gráfica da rede neural e controla o *kernel* durante a simulação - adicionalmente, a interface gráfica pode ser utilizada para a criação, manipulação e visualização das redes neurais artificiais. Cabe notar que o SNNS disponibiliza vários algoritmos de aprendizado, e também possibilita a implementação de várias topologias de redes neurais artificiais de diferentes tipos. Exemplos da interface gráfica para o usuário do SNNS podem ser observadas na Figura 4.8, abaixo

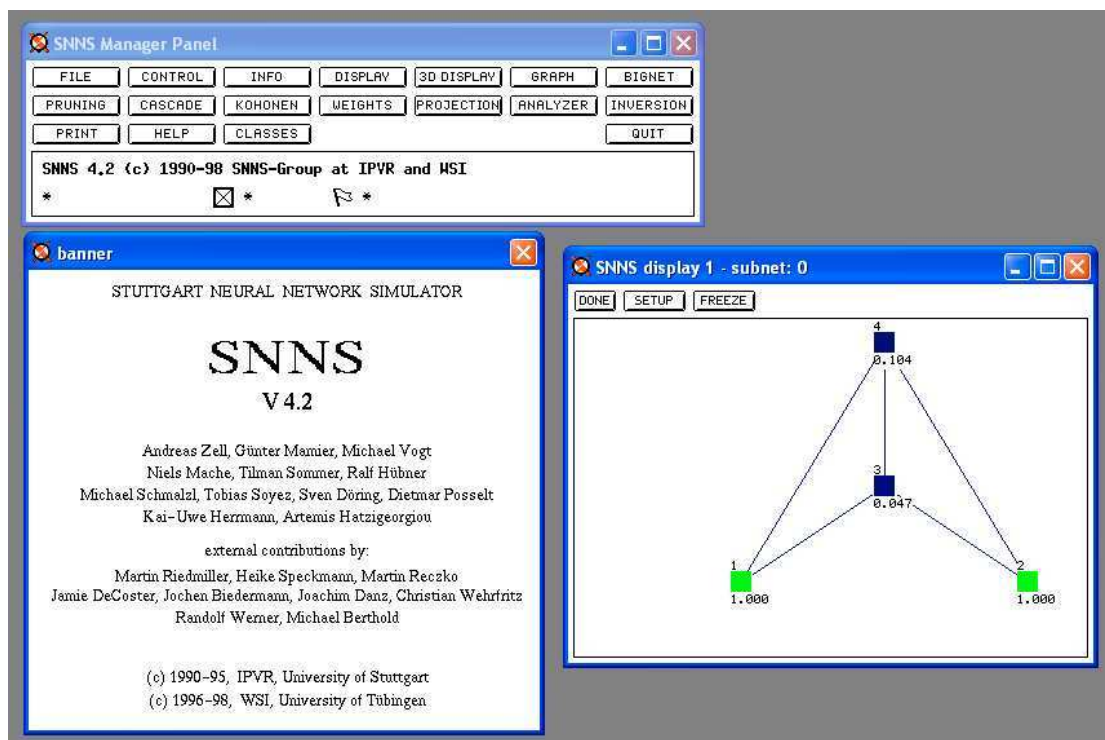


Figura 4.8 Exemplos da Interface Gráfica do SNNS

#### 4.4 Considerações Finais Sobre Este Capítulo

Este capítulo apresentou os fundamentos das redes neurais artificiais. As redes neurais artificiais foram adotadas para este trabalho pois constituem-se numa ferramenta robusta

capaz de reconhecer os padrões para os quais foi treinada - mesmo que estes padrões estejam corrompidos por ruídos, ou incompletos.

Para o projeto e implementação da RNA a ser utilizada no RIST-Br, serão utilizados: o simulador SNNS, apresentado na seção 4.3; o algoritmo de aprendizado RPROP, apresentado na seção 4.2.5.2; e uma base de dados (que será construída a partir de uma base de imagens) dividida em dois conjuntos: o conjunto de treinamento e o conjunto de validação da RNA (conforme a técnica da validação cruzada, apresentada na seção 4.2.5.3).

## **5 METODOLOGIA ADOTADA PARA A IMPLEMENTAÇÃO DO RECONHECEDOR INTELIGENTE DE SINAIS DE TRÂNSITO BRASILEIROS (RIST-Br)**

Neste capítulo será apresentada a metodologia adotada para a execução deste trabalho. Em primeiro lugar, será apresentada uma visão geral do mecanismo de atenção visual para a detecção de objetos de interesse, com destaque para: o mecanismo de atenção visual em seres humanos; e alguns mecanismos de atenção visual desenvolvidos para as máquinas. A atenção visual mostrará sua relevância em uma das fases iniciais do reconhecimento inteligente de sinais de trânsito implementado para esta monografia.

