ALEXANDRE DA SILVA SIMÕES

**SEGMENTAÇÃO DE IMAGENS POR
CLASSIFICAÇÃO DE CORES : UMA
ABORDAGEM NEURAL**

Dissertação apresentada à Escola
Politécnica da Universidade de São Paulo
para a obtenção do título de Mestre em
Engenharia.

São Paulo
2000.

ALEXANDRE DA SILVA SIMÕES

**SEGMENTAÇÃO DE IMAGENS POR
CLASSIFICAÇÃO DE CORES : UMA
ABORDAGEM NEURAL**

Dissertação apresentada à Escola
Politécnica da Universidade de São Paulo
para a obtenção do título de Mestre em
Engenharia.

Área de Concentração:
Sistemas Digitais

Orientadora:
Profª. Dra. Anna Helena Reali Costa

São Paulo
2000.

Simões, Alexandre da Silva

Segmentação de imagens por classificação de cores: uma abordagem neural. São Paulo, 2000.

132 p.

Dissertação (Mestrado) - Escola Politécnica da Universidade de São Paulo. Departamento de Engenharia Elétrica.

1. Visão Computacional 2. Redes Neurais Artificiais 3. Processamento de Imagens - I. Universidade de São Paulo. Escola Politécnica. Departamento de Engenharia Elétrica. II.t

A todos aqueles que acreditam que um dia não haverá fronteiras, injustiças nem diferenças entre os homens, e àqueles que acreditam que esse mundo livre - tal como desejamos - é fruto do compartilhamento de todo o conhecimento pelo simples engrandecimento da humanidade.

Este trabalho foi produzido inteiramente utilizando *software* livre.

AGRADECIMENTOS

Aos meus pais e a Iowana, incentivadores maiores deste trabalho.

Aos amigos da Escola Politécnica: Mikio, André, Jomi, Márcia e aos inúmeros outros que colaboraram para tornar possível o presente trabalho.

Aos professores da Escola Politécnica: Prof. Dr. Jaime Simão Sichman, Prof. Dr. Edson Satoshi Gomi, Profa. Dra. Denise Consonni e Prof. Dr. Geraldo Lino de Campos, pela ajuda prestada das mais diversas maneiras além do simples dever do ensino. Aos técnicos e funcionários da Escola Politécnica: Elizabeth e especialmente a Fátima, que constantemente desdobrou-se sem medir esforços para que as atividades propostas fossem cumpridas.

Um agradecimento especial aos companheiros Reinaldo Augusto da Costa Bianchi e Leonardo Azevedo Scardua e ao amigo e professor Dr. André Riuty Hirakawa, que mais do que colaboradores, foram co-responsáveis por este trabalho, e sem os quais essa realização não teria sido possível.

Da mesma forma ao Prof. Dr. Guido Stolfi que pacientemente me forneceu o conhecimento necessário para o desenvolvimento de minha pesquisa.

À CAPES que, cumprindo seu papel institucional possibilitou esta pesquisa.

A Anna Helena Reali Costa, que aceitou o desafio de receber um jovem desconhecido e ensiná-lo a ser mestre.

RESUMO

A presente dissertação aborda a segmentação de imagens coloridas através do processo de classificação de cores, isto é, a segmentação de imagens baseada no atributo cor dos pixels.

O problema a ser tratado diz respeito à obtenção de uma classificação de cores tão próxima à humana quanto o possível. Em outros termos, busca-se uma classificação robusta à variação de grandezas como a iluminação ou brilho da cor, além de ser tolerante a erros no processo de amostragem. Tal problemática é encontrada em diversas situações práticas que sofram influência do ambiente, sobretudo no domínio de aplicação: o futebol de robôs.

Com relação a tal problema, diversas questões permanecem abertas, tais como a forma de representação de cores e o tipo de classificador capaz de maximizar o desempenho da classificação. As modelagens clássicas, de forma geral, têm se mostrado inadequadas nesse contexto, estimulando-nos a buscar novas soluções. Assim sendo, apresentamos um classificador utilizando uma das técnicas que tem mostrado grande aplicabilidade nesse âmbito: as redes neurais artificiais.

A aplicabilidade da técnica esbarra na obtenção de uma generalização adequada por parte da rede para o problema proposto, o que implica na necessidade de uma metodologia para o fornecimento de exemplos na fase de treinamento da rede.

Desta forma, além da modelagem e implementação do classificador, buscamos proceder uma investigação a cerca de sua generalização em contextos diversos para situações não presentes no universo de treinamento, visando determinar o conjunto de fatores (sistema de representação de cores, metodologia de fornecimento de exemplos e arquitetura de rede) que maximizem o desempenho do classificador.

ABSTRACT

The present work approaches the segmentation of colored images through the process of color classification, i.e., the segmentation of images based on the color attribute of pixels.

We look for a color classification as close as possible of human classification. In other words, we look for a robust classification with respect to the variation of illumination and color brightness, which tries to be tolerant to errors in the sampling process. We may find such kind of problems in various practical situations, for instance, situations that are influenced from the environment in the application domain: the robotic soccer.

With regard to this problem, there are still diverse questions that remain unsolved, such as color representation form and type of classifier which maximizes the classification performance. In fact, classic models have shown to be inadequate in this context, in general, stimulating us to investigate new solutions. In our work, we present a classifier using one technique that has shown great applicability in this scope: the artificial neural networks.

In order to obtain a correct generalization in the network, we faced the necessity to build a methodology to supply examples in the training phase of the network.

In short, we model and implement a classifier, while searching to assess about its generalization power in different contexts and in the universe of training, so as to determine the set of factors (system of representation of colors, methodology of supply of examples and architecture of network) that maximizes its performance.

SUMÁRIO

LISTA DE TABELAS

LISTA DE FIGURAS

RESUMO

1. INTRODUÇÃO.....	1
1.1 O PROBLEMA DA SEGMENTAÇÃO BASEADA EM CORES.....	2
1.2 OBJETIVOS.....	3
1.3 MOTIVAÇÃO E RESULTADOS ESPERADOS.....	3
1.4 ORGANIZAÇÃO DO TRABALHO.....	4
2. IMAGEM.....	5
2.1 O SISTEMA VISUAL HUMANO.....	5
2.2 FÍSICA DA FORMAÇÃO DA IMAGEM.....	8
2.3 MODELAGEM MATEMÁTICA DA IMAGEM.....	10
2.4 IMAGEM DIGITAL.....	11
2.5 REVISITANDO O CONCEITO DE IMAGEM.....	14
3. CORES.....	15
3.1 DEFINIÇÕES E DEPENDÊNCIAS.....	15
3.2 GRANDEZAS FOTOMÉTRICAS.....	17
3.3 REPRESENTAÇÃO DE CORES.....	19
3.4 PADRONIZAÇÕES DE CORES.....	23
3.5 INTERFACES.....	30
4. SEGMENTAÇÃO.....	34
4.1 MÉTODOS PARA SEGMENTAÇÃO.....	34
4.2 A IMPOSIÇÃO DE LIMIARES.....	35
4.3 ALGORITMOS BASEADOS EM VIZINHANÇA.....	37
4.4 A CLASSIFICAÇÃO DE PIXELS.....	38
5. REDES NEURAIS ARTIFICIAIS.....	40
5.1 HISTÓRICO.....	40
5.2 O NEURÔNIO.....	43
5.3 OS TIPOS DE RNAs.....	47
5.4 O PERCEPTRON MULTICAMADA.....	49
5.5 JUSTIFICATIVAS PARA A ADOÇÃO DO MLP E DA RETROPROPAGAÇÃO DO ERRO.....	61
6. REVISITANDO O PROBLEMA.....	63
7. TRABALHOS CORRELATOS.....	69
7.1 COGNACHROME.....	69
7.2 A CLASSIFICAÇÃO DE CORES POR RNAs DE RAE-RITTER-LITTMAN.....	72
7.3 A CLASSIFICAÇÃO DE CORES POR APRENDIZADO SUPERVISIONADO DE VELOSO-UTHER.....	77
8. IMPLEMENTAÇÃO DO CLASSIFICADOR.....	80
8.1 DOMÍNIO DE APLICAÇÃO.....	80
8.2 O CLASSIFICADOR ADOPTADO.....	82
8.3 IMPLEMENTAÇÕES BÁSICAS E DE SUPORTE.....	84
9. METODOLOGIA.....	86
9.1 APRESENTAÇÃO DE EXEMPLOS.....	86
9.2 CONVERGÊNCIA NO TREINAMENTO DA RNA.....	87
9.3 VALIDAÇÃO DA RNA.....	89
9.4 TREINAMENTO.....	91
9.5 ORGANIZAÇÃO DOS RESULTADOS.....	93

9.6 CONFIGURAÇÕES UTILIZADAS.....	97
10. DADOS E RESULTADOS.....	98
10.1 TREINAMENTO COM CORES CONTÍNUAS.....	98
10.2 TREINAMENTO COM CORES EM DIFERENTES SATURAÇÕES.....	108
10.3 TREINAMENTO COM CORES EM DIFERENTES INTENSIDADES.....	116
11. CONCLUSÃO.....	123
12. REFERÊNCIAS BIBLIOGRÁFICAS.....	128
APÊNDICE A: VALORES DE MÉDIA, DESVIO PADRÃO E CURVA NORMAL DE DISTRIBUIÇÃO PARA CADA GRANDEZA USADA NA REPRESENTAÇÃO DAS CORES DAS IMAGENS DE CORES CONTÍNUAS E EM DEGRADÊ UTILIZADAS NOS EXEMPLOS.	
APÊNDICE B: VALORES DOS PONTOS E SUAS RESPECTIVAS CLASSIFICAÇÕES FORNECIDAS ÀS RNAS DURANTE SEU TREINAMENTO SUPERVISIONADO.	
APÊNDICE C: EXEMPLOS USADOS PARA OS TREINAMENTOS DAS RNAS E SUAS RESPECTIVAS MÉDIA, DESVIO PADRÃO E CURVAS NORMAIS DE DISTRIBUIÇÃO PARA CADA GRANDEZA UTILIZADA NA REPRESENTAÇÃO DE CORES.	

LISTA DE TABELAS

TABELA 2.1 - AS VÁRIAS MEDIDAS DE DISTÂNCIA ENTRE PONTOS EM UMA IMAGEM DIGITAL.....	14
TABELA 3.1 - LUZES MONOCROMÁTICAS.....	16
TABELA 3.2 - AS GRANDEZAS RADIOMÉTRICAS, SUAS UNIDADES USUAIS E DEFINIÇÕES.....	17
TABELA 3.3 - GRANDEZAS RADIOMÉTRICAS E FOTOMÉTRICAS E SUAS UNIDADES USUAIS	19
TABELA 3.4 - COMPRIMENTOS DE ONDA PARA O SISTEMA CIE-RGB.....	24
TABELA 3.5 - COORDENADAS DE CROMATICIDADE DO SISTEMA XYZ	26
TABELA 5.1 - FUNÇÕES DE ATIVAÇÃO DE UM NEURÔNIO.....	47
TABELA 7.1 - A TABELA DE LIMIARES DO CLASSIFICADOR DE CORES.....	78
TABELA 8.1 - PADRÕES DE COR PARA O FUTEBOL DE ROBÔS.....	82
TABELA 9.1 - DISTRIBUIÇÃO DOS EXEMPLOS FORNECIDOS ÀS RNAS.....	95
TABELA 10.1 - LIMITES PARA A VARIÁVEL V.....	101
TABELA 10.2 - LIMITES PARA A VARIÁVEL U.....	101

LISTA DE FIGURAS

FIGURA 2.1 - REPRESENTAÇÃO ESQUEMÁTICA DO GLOBO OCULAR E SEUS PRINCIPAIS COMPONENTES.....	6
FIGURA 2.2 - FUNÇÃO DE RESPOSTA ESPECTRAL PARA CADA UM DOS TRÊS TIPOS DE CONES DA RETINA HUMANA	7
FIGURA 2.3 - AS CAMADAS CELULARES NA RETINA HUMANA	8
FIGURA 2.4 - A GEOMETRIA DA REFLECTÂNCIA.....	9
FIGURA 2.5 - A COLEÇÃO DE FUNÇÕES DELTA APLICADA SOBRE UMA IMAGEM BIDIMENSIONAL.....	11
FIGURA 2.6 - UMA RETA (GRANDEZA CONTÍNUA) E SUA REPRESENTAÇÃO EM UM DOMÍNIO DISCRETO....	13
FIGURA 2.7 - CONTORNOS EQUIDISTANTES PARA DIFERENTES MÉTRICAS	14
FIGURA 3.1 - FUNÇÃO DE LUMINOSIDADE DE UM OBSERVADOR PADRÃO AOS DIFERENTES COMPRIMENTOS DE LUZ VISÍVEL.....	18
FIGURA 3.2 - O PLANO DE MAXWELL E O SPECTRUM LOCUS.....	22
FIGURA 3.3 - DIAGRAMA DE CROMATICIDADE R-G.....	25
FIGURA 3.4 - OS EIXOS XYZ SOBRE O DIAGRAMA RGB.....	26
FIGURA 3.5 - DIAGRAMA DE CROMATICIDADE DO PADRÃO CIE-XYZ	27
FIGURA 3.6 - O SÓLIDO DE COR DO SISTEMA HS.....	30
FIGURA 3.7 - DIAGRAMA REPRESENTATIVO DO TRIÂNGULO DE CROMATICIDADE PARA O RGB DE DOIS MONITORES DISTINTOS.....	33
FIGURA 4.1 - DETERMINAÇÃO DE LIMAR PARA HISTOGRAMA EM TONS DE CINZA.....	36
FIGURA 4.2 - OS LIMIARES APLICADOS A UM ESPAÇO TRIDIMENSIONAL DE CORES.....	36
FIGURA 5.1 - O NEURÔNIO BIOLÓGICO.....	44
FIGURA 5.2 - BOMBA DE SÓDIO E POTÁSSIO: A PROPAGAÇÃO DO IMPULSO ELÉTRICO EM UM NEURÔNIO...	44
FIGURA 5.3 - DISPARO DE UM NEURÔNIO.....	45
FIGURA 5.4 - MODELO NÃO-LINEAR DE UM NEURÔNIO.....	46
FIGURA 5.5 - A REDE NEURAL MULTICAMADA.....	50
FIGURA 5.6 - DIFERENTES GENERALIZAÇÕES PARA UM EXEMPLO "A" AUSENTE DO UNIVERSO DE TREINAMENTO {B,C,D}.....	59
FIGURA 5.7 - CURVA TÍPICA DO ERRO QUADRÁTICO DA VALIDAÇÃO CRUZADA ONDE É APLICÁVEL A SUSPENSÃO PRÉVIA DO TREINAMENTO	61
FIGURA 6.1 - AMOSTRAGEM DE CORES POR UMA CÂMERA.....	67
FIGURA 6.2 - A FIGURA 6.1 NO ESPAÇO DE CORES.....	67
FIGURA 6.3 - VISTAS LATERAIS DA FIGURA 6.2.....	67
FIGURA 6.4 - A IMAGEM 6.1 ILUMINADA COM LUZ INCANDESCENTE A	68
FIGURA 6.5 - ESPAÇO DE CORES PARA AS FIGURA 6.4 A, B E C	68
FIGURA 7.1 - DIAGRAMA ESQUEMÁTICO DE FUNCIONAMENTO DO SISTEMA COGNACHROME.....	69
FIGURA 7.2 - O CLASSIFICADOR DE CORES DO SISTEMA COGNACHROME.....	70
FIGURA 7.3 - DIAGRAMA ESQUEMÁTICO DO RECONHECIMENTO DE FACES.....	73
FIGURA 7.4 - A APROXIMAÇÃO LINEAR DA LLM.....	75
FIGURA 7.5 - OS CONJUNTOS DE PIXELS DE UMA IMAGEM.....	76
FIGURA 7.6 - A CLASSIFICAÇÃO DE CORES NO SISTEMA YUV.....	77
FIGURA 7.7 - O SISTEMA VISUAL DE VELOSO & UTHER.....	78
FIGURA 8.1 - DIAGRAMA ESQUEMÁTICO DO AMBIENTE DE FUTEBOL DE ROBÔS MÓVEIS AUTÔNOMOS.....	81
FIGURA 8.2 - A REDE NEURAL PROPOSTA.....	83
FIGURA 9.1 - UMA PALETA DE CORES IMPRESSA AMOSTRADA POR UMA CÂMERA.....	86
FIGURA 9.2 - OS ERROS QUADRÁTICOS DOS NEURÔNIOS DE SAÍDA.....	88

FIGURA 9.3 - OS ERROS QUADRÁTICOS DE SAÍDA DOS NEURÔNIOS	89
FIGURA 9.4 - REPRESENTAÇÃO ESQUEMÁTICA DOS CONJUNTOS DE VALIDAÇÃO UTILIZADOS.....	90
FIGURA 9.5 - OS PIXELS DA IMAGEM - SUBCONJUNTOS DE TREINAMENTO E VALIDAÇÃO.....	91
FIGURA 9.6 - AS IMAGENS DA PALETA CONTÍNUA	93
FIGURA 10.1 - RESPOSTA À VARIAÇÃO DO NÚMERO DE NEURÔNIOS DA CAMADA OCULTA PARA 25 EXEMPLOS.....	98
FIGURA 10.2 - RESPOSTA À VARIAÇÃO DO NÚMERO DE EXEMPLOS PARA 6 NEURÔNIOS NA CAMADA OCULTA.	98
FIGURA 10.3 - AS RESPOSTAS DA RNA TREINADA COM 25 EXEMPLOS DA IMAGEM CONTÍNUA, USANDO 10 NEURÔNIOS NA CAMADA OCULTA APÓS CONVERGÊNCIA	98
FIGURA 10.4 - A PARTIÇÃO DO ESPAÇO DE CORES PARA O SISTEMA RGB TREINADO COM 25 EXEMPLOS DA IMAGEM CONTÍNUA UTILIZANDO 10 NEURÔNIOS NA CAMADA OCULTA	99
FIGURA 10.5 - RESPOSTA À VARIAÇÃO DO NÚMERO DE NEURÔNIOS DA CAMADA OCULTA COM 25 EXEMPLOS.....	102
FIGURA 10.6 - RESPOSTA À VARIAÇÃO DO NÚMERO DE EXEMPLOS COM 8 NEURÔNIOS NA CAMADA OCULTA.	102
FIGURA 10.7 - AS RESPOSTAS DA RNA COM 8 NEURÔNIOS NA CAMADA OCULTA TREINADA PARA 45 EXEMPLOS APÓS CONVERGÊNCIA	102
FIGURA 10.8 - A PARTIÇÃO DO ESPAÇO DE CORES PARA O SISTEMA YUV TREINADO 45 EXEMPLOS DA IMAGEM CONTÍNUA E UTILIZANDO 8 NEURÔNIOS NA CAMADA OCULTA	103
FIGURA 10.9 - OS LIMITES DO SISTEMA HSV UTILIZADOS.....	105
FIGURA 10.10 - RESPOSTA À VARIAÇÃO DO NÚMERO DE NEURÔNIOS NA CAMADA OCULTA UTILIZANDO 25 EXEMPLOS.....	105
FIGURA 10.11 - RESPOSTA À VARIAÇÃO DO NÚMERO DE EXEMPLOS UTILIZANDO 12 NEURÔNIOS NA CAMADA OCULTA.....	105
FIGURA 10.12 - AS RESPOSTAS DA RNA TREINADA COM 45 EXEMPLOS DA IMAGEM CONTÍNUA, UTILIZANDO 15 NEURÔNIOS NA CAMADA OCULTA APÓS CONVERGÊNCIA	105
FIGURA 10.13 - A DIVISÃO DO ESPAÇO DE CORES PARA O SISTEMA HSV TREINADO COM PALETA CONTÍNUA	106
FIGURA 10.14 - A APRESENTAÇÃO DOS EXEMPLOS NO FORMATO HSV.....	108
FIGURA 10.15 - RESPOSTA À VARIAÇÃO DO NÚMERO DE NEURÔNIOS NA CAMADA OCULTA PARA 25 EXEMPLOS.....	108
FIGURA 10.16 - RESPOSTA À VARIAÇÃO DO NÚMERO DE EXEMPLOS PARA 12 NEURÔNIOS NA CAMADA OCULTA.....	108
FIGURA 10.17 - AS RESPOSTAS DA RNA TREINADA COM 45 EXEMPLOS COM 9 NEURÔNIOS NA CAMADA OCULTA APÓS CONVERGÊNCIA	109
FIGURA 10.18 - AS RESPOSTAS DA RNA TREINADA COM 25 EXEMPLOS COM 12 NEURÔNIOS NA CAMADA OCULTA APÓS CONVERGÊNCIA.....	109
FIGURA 10.19 - A PARTIÇÃO DO ESPAÇO DE CORES PARA O SISTEMA RGB TREINADO COM CORES COM VARIAÇÃO DE SATURAÇÃO, APÓS CONVERGÊNCIA.....	109
FIGURA 10.20 - RESPOSTA À VARIAÇÃO DO NÚMERO DE NEURÔNIOS NA CAMADA OCULTA PARA 25 EXEMPLOS.....	112
FIGURA 10.21 - RESPOSTA À VARIAÇÃO DO NÚMERO DE EXEMPLOS PARA 10 NEURÔNIOS NA CAMADA OCULTA.....	112
FIGURA 10.22 - AS RESPOSTAS DA RNA TREINADA COM 50 EXEMPLOS DA IMAGEM DEGRADÊ USANDO 10 NEURÔNIOS NA CAMADA OCULTA APÓS CONVERGÊNCIA	112
FIGURA 10.23 - A PARTIÇÃO DO ESPAÇO DE CORES YUV PARA O CASO DO TREINAMENTO DE CORES EM DIFERENTES SATURAÇÕES.....	113
FIGURA 10.24 - RESPOSTA À VARIAÇÃO DO NÚMERO DE NEURÔNIOS NA CAMADA OCULTA PARA 25 EXEMPLOS.....	114

FIGURA 10.25 - RESPOSTA À VARIAÇÃO DO NÚMERO DE EXEMPLOS PARA 18 NEURÔNIOS NA CAMADA OCULTA.....	114
FIGURA 10.26 - AS RESPOSTAS DA RNA TREINADA COM 25 EXEMPLOS DA IMAGEM EM DIFERENTES SATURAÇÕES USANDO 18 NEURÔNIOS NA CAMADA OCULTA APÓS CONVERGÊNCIA	114
FIGURA 10.27 - A DIVISÃO DO ESPAÇO DE CORES HSV PARA O CASO DO TREINAMENTO COM CORES EM SATURAÇÃO VARIADA.....	115
FIGURA 10.28 - RESPOSTA À VARIAÇÃO DO NÚMERO DE NEURÔNIOS NA CAMADA OCULTA PARA 25 EXEMPLOS.....	116
FIGURA 10.29 - RESPOSTA À VARIAÇÃO DO NÚMERO DE EXEMPLOS PARA 20 NEURÔNIOS NA CAMADA OCULTA.....	116
FIGURA 10.30 - A RESPOSTA DA RNA COM 5 EXEMPLOS EM DIFERENTES INTENSIDADES LUMINOSAS E 20 NEURÔNIOS NA CAMADA OCULTA APÓS CONVERGÊNCIA PARA O CASO CONTÍNUO.....	116
FIGURA 10.31 - AS RESPOSTAS DA RNA PARA A PALETA DEGRADÊ QUANDO TREINADA EM DIFERENTES INTENSIDADES LUMINOSAS E 20 NEURÔNIOS NA CAMADA OCULTA APÓS CONVERGÊNCIA	117
FIGURA 10.32 - A DIVISÃO DO ESPAÇO DE CORES RGB PARA O CASO DO TREINAMENTO EM DIFERENTES INTENSIDADES, USANDO REDE DE 20 NEURÔNIOS E 45 EXEMPLOS.....	117
FIGURA 10.33 - RESPOSTA À VARIAÇÃO DO NÚMERO DE NEURÔNIOS NA CAMADA OCULTA PARA 45 EXEMPLOS.....	118
FIGURA 10.34 - RESPOSTA À VARIAÇÃO DO NÚMERO DE EXEMPLOS PARA 11 NEURÔNIOS NA CAMADA OCULTA.....	118
FIGURA 10.35 - AS RESPOSTAS DA RNA TREINADA COM 45 EXEMPLOS SOB DIFERENTES LUMINOSIDADES COM 11 NEURÔNIOS NA CAMADA OCULTA	119
FIGURA 10.36 - ANÁLISE DA PARTIÇÃO DO ESPAÇO DE CORES YUV PARA O CASO DO TREINAMENTO EM INTENSIDADE LUMINOSA VARIADA.....	119
FIGURA 10.37 - RESPOSTA À VARIAÇÃO DO NÚMERO DE NEURÔNIOS NA CAMADA OCULTA COM 45 EXEMPLOS.....	120
FIGURA 10.38 - RESPOSTA À VARIAÇÃO DO NÚMERO DE EXEMPLOS COM 17 NEURÔNIOS NA CAMADA OCULTA.....	120
FIGURA 10.39 - RESPOSTA DA RNA TREINADA COM 45 EXEMPLOS EM DIFERENTES INTENSIDADES LUMINOSAS COM 18 NEURÔNIOS NA CAMADA OCULTA PARA O CASO CONTÍNUO.....	120
FIGURA 10.40 - RESPOSTA DAS RNAS TREINADAS EM DIFERENTES INTENSIDADES LUMINOSAS COM 17 NEURÔNIOS NA CAMADA OCULTA	121
FIGURA 10.41 - A DIVISÃO DO ESPAÇO DE CORES HSV PARA O CASO DO TREINAMENTO EM INTENSIDADE VARIADA, PARA UMA REDE COM 18 NEURÔNIOS NA CAMADA OCULTA E 45 EXEMPLOS.....	121
FIGURA 11.1 - PORCENTAGEM MÉDIA DE ACERTO DAS RNAS PARA CADA SISTEMA DE REPRESENTAÇÃO DE CORES PARA CADA FORMA DE TREINAMENTO.....	124

1. Introdução

A Inteligência artificial é *"(...) o estudo da computação que torne possível a percepção, raciocínio e ação"*.

WINSTON (1992).

Winston deixa claro em sua definição a percepção como um dos elementos fundamentais na tarefa de permitir interações entre os computadores e o mundo *real*, em contrapartida ao mundo idealizado (ou mundo fechado), outrora extensivamente utilizado como plataforma para muitos estudos da inteligência artificial (IA).

A *visão computacional* (VC) tem por objetivo a interpretação automática de imagens e constitui uma das principais formas de percepção. Uma importante sub tarefa em VC é a segmentação de imagens, ou, mais precisamente, a extração de parâmetros que permitam associar regiões da imagem com objetos da cena [BALLARD & BROWN, 1982]. Tal segmentação é realizada com base nos diversos atributos e características da imagem. Tipicamente, faz-se uso das características de descontinuidade e similaridade de regiões [GONZALES & WOOD, 1992].

As cores presentes na imagem são também atributos de grande importância em sua análise, visto que *"as sensações visuais causadas pela luz colorida são muito mais ricas que aquelas causadas pela luz acromática"* [FOLEY et. al., 1990, p.574]. A análise baseada em cores tem se mostrado importante em um sem número de aplicações [RAE & RITTER, 1998] [BIANCHI, 1998] [BIANCHI et. al., 1998] [JUNGWIRTH, 1998].

Contudo, as modelagens clássicas existentes na literatura para o tratamento de cores em imagens [BALLARD & BROWN, 1982] [MARR, 1982] têm se mostrado falhas em muitos aspectos, sobretudo naqueles onde as condições do ambiente - particularmente a iluminação - não são as ideais ou onde estas nem sempre permanecem estáveis, caso típico do futebol de robôs móveis autônomos [REALI COSTA et. al., 1999], ou dos robôs manipuladores [BIANCHI, 1998], projetos de interesse do grupo de trabalho do Laboratório de Técnicas Inteligentes do Departamento de Engenharia de Computação e Sistemas Digitais da Escola Politécnica da Universidade de São Paulo.

Nesse sentido, a busca por algoritmos capazes de interpretar cores corretamente sob diferentes iluminações tem levado muitos pesquisadores à aplicação de diferentes técnicas e abordagens. Na busca de novas abordagens, as *redes neurais artificiais (RNAs)* surgem como uma alternativa às técnicas clássicas e têm sido extensivamente utilizadas em situações nas quais a modelagem do problema é difícil ou praticamente impossível. Seu inerente poder de generalização tem possibilitado sua utilização nas mais diversas situações, tais como o reconhecimento de padrões, classificação, controle, entre outros.

1.1 O problema da segmentação baseada em cores

O conceito da palavra "cor", embora aparentemente imediato, envolve muitas noções de ordem física e psicológica, o que dificulta enormemente sua modelagem.

A teoria da ondulatória explica a luz através de seu modelo dual, ou seja, os *fótons* (partículas que compõe a luz) em movimento determinam uma onda cuja intensidade em cada ponto é igual à probabilidade de se encontrar o fóton nesse ponto. Dessa maneira a luz se comporta ora como onda eletromagnética e ora como partícula. Entende-se por cor [GOMES & VELHO, 1994] o estímulo percebido pelo cérebro transportado na forma de impulsos elétricos gerados pela retina humana quando atingida por fótons.

Embora os comprimentos de onda da faixa de luz visível sejam grandezas contínuas ao longo de toda a faixa do espectro, quando um ser humano observa tais comprimentos de onda, estes são imediatamente agrupados em classes, fazendo uso de noções psicológicas e exemplos prévios de cada indivíduo na separação entre cada uma das cores visíveis, ou, na forma como é usualmente colocado, na *classificação de cores*.

Transportando tal problema ao universo computacional, o processo de classificação de cores pode ser visto como a aplicação de funções altamente não-lineares a uma função de intensidades luminosas sobre os diversos *pixels* que compõem uma imagem digital, visando adequá-los a parâmetros conhecidos (classes).

Nas aplicações que objetivam implementar tal classificação, é tarefa inicial da VC separar algumas cores - que representem objetos de interesse nas imagens amostradas - das demais cores, que trataremos simplesmente como *fundo* da imagem. Contudo, muitas

questões permanecem abertas no trato com cores, tais como a forma de representação das cores e as ferramentas ideais capazes de promover a separação desejada.

1.2 Objetivos

São os objetivos da presente dissertação a *modelagem e implementação* de um sistema de classificação de cores baseado em redes neurais artificiais (RNAs), além da *análise* de sua atuação em relação à robustez quanto a variações na iluminação do ambiente e na saturação das cores apresentadas.

1.3 Motivação e resultados esperados

A presente dissertação foi motivada pelas dificuldades encontradas pelo grupo de trabalho com a plataforma do futebol de robôs móveis autônomos [REALI COSTA et. al., 1999]. Nessa plataforma - detalhada ao longo do trabalho - é realizada a segmentação de imagens utilizando o atributo cor. Embora nesse ambiente as cores relevantes sejam conhecidas *a priori*, grandezas como a intensidade e a tonalidade das cores envolvidas são totalmente ignoradas. Sob esta óptica, a segmentação pode ser vista como a classificação dos pixels presentes na imagem como pertencentes a uma das classes (cores) envolvidas.

Uma vez que as técnicas clássicas não se mostram adequadas para o problema proposto, buscamos investigar a utilização das RNAs como classificadoras para este domínio, visto que estas apresentam algumas vantagens com relação a outras ferramentas, tais como:

- i) capacidade de aproximar qualquer função contínua [HECHT-NIELSEN, 1990];
- ii) capacidade de aprender através de exemplos;
- iii) capacidade de generalização para situações em que a rede não foi treinada.

Assim, as RNAs apresentam-se como uma ferramenta bastante propícia ao trabalho proposto. Em nossa aplicação, seguindo algumas das tendências atuais no trabalho com cores [JUNGWIRTH, 1998] [LITTMAN & RITTER, 1998], buscamos utilizar um classificador por RNAs como uma camada responsável pela segmentação da

imagem em um sistema de visão computacional, com o intuito de conseguir uma identificação de cores tão próxima ao comportamento humano quanto possível, superando as respostas de sistemas que se utilizam de técnicas clássicas.

Assim, uma vez que a RNA seja treinada para o reconhecimento das cores importantes no domínio adotado, espera-se que esta seja capaz de generalizar para situações adversas (entenda-se diferentes iluminações e saturações de cores) ou para situações não presentes na etapa de treinamento, proporcionando uma melhor classificação de cores.