Em seguida, será apresentado o experimento prático através do qual o autor desta monografia desenvolveu parte da metodologia aqui descrita. Nesse experimento, foi definida a metodologia a ser adotada para o reconhecimento inteligente de padrões - metodologia essa que demonstra ter um papel fundamental na implementação desenvolvida para o reconhecimento inteligente de sinais de trânsito.

Finalmente, serão apresentadas as fases do reconhecimento de sinais de trânsito, adotadas para o desenvolvimento do sistema de Reconhecimento Inteligente de Sinais de Trânsito Brasileiros (RIST-Br), sobre o qual trata o Capítulo 6.

### ***5.1 A Atenção Visual para a Detecção de Objetos de Interesse***

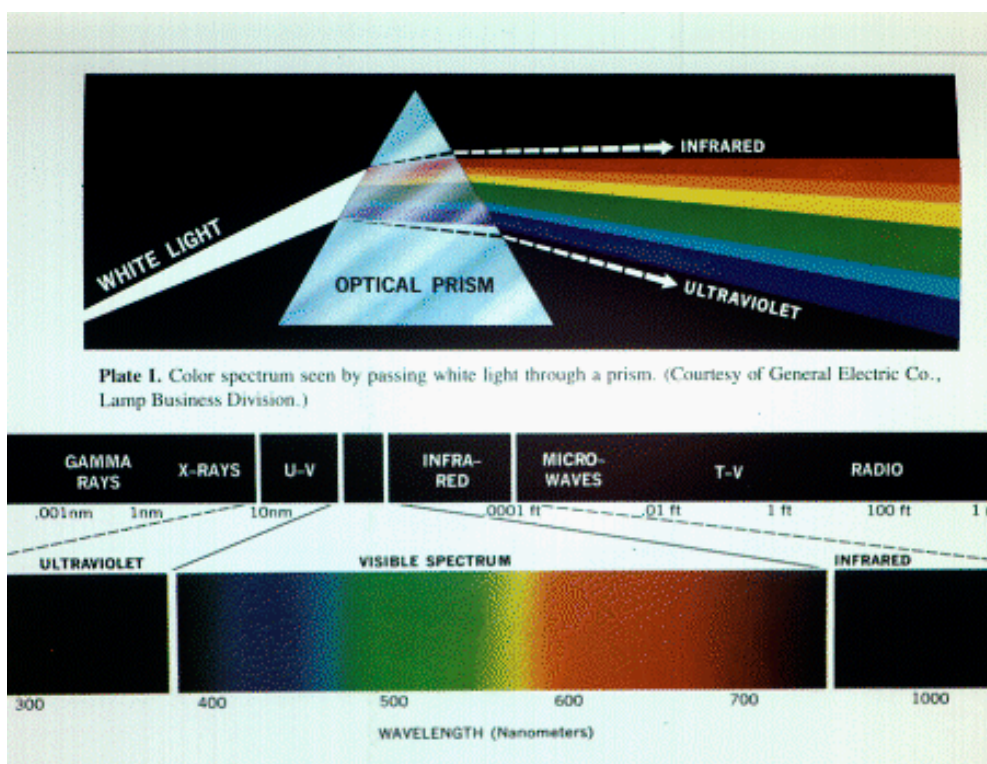
A atenção visual é a capacidade de detecção rápida de áreas importantes da imagem capturada pelo sistema, seja ele biológico ou artificial [Rodrigues 2002]. No caso de sistemas biológicos, como o cérebro humano, a atenção visual pode ser vista como um método que reduz a quantidade de informações de entrada até um nível que possa ser facilmente tratado pelas redes neurais do cérebro [Rodrigues 2002]. No caso de máquinas, a atenção visual pode ser considerada um mecanismo capaz de tornar o problema do processamento e reconhecimento de imagens mais facilmente tratável - reduzindo a complexidade computacional envolvida [Culhane 1992].

#### **5.1.1 Atenção Visual em Seres Humanos**

A seguir, serão apresentadas noções básicas sobre a anatomia (que estuda a localização dos órgãos) e a fisiologia (que estuda o funcionamento dos órgãos) do olho humano. Em seguida, será apresentado o mecanismo dos movimentos oculares (o qual permite que os olhos rastreiem uma cena e fixem sua posição na direção de uma região de interesse), que é, em grande parte, responsável pelo mecanismo de atenção visual em seres humanos.

### 5.1.1.1 O Olho Humano

O olho humano é um receptor sensível a uma faixa muito estreita do espectro de radiações eletromagnéticas. Essa faixa do espectro cuja radiação pode ser detectada pela visão humana é denominada de "luz visível". O espectro eletromagnético, por sua vez, estende-se desde as ondas de maior comprimento e menor energia (frequência), tais como as ondas de rádio, até as ondas de menor comprimento e maior energia (frequência), tais como os raios gama [Silverthorn 2003]. A Figura 5.1, abaixo, ilustra a decomposição da luz branca por um prisma óptico (*optical prism*), e também apresenta o espectro eletromagnético.



**Figura 5.1** Decomposição da Luz Branca, e Espectro Eletromagnético [Foley 1997]

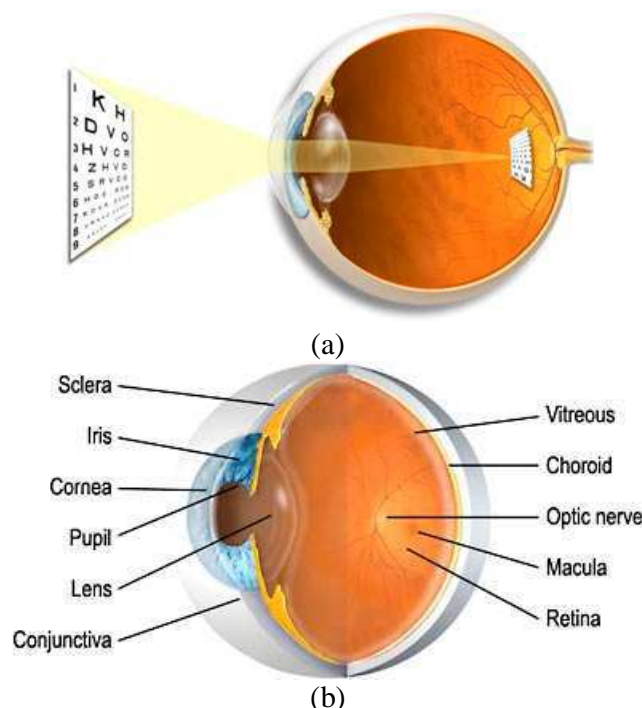
A "luz visível" (*visible spectrum*, na Figura 5.1) se estende da cor vermelha (comprimento de onda de 750 nanômetros (nm), de energia - ou frequência - mais baixa) até a cor violeta (comprimento de onda de 400 nm, de energia - ou frequência - mais alta), e é composta por ondas com uma frequência de  $4,0 \times 10^{14}$  a  $7,5 \times 10^{14}$  ondas por segundo (hertz).

Nas extremidades do espectro de luz visível encontram-se as ondas de luz infravermelha e ultravioleta. Apesar de invisíveis aos olhos humanos, alguns animais conseguem enxergar a

luz ultravermelha e a ultravioleta. Abelhas, por exemplo, utilizam a luz ultravioleta para localizar o pólen e o néctar [Silverthorn 2003].

O olho humano detecta a luz visível de maneira análoga a uma câmera. Basicamente, a luz que chega ao olho atravessa a córnea e é focalizada pelo cristalino em uma superfície interna do globo ocular que é sensível à luz - a retina. A quantidade de luz que entra no olho é regulada pela abertura frontal do olho, chamada de pupila. Tendo atingido a retina, a energia luminosa é transformada em energia elétrica, que segue pelas vias neurais para o processamento da imagem pelo cérebro [Silverthorn 2003].

Abaixo, podem ser observados: Figura 5.2 (a) o caminho percorrido pela luz que entra no olho humano; e (b) algumas das estruturas do olho humano: a íris (*iris*, a parte colorida externa frontal do olho); a córnea (*cornea*, que delimita a câmara cheia de líquido - o humor aquoso - que fica diante do cristalino); a pupila (*pupil*, que regula a entrada da luz no olho); o cristalino (*lens*, são as lentes que focalizam a luz sobre a retina); a conjuntiva (*sclera*, a parte externa branca do olho); o humor vítreo (*vitreous*, líquido transparente que preenche o globo ocular); a coróide (choroid, localizada entre a retina e a conjuntiva, possui vasos sanguíneos que nutrem o olho); o nervo ótico (*optic nerve*, que conduz os sinais elétricos gerados pela retina até o cérebro); a fóvea (*macula*, que é a região da retina onde a visão humana é mais aguçada); e, finalmente, a retina (*retina*, responsável por transformar os sinais luminosos em sinais elétricos) [Silverthorn 2003].



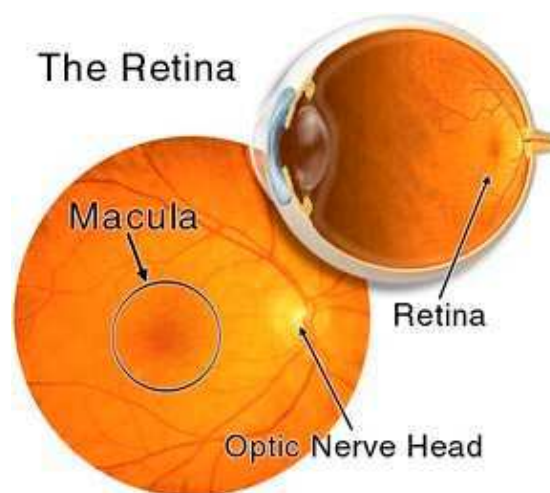
**Figura 5.2 (a) O Caminho da Luz até a Retina no Olho Humano;  
(b) As Principais Estruturas do Olho Humano [Erickson 2004]**



Cabe notar que, na região da retina chamada de disco ótico (região por onde o nervo ótico e os vasos sanguíneos saem do olho), existe o ponto cego (os sinais luminosos não são transformados em sinais elétricos) [Silverthorn 2003].