1.4 Organização do trabalho

O presente trabalho está dividido da seguinte maneira: no capítulo 2 que se segue, os conceitos básicos referentes à imagem são apresentados, incluindo sua digitalização e visualização. O capítulo 3 é dedicado ao estudo das cores e sua problemática. A seguir, no capítulo 4, abordam-se as formas de segmentação de imagens baseada em cores, destacando as RNAs como uma abordagem interessante, que serão melhor discutidas no capítulo 5. No capítulo 6 o problema a ser tratado é revisitado segundo conceitos apresentados nos capítulos anteriores e, no capítulo 7, são descritos alguns trabalhos correlatos desenvolvidos. No capítulo 8 detalham-se os experimentos realizados e a metodologia empregada. No capítulo 9 apresentam-se os dados e avaliações dos resultados obtidos e no capítulo 10, os detalhes da implementação proposta. Finalmente, no capítulo 11 apresentamos as conclusões da presente dissertação.

2. Imagem

"A formação de uma imagem ocorre quando um sensor registra radiação que interagiu com um objeto físico".

BALLARD; BROWN (1982), p.17

Para a compreensão do que vem a representar a palavra "imagem", no sentido da afirmação de BALLARD; BROWN (1982), são necessários entendimentos de alguns conceitos relacionados a esta, tais como:

- i) o sistema visual humano;
- ii) a física da formação da imagem;
- iii) a modelagem matemática de uma imagem;
- iv) a modelagem matemática de uma imagem discreta.

Desta forma, o presente capítulo tem por objetivo a exposição dos conceitos básicos ligados à imagem no que diz respeito à sua visualização e modelagem, além de estabelecer a relação entre a imagem abstrata (no mundo físico) e a imagem computacional.

2.1 O sistema visual humano

O sistema visual humano é um mecanismo natural de interface com o ambiente. Visto que a Inteligência Artificial é o estudo da computação que torne possível - entre outras tarefas - a percepção, o sistema visual humano (assim como muitos outros sistemas biológicos) são uma contínua fonte de novos modelos, o que justifica cada vez mais o aprofundamento de seu estudo.

Dada a sua complexidade, no entanto, o sistema visual é *"o mais estudado e o menos entendido entre os sistemas sensórios"* [KOVÁCS, 1997, p.94]. Contudo, sabidamente, no caminho descrito pelo processamento da informação visual estão estruturas como o globo ocular, a retina e, diretamente, o sistema nervoso central (SNC).

Pode-se fazer a analogia do globo ocular (visto na figura 2.1) com uma câmara dotada de capacidade de focalização e com abertura controlável. Tal estrutura tem por objetivo levar as imagens externas a regiões da retina fazendo uso de uma seqüência de movimentos biológicos, dentre eles os sacádicos [GREGORY, 1990]. O estudo de tais movimentos [FERREIRA & LISBERGER, 1997] [BRUCE & GOLDEMBERG, 1985] [BOMN & LUCCHETTI, 1997] [HONDA, 1997] [FERREIRA & LISBERGER, 1995] e sua simulação computacional [KINSER, 1999] têm se tornado grandes atrativos para a comunidade da IA.

Figura 2.1 - Representação esquemática do globo ocular e seus principais componentes [KOVÁCS, 1997, p.96].

A retina, segundo muitos autores, é definida como uma extensão do SNC responsável pela interpretação preliminar das imagens. Nela podem ser encontradas, além de algumas camadas de neurônios, duas estruturas principais: os *cones* e os *bastonetes*. Tais estruturas foram estudadas preliminarmente em 1866 por *Max Schultze*, que concluiu que os cones são as estruturas responsáveis pela visão colorida, enquanto que os bastonetes são responsáveis pela acuidade visual (ou visão noturna), isto é, responsáveis pela resolução espacial da imagem. Nos cones podem ser encontrados fotopigmentos que respondem aos espectros de luz definidos em três comprimentos de onda distintos: 445, 535 e 570 nm. Tais comprimentos de onda correspondem às cores vermelho, verde e azul. Um fato de grande relevância aqui é que a sensibilidade aos diferentes comprimentos de onda não é - fisicamente - idêntica. Tal fato pode ser observado na

figura 2.2, que mostra a fração percentual da luz absorvida pelos cones para os diferentes comprimentos de onda do espectro visível.

Embora os cones sejam encontrados em toda a retina humana, esta possui uma região espacialmente alinhada com o globo ocular denominada *fóvea* (vide figura 2.1), que é especialmente rica em cones, e capaz de fornecer, portanto, a máxima acuidade visual e resolução. No restante da retina, predomina a visão periférica, ou seja, trata-se de uma região dominada pelos bastonetes.

Figura 2.2 - Função de resposta espectral para cada um dos três tipos de cones da retina humana [FOLEY et. al., 1990, p.577].

As camadas neurais da retina, mencionadas anteriormente e apresentadas na figura 2.3, são classificadas como: *horizontais*, *amacrinas*, *bipolares* e *ganglionares*. Seu mecanismo de funcionamento ainda não é totalmente entendido. Assume-se que depois de capturada e pré-processada de alguma maneira pelas células retinais, a informação visual é levada a diferentes estruturas do SNC [GREGORY, 1990].

Nos trabalhos com cores, duas propriedades do mecanismo ocular - que envolvem a interpretação de cores - apresentam-se relevantes: **i)** a *adaptatividade ao brilho* e **ii)** a *adaptatividade à cor*.

A *adaptatividade ao brilho* é descrita por NOVAKOVSKY (1975). Segundo ele, superfícies com brilho acima do brilho médio do campo visual total apresentam-se como

brilhantes, e aquelas abaixo da média apresentam-se escuras. Em um processo similar, a *adaptatividade à cor* faz com que uma fonte de cor brilhante tenda a reduzir a sensibilidade dos receptores visuais a esta cor, e ampliar a sensibilidade dos receptores para os quais o estímulo é fraco. Exemplificando: se observarmos por algum tempo um objeto azul brilhante, e, em seguida um objeto *ciano* (cor entre o azul e o verde), esta nos parecerá verde, dado que a sensibilidade ao azul estará reduzida.

Figura 2.3 - As camadas celulares na retina humana [KOVÁCS, 1997].

2.2 Física da formação da imagem

O entendimento de como uma imagem se comporta aos olhos de um observador inevitavelmente passa pelo entendimento da física da formação desta imagem, que, por sua vez, nos leva à necessidade da descrição matemática desses sistemas físicos.

Os objetos do mundo físico são usualmente classificados - segundo a sua emissão de luz - como *emissivos* ou *radiantes* (aqueles que emanam algum fluxo de energia própria) ou como *reflexivos* ou *irradiantes* (aqueles que meramente refletem a luz emanada por outros corpos). Assim, seja uma fonte de luz distante incidindo sobre uma pequena parte de uma superfície, como o disposto na figura 2.4. Denomina-se *reflectância* da superfície a fração de um dado fluxo de energia incidente, refletido em uma certa direção [BALLARD & BROWN, 1982].

A *função de reflectância* é uma medida de como se comporta a energia emanada de um ponto, com relação ao fluxo incidente. Usualmente, esta é uma função dos três ângulos (i , e , g), com o ângulo incidente i formado pelo vetor de iluminação e o vetor normal da superfície, ângulo de emissão e formado pelo vetor normal N e vetor de observação O , e o ângulo de observação g , pelo vetor de iluminação e vetor de luz refletida (vetor de observação).

Figura 2.4 - A geometria da reflectância.

Os diversos tipos de superfícies terão, portanto, diferentes complexidades de função de reflectância. Superfícies *lambertianas* (puramente foscas) terão o mesmo brilho em todas as direções. Superfícies *especulares* (puramente reflexivas) terão reflexão na direção em que o ângulo de observação for igual ao ângulo de incidência, no plano formado pelos vetores envolvidos.

Em termos formais, podemos dizer que a radiância de um ponto será dada por:

$$L = \int r \cdot dE \quad (2.1)$$

onde:

L é a radiância da imagem
r é a função de reflectância
E é o fluxo incidente

A intensidade de luz refletida por um ponto $L(p)$ é, portanto, uma função do fluxo de luz incidente sobre tal elemento de área - $E(p)$ - e da função de reflectância da superfície (r). Ocorre que, em sistemas reais, a luz incidente é extremamente complexa de

ser modelada. Fontes de iluminação não pontuais, múltiplas iluminações ou mesmo reflectâncias de objetos que incidem sobre outros objetos podem complicar de modo considerável a análise matemática desses sistemas.

É importante destacar que, embora a modelagem de uma imagem tenha complexidade crescente na mesma proporção em que aumenta a complexidade da iluminação, o processo inverso (de recriação do modelo do mundo a partir de imagens), tarefa da visão computacional, torna-se proibitivamente complexo.

2.3 Modelagem matemática da imagem

Da modelagem matemática do mundo físico é sabido que a luz refletida de um ponto de uma superfície pode ser descrita como uma função da luz nele incidente e da reflectância desse ponto. Pretende-se agora estabelecer como tal função compõe o que venha a ser uma imagem.

Define-se como *função imagem* a abstração fundamental de uma imagem [BALLARD & BROWN, 1982], ou, em outros termos, sua representação matemática. Em geral, esta função toma como argumentos duas coordenadas espaciais e, definida sobre um ponto genérico $p(x,y)$, é dada por:

$$I\{p\} = f\{x,y\} \quad (2.2)$$

onde:

$I(p)$ é a função imagem para o ponto p; (x,y) é a coordenada espacial do ponto p; $f(.)$ é a função que mapeia o ponto p no universo de cor.	
---	--

Em termos de um mapeamento monoespectral, podemos assumir que $f(.)$ representa a função do ponto mapeando-o em um universo de tons de cinza, ou seja, trata-se do mapeamento das intensidades luminosas $L(p)$ vistas por um observador. Já o mapeamento colorido - mais complexo - será melhor abordado nos capítulos subsequentes.

2.4 Imagem digital

Um caso particular da função imagem é a *função imagem digital*. Nesta, os argumentos e o valor da função são sempre inteiros. Para realizar o processo de formação da imagem digital a partir da imagem contínua, utiliza-se a ferramenta matemática da *função delta* [BALLARD & BROWN, 1982] [GOMES & VELHO, 1994]. Tal função aplicada sobre um ponto p pode ser definida como:

$$\int_{-\infty}^{\infty} \delta(p) dp = 1 \quad (2.3)$$

$$\text{onde: } \delta(p) = \begin{cases} 0 & \text{quando } p \neq 0 \\ \text{não definido} & \text{quando } p = 0. \end{cases}$$

Com a aplicação desta ferramenta matemática nas dimensões usuais, ou seja, a multiplicação da imagem bidimensional pela função $\delta(x)$ também bidimensional (conhecida como "pente" e exemplificada na figura 2.5), pode-se obter um conjunto de amostras desta imagem, gerando uma imagem discreta. Destaque-se que neste caso $p \in \mathbb{Z}^2$.

A modelagem do mundo físico através da imagem discreta, como o proposto, possibilita uma forma finita de representação de grandezas contínuas. Contudo, devido ao processo de digitalização, diversos problemas são introduzidos no universo do mapeamento mundo-imagem, tais como *amostragem*, *quantização*, conceito de *distância* entre pontos, etc.

Figura 2.5 - A coleção de funções delta aplicada sobre uma imagem bidimensional.

2.4.1 Amostragem e quantização

Para ser conveniente para o processamento computacional, uma função imagem precisa ser digitalizada espacialmente e em amplitude [GONZALEZ & WOODS, 1992]. Nesse sentido, estabelecem-se dois conceitos da imagem digital: *amostragem* é a discretização do domínio da imagem (usualmente das coordenadas espaciais da imagem), e *quantização* é a digitalização das amplitudes (ou imagens) da função imagem [NALWA, 1993].

A *amostragem* da imagem envolve dois pontos de destaque: o *intervalo espacial de amostragem* e o *padrão espacial dos pontos de amostragem* [BALLARD & BROWN, 1982]. No primeiro tópico, usualmente o problema pode ser reduzido a uma curva de resolução da imagem *versus* custo computacional. Evidentemente, busca-se utilizar imagens que preservem a fidelidade da visão humana, com a menor resolução espacial possível. Claramente, intervalos de amostragem excessivamente grandes comprometerão o resultado final da imagem observada, levando ao problema computacional da subamostragem da imagem contínua [NALWA, 1993], fazendo com que uma superfície suave, tipicamente, apareça "dentada".

De forma a adotar um critério para o problema da amostragem, faz-se uso do *teorema da amostragem de Shannon* [SHANNON, 1948] que, em sua essência, define como deve ser uma amostragem mínima para garantir a representatividade da imagem. Em muitos casos tal problema leva à utilização de amostragens não-homogêneas no campo visual, similar ao encontrado na região da fóvea, na retina, discutida anteriormente.

Passemos agora à abordagem do segundo tópico: o *padrão espacial dos pontos de amostragem*. Uma vez que no plano da imagem digital as amostragens são representadas por pontos, podemos encontrar na literatura a definição da menor unidade de particionamento da imagem: o *pixel* (composição de *picture element*). Com relação à forma do pixel, define-se um *padrão espacial* no qual o plano da imagem é definido. Tal padrão define um conjunto finito de células sobre as quais o nível de cor é constante. Embora o formato mais comum do gradeamento seja o retangular, outros padrões podem ser assumidos, tais como triangular ou hexagonal.

Já a *quantização* envolve a problemática de com quantos níveis de intensidade uma dada imagem será mapeada. Novamente, a discretização impõe um compromisso entre a eficiência computacional dos sistemas e a visualização das imagens. Particularmente no domínio dos trabalhos com cores, uma baixa quantização implica na perda de informações importantes do mundo físico.

2.4.2 Geometria e conectividade

O processo de amostragem e representação de grandezas no mundo físico por meio de valores inteiros, invariavelmente, impõe alguma perda de informação. Essa problemática é particularmente visível quando tratamos do problema da subamostragem, já abordado. Todavia, por maior que seja a resolução da amostragem espacial, sempre existirá algum nível de subamostragem em uma imagem digital - implicando em perda da informação geométrica do universo amostrado - já que esta é representada, em última instância, por *pixels*, ou seja, entidades discretas. Tal fato é representado pela figura 2.6.

Uma das implicações deste problema (conhecido como "*paradoxo da conectividade*") diz respeito ao problema estrutural da adoção da vizinhança de um ponto. A distância entre dois pontos de uma imagem é de grande relevância para muitos algoritmos de visão computacional. Os medidas mais comuns para essa grandeza para dois pontos $x(x_1, y_1)$ e $y(x_2, y_2)$ são apresentadas na tabela 2.1.

Figura 2.6 - Uma reta (grandeza contínua) e sua representação em um domínio discreto.

Nomenclatura	Distância
Euclidiana	$d\{x,y\} = \sqrt{\{x_1 - x_2\}^2 + \{y_1 - y_2\}^2}$ (2.4)
Quarteirão (ou D4)	$d\{x,y\} = x_1 - x_2 + y_1 - y_2 $ (2.5)
Tabuleiro de xadrez (ou D8)	$d\{x,y\} = \max\{ x_1 - x_2 , y_1 - y_2 \}$ (2.6)

Tabela 2.1 - As várias medidas de distância entre pontos em uma imagem digital.

Tais distâncias podem ser visualizadas abaixo, na figura 2.7:

Figura 2.7 - Contornos eqüidistantes para diferentes métricas (a) Euclidiana (com aproximação para inteiro de número decimal) (b) D4 (c) D8 (como em [BALLARD & BROWN, 1982,p.40]).

Definidos tais padrões, a adoção de um deles (ou de outros) permanece a cargo do algoritmo envolvido, isto é, do padrão que maximiza sua performance.

2.5 Revisitando o conceito de imagem

Uma vez entendido o processo de formação e digitalização da imagem, utilizaremos o termo "*imagem*" no decorrer do presente texto, referindo-se a *uma função bidimensional de intensidades luminosas*, particularmente tratando-se de intensidades luminosas coloridas no espaço discreto.