Outra formação importante da retina é a área da fóvea (*macula*) (Figura 5.3). A luz que atinge a retina deve atravessar várias camadas de células da própria retina até chegar na camada de células fotorreceptoras. Porém, na fóvea, as camadas de neurônios sobrepostas são empurradas para longe dessa pequena área, e a luz, assim, pode atingir diretamente os fotorreceptores, nessa região. Por isso, a fóvea é a região da retina onde a visão humana é mais nítida - e é para onde a luz que entra no olho humano é focalizada pelo cristalino (Figura 5.2 (a)).

A Figura 5.3, abaixo, mostra a retina do olho humano. Observar as regiões da fóvea (*Macula*) e a do início do nervo ótico (*Optic Nerve Head*).



**Figura 5.3 A Retina do Olho Humano [Erickson 2004]**

#### **5.1.1.2 Os Movimentos Oculares em Seres Humanos**

Para vasculhar a cena em busca de pontos de interesse, e para finalmente manter a atenção em determinada região da imagem captada, os olhos humanos utilizam três tipos de movimentos: as sacadas (*saccades*); a perseguição suave; e as microsacadas.

As sacadas são movimentos rápidos, e em geral de grande amplitude. O rápido movimento dos olhos durante as sacadas impede o processamento visual devido à perda de definição da imagem. A perseguição suave consiste em uma movimentação dos olhos de maneira que os mesmos possam acompanhar objetos que estejam em movimento com relação à pessoa. O objetivo da movimentação suave é o de manter o objeto de interesse focalizado sobre a

fóvea, na retina (Figura 5.3). Durante uma perseguição suave, podem ocorrer perdas momentâneas da focalização do objeto sobre a fóvea. Nesses instantes de mau posicionamento da imagem sobre a fóvea, o terceiro tipo de movimento ocular - as microsacadas - entra em ação para reposicionar os olhos de maneira que o objeto de interesse volte a ser focalizado sobre a fóvea. Em seguida à recuperação do foco, os olhos voltam a se mover de acordo com a perseguição suave, para continuarem acompanhado o objeto escolhido [Caetano 2000].

Assim, para que a atenção visual de uma pessoa seja levada para alguma região de uma cena, ocorrerá primeiro uma sacada, seguida (se houver movimento relativo entre a pessoa e a região de interesse) por uma perseguição suave, por sua vez pontuada por microsacadas - para reposicionamentos da imagem sobre a fóvea.

### 5.1.1.3 O Mecanismo Atencional Humano

Nos seres humanos, estudos mostram a existência de dois mecanismos atencionais para a visão. O primeiro mecanismo é voluntário, e diz respeito à orientação proposital dos olhos em busca de algum objeto na cena, de acordo com algum objetivo pessoal. Esse primeiro mecanismo é denominado controle atencional endógeno. O segundo mecanismo atencional é involuntário, e é causado por eventos que provocam estímulos externos ao organismo. Esse segundo mecanismo é chamado de controle atencional exógeno. [Caetano 2000].

Já [Rodrigues 2002] cita que os mecanismos atencionais, em seres humanos, podem ser divididos em sistemas *botton-up* e sistemas *top-down*. Nos sistemas *botton-up*, a informação para a atenção visual baseia-se na própria imagem de entrada, e o processamento ocorre com algo grau de paralelismo, rapidez e de maneira independente do esforço consciente. Por sua vez, nos sistemas *top-down*, a atenção é guiada por um conjunto de modelos armazenados, e pouco se sabe sobre a estrutura neural do cérebro que guia a natureza cognitiva dessa atenção.

### 5.1.2 Atenção Visual em Máquinas

Como mostra o Capítulo 3, nas Figura 3.3 e 3.5, existem diversas formas de capturar uma imagem através de dispositivos eletrônicos, com diferentes tipos de sensores sensíveis à luz. Embora a retina humana receba uma enorme quantidade de informações do mundo exterior, o cérebro trata com maior atenção a informação recebida pela retina sobre a fóvea - e assim o faz para que o processamento da imagem captada pelo olho seja mais eficiente. Surge desse fato, então, a motivação para lidarmos com a grande quantidade de informações provenientes de um sensor eletrônico de maneira análoga à maneira como o cérebro lida com as informações visuais que atingem a retina. Assim, poderemos reduzir a complexidade computacional e aumentar a eficiência de um sistema de processamento e reconhecimento de imagens.

Problemas em visão computacional cujas soluções não incluam algum sistema de atenção visual têm se mostrado computacionalmente intratáveis. Isso decorre da grande quantidade de informação a ser processada a partir da captura de uma imagem. Para que o problema seja resolvido em um período de tempo aceitável (muitas vezes, em tempo real), a atenção do sistema deve ser voltada apenas para as regiões da imagem de interesse para a resolução do problema, excluindo-se todas as demais regiões irrelevantes [Culhane 1992].

Existem várias técnicas para o desenvolvimento de mecanismos de atenção visual para máquinas. [Rodrigues 2002] cita as seguintes: (i) computação de características visuais primitivas - onde características básicas da imagem, tais como suas cores ou suas texturas (ver também [Bender 2003]), são utilizadas para a fase de segmentação - [Bender 2003] também menciona a utilização da forma dos objetos para a extração de características da imagem; (ii) representação piramidal - onde diversas imagens filtradas são geradas a partir da imagem original, para realçar detalhes ou suprimir ruídos; (iii) mapa de saliência - onde uma região da imagem (que será o foco da atenção) é definida como saliente se for diferente das demais regiões segundo algum aspecto determinado; (iv) inibição do retorno - através do uso de redes neurais artificiais, regiões de saliência são encontradas na imagem, mas para que o mecanismo possa continuar buscando outras regiões de saliência, os neurônios são inibidos periodicamente; e, por fim, (v) alguns modelos de atenção *bottom-up* - os quais baseiam-se, em sua maioria, na utilização dos mapas de saliência.

Neste trabalho, para o mecanismo atencional do Reconhecedor Inteligente de Sinais de Trânsito Brasileiros (RIST-Br), será adotada a técnica da computação de características visuais primitivas, e o atributo escolhido para a fase de segmentação (que será apresentada com maiores detalhes na seção 5.3.2, deste capítulo, e na seção 6.2.2, do Capítulo 6) da imagem será a cor dos sinais de trânsito. Essa escolha permite que seja explorada a vantagem atencional que as cores dos sinais de trânsito oferecem, pois as mesmas servem para destacar os sinais de trânsito dos demais objetos de uma cena.

## **5.2 Experimento Sobre o Reconhecimento Inteligente de Padrões**

Um trabalho prático voltado para o reconhecimento inteligente de padrões foi desenvolvido pelo autor desta monografia sob a Disciplina de Redes Neurais Artificiais, da Universidade do Vale do Rio dos Sinos - UNISINOS, ministrada pelo orientador deste trabalho - Prof. Fernando Santos Osório. Esse trabalho prático foi elaborado com a utilização do simulador de redes neurais SNNS (*Stuttgart Neural Network Simulator*), descrito brevemente no Capítulo 4 (seção 4.3) e da base de imagens COIL (*Columbia Object Image Library*) [COIL 2003], visando aprender a reconhecer objetos presentes nessa base de imagens.

### **5.2.1 Objetivos do Experimento**

O objetivo deste experimento foi usar uma rede neural artificial para reconhecer uma imagem em cores presente na base de dados COIL dentre imagens de diversos objetos



conjunto de validação para a rede neural, com 20 padrões. Vale observar que em cada imagem da base COIL-100 há apenas um objeto.

### 5.2.3 Pré-processamento das Imagens do Experimento

As imagens originalmente possuem o formato de 128x128 pixels (24 bits/pixel), na base de dados COIL-100. A partir destas imagens foram obtidos dados retirados do histograma da imagem, com a utilização do Paint Shop Pro [Jasc 2004] (pode ser usado também o MatLab). As informações foram obtidas uma de cada vez (isto é, sem combinar *Red* com *Green*, por exemplo). Os procedimentos adotados para as 60 imagens selecionadas foram os seguintes:

1. Foram obtidos dados do histograma conforme informações da Tabela 5.1;

Nome	Red		Green		Blue		Luminance	
Original	Maximum	Average	Maximum	Average	Maximum	Average	Maximum	Average
obj01__090	4012	122	4119	92	4063	27	4136	93
obj10__270	469	104	476	94	480	89	422	96
obj100__150	6971	67	7134	33	7105	25	7071	42
obj12__045	1581	91	1268	72	1296	50	1428	75
obj12__170	1446	99	1496	78	1387	51	1662	81
obj15__170	9103	56	9243	42	8920	23	9498	44
obj15__295	4819	82	5157	58	4628	22	4879	61
obj16__225	1783	84	1908	69	1900	66	2010	73
obj18__080	1554	91	1525	88	1328	84	1464	88
obj18__355	1676	89	1661	85	1509	81	1628	85