Entendido o processo básico de formação da imagem, faz-se necessário no contexto da segmentação de imagens baseada em cores - proposição deste trabalho - uma melhor compreensão do termo "cor" e sua problemática, visto que, ao contrário do que se possa pensar, não se trata de um conceito imediato. Tal assunto é o alvo do capítulo que se segue.

3. Cores

"As sensações visuais causadas pela luz colorida são muito mais ricas que aquelas causadas pela luz acromática".

[FOLEY et. al, 1990, p.574].

Historicamente, o trabalho com cores tem se mostrado de difícil modelagem. Isso decorre do fato de que entendemos como cor não ser um atributo físico, mas uma sensação. Tal problemática nos leva à necessidade de um estudo mais aprofundado nesse conceito cognitivo e também nos aspectos eletromagnéticos da luz.

Buscamos, neste capítulo, portanto, apresentar: **i)** as definições e dependências do conceito de cor, **ii)** as formas de se medir a grandeza cor, **iii)** as representações de cores, **iv)** os sistemas mais relevantes para a representação de cores. Por fim explanaremos a respeito dos dispositivos de interface homem-máquina.

3.1 Definições e dependências

Da teoria da ondulatória é sabido que um raio luminoso é constituído por partículas chamadas *fótons*. A frequência e velocidade do movimento dos fótons determinam o comprimento de uma onda eletromagnética que emana dessas partículas, conhecida genericamente como luz [GOMES & VELHO, 1994]. Matematicamente, pode-se definir o comprimento de uma onda como:

$$\lambda = \frac{v}{f} \quad (3.1)$$

onde:	λ é o comprimento da onda; v é a velocidade do fóton; f é a frequência de oscilação do fóton.
-------	---

O olho humano é um mecanismo *físico* de recepção da luz. Tipicamente, este excita-se com radiações eletromagnéticas entre 380 nm a 780 nm, espectro conhecido como *luz visível*. Já a luz, por sua vez, possui uma composição que depende de propriedades da radiação (ou irradiância) de um corpo, já discutidas anteriormente.

A radiação composta por um único comprimento de onda (ou de uma faixa bem estreita) é classificada como *luz monocromática* (ou espectral). As luzes monocromáticas captadas pelo olho humano são mostradas na tabela 3.1. As demais cores podem ser produzidas pela somatória de ondas com diferentes comprimentos, não se tratando, portanto, de luzes monocromáticas - exceção ao preto (ausência de cor).

Comprimento de onda (λ)	Cor
780 a 605 nm	Vermelho
605 a 590 nm	Laranja
590 a 560 nm	Amarelo
560 a 500 nm	Verde
500 a 470 nm	Ciano
470 a 430 nm	Azul
430 a 380 nm	Violeta

Tabela 3.1 - Luzes monocromáticas.

Contudo, segundo NOVAKOVSKY (1975), a percepção da cor pelos seres humanos dá-se não somente por fatores *físicos*, mas também *psicológicos*. As células sensíveis do olho são capazes de identificar a composição espectral das luzes nele incidentes, transmitindo-nos - após a interação dessa amostragem com o cérebro - a sensação que conhecemos como cor. A passagem da informação visual pelo SNC faz com que uma função de ordem cognitiva passe a atuar sobre a amostragem da onda eletromagnética realizada pelo sistema visual. "*É a presença ou ausência da luz que nos dá a sensação de cor. A luz é um fenômeno físico, mas a cor depende da interação da luz com o sistema visual, sendo portanto, um fenômeno psicofísico*" [GOMES & VELHO, 1994, p.63].

Sejam, por exemplo, as ondas eletromagnéticas monocromáticas visíveis e as ondas produzidas pelas somatórias destas. Observando seu espectro de cores - por exemplo - não são estabelecidos com exatidão os limites para todas as cores perceptíveis,

ou seja, envolvem-se noções psicológicas e exemplos prévios de cada indivíduo na rotulação de cor às diferentes ondas visíveis, isto é, na *classificação* das cores. Assim podemos entender que, uma vez observada pelos mecanismos sensórios, a forma de onda é classificada pelo cérebro segundo algum critério de ordem *psicológica*, ou seja, de forma subjetiva.

Para a proposição de modelagens matemáticas para tal classificação, então, faz-se necessário um maior domínio dos conceitos físicos de uma cor e sua amostragem pelo sistema sensório humano, alvos da próxima seção.

3.2 Grandezas fotométricas

Uma radiação eletromagnética genérica pode ser quantificada fazendo-se uso de uma coleção de grandezas, denominadas *radiométricas*. Usualmente - no domínio dos trabalhos com cores - uma onda é referenciada em termos de sua potência (ou fluxo de energia), isto é, uma medida que é função da intensidade da onda em questão. A partir deste conceito básico, seguem-se algumas medidas usuais, apresentadas na tabela 3.2.

Grandeza	Unidade	Descrição
Unidades Básicas:		
Energia	J	Quantidade de radiação da onda.
Potência	W	Fluxo de energia por unidade de tempo (J/seg).
Superfícies Emissivas:		
Emitância (ou excitância)	W/m ²	Fluxo de energia emitida por unidade de tempo por unidade de área.
Energia radiante	W/srad	Fluxo de energia emitida por unidade de tempo em uma dada direção
Radiância	W/m ² .srad	Fluxo de energia emitida por unidade de tempo por unidade de área em uma dada direção.
Superfícies reflexivas:		
Irradiância	W/m ²	Fluxo de energia refletida de um ponto da superfície por unidade de área.

Tabela 3.2 - As grandezas radiométricas, suas unidades usuais e definições.

Embora tais grandezas possam ser utilizadas genericamente com qualquer radiação, estas não são particularmente adotadas onde existe interação humana. Isso decorre do fato do olho humano não possuir igual sensibilidade a todos os diferentes comprimentos de onda das radiações. A curva apresentada na figura 3.1 é denominada *função de luminosidade para o observador padrão*. Esta curva estabelece a luminosidade relativa (V_λ) - isto é, a sensação de brilho - aos diferentes comprimentos de onda por um indivíduo considerado normal, ou seja, sem qualquer tipo de anomalia visual. Segundo FOLEY et. al. (1990), existem evidências de que esta curva (obtida experimentalmente) seja apenas a soma das três curvas de sensibilidade dos cones humanos, mostradas na figura 2.2.

Figura 3.1 - Função de luminosidade (V_λ) de um observador padrão aos diferentes comprimentos de luz visível [NOVAKOVSKY, 1975].

Uma vez observada tal curva, pode-se perceber que o comprimento de onda que corresponde aproximadamente à cor verde-amarelada (555 nm) é fortemente captada pelo olho humano, enquanto que cores como o vermelho e azul sofrem atenuações. Assim, torna-se claro que as unidades *radiométricas* anteriormente adotadas não terão validade no trato de sistemas de radiação visível observados pelo olho humano. Nesse sentido, estabeleceram-se novas unidades de quantificação, denominadas *fotométricas*.

Tais grandezas fazem uso das medidas anteriormente apresentadas, com uma adequação à curva do observador padrão. Assim, tomando-se por exemplo uma cena que recebe uma certa iluminação do ambiente - situação que nos interessa mais diretamente - é possível demonstrar [NOVAKOVSKY, 1975] [GOMES & VELHO, 1994] que, para a interpretação humana, da irradiância de um ponto, denominada *iluminância*, para uma radiação distribuída continuamente pelo espectro de frequência dá-se por:

$$L = 683 \cdot \int_{\lambda_1}^{\lambda_2} V_{\lambda} I_{\lambda} d\lambda \quad (3.2)$$

onde:

L é a iluminância;
 I_{λ} é a irradiância da superfície;
 V_{λ} é a função de sensibilidade do observador padrão;
 λ_1 é o menor comprimento de onda do espectro incidente;
 λ_2 é o maior comprimento de onda do espectro incidente;
683 constante.

Utilizando a mesma metodologia para os demais conceitos apresentados na tabela 3.2, é possível estabelecer uma correlação entre grandezas fotométricas e grandezas radiométricas, que é mostrada na tabela 3.3.

Características Radiométricas		Características Fotométricas	
Grandeza	Unidade	Grandeza	Unidade
Fluxo de energia	Watt	Fluxo luminoso	lm (Lumen)
Emitância radiante	W/m ²	Emitância luminosa	lm/m ² (Lux)
Irradiância	W/m ²	Iluminância	lm/m ² (Lux)
Intensidade radiante	W/srad	Intensidade luminosa	lm/srad (Candela)
Radiância	W/m ² .srad	Luminância	lm/m ² .srad

Tabela 3.3 - Grandezas radiométricas e fotométricas e suas unidades usuais [NOVAKOVSKY, 1975].

3.3 Representação de cores

A necessidade humana de manipulação das cores levou à necessidade de representá-las. Entretanto, uma vez que a cor é uma atribuição subjetiva, a definição de grandezas que a representem não é uma tarefa simples. Dessa forma, as cores têm sido descritas na literatura [FOLEY et. al., 1990] [NOVAKOVSKY, 1975] [GONZALEZ & WOODS, 1990] [GOMES & VELHO, 1994] de duas formas distintas: i) com *medidas*

perceptíveis, isto é, medidas de caráter mais informal utilizadas com o intuito de descrever as cores de forma geral e **ii)** com *medidas físicas*, ou seja, medidas científicas que buscam embasar sua representação com a chamada ciência da colorimetria.

3.3.1 Medidas perceptíveis

As *medidas perceptíveis* procuram descrever atributos genéricos relativos à forma da onda em questão. São elas: *luminância*, *saturação* e *tonalidade (hue)*.

A *luminância* é uma medida do nível da quantidade de energia percebida por um observador de uma fonte de luz. Na prática, esta grandeza está ligada ao conceito genericamente conhecido como "brilho" da cor.

A *tonalidade* da cor é uma comparação (subjetiva) entre a cor em observação e exemplos previamente conhecidos. De maneira genérica, trata-se da identificação da cor (ou comprimento de onda) dominante, isto é, a determinação de uma luz monocromática que corresponda (aproximadamente) à cor observada.

A *saturação* refere-se ao grau de pureza de uma cor, ou seja, alguma graduação que represente uma cor qualquer como uma grandeza entre uma onda monocromática (espectro de frequência com um único fasor na frequência da onda) e a luz branca (que contém todas as cores). Assim, as cores monocromáticas têm saturação máxima e a luz branca, mínima (zero).

As grandezas de tonalidade e saturação juntas correspondem à grandeza conhecida como *crominância*. Em termos simples, a crominância pode ser definida como a parte da cor que é invariante ao brilho.

É importante se destacar que a grandeza conhecida como "brilho" é de caráter cognitivo, e, portanto, não possui definição uniforme na literatura. Alguns autores admitem que "*o brilho é um descritor subjetivo e é praticamente impossível de ser medido*" [GONZALES & WOODS, 1992, p.222]. FOLEY et. al. (1990) assume que o brilho é uma grandeza que está ligada à luminância de objetos emissivos. Já segundo GOMES; VELHO (1994), os modelos de interface introduzem uma grandeza que varia monotonicamente com a luminância, que é chamada genericamente de brilho da cor. Contudo, há uma concordância em se tratar de uma grandeza subjetiva, que está ligada de alguma forma ao conceito de luminância da imagem. No escopo do presente trabalho,

adotaremos este conceito como designado à palavra "brilho", utilizando-a tanto para objetos emissivos quanto para reflexivos.

3.3.2 Medidas físicas

Todas as cores visíveis podem ser obtidas pela soma de ondas monocromáticas do espectro. Assim, uma notação intuitiva para representação de cores está ligada à somatória dessas ondas elementares (espectrais) para a produção das ondas desejadas. No entanto, essa notação incorre em somatórias de quantidades variáveis de comprimentos de onda. Seria desejável, então, alguma notação onde o número de componentes fosse fixo.

Em um primeiro momento, poderia-se pensar na representação de ondas como funções de algumas ondas monocromáticas adotadas *a priori*; no entanto, tal reconstrução não é fisicamente possível, já que sabidamente a superposição de ondas com frequências diferentes sempre produz uma onda complexa (não-senoidal) [RESNIC & HALLIDAY, 1991]. Em outras palavras, a somatória de duas ondas puras sempre produz uma onda (no caso uma cor) menos saturada que uma onda monocromática, não sendo portanto possível reproduzir a maioria das ondas monocromáticas - e as cores que são combinações destas - dessa forma.

Contudo, a reconstrução física da onda não é a única forma de se transmitir a sensação de cor. Sucessivos estudos de Isaac Newton (1666), H. Grassman (1853) e Maxwell (1860) provaram que, para conseguir a *equivalência de uma cor* é condição necessária e suficiente proceder à soma de três ondas de comprimentos distintos, tais que duas delas não produzam a terceira. Desta forma, a equivalência de uma cor fica associada à reprodução dos estímulos observados nos cones humanos [NOVAKOVSKY, 1975].

Assim, desenvolveu-se o paradigma da *representação tricromática*. Nesse sistema, então, podemos ver as ondas primárias como três eixos ortogonais de um sistema de coordenadas, onde as demais cores podem ser representadas como uma tripla, ou seja, uma combinação das três cores dos eixos.

Alguns outros aspectos seguem desse paradigma. Consideremos um espaço tridimensional cujos eixos são três cores arbitrárias tomadas como referência X, Y e Z.

Denomina-se *plano de cromaticidade* (ou *plano de Maxwell*) o plano de equação $X+Y+Z=1$, conforme o apresentado na figura 3.2.

Figura 3.2 - O plano de Maxwell e o *spectrum locus* [NOVAKOVSKY, 1975].

A grande particularidade do plano de Maxwell é a possibilidade de representação de qualquer cor C do espaço de cores visíveis com uma projeção C' sobre ele (usualmente conhecida como cromaticidade) e uma grandeza escalar $(C-C')$ representativa da intensidade luminosa da cor. Em outros termos, o plano de Maxwell é um plano *padrão* de representação de cores.

As cores espectrais podem ser representadas em função das cores escolhidas como eixos do sistema e estarão representadas no plano de Maxwell por pontos em uma curva denominada *spectrum locus* (local do espectro), que usualmente se apresenta em forma de uma ferradura. Interna à curva, apresentar-se-ão as possíveis combinações das ondas monocromáticas, resultando em um conjunto de todas as cores possíveis, a menos da variação de luminosidade, ou seja, em termos de cromaticidade.

Uma outra particularidade do plano de Maxwell é a redundância de uma variável. Seja uma cor C (C_1, C_2, C_3) expressa nesse sistema. As coordenadas do ponto C' (C'_1, C'_2, C'_3) - projeção do ponto C sobre o plano de Maxwell - podem ser expressas por:

$$C'1 = \frac{C1}{\{C1+C2+C3\}} \quad ; \quad C'2 = \frac{C2}{\{C1+C2+C3\}} \quad ; \quad C'3 = \frac{C3}{\{C1+C2+C3\}} \quad (3.3)$$

Uma vez que o ponto C' encontra-se sobre o plano de Maxwell, uma de suas componentes pode ser expressa em termos das outras duas. Por exemplo:

$$C'3 = 1 - C'1 - C'2 \quad (3.4)$$

Nesse sentido, usualmente, adota-se na literatura a projeção do triângulo de Maxwell e do *spectrum locus* sobre algum dos planos laterais (usualmente o r-g), gráfico que é denominado *diagrama de cromaticidade*. É importante observar que a posição do triângulo de Maxwell e o seu formato projetado (e portanto o formato do *spectrum locus*) dependerão da escolha das cores representadas nos eixos ortogonais.

Uma vez que a escolha das cores dos eixos não é uma tarefa trivial, diversos padrões foram propostos, sendo os principais apresentados nas seções seguintes.

3.4 Padronizações de cores

O órgão padronizador internacional na área de colorimetria é o CIE - Comissão Internacional de Iluminação (*Commission Internationale de l'Eclairage*). A maioria de suas padronizações foi estabelecida na década de 30 e permanece até hoje (tais como o RGB, XYZ e YUV). Contudo, alguns outros padrões estabeleceram-se mundialmente no trabalho com cores mesmo sem as padronizações do CIE (caso do HSI), adotando até mesmo outras formas de representação que não a tricromática. Alguns padrões serão apresentados a seguir.

3.3.1 CIE-RGB

Uma vez que o sistema visual humano possui células capazes de detectar comprimentos de onda de três tamanhos distintos (vermelho, verde e azul), nas faixas

alta, média e baixa do espectro - respectivamente - é imediata a utilização de comprimentos de onda em faixas similares para a composição da tripla de representação. Embora com o passar dos anos os comprimentos de onda adotados tenham se alterado consideravelmente desde a proposição inicial de Maxwell (1860), atualmente definem-se os comprimentos de onda para o sistema RGB como os mostrados na tabela 3.4. Observe-se que não se tratam rigorosamente dos mesmos comprimentos de onda observados pelos cones humanos (570, 535 e 445 nm).

Comprimento de onda (λ)	Cor
700 nm	Vermelho
546 nm	Verde
435,8 nm	Azul

Tabela 3.4 - Comprimentos de onda dos picos das curvas de sensibilidade das cores para o sistema CIE-RGB.

Definidas as cores primárias, o *diagrama de cromaticidade r-g* do sistema RGB pode ser visto na figura 3.3.

Uma vez que agora estamos tratando de um sistema descrito apenas em termos de três cores, a determinação da luminância deste não mais requer uma integração de todo o espectro de cores sobre a curva do observador padrão, mas sim a observação das potências aparentes relativas a esses três comprimentos de onda. Então, em termos de sua variação de luminância, é possível demonstrar [NOVAKOVSKY, 1975] [GOMES & VELHO, 1994] que o sistema CIE-RGB - para as primárias adotadas - pode ser descrito como:

$$L = 0,177R + 0,813G + 0,011B \quad (3.5)$$

Podemos dizer, então, que a variação da luminosidade do sistema RGB dá-se no sentido da reta desenhada pela equação, não coincidindo portanto com nenhum dos eixos do sistema.

Figura 3.3 - Diagrama de cromaticidade r-g [GOMES & VELHO, 1994].

Nesse ambiente, destacamos alguns pontos importantes da representação RGB [GOMES & VELHO, 1994]:

- i) nem todas as cores do espectro visível estão representadas no espaço positivo do sistema, ou seja, algumas cores podem requerer quantidades negativas de alguma primária (como pode ser visto na figura 3.3);
- ii) grandezas fotométricas de interesse (como a luminância) não são de obtenção imediata, necessitando do cálculo da combinação linear (eq. 3.5) ou da integração sobre a curva do observador padrão;
- iii) a figura resultante não faz uma utilização ótima do espaço, acarretando perda de precisão e homogeneidade na distinção de cores.

3.3.2 CIE-XYZ

Dados os problemas apresentados na representação anterior, a CIE definiu, em 1931, o sistema XYZ. As três primárias desse sistema foram escolhidas tendo-se em vista a utilização de coordenadas somente positivas e buscando um melhor aproveitamento do espaço de cores. Dessa forma, adotaram-se cores ótimas nesse sentido que, todavia, não correspondem a cores do espectro visível. Tomemos o plano de luminância nula:

$$0,177R + 0,813G + 0,011B = 0 \quad (3.6)$$

Esse plano intercepta o plano de Maxwell em uma reta, conhecida como reta de luminância nula (pode-se obter a projeção dessa reta no plano r - g fazendo-se $B=1-R-G$). De maneira a fazer com que apenas uma das componentes (Y) tenha influência da luminância, adotam-se duas das três primárias (X e Z) sobre a reta de luminância nula. A reta YX é definida de forma a se obter o maior número possível de cores espectrais com algumas coordenadas de cromaticidade nulas. A última reta do triângulo é obtida de forma a minimizar a área do triângulo, visto na figura 3.4.

O novo triângulo adotado margeia o diagrama RGB, mantendo internas a si todas as cores daquele sistema, ocasionando que todas as cores sejam compostas de adições entre as primitivas. Da forma como foi definido, o eixo Y é o único que carregará informação de luminância, já que os demais eixos (cores não visíveis) não contribuem com nenhuma luminância.

Figura 3.4 - Os eixos XYZ sobre o diagrama RGB [NOVAKOVSKY, 1975].

As coordenadas de cromaticidade adotadas, então, no sistema XYZ são mostradas na tabela 3.5.

<i>Cor</i>	<i>r</i>	<i>g</i>	<i>b</i>
X	1,2750	-0,2778	0,0028
Y	-1,7393	2,7673	-0,0280
Z	-0,7431	0,1409	1,6022

Tabela 3.5 - Coordenadas de cromaticidade do sistema XYZ [NOVAKOVSKY, 1975].

O novo diagrama de cromaticidade, conhecido como diagrama de cromaticidade CIE-XYZ é apresentado na figura 3.5. Como pode ser visto nesta figura, as cores no diagrama de cromaticidade estão melhor distribuídas ao longo do espaço de cores do que no sistema RGB.

Figura 3.5 - Diagrama de cromaticidade do padrão CIE-XYZ [GOMES & VELHO, 1994].

3.3.3 CIE-YUV

Em sistemas televisivos, particularmente, é usual a utilização da decomposição cromaticidade-luminância no espaço de cores. Um sistema usual nesse contexto é o sistema YUV. Seja um eixo Y representativo da luminância (definida em termos das componentes RGB na equação 3.5, pg. 24). Um modo simples de obter a decomposição cromaticidade-luminância é proceder a subtração da componente Y das outras componentes do sistema, obtendo (R-Y), (G-Y) e (B-Y). Uma vez que a quantidade de informação da componente verde é grande no eixo da luminância, adotaram-se os eixos (R-Y) e (B-Y), dando origem aos eixos U e V, a menos de alguma transformação linear que visa uma melhor distribuição dessas coordenadas pelo espaço. Assim, usualmente:

$$U = 0,494 \cdot \{B - Y\} \quad (3.7)$$

$$V = 0,877 \cdot \{R - Y\} \quad (3.8)$$

Os eixos U e V são conhecidos respectivamente como os eixos azul-vermelho e magenta-verde. Uma observação de grande relevância no sistema YUV é que as componentes primárias de cor utilizadas não são rigorosamente as mesmas que as do sistema CIE-RGB. Tal fato decorre da necessidade de compensação da não-linearidade dos monitores. Nesses casos, o cálculo da luminância do sistema é alterada, e passa a ser dado por [GOMES & VELHO, 1994]:

$$Y = 0,299R + 0,587G + 0,114B \quad (3.9)$$

Onde R, G e B representam respectivamente as cores primárias para o sistema de vídeo.

3.3.4 A família HS

Embora o modelo RGB seja bastante utilizado computacionalmente, a determinação ou classificação de algumas cores nesse sistema por parte dos usuários é uma tarefa absolutamente não trivial. Para sanar esse problema, define-se uma nova forma de representação que busca aliar o espaço de cores a coordenadas de discernimento mais intuitivo por parte do usuário, utilizando-se das grandezas perceptíveis da cor [GOMES & VELHO, 1994].

A família de modelos HS apresenta três coordenadas, não se tratando, no entanto, de uma representação ortogonal. Os diversos comprimentos de onda dispõem-se na forma de um círculo, que tem sua saturação variando ao longo do raio do círculo. Assim, a dupla tonalidade (*hue*) - usualmente um ângulo - e saturação (*saturation*) - usualmente um escalar - é utilizada de modo a descrever o plano da crominância. A trinca é completada com uma medida da iluminação do sistema, ortogonal ao plano da crominância. Uma vez que tal medida não é de consenso, dá-se origem a toda uma família de padrões.

Assim, seja uma cor C de coordenadas (R, G, B). O sistema HSV (*hue-saturation-value*), datado de 1978, por exemplo, define a componente ortogonal como sendo o valor (*value*) da cor V(C), como:

$$V(C) = \max \{R, G, B\} \quad (3.10)$$

O sistema HSL (*hue-saturation-lightness*), de 1979, define a luminosidade (*lightness*) da cor L(C) como:

$$L\{C\} = \frac{\max\{R, G, B\} + \min\{R, G, B\}}{2} \quad (3.11)$$

O sistema HSI (*hue-saturation-intensity*), de 1975, define a intensidade (*intensity*) da cor I(C) como:

$$I\{C\} = R + G + B \quad (3.12)$$

Independente da coordenada da iluminação do sistema, as demais coordenadas (HS) são definidas da mesma forma para toda a família. BALLARD; BROWN (1982) apresentam as conversões entre a família de sistemas HS e o RGB:

$$S = 1 - \left(\frac{3 \cdot \min\{R, G, B\}}{\text{intensidade}} \right) \quad (3.13)$$

$$H = \cos^{-1} \left(\frac{\frac{1}{2} [\{R - G\} + \{R - B\}]}{\{\sqrt{R - G}\}^2 + \{R - B\} \{G - B\}^{\frac{1}{2}}} \right) \quad (3.14)$$

Uma vez que H é uma medida de ângulo, observe-se que, caso $B > G$: $H = 2\pi - H$. O sólido de cores formado por este espaço é mostrado na figura 3.6.

Figura 3.6 - O sólido de cor do sistema HS [GOMES & VELHO, 1994].

3.5 Interfaces

Uma vez analisados os aspectos básicos dos trabalhos com imagens abstratas e digitais, além da temática das cores, trataremos, na presente seção, da classe de dispositivos digitais de aquisição e representação de imagens, que realizam o interfaceamento com o mundo físico.

3.5.1 Dispositivos de aquisição

Segundo GONZALES; WOODS (1992), dois elementos são necessários e indispensáveis na *aquisição* de imagens digitais por qualquer dispositivo: **i)** um elemento físico que seja sensível às diferentes bandas de comprimento de onda eletromagnética (tipicamente da luz visível, mas, genericamente, qualquer faixa de onda), responsável pela amostragem do universo e a conversão dessas amostragens em impulsos elétricos e **ii)** um dispositivo digitalizador, responsável pela conversão do impulso elétrico para quantificações digitais.

Os processos pelos quais as intensidades são capturadas, entretanto, podem variar consideravelmente. Em sistemas de *raio-x*, por exemplo, entre um canhão emissor de raios e um conjunto de receptores sensíveis a tal raio, interpõe-se um alvo, constituído de material capaz de absorver parcialmente os raios-x. Nos receptores, obter-se-á um mapeamento da absorção dos raios pelo alvo. Tal sistema é um exemplo típico de elementos sensíveis a ondas não-visíveis.

A *câmera de vídeo* - que nos interessa mais diretamente no presente contexto - é um dispositivo que amostra a luz visível tipicamente fazendo uso de elementos sensíveis a particulares comprimentos de onda. Tal dispositivo possibilita a utilização de diversas tecnologias. Uma delas - denominada *vidicon* - utiliza fototransistores para amostrar uma imagem. Outro tipo - que abrange praticamente a totalidade das câmeras atuais - faz uso da tecnologia conhecida como CCD, à qual dedicaremos algum estudo.

O CCD (*charge-coupled device* - dispositivo de dupla descarga) utiliza-se de uma coleção de elementos denominados fotossensíveis, que produzem uma tensão de saída - na forma de cargas armazenadas - proporcional à intensidade de algum comprimento de onda incidente. Trata-se de elementos com grande resolução e alta velocidade de aquisição (da ordem de 1/10.000 seg.) [GONZALEZ & WOODS, 1990].

É possível dispor os fotosensores de duas maneiras: **i)** em linha ou **ii)** em bloco. A arquitetura com captura em linha utiliza apenas uma coluna de tais elementos e produz uma imagem bidimensional pela variação da posição relativa entre a coluna e um objeto a ser amostrado. Tipicamente *scanners* (digitalizadores de imagem) utilizam-se desta tecnologia. Nesses sistemas, o incremento do movimento da coluna dá origem às diversas

resoluções possíveis. Na segunda arquitetura (em bloco ou matricial), toda uma matriz desses componentes é utilizada. Fazem uso dessa arquitetura as câmeras de vídeo.

Embora os elementos fotossensíveis tenham grande velocidade de aquisição de dados, a manipulação desses sinais analógicos até sua emissão na forma de um sinal de vídeo fazem com que tais dispositivos funcionem, tipicamente, na faixa de 1/30 seg.

Um outro aspecto - de suma importância - com relação às câmeras diz respeito a sua curva de sensibilidade. Idealmente, uma câmera que amostra o ambiente e o retorne em coordenadas RGB deverá possuir elementos sensíveis aos comprimentos de onda definidos pelo sistema CIE-RGB. Contudo, fisicamente, tal sensibilidade não é idêntica para todos os dispositivos. Isso implica que, para cada câmera, o cálculo - por exemplo - da luminância será particular e observado sobre suas próprias primárias.

3.5.2. Dispositivos de representação

Tradicionalmente, os monitores (de TV ou vídeo) constituem a principal classe de dispositivos representadores de imagem, e, portanto, de cores. O sistema de vídeo como um todo é constituído de quatro componentes: **i)** um monitor propriamente dito, **ii)** um controlador de vídeo, **iii)** uma memória de exibição e **iv)** um conversor digital-analógico.

O *monitor de vídeo* consiste de um tubo de raios catódicos, uma tela e um canhão que produz um feixe de elétrons. Em pontos da tela (ou mais precisamente a cada pixel) encontram-se camadas de fósforo, que, ao serem atingidos pelo feixe de elétrons, passam a emanar radiação eletromagnética visível por algum tempo. Usualmente, utilizam-se fósforos com as radiações correspondentes aos níveis vermelho, verde e azul (RGB). O *controlador de vídeo* é o dispositivo que tem por finalidade controlar o movimento do feixe de elétrons sobre a tela, produzindo uma varredura desta. O padrão de varredura é variável, dependendo do dispositivo [GOMES & VELHO, 1994]. A *memória de exibição* armazena dados da imagem digital, que, através do *conversor digital-analógico* são convertidos em tensões a serem aplicadas ao canhão de elétrons.

Devido às limitações físicas dos dispositivos de emissão, tais como os fósforos coloridos (que não são capazes de representar todas as cores do espectro visível), um monitor de vídeo terá seu espaço de cores limitado, ou seja: "(...) o sólido de cor desse sistema será um subconjunto limitado do espaço gerado pelas primárias R, G e B. Isso

porque cada cor primária do espaço de cor do monitor assume um valor de intensidade máxima" [GOMES & VELHO, 1994, p. 109].

Dessa forma, cada monitor terá o seu próprio espaço RGB que, por sua vez, será um novo subconjunto do sistema RGB padronizado pelo CIE. Graficamente, então, esse espaço pode ser expresso por um triângulo sobre o espaço de cromaticidade (figura 3.7).

Figura 3.7 - Diagrama representativo do triângulo de cromaticidade para o RGB de dois monitores distintos [GOMES & VELHO, 1994].

4. Segmentação

"Na visão computacional, o agrupamento de partes de uma imagem genérica em unidades que sejam homogêneas com respeito a uma ou mais características resulta em uma imagem segmentada".

BALLARD & BROWN (1982), p. 116.

Segundo BALLARD; BROWN (1982), a idéia da segmentação tem suas raízes nos estudos psicológicos de Gestalt, que concluiu que os seres humanos desenvolvem agrupamento visual baseado em características como proximidade, similaridade e continuidade. Aplicados à teoria da visão computacional, tais estudos resultam em um processo denominado segmentação, pelo qual entende-se a extração de parâmetros que permitam associar regiões da imagem com objetos na cena [BALLARD & BROWN, 1982]. Nesse sentido, busca-se particionar a imagem digital em regiões disjuntas, tais que um *pixel* pertença apenas a uma das regiões.

4.1 Métodos para segmentação

Segundo GONZALES; WOODS (1992) - em acordo com o proposto por BALLARD; BROWN (1982) "*os algoritmos para segmentação são usualmente baseados em duas propriedades das imagens: i) descontinuidade e ii) similaridade*" [GONZALEZ & WOODS, 1992, p.331].

Na *segmentação baseada na descontinuidade* a partição é baseada em alterações bruscas nos níveis da função imagem. Tais descontinuidades em uma imagem digital são decorrentes de diversas situações na cena como, por exemplo: descontinuidade da normal das superfícies, descontinuidade em profundidade, descontinuidade na reflectância da superfície e descontinuidade de iluminação.

Em muitos casos, esta informação é grandemente relevante para a segmentação de imagens. Para tanto, definem-se na literatura diversos operadores que, atuando sobre as imagens digitais, visam determinar as descontinuidades nelas presentes - usualmente

realizando operações entre pixels (tal como a diferença e derivada) - fornecendo as bordas de uma imagem a camadas superiores do sistema de VC, que visam determinar a conexão entre tais bordas, concluindo a etapa de segmentação.