**Tabela 5.1 Dados do Histograma de Cores**

2. Foi reduzida a resolução da imagem de 128x128 para 32x32 pixels (Figura 5.6);



**Figura 5.6 Redução da Resolução e Detecção de Bordas**

3. Foi aplicado um filtro para encontrar bordas (Figura 5.6);

4. Foram reduzidas as cores da imagem para apenas duas cores (1 bit/pixel - Nearest color / Floyd-Steinberg - disponível no software de edição gráfica Paint Shop Pro[Jasc 2004]) (Figura 5.6);
5. As imagens foram exportadas no formato PBM - "Portable Bit Map", que é um formato ASCII fácil de ser adaptado para o formato adotado pelo SNNS (Simulador Neural).
6. Foi gerado o arquivo SNNS usado no treinamento da Rede Neural (Figura 5.7).

```

SNNS pattern definition file V4.2
generated at Mon Sep 29 15:32:00 2003

# f01-treino-img-01-40.pat

No. of patterns : 40
No. of input units : 72
No. of output units : 1

# Entrada 1:

# 32 valores - Somatorios de pixels por linha:
0 0 0 0 0 16 10 14 9 11 10 12 6 16 3 4
3 2 4 7 8 1 1 1 1 24 0 0 0 0 0 0

# 32 valores - Somatorios de pixels por coluna:
0 0 21 3 2 11 1 8 5 2 7 8 7 4 3 5
4 7 6 8 4 4 3 6 2 6 8 3 4 6 4 1

# 8 valores - Vermelho, Verde, Azul e Luminancia (Maximo e Media de cada um):
4012 122 4119 92 4063 27 4136 93

# Saida 1:
0

```

**Figura 5.7 Trecho do Arquivo de Treinamento Usado no SNNS**

Como pode ser observado na Figura 5.6, a imagem (após terem sido obtidas as suas informações de cores e após terem sido detectadas as suas bordas) foi reduzida a uma matriz de 32x32 pixels, com 1 bit/pixel. Com o auxílio de um *software* elaborado pelo autor, foi efetuado o somatório dos pixels “ativados” na imagem – linha a linha e coluna a coluna. Esses resultados, para uma dada imagem, podem ser observados na Figura 5.7 acima, que mostra uma parte do arquivo usado no SNNS para treinamento e validação da rede neural.

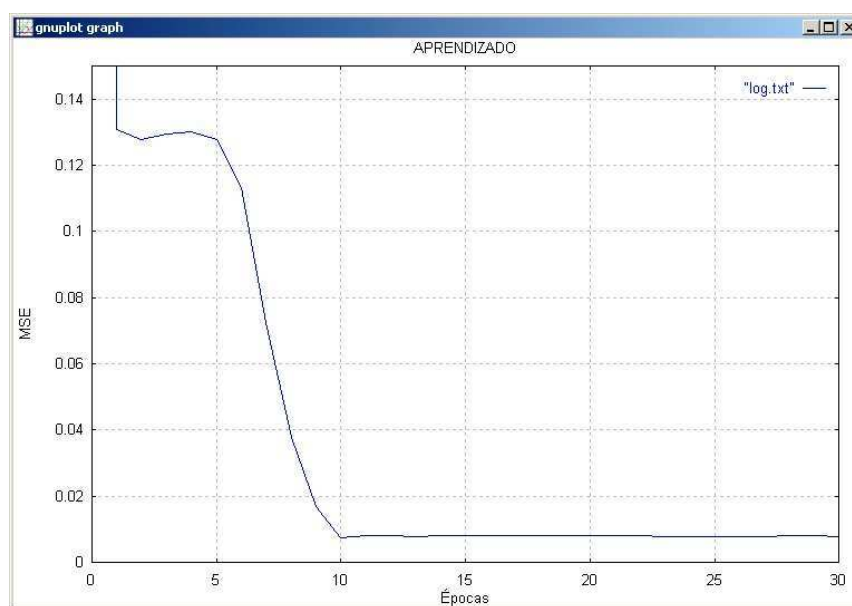
Esses 64 somatórios, juntamente com as oito informações sobre as cores da imagem (Tabela 5.1, acima) resultaram nos 72 valores representativos de uma imagem. Assim, ao invés da rede neural ser projetada com 49.152 entradas (128x128 pixels x 3 bytes (1 byte/componente de cor)), a rede foi especificada com apenas 72 entradas, o que resulta numa redução de 99,85% no volume de informações necessário para representar cada imagem das bases de treino e validação para a rede neural artificial.

A partir da base de dados das imagens pré-processadas, foi possível aplicar o algoritmo de treinamento da rede neural (*Resilient Back Propagation*, descrito no Capítulo 4, seção 4.2.5.2), visando o reconhecimento dos padrões em questão (imagens do pato). A rede

neural, totalmente conectada, foi projetada com 72 neurônios na camada de entrada, 18 neurônios na única camada oculta, e um neurônio na camada de saída.