Já a *segmentação baseada em similaridade* busca agrupar regiões com características semelhantes. Dentre as possíveis características utilizáveis, encontram-se a textura e as cores - característica que nos interessa mais diretamente - caso que, portanto, passaremos a abordar com maior profundidade.

No contexto da segmentação por similaridade, as abordagens tradicionais que fazem uso de cores dividem-se em dois grupos: **i)** as que fazem uso da informação topológica da imagem - tais como o *crescimento de região* e a *divisão e agrupamento de região (split and merging)* - e **ii)** as que não fazem uso da informação topológica, caso típico da *imposição de limiares*. Tais técnicas serão alvos das próximas seções.

4.2 A imposição de limiares

A técnica da imposição de limiares (*thresholds*) foi primeiramente adotada para a segmentação de imagens em tons de cinza e, posteriormente, generalizada para imagens multicromáticas. Tradicionalmente, faz uso da representação da imagem em um espaço de cores e da similaridade entre os valores dos pixels para agrupamento.

Tratando-se de imagens monocromáticas, por esta abordagem, um *pixel* $p(x,y)$ em uma imagem $f(x,y)$ seria classificado como pertencente a uma região:

$$\begin{cases} \mathbf{R}_1, & \text{caso } f(x,y) < \mathbf{T} \\ \mathbf{R}_2, & \text{caso contrário.} \end{cases}$$

onde $\mathbf{T}$ é o *threshold* (ou limiar) para a classificação.

Tradicionalmente, a aplicação dos limiares está ligada a um estudo do histograma da imagem, e a determinação de um nível capaz de separar duas regiões, conforme ilustra a figura 4.1.

Figura 4.1 - Determinação de limiar para histograma em tons de cinza.

Para as imagens coloridas, o paradigma da imposição de limiares evoluiu para um espaço multidimensional. Dada a representação das cores de *pixels* de uma imagem em algum espaço tridimensional, "*espera-se determinar parâmetros de cores para os quais limiares possam processar a segmentação*" [BALALRD & BROWN, 1982, p.153]. Nesse sentido, como pode ser visto na figura 4.2, para pertencer a um grupo R, a representação da cor de um *pixel* precisa estar interna aos limites das três faixas das primárias de cores (A, B e C).

Figura 4.2 - Os limiares aplicados a um espaço tridimensional de cores.

O método dos limiares é considerado de grande utilização em situações simples, mas bastante primitivo [BALLARD & BROWN, 1982]. Tipicamente, pixels bem divididos de um histograma - ou espaço de cores - podem não estar conectados espacialmente. Outras situações como a presença de ruído ou mesmo fundos de imagem

que apresentam tonalidades variáveis podem incorrer em segmentações inapropriadas. Duas observações são, portanto, importantes no contexto dos trabalhos com *limiares* aplicado ao reconhecimento de cores:

- i) Nenhuma consideração sobre a localização espacial do *pixel* e topologia da imagem são utilizadas em sua classificação - apenas o espaço de cores é analisado;
- ii) Nem sempre existe um limiar ótimo para segmentar regiões.

Tais limitações restringem fortemente a utilização do método.

4.3 Algoritmos baseados em vizinhança

Um dos algoritmos que proporciona a segmentação pela análise da vizinhança é o *crescimento de região*. Trata-se de um processo pelo qual *pixels* com características semelhantes em uma vizinhança são agregados em regiões maiores, sucessivamente, partindo-se de um *pixel* denominado a "semente" da região.

GONZALES; WOODS (1992) propõem que, para a segmentação de imagens monocromáticas por este processo, a propriedade P - utilizada para classificar um *pixel* em uma dada região - seja o valor absoluto da diferença entre o nível de cinza de um pixel e o nível de cinza do pixel tomado como semente. Em outras palavras, todo *pixel* de uma vizinhança que possuir a uma diferença menor que um limite T do *pixel* semente, será enquadrado na região deste *pixel*. De maneira análoga, para imagens coloridas, podemos utilizar um limiar tricromático.

Observe-se que, embora faça uso da imposição de limiares, este método leva em consideração a informação da vizinhança do ponto, o que é desconsiderado pelo método anterior.

Um método alternativo ao anterior é o da *divisão e agrupamento* de regiões. A alternativa apresentada é inicialmente subdividir uma imagem em um conjunto arbitrário de regiões, e, então proceder à análise das mesmas, dividindo-as novamente em partes menores (caso não atendam a determinados critérios), ou agrupando-as com regiões vizinhas.

4.4 A classificação de pixels

Tanto os algoritmos baseados em análise de descontinuidades quanto os que efetuam o agrupamento de regiões freqüentemente utilizam-se da comparação de níveis de intensidade luminosa dos pixels da imagem para a tomada de decisões, e, portanto, classicamente, da imposição de limiares. Contudo, tratando-se de imagens segmentadas por cores, nem sempre a diferença entre o valor luminoso de dois pontos é suficiente para proceder segmentações, visto que as noções cognitivas de cor não são homogeneamente distribuídas pelo espectro de cores. Tratando-se do paradigma da imposição de limites, isto implica que as regiões ótimas de cada cor no espectro de cores não são necessariamente separáveis por sólidos de tamanhos iguais no espaço tricromático de cores, e, mais que isso, não são necessariamente separáveis por sólidos regulares.

Um algoritmo de segmentação mais cuidadoso com esse conceito deveria estar apto a realizar a classificação de cores antes da fase de segmentação propriamente dita. Uma vez realizada tal análise, a imagem classificada poderia ser submetida aos algoritmos acima descritos. Tal camada (descrita aqui como um pré-processamento) pode também ser vista como uma classe de segmentação, em situações onde apenas a informação de cores é suficiente para a segmentação.

O método da imposição de limiares, nesse contexto, pode, portanto, ser considerado um método de classificação - já que subdivide o espaço de entrada em regiões discretas que podem ser interpretadas como classes. Contudo, na esperança de obter melhores classificações, ferramentas mais elaboradas podem ser utilizadas, em especial ferramentas do âmbito da inteligência artificial, onde o problema de classificação é de grande relevância. Uma abordagem natural e com grande habilidade de generalização é a das *redes neurais artificiais* (RNAs).

Diversos pesquisadores, nos últimos anos, têm utilizado a modelagem das RNAs em sistemas de visão computacional. Sua capacidade de responder a situações previamente treinadas e de generalizar para situações não presentes em seu universo de treinamento têm estimulado as pesquisas na área [LITTMAN & RITTER, 1997] [VELOSO & UTHUR, 1998] [MEWTON LABS, 1998] [MOREIRA & COSTA, 1996]. Mais que isso, algumas das novas classes de redes neurais apresentam respostas que se assemelham fortemente às de células encontradas no sistema de visão humano, abrindo

um novo leque no trabalho com imagens por esse paradigma [LINDBLAD & KINSER, 1999] [JOHNSON, 1994] [BOUVEZ, 1993]. Outros pesquisadores têm se dedicado ao uso das RNAs na simulação de movimentos oculares, como os sacádicos [KINSER, 1999]. Uma visão geral da problemática que envolve a modelagem por RNAs para interpretação de imagens pode ser encontrado em [DOHERTY & AL-SHAYKU, 1993], [BARNARD & CASANET, 1991] e [MAXWELL & GILES, 1998].

A proposição da segmentação por cores como a separação de coleções de pontos em um espaço n -dimensional leva-nos a um caso típico de utilização das redes neurais artificiais, objeto de maior detalhamento no capítulo que se segue.

5. Redes Neurais Artificiais

"Na nossa busca de construir máquinas inteligentes, só temos um modelo que ocorre naturalmente: o cérebro humano. Uma idéia óbvia para IA, então, é simular o funcionamento do cérebro diretamente em um computador".

RICH & KNIGHT (1991)

Uma das teorias que se seguem da idéia apresentada por Rich - normalmente denominada *conexionista* - recebe inspiração da maneira como o cérebro humano funciona, tendo como unidade básica de processamento uma entidade - de *hardware* ou *software* - inspirada no funcionamento do neurônio biológico.

As seções que se seguem abordam os principais aspectos dessa abordagem, apresentando: **i)** um panorama histórico da técnica, **ii)** os aspectos biológicos, **iii)** a modelagem matemática do neurônio, e **iv)** a arquiteturas de RNA que será usada como classificador nesse trabalho.

Destaca-se entretanto que, devido à grande diversidade de arquiteturas encontradas na literatura, apenas as de maior importância ou de alguma forma relevantes ao trabalho proposto serão abordadas, em especial o perceptron multicamada. Um panorama geral de outras arquiteturas pode ser encontrado em algumas referências bibliográficas da área [HAYKIN, 1999] [CABRAL, 1999].

5.1 Histórico

No ano de 1894 o espanhol Ramón y Cajal em sua publicação "A retina dos vertebrados" (*Die Retine der Wirbeltiere*), identificava anatomicamente uma célula nervosa. Quase meio século depois, McCULLOCH; PITTS (1943) publicaram seu célebre artigo "Um cálculo lógico das idéias provenientes da atividade nervosa" (*A Logical Calculus of the Ideas Immanent in Nervous Activity*), levando ao surgimento das "redes neurais artificiais". Especulava-se, então, sobre ser booleana a natureza da inteligência. Contribuindo ainda mais para tal especulação, os mesmos autores

implementaram um *discriminador linear* extremamente simples, capaz de separar duas coleções de entradas booleanas (que fossem linearmente separáveis em duas coleções) tomando como critério seus valores de saída, ou seja, tal discriminador era capaz de implementar as funções booleanas básicas AND, OR e NOT. Era o primeiro modelo matemático e computacional para o neurônio biológico.

O passo seguinte no desenvolvimento das RNAs foi dado por ROSENBLATT (1962), que construiu uma rede com os discriminadores lineares de Mccluch e Pitts dispostos em camadas, batizada *perceptron*. Necessitava-se, entretanto, de uma forma de treinar esta rede, que apresentava um número muito maior de pesos sinápticos. O aprendizado nas redes neurais nasceu dos estudos de HEBB (1949). O biólogo, que estudava o comportamento de animais, propôs que o aprendizado em um sistema neural complexo poderia ser expresso na forma da alteração dos pesos sinápticos *locais* em função dos erros detectáveis localmente. Nesse momento histórico, a teoria das RNAs já se tornava estabelecida como área de pesquisa científica.

Frank ROSENBLATT (1962) empregou o princípio de Hebb em um discriminador linear, enunciando a “*lei do aprendizado do perceptron*” e elaborou uma modelagem que permitia que os estímulos da rede pudessem assumir valores contínuos, embora o neurônio utilizado ainda tivesse a função de ativação não-linear de McCluch. Contudo, Rosenblatt não conseguiu estender tal lei ou sua prova para redes, o que era o grande limitador da utilização prática da modelagem até então.

Paralelamente a Rosemblat, WIDROW (1962) desenvolveu um modelo neural linear, batizado ADALINE (*Adaptive Linear Element*). O grande valor desse modelo foi a proposta de Widrow para a aproximação linear através do cálculo do gradiente de uma função de erro quadrático, que, em sua homenagem, foi batizada “*regra Δ de Widrow*”. Posteriormente, esse método foi generalizado para uma rede neural multidimensional, que, por se tratar de múltiplas ADALINES, ficou conhecida como MADALINE.

Seguiu-se durante a década de 70 uma fase de pouca produção de pesquisas na área. MINSKY; PAPPERT (1969) demonstravam as limitações das teorias desenvolvidas. As redes de uma camada falhavam na tarefa de resolver problemas extremamente simples, como a operação XOR.

Embora em menor escala de produção destacam-se nesse período o trabalho de KOHONEN (1972) desenvolvendo os mapas auto-organizados ("*self-organized maps*"), uma nova classe de redes neurais que não necessita de supervisão para o aprendizado, e o trabalho de James ANDERSON (1972) desenvolvendo as memórias associativas.

Na década de 80 ocorreu um novo impulso nas pesquisas em redes neurais. O desenvolvimento das memórias associativas de HOPFIELD (1982) inaugura essa fase. Tratava-se de uma rede neural capaz de associar modelos apresentados a ela a modelos presentes em uma memória prévia.

Alguns anos depois Rummelhart, Hinton e Williams [RUMMELHART et. al., 1986] desenvolveram o algoritmo que ficou conhecido como a "*retropropagação do erro*" (*error backpropagation*), basicamente uma generalização da regra Δ de Widrow para uma rede ADALINE que tinha uma função de ativação *semilinear*. O resultado obtido foi o de uma rede com grande poder de generalização quando treinada adequadamente, que poderia aproximar qualquer função contínua. Este trabalho é considerado pela comunidade mundial como o grande impulsionador da teoria das RNAs.

No início da década de 80, graças ao aparecimento de computadores mais rápidos e baratos e ao elevado número de trabalhos, as RNAs deram o seu grande salto. Os antes escassos trabalhos que vislumbravam as aplicabilidades desta teoria tiveram um crescimento espantosamente rápido, atraindo os olhos da comunidade científica mundial. Seu domínio de aplicações atingiu a engenharia, física, psicologia. As conferências multiplicaram-se enormemente e um sem número de publicações tornou-se disponível em todo o mundo. As aplicações vislumbradas anteriormente como o reconhecimento da fala e da visão voltavam à tona.

Em 1980, K. FUKUSHIMA (1988), dos laboratórios NHK em Tóquio, desenvolveram uma série de redes neurais especializadas para o reconhecimento visual de caracteres. As primeiras tentativas com a teoria dos "*self-organized networks*", em uma rede batizada *neocognitron* apresentou os problemas (hoje clássicos) de falhas no reconhecimento de caracteres distorcidos rotacionalmente ou em localizações espaciais diversas.

Persistia ainda a questão do número de neurônios e camadas necessárias para uma dada implementação. HECHT-NIELSEN (1990) baseado nos trabalhos do russo Kolmogorov, e GIROSI; POGGIO (1990) mostraram que qualquer função contínua pode

ser aproximada com uma dada precisão (estabelecida a priori) por uma rede com uma única camada oculta.

Com base nesse estudo, PARK; SANDBERG (1991) desenvolveram uma nova classe de redes neurais, conhecidas como *redes de base radial*. Trata-se de uma rede dotada de “campos receptivos” para seus neurônios, encontrando semelhança em neurônios biológicos tipicamente encontrados em áreas relativas à interpretação da visão e audição.

Nestes últimos anos, a diversidade de arquiteturas das RNAs cresceu enormemente, expandindo suas aplicações na mesma proporção e impondo novos domínios à teoria. Podemos destacar sua aplicação em processamento de sinais (quer seja processamento de dados, voz ou imagens), reconhecimentos de padrões diversos, sistemas de controles e qualquer outro sistema onde a modelagem seja difícil.

5.2 O Neurônio

5.2.1 O Neurônio biológico e sua ativação

Com o avanço dos estudos fisiológicos do sistema nervoso iniciado no século 19, o neurônio foi reconhecido como o dispositivo elementar do processamento cerebral e este passou a ser alvo de extensivos estudos, culminando nos dias atuais em que seu funcionamento é conhecido com alguma profundidade.

A anatomia da célula nervosa (figura 5.1) caracteriza-se pela presença de um corpo celular (*soma*) - responsável pelos processos metabólicos da célula - e pelas projeções deste corpo (*dendritos* e o *axônio*). Tal célula destaca-se pela capacidade de transmissão de impulsos elétricos, possível graças às propriedades peculiares de sua membrana celular.

As entradas neurais provêm das conexões sinápticas, efetuadas pelas ramificações dos dendritos de um neurônio em contato com os axônios de outros neurônios. O neurônio, após receber tais entradas (aqui referidas como impulsos nervosos), processa-as de alguma forma e produz uma saída em seu axônio.

Figura 5.1 - O neurônio biológico [RUSSEL & NORVIG, 1995].

O impulso nervoso propaga-se segundo o processo biológico conhecido como a *bomba de sódio e potássio*, que pode ser visto na figura 5.2. A difusão de íons pela membrana possibilita que esta apresente uma polarização entre suas regiões interna e externa da ordem de alguns milivolts. Nesses termos, caracterizamos o impulso nervoso como sendo a propagação de uma despolarização, durante algum tempo, ao longo da membrana.

Figura 5.2 - Bomba de sódio e potássio: a propagação do impulso elétrico em um neurônio.

A formação do potencial elétrico em um neurônio foi alvo de estudo de diversos pesquisadores, dentre os quais se destacam HODGKING; HUXLEY (1952), agraciados com o nobel da fisiologia de 1963 pelo estudo da despolarização de uma membrana *in vitro*.

KOVÁCS (1996) apresenta resumidamente a formação de tal potencial em um neurônio. Segundo ele, este ocorre sempre que a membrana recebe estímulos suficientes para cruzar um *limiar de disparo*. Ao receber tal estímulo, a célula se excitará por tipicamente alguns microsegundos (T_n), emitindo nesse período forte descarga; depois entrará em um período de *refração absoluta* (T_R), em que a célula não produzirá novo disparo, chegando a um período de *refração relativa* (T_r), em que um novo disparo será produzido somente sob uma excitação bastante maior que o valor do limiar de disparo e finalmente retornará ao seu estado de repouso. Um potencial de ação pode ser visto na figura 5.3.

Figura 5.3 - Disparo de um neurônio [KOVÁCS, 1996].

5.2.2 A modelagem matemática

Com a base descrita acima, podemos entender um neurônio como um dispositivo que tem muitas entradas e apenas uma saída. Nesse sentido, Rummelhart, Hinton e Willians [RUMMELHART et. al., 1986] propuseram um modelo não-linear para o neurônio biológico, que é amplamente aceito pela comunidade científica [KOVÁCS, 1996] [HAYKIN, 1999] [CABRAL, 1999] [RICH & NIGHT, 1991]:

Figura 5.4 - Modelo não-linear de um neurônio.

onde:

- X é o vetor de entrada do neurônio k ;
- x_i é a excitação de entrada na sinapse i ;
- y_k é a resposta (ou saída) do neurônio k ;
- w_{ki} é o peso sináptico da entrada i do neurônio k ;
- w_{k0} é o peso sináptico da entrada de polarização do neurônio k ;
- $g(.)$ é a função de ativação do neurônio.

Esse modelo é basicamente um perceptron de Rosenblatt, com uma função de ativação semilinear. O perceptron modela um neurônio processando uma soma ponderada de suas entradas e submetendo o resultado a uma camada de processamento de limiares ($g(.)$). No caso dos perceptrons, essa camada consiste basicamente de uma função degrau, mapeando as saídas nos níveis binários 0 ou 1 - naturalmente uma função altamente não-linear. Rummelhart, Hinton e Williams propuseram a utilização de funções semilineares visto que, no desenvolvimento de seu método de aprendizado, fazem uso da derivada da função adotada. Esta alteração permite o treinamento dos neurônios e torna uma rede neural, como será visto adiante, capaz de mapear qualquer função contínua.

Os principais tipos de função de ativação são mostrados na tabela 5.1, juntamente com suas respectivas derivadas, que serão úteis no contexto das próximas seções.

Tipo da função	Equação	Derivada
Reta	$g(v) = \alpha.v$	$g'(v) = \alpha$
Degrau	$g(v) = h(v)$	-
Sigmóide	$g(v) = (1+\exp(\beta v))^{-1}$	$g'(v) = \beta.(1-g(v)).g(v)$
Tangente Hiperbólica	$g(v) = (1-\exp(-2v))/(1+\exp(2v))$	$g'(v) = \beta.(1-g(v))^2$

Tabela 5.1 - Funções de ativação de um neurônio.

onde:

- $g(.)$ é a função de ativação;
- $g'(.)$ é a derivada da função de ativação;
- $h(.)$ é a função degrau unitário;
- α e β são constantes;

5.3 Os tipos de RNAs

A abordagem conexionista das RNAs abre um amplo leque de formas de conexão entre as unidades de processamento, isto é, os neurônios. Isso abrange o número de camadas presentes na rede, a forma de conexão entre tais unidades, a forma de treinamento, as funções de ativação presentes em cada camada, etc.

Devido à grande bibliografia disponível sobre o tema redes neurais artificiais, até mesmo sua classificação gera alguma controvérsia. FAUSETT (1994), por exemplo, define *arquitetura de uma rede neural* como a disposição dos neurônios em camadas e as conexões entre as camadas. Em um sentido mais amplo, outros pesquisadores utilizam a notação arquitetura na denominação de todo um conjunto de características de uma rede, englobando sua forma de treinamento, finalidade, etc.

PHAN(1995) define dois critérios básicos para a classificação das RNAs:

- i) Quanto à estrutura;
- ii) Quanto à forma de treinamento.

5.3.1 A estrutura das redes

Com relação às estruturas, uma rede neural pode ser definida por duas maneiras distintas:

- **Redes "Feedforward":** A propagação do processamento neural é feita em camadas sucessivas, ou seja, neurônios dispostos em camadas terão seus sinais propagados seqüencialmente da primeira à última camada, de forma unidirecional. O exemplo típico desse tipo de rede seria o perceptron multicamadas (MLP);
- **Redes recorrentes:** As saídas de um (ou todos) os neurônios podem ser realimentadas a neurônios de camadas precedentes (tipicamente da primeira). Esse tipo de rede é classificado como memória dinâmica. O exemplo típico dessa rede é a rede de Hopfield.

5.3.2 O treinamento das redes

Haykin propõe a seguinte definição para o aprendizado no contexto neural: *"(...) é o processo pelo qual parâmetros livres de uma RNA são adaptados através de um processo de estímulos pelo ambiente no qual a rede está envolvida. O tipo de aprendizado é determinado pela maneira na qual os parâmetros mudam"* [HAYKIN, 1999, p.50].

A forma de treinamento, então, diz respeito a como são atualizados os valores dos pesos sinápticos durante o aprendizado da rede. Nesse contexto, podemos descrever:

1. Redes com pesos fixos;
2. Redes com treinamento não-supervisionado;
3. Redes com treinamento supervisionado;
4. Redes com aprendizado por reforço.

Nas *redes com pesos fixos* os pesos de alguns (ou todos) neurônios são fixados em um dado valor, ou fixados ao peso (variável) de outro neurônio. Esse tipo de rede tipicamente é aplicada a soluções de problemas particulares e tem aplicação tradicionalmente ligada à invariância espacial: se considerarmos por exemplo uma imagem de entrada que deva ser classificada por uma RNA independentemente de sua rotação sobre seu centro, poderíamos adotar uma rede onde os pesos dos neurônios equidistantes do centro mantivessem um mesmo peso sináptico [BARNARD & CASASANT, 1991].

Nas *redes com treinamento não supervisionado* não existe a apresentação de mapeamentos entrada-saída à rede: caberá exclusivamente a ela a tarefa de realizar a classificação, com base na informação de número de classes e topologia da rede. É o caso das redes auto-organizadas de Kohonen.

Nas *redes com treinamento supervisionado*, tipicamente, uma sequência de padrões de entrada associados a padrões de saída é apresentada à rede. Esta utiliza as comparações entre a sua classificação para o padrão de entrada e a classificação correta dos exemplos para recalibrar seus pesos. Enquadram-se nesse contexto a maioria das redes utilizadas, como o MLP.

O *aprendizado por reforço* refere-se à utilização de alguma função heurística (definida a priori) que descreva o quão boa é a resposta da rede a uma dada entrada, e da busca da maximização de tal função [PHAN, 1995]. Em outras palavras, não é fornecido à rede o mapeamento direto entrada-saída, mas sim uma recompensa (ou penalização) decorrente da saída gerada pela rede à entrada apresentada. Tal reforço é utilizado na recalibração dos pesos da rede.

5.4 O Perceptron multicamada

5.4.1 Introdução

O *perceptron* foi desenvolvido por ROSENBLATT (1962). Tratava-se de uma rede baseada no discriminador linear de McCULLOCH (1943), que utilizava o princípio do aprendizado de HEBB (1949) para aprender através de exemplos. WIDROW (1962), simultaneamente a Rosenblatt, aprimorou um sistema de treinamento que fazia uso do erro quadrático, embora ainda o tenha aplicado a sistemas lineares.

Rummelhart, Hinton e Williams [RUMMELHART et. al., 1986], anos mais tarde, desenvolveram um princípio de aprendizado para redes com funções de ativação semilineares (vistas adiante) baseadas nas redes desenvolvidas por Widrow, que foi denominado "*error backpropagation*", ou a *retropropagação do erro*. A arquitetura e o treinamento do *Multilayer perceptron* (MLP) ou perceptron multicamada, serão abordadas a seguir.

5.4.2 Arquitetura do MLP

O MLP é uma rede com a topologia apresentada na figura 5.5.

Figura 5.5 - A rede neural multicamada.

Os MLPs podem ser entendidos como uma coleção de neurônios dispostos em k camadas, usualmente classificadas como *camada de entrada* (primeira camada da rede), *camada de saída* (última camada da rede) e *camadas ocultas* (todas as demais). Tais camadas são ainda referenciadas da camada 0 (camada de entrada) à camada n (camada de saída). As j dimensões do estímulo são apresentadas aos j neurônios da camada de entrada. As saídas de cada neurônio são propagadas como entradas para todos os neurônios da camada seguinte, sucessivamente, até a resposta final da rede, de dimensão l igual ao número de neurônios da camada de saída.

Por N_{kj} entende-se o neurônio disposto na j -ésima posição da camada k e por w_{kja} entende-se o peso sináptico associado à entrada a do neurônio na posição j da camada k .

5.4.3 Treinamento: a retropropagação do erro

Segundo Kovács "(...) a retropropagação do erro é o mais poderoso procedimento que se conhece até hoje para o treinamento de redes neurais" [KOVÁCS, 1996, p.64]. Uma vez que a retropropagação do erro foi uma generalização do trabalho

de Widrow para funções de ativação não-lineares, e que este, por sua vez, foi uma evolução do trabalho de Hebb, abordaremos nas seções seguintes esses três tópicos.

5.4.3.1 O princípio de Hebb

Seja um discriminador linear (basicamente o modelo apresentado na figura 5.4 - pg. 46 - utilizando a função de ativação do tipo degrau unitário, mostrada na tabela 5.1 pg. 47). Hebb propôs, em 1949, um princípio pelo qual o aprendizado de um sistema poderia ser reduzido a um processo puramente local, ou seja, a atualização de um peso sináptico é vista como uma função apenas das entradas daquela sinapse e da saída atual do neurônio.

Seja o sinal de erro sobre um neurônio N definido como:

$$e_N\{n\} = y_N^d\{n\} - y_N\{n\} \quad (5.1)$$

onde:

- $e_N(n)$ é o sinal de erro na iteração n ;
- $y_N^d(n)$ é a saída desejada do neurônio N na iteração n ;
- $y_N(n)$ é a saída efetiva do neurônio N na iteração n .

O aprendizado do neurônio N , então, é dado por:

$$w_i\{n+1\} = w_i\{n\} + \Delta w_i \quad (5.2)$$

sendo: $\Delta w_i(n+1) = \zeta \cdot e_N(n) \cdot x_i(n)$

onde:

- n é a iteração;
- w_i é o i -ésimo peso sináptico do neurônio N ;
- ζ é uma constante que representa o passo do aprendizado;
- $e_N(n)$ é o erro associado ao neurônio N na iteração n ;
- x_i é o valor do i -ésimo elemento do vetor de entrada do neurônio N .

Nesses termos, a alteração do peso de uma sinapse de um neurônio fica condicionado apenas à proporção do erro do neurônio relativo à própria sinapse. Contudo, o equacionamento de Hebb apresenta alguns pontos de relevância:

- i) somente é válido para o discriminador linear (contínuo ou discreto);
- ii) somente é válido para um único discriminador linear, visto que Roseblatt não conseguiu estender tal princípio para redes.

Assim, embora o equacionamento de Hebb tenha apresentado os princípios básicos do aprendizado, fazia-se necessária alguma ferramenta mais poderosa para o treinamento de redes de perceptrons.

5.4.3.2 A regra de Widrow

Seja um conjunto de M exemplos (mapeamentos entrada-saída) apresentados a um neurônio N . Define-se o erro quadrático [RUGGIERO & LOPES, 1988] aplicado sobre tal conjunto como sendo:

$$E = \sum_{m=1}^M \{y_m - y_m^d\}^2 \quad (5.3)$$

onde:

E é o erro quadrático;
 y_m^d é a saída desejada para o exemplo m ;
 y_m é a saída efetiva para o exemplo m ;
 M é o número de exemplos.

O gráfico da função apresenta-se como uma curva convexa, que, portanto, possui um ponto de mínimo global. Utilizando-se tal equacionamento, o problema do aprendizado de redes pode ser reduzido a um problema de obtenção do ponto de mínimo da curva de erro quadrático. A obtenção deste ponto para um conjunto de mapeamentos passa a ser visto como a determinação de um conjunto ótimo de pesos W . Assim, é possível expressar a saída y_m da rede como a função:

$$y_m = f\{W; X\} \quad (5.4)$$

onde:

- y_m é a saída efetiva da rede para o exemplo m ;
- W é um conjunto de pesos sinápticos para o neurônio N ;
- X é a entrada para o exemplo m ;
- $f(.)$ é uma função linear de ativação do neurônio N .

Widrow mostrou (para a rede ADALINE) que, partindo de um ponto arbitrário, pode-se caminhar pela superfície da curva $E(w)$ em direção ao ponto de mínimo global, atualizando o conjunto de pesos sinápticos constantemente. Este princípio, denominado "regra delta de Widrow" é dado por:

$$w_N\{n+1\} = w_N\{n\} - \zeta \frac{d}{dt} E\{w_N\{n\}\} \quad (5.5)$$

onde:

- n é a iteração;
- $w_N(n)$ é o conjunto de pesos sinápticos do neurônio N na iteração n ;
- ζ é o passo que se dará na direção contrária ao gradiente;
- $E(w_N(n))$ é o erro quadrático em função dos pesos sinápticos do neurônio N na iteração n ;
- $d(.)/dt$ é a derivada do argumento com relação ao tempo.

Uma importante observação sobre a regra delta de Widrow é que esta permite generalização para múltiplos ADALINE, tornando possível treinar toda uma rede MADALINE. Contudo, tal regra não faz qualquer menção à utilização de funções de ativação diferentes das lineares (semilineares ou não-lineares), o que faz necessária a definição de uma nova ferramenta.

5.4.3.3. A retropropagação do erro

Conforme o mencionado anteriormente, o algoritmo de aprendizado da retropropagação do erro corresponde à generalização do algoritmo de Widrow para funções de ativação semilineares (modelo visto na figura 5.4 com as funções vistas na tabela 5.1).

Dessa maneira, redefine-se o erro quadrático sobre um neurônio N como:

$$E\{W\} = \sum_{m=1}^M \{g\{W.x_m\} - y_m^d\}^2 \quad (5.6)$$

onde:

- W é o vetor de pesos sinápticos para o neurônio N;
- E(w) é o erro quadrático em função dos pesos sinápticos;
- M é o número de exemplos;
- X_m é o vetor de entrada para o neurônio N para o exemplo m;
- y_m^d é a saída desejada para o exemplo m;
- g(.) é a função de ativação semilinear.

É possível mostrar [KOVÁCS, 1996] que tal função de erro quadrático leva a uma atualização de pesos sinápticos para um neurônio N expressa por:

$$W\{n+1\} = W\{n\} + \zeta \sum_{m=1}^M \{y_m^d - y_m\} \frac{dg\{v\}}{dv} . x_i \quad (5.7)$$

onde:

- M é o número de exemplos;
- W(n) é o vetor de pesos sinápticos na iteração n;
- X_m é o vetor de entradas do exemplo m fornecido ao neurônio N;
- y_m é a saída para o exemplo m;
- y_m^d é a saída desejada para o exemplo m;
- ζ é o passo de aprendizado;
- g(v) é a função de ativação semilinear do neurônio N.