#### 5.2.4 Aprendizado Neural no Experimento

Com o uso do “Batchman” – ferramenta disponível no pacote do SNNS - diversos experimentos foram realizados, de forma a medir o desempenho do sistema no reconhecimento dos padrões apresentados. Foram criadas 2 bases de dados a partir das 60 imagens tratadas, uma base de aprendizado (usada no ajuste dos pesos da rede neural – 2/3 dos dados) e uma base de teste (1/3 dos dados), que permite realizar a validação do aprendizado, medindo o grau de generalização obtido pela rede neural para exemplos não apresentados no processo de aprendizado. A Matriz de Confusão, mostrada abaixo, demonstra que a rede neural conseguiu reconhecer corretamente, 100% das vezes, a imagem do pato dentre os 20 padrões apresentados durante a validação. Adicionalmente, a rede também identificou corretamente, 100% das vezes, os padrões correspondentes a objetos “não-pato”, apresentados entre os 20 padrões da validação.



**Figura 5.8 Evolução do Erro MSE (*Mean Square Error* - Erro Médio Quadrado) na Saída da Rede Neural Artificial**

Classe (0: não-pato; 1: pato); n = 20	Prevista: "é pato!"	Prevista: "não é pato!"
Encontrada "é pato!"	3 Verdadeiros Positivos (reconheceu pato!)	ZERO Falsos Positivos (achou pato onde não existe!)
Encontrada "não é pato!"	ZERO Falsos Negativos (pato não reconhecido!)	17 Verdadeiros Negativos (reconheceu não-pato!)

**Tabela 5.2 Matriz de Confusão para a base de Validação**

### 5.2.5 Conclusões sobre o Experimento

A partir da aplicação das técnicas de pré-processamento em uma base de dados de imagens e de uma rede neural artificial treinada com estes dados, foi possível reconhecer o padrão-alvo (a imagem de um pato).

Os resultados do experimento prático apresentado comprovaram que a metodologia desenvolvida para o mesmo poderia ser aplicada ao reconhecimento inteligente de sinais de trânsito desenvolvido nesta monografia.

## 5.3 *As Fases do Reconhecimento de Sinais de Trânsito*

Esta seção apresenta, de maneira geral, a metodologia adotada para a implementação do Reconhecedor Inteligente de Sinais de Trânsito Brasileiros (RIST-Br), que é descrito no Capítulo 6.

### 5.1.1 A Captura das Imagens de Entrada

Foi escolhido, para este trabalho, o método de captura de imagens através de uma câmera fotográfica digital. As imagens capturadas são, em sua maioria, de cruzamentos de ruas, onde podem ser encontradas placas como a de Parada Obrigatória, ou a de Sentido Obrigatório [DNIT 2004]. Esse método de captura foi selecionado por dois motivos principais: o sistema de RIST-Br em uma primeira etapa não foi projetado para operar em tempo real, podendo futuramente vir a ser implementado para operar dessa maneira; e imagens de boa qualidade podem ser obtidas, o que facilita o processamento por parte do sistema.

### 5.3.1 O Pré-processamento das Imagens de Entrada

Em um primeiro passo, as imagens capturadas, no modelo de cores RGB, são convertidas para o modelo de cores HSI (adotado devido às suas características semelhantes às apresentadas pelo modelo HSV no Capítulo 3, seção 3.2.2) na forma de três arquivos de imagem separados: um arquivo para cada componente do modelo HSI. Apenas o arquivo correspondente à componente H (*hue* - cor) de cada imagem é utilizado para o passo seguinte do pré-processamento: a binarização da imagem segundo o modelo atencional adotado na próxima seção.

### 5.3.2 A Segmentação das Imagens de Entrada

Como mecanismo atencional, será utilizada a técnica da computação de características visuais primitivas [Rodrigues 2002], na qual a característica selecionada foi a cor. A cor foi



selecionada por ser um elemento característico muito importante da sinalização do trânsito, servindo para destacar o sinal dos outros elementos da imagem e atrair a atenção do sistema atencional humano. Além disso, segundo [Bender 2003], a cor é relativamente independente com relação à resolução, ao tamanho e à orientação da imagem. Conforme foi comentado na seção anterior, o arquivo de imagem correspondente à componente H será binarizado de acordo com a cor que está sendo procurada na imagem. Finalmente, as subimagens candidatas são extraídas - através de uma varredura da imagem que utiliza uma "cabeça de leitura" (uma área de 200x200 pixels) - para servirem à próxima fase do processo: a de representação e descrição.

### 5.3.3 A Representação e a Descrição das Subimagens Candidatas

A representação das subimagens candidatas se dará através de informações de seus respectivos histogramas de cores, no modelo RGB, e da disposição de seus pixels em uma matriz de resolução menor que a da subimagem original.

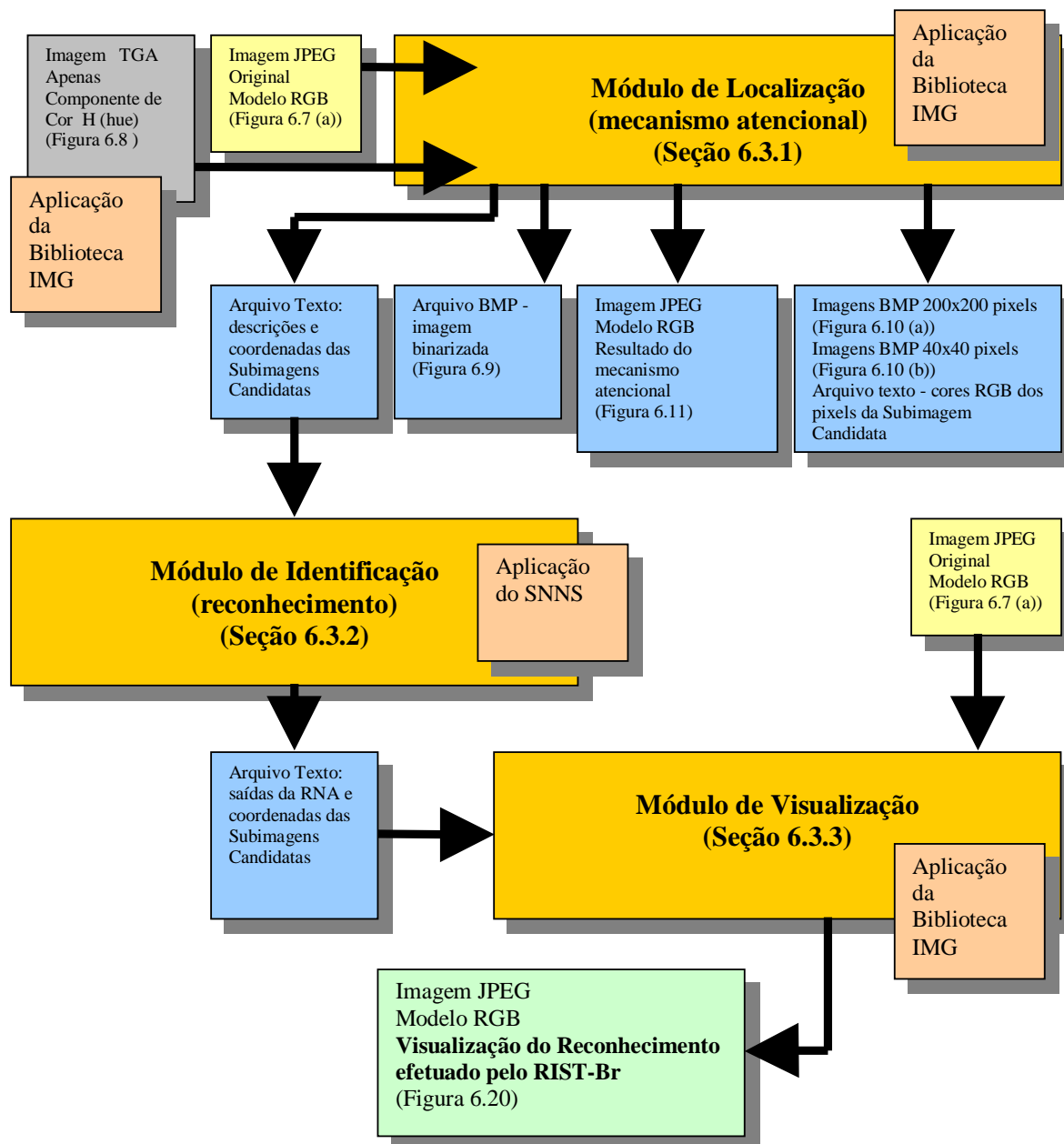
A descrição das subimagens, por sua vez, será feita através de valores numéricos correspondentes aos somatórios dos *pixels* ativos em cada coluna e em cada linha da subimagem já representada, adicionalmente às informações de seus respectivos histogramas de cores, que permanecem as mesmas do passo de representação.

### 5.3.4 O Reconhecimento dos Sinais de Trânsito nas Subimagens Candidatas

Para o reconhecimento inteligente dos sinais de trânsito, foi escolhida a técnica das redes neurais artificiais (Capítulo 4). Será utilizada uma rede do tipo perceptron de múltiplas camadas, totalmente conectada, com aprendizado do tipo propagação resiliente (*resilient propagation* - RPROP). Essa rede neural receberá um vetor de entrada com os valores numéricos do passo de descrição, da seção anterior, e fornecerá como resposta um vetor de saída com os valores que indicarão a classe da subimagem candidata apresentada para a rede (na forma do vetor de entrada).

## 5.4 Considerações Finais Sobre Este Capítulo

Neste capítulo foi apresentada a metodologia adotada pelo autor para o desenvolvimento do Reconhecedor Inteligente de Sinais de Trânsito Brasileiros - RIST-Br. O diagrama da Figura 5.9, a seguir, que será reapresentado na seção 6.3.3, do Capítulo 6, mostra uma visão geral do sistema RIST-Br. Os módulos desse diagrama são apresentados na seção 6.3, do Capítulo 6.



**Figura 5.9 Visão Geral do RIST-Br**