As derivadas das diferentes funções de ativação para os neurônios podem ser encontradas na tabela 5.1 (pg. 47). A equação 5.7 é conhecida como *a regra delta generalizada*, e é aplicável a um neurônio isoladamente ou a um grupo de neurônios dispostos em camadas. Com base no equacionamento acima apresentado, define-se o

algoritmo da retropropagação do erro como a aplicação sequencial das seguintes regras [HAYKIN, 1999]:

1. *Inicialização*: tipicamente, iniciar os pesos sinápticos com uma distribuição uniforme e média zero. Calibrar a faixa de operação dos neurônios de saída à tabela de treinamento;
2. *Apresentação dos exemplos de treinamento*: para cada exemplo do conjunto de treinamento, proceder os passos 3 e 4, respectivamente;
3. *Determinação da saída da rede*: Seja um vetor de entrada $X(n)$ aplicado sobre os neurônios da camada de entrada da rede. Computar os campos locais introduzidos (vide passo 3.1) e a função de ativação (vide passo 3.2) sucessivamente a todos os neurônios da rede:

3.1 O campo local de um neurônio N é dado como:

$$v_N^k\{n\} = \sum_{m=0}^M w_{Nm}^k\{n\} \cdot y_m^{k-1}\{n\} \quad (5.8)$$

onde:

k é a camada em que se encontra o neurônio N ;
 n é a iteração;
 M é o número de neurônios da camada $k-1$;
 $v_N^k\{n\}$ é o campo local no neurônio N na camada k ;
 $w_{Nm}^k\{n\}$ é o m -ésimo peso sináptico do neurônio N (na camada k) que atua na sinapse onde está conectado o neurônio m (da camada $k-1$), na iteração n ;
 $y_m^{k-1}\{n\}$ é o sinal de saída do neurônio m da camada anterior ($k-1$), na iteração n ;
 $y_0^{k-1}\{n\}$ é a polarização do neurônio N , usualmente 1;
 $y_N^0\{n\}$ é N -ésimo elemento do vetor de entrada $X_N(n)$;

3.2 A função de ativação de um neurônio sobre seu campo local é dada por:

$$y_N^k = g\{v_N^k\{n\}\} \quad (5.9)$$

onde:

- n é a iteração;
- y_N^k é a saída para o neurônio N da camada k ;
- $g(.)$ é a função de ativação semilinear do neurônio N ;
- $v_N^k\{n\}$ é o campo local do neurônio N na iteração n .

4. Retropropagação: para os neurônios da última camada da rede:

4.1 Calcular o vetor de erro do neurônio, dado por:

$$e_N^K\{n\} = y_N^d - y_N^K \quad (5.10)$$

onde:

- n é a iteração;
- K é a última camada da rede neural;
- $e_N^K\{n\}$ é o erro do N -ésimo neurônio de saída da rede na iteração n ;
- y_N^d é a saída desejada para a o neurônio N ;
- y_N^K é a saída efetiva do neurônio N da camada K ;

4.2 Calcular o gradiente local da rede, definido para um neurônio da camada K de saída (equação 5.11) e para um neurônio da camada oculta (equação 5.12), dados por:

$$\delta_N^k = \left(e_N^K \right) \left(g \{ v_N^K \{ n \} \} \right)' \quad (5.11)$$

$$\delta_N^k = \left(\sum_M \delta_N^{k+1} \{ n \} \cdot w_{MN}^{k+1} \right) \left(g \{ v_N^k \{ n \} \} \right)' \quad (5.12)$$

onde:

- k é a camada onde se encontra o neurônio N sob análise;
- K é a última camada da rede;
- δ_N^k é o gradiente do neurônio N da camada k ;
- e_N^k é o erro relativo ao neurônio N da camada k ;
- $v_N^k\{n\}$ é o campo local do neurônio N da camada k na iteração n ;
- $(g(.))'$ é a diferenciação da função de ativação com respeito ao argumento;
- M é o conjunto de neurônios da camada $k+1$;
- w_{MN}^{k+1} é o N -ésimo peso sináptico do neurônio M na camada $k+1$ (que multiplica a saída do neurônio N da camada k).

4.3 Atualizar os pesos sinápticos:

$$w_{Ni}^k\{n+1\} = w_{Ni}^k\{n\} + \zeta \left(\delta_N^k\{n\} \cdot y_i^{k-1}\{n\} \right) \quad (5.13)$$

onde:

- n é a iteração;
- $w_{Ni}^k\{n\}$ é o peso sináptico associado à entrada i do neurônio N da camada k na iteração n ;
- ζ é o passo de aprendizado;
- $\delta_N^k\{n\}$ é o gradiente local do neurônio N da camada k na iteração n ;
- $y_i^{k-1}\{n\}$ é a saída efetiva do i -ésimo neurônio da camada $k-1$ na iteração n .

5. Iteração: repetir os passos 3 e 4 apresentando novos exemplos à rede, até que um critério de parada seja atingido (tipicamente um erro quadrático menor que um limiar).

A equação 5.13 é a versão discreta da retropropagação do erro, apresentada na sua forma contínua na equação 5.7 (pg. 54). Observando-se os passos do algoritmo, fica caracterizado o termo "retropropagação do erro": "*partindo-se de um erro de saída,*

caminha-se com esse erro retrocedendo da saída em direção à entrada" [KOVÁCS, 1996, p.78].

5.4.4 Considerações sobre o MLP

Uma das questões que emerge quando dos trabalhos com o MLP diz respeito à arquitetura de rede a ser utilizada. HEICH-NIELSEN (1990), readaptando à realidade das RNAs um teorema enunciado pelo russo Kolmogorov (em 1957), apresenta uma solução a esta questão: *"Dada uma função f contínua arbitrária, existe sempre uma implementação exata¹ para f com uma rede neural de três camadas, sendo a camada de entrada um vetor de dimensão n , a camada oculta composta por $(2n+1)$ neurônios, e a camada de saída com m neurônios representando as m componentes do vetor"* [KOVÁCS, 1996, p. 107]. Um MLP é, portanto, um aproximador de qualquer função contínua.

O próprio algoritmo impõe, entretanto, algumas ressalvas à modelagem:

- ✦ No caminho percorrido sobre a curva do erro quadrático, mesmo quando a rede estabiliza sua curva de erro próxima de zero, não é possível determinar se encontramos o mínimo global da função ou se trata de algum mínimo local;
- ✦ No caminho percorrido sobre a curva de erro, o passo de caminhada é sempre uma questão em aberto na abordagem. Passos excessivamente grandes inviabilizam a convergência para alguns pontos da curva e passos demasiadamente pequenos podem implicar em um tempo excessivo de computação;
- ✦ O teorema de Kolmogorov é basicamente um teorema de existência, não fornecendo nenhum procedimento de como se chegar a tal rede. Em termos práticos, a utilização de uma arquitetura de rede com complexidade subestimada poderá levar à não-convergência para uma solução, enquanto redes superdimensionadas poderão incorrer em um esforço computacional desnecessário;
- ✦ Uma vez construída uma rede com uma complexidade qualquer, não existe uma metodologia que garanta, com exatidão, o número de exemplos a ser utilizado e sua representatividade, para a obtenção de uma generalização desejada para pontos não presentes no universo de treinamento da rede.

¹ A implementação exata da função f é assegurada para funções de ativação genéricas, isto é, não pode ser garantida - *a priori* - para funções de ativação sigmóide ou tangente hiperbólica.

Uma das ferramentas usualmente utilizadas no contexto dos dois últimos problemas apresentados é a ferramenta estatística da validação cruzada, que será abordada a seguir.

Um outro fator que tem grande influência sobre a generalização a ser atingida pela rede - e ainda não mencionada - é o ponto de suspensão no treinamento, que será também abordado a seguir.

5.4.4.1 A validação cruzada

Para um dado conjunto de treinamento, existem diversas configurações finais de pesos sinápticos capazes de mapear as entradas apresentadas nas saídas desejadas. Cada uma das quais fará uma interpretação diferente do mapeamento, resultando, portanto, em diferentes generalizações por parte da rede. Um diagrama esquemático de tal situação é mostrado na figura 5.6.

Figura 5.6 - Diferentes generalizações para um exemplo A ausente do universo de treinamento {B,C,D}.

Nesse sentido, existe um claro compromisso entre a generalização desejada e a representatividade e o número de exemplos a serem apresentados à rede na fase de treinamento. Pela abordagem conhecida como validação cruzada (*cross-validation*) [HAYKIN, 1999], um conjunto de exemplos inicial é aleatoriamente particionado em dois conjuntos: conjunto de *treinamento* e conjunto de *teste* e o conjunto de treinamento é novamente dividido no subconjunto de *estimação* e *validação*.

Admita-se que estamos tomando diversas RNAs com complexidade crescente. O *subconjunto de estimação* é utilizado nesse contexto para a estimativa dos pesos sinápticos de cada uma das redes, ou seja, a geração de modelos (redes) candidatos à generalização desejada. Uma vez gerados os modelos, o *subconjunto de validação* será usado para testar a generalização de cada um deles. Contudo, segundo HAYKIN (1999), existe uma distinta possibilidade que o modelo com a melhor performance possa terminar envolvendo o conjunto de validação, sem contudo corresponder à generalização desejada. Uma arma contra tal possibilidade é a utilização do conjunto de teste, diferente do conjunto de validação.

Supondo-se que um novo modelo tenha sido gerado (isto é, um MLP com L camadas, N neurônios por camada e um conjunto W de pesos sinápticos), se, a cada i iterações da função de treinamento verificarmos a resposta da rede ao conjunto de validação, obteremos - tipicamente - uma curva como a que pode ser vista na figura 5.7. No ponto n , a generalização da rede será considerada a melhor possível para os exemplos contidos no conjunto de validação, e, portanto, a rede que implementa a função mais próxima da procurada.

A abordagem descrita é uma das usuais para a validação cruzada, descrita por HAYKIN (1999) como *hold out method* (método da resistência). Uma outra abordagem - indicada para quando existe escassez de exemplos - faz uso da alternância dos conjuntos e dos testes empíricos para todas as combinações destes - dando maior sentido ao nome validação cruzada - implicando, entretanto, em um custo computacional de maiores proporções.

5.4.4.2 O ponto de suspensão de treinamento

Sejam as curvas de erro quadrático dos conjuntos de treinamento (com N elementos) e validação de um MLP com W parâmetros livres. Segundo AMARI et al. (1996), existem dois tipos distintos de comportamento para tais curvas:

a) *Modo não-assintótico*: Nesse comportamento o treinamento exaustivo (ininterrupto) melhora o desempenho (generalização) da rede;

b) *Modo assintótico*: Para este comportamento, o performance da generalização é melhorada cessando-se o treinamento antes da convergência para os exemplos do arquivo de treinamento. Tal comportamento pode ser visto esquematicamente na figura 5.7.

Segundo AMARI et. al. (1996), tipicamente, redes com $N < W$ respondem de maneira não-assintótica, e redes com $N > 30W$ respondem de maneira assintótica.

Figura 5.7 - Curva típica do erro quadrático da validação cruzada onde é aplicável a suspensão prévia do treinamento (como em [HAYKIN, 1999, p.216]).

5.5 Justificativas para a adoção do MLP e da retropropagação do erro

O algoritmo da retropropagação do erro permite o aprendizado dos pesos de uma rede MLP *feedforward* através do fornecimento de exemplos de entradas e respectivas saídas desejadas (aprendizado supervisionado).

Este algoritmo emprega o declínio do gradiente como método de busca para a minimização do erro quadrático entre os valores de saídas e valores desejados para estas saídas.

Entretanto, em redes multicamadas, a superfície de erro pode apresentar múltiplos mínimos locais. Assim, o método de declínio do gradiente só garante convergência para um mínimo local. Apesar deste inconveniente, o algoritmo da retropropagação do erro tem demonstrado, na prática, excelentes resultados na maioria das aplicações reais, sendo sua adoção amplamente difundida. Métodos de validação cruzada podem ser

usados para estimar um ponto apropriado de parada na busca por declínio do gradiente, minimizando o referido inconveniente

A habilidade de redes multicamadas em automaticamente descobrir representações úteis em suas camadas escondidas é uma de suas características-chave [MITCHEL, 1997]. Quanto mais camadas são usadas, propriedades mais complexas podem ser descobertas. No entanto, redes *feedforward* contendo três camadas são capazes de aproximar qualquer função com uma precisão arbitrária, dado um número suficiente de neurônios em cada camada fazendo com que sejam uma ótima escolha para o aprendizado de funções discretas e contínuas, cuja forma geral é desconhecida a priori [MITCHEL, 1997].

Além disso, é sabido que MLPs usando retropropagação do erro se adequam muito bem a problemas nos quais os dados de treinamento correspondem a dados ruidosos de sensores, como aqueles vindos de microfones e câmeras [MITCHEL, 1997]; isto é, os exemplos de treinamento podem conter erros.

Apesar do relativamente longo tempo de treinamento das RNAs, outra característica importante dos MLPs é o fato de possibilitarem a obtenção bastante rápida das respostas quando os mesmos já estiverem treinados.

Devido a estas propriedades apresentadas pelos MLPs com treinamento através da retropropagação do erro e das suas inúmeras aplicações com sucesso encontradas na literatura, adotou-se esta abordagem para o classificador desejado.

6. Revisitando o problema

Uma vez analisados os sistemas de cores, os métodos de segmentação e alguns métodos de separação de coleções de pontos, busca-se, nesta seção, promover uma revisitação ao problema a ser abordado.

Podemos assumir que, com relação aos parâmetros tomados como referência para a segmentação de imagens baseada em cores, os métodos dividem-se em dois grupos principais: *i) os que fazem uso de informações topológicas da imagem*, e *ii) os que fazem uso (apenas) de informações do plano de cores*. Nosso trabalho encaixa-se nesta segunda categoria. Assim, o problema da classificação das cores fica reduzido a um problema de determinação de bordas ótimas para a divisão do espaço de cores em regiões que representem cores distintas.

Seja, então, a imagem da figura 6.1 tomada como exemplo. Trata-se de uma imagem em que são amostradas cinco cores (azul, amarelo, laranja, rosa e preto) com uma iluminação fluorescente aproximadamente homogênea a 400 Lux. No espaço de cores RGB, esta figura encontra-se como a apresentada nas figuras 6.2 (vista tridimensional) e 6.3 (planos laterais). Esta, entretanto, é uma situação ideal de análise, onde a iluminação é homogênea e corresponde aproximadamente à luz branca, isto é, os percentuais das cores primárias não são alterados pela iluminação. De maneira a facilitar a localização de cada região no espaço de cores, os *pixels* da imagem foram identificados com sua própria cor, ou seja, as coordenadas cartesianas são também as cores do pixel.

Na figura 6.5a, podemos ver a figura 6.4a em seu espaço de cores. Trata-se da figura anterior, iluminada com luz incandescente de maneira não-homogênea a aproximadamente 100 Lux. Pode-se observar que, devido à alta incidência da componente vermelha da luz incandescente, as cores da imagem arrastam-se demasiadamente pela componente *Red*. Além disso, a baixa iluminação praticamente une as coleções de pontos, dificultando enormemente sua separação no espaço de cores. As demais imagens (6.5b e 6.5c - respectivamente figuras 6.4b 6.4c no espaço de cores) traçam um perfil de como caminham pelo espaço as cores dos *pixels* com diferentes luminosidades, respectivamente 500 e 1000 Lux.

O domínio de interesse que motivou o trabalho consiste no futebol de robôs. Trata-se de um domínio dinâmico, que exige atuação em tempo real e que propicia o

desenvolvimento e teste de técnicas e métodos em visão computacional, inteligência artificial, robótica, controle e sistemas multi-agentes. Velocidade, habilidade no controle da bola, decisões precisas e rápidas e a capacidade de cooperação entre os robôs são determinantes no resultado dos jogos.

Neste domínio, nas categorias Mirobot da FIRA e *small-size* da RoboCup, jogos são realizados entre dois times de três robôs usando uma bola de golf laranja. No topo de cada robô é colocado um rótulo - azul para um time e amarelo para outro - através do qual um sistema de visão global pode identificar e localizar os robôs. As tarefas de visão são realizadas por um computador central, que recebe imagens de uma câmera colocada sobre o campo. A informação resultante é usada pelo módulo de estratégia do time, o qual comanda os robôs através de comunicação sem fio.

Do ponto de vista do módulo de VC, o domínio impõe os seguintes problemas e restrições:

- Identificação dos elementos de interesse (bola e jogadores) através de suas cores;
- Atuação em tempo real (pelo menos 30 quadros por segundo);
- Iluminação inconstante em toda a área do campo;
- Não se sabe, a priori, a saturação das cores de interesse no domínio (tom de laranja, azul, amarelo, etc.).

Assim, em relação ao problema da classificação de cores, pode-se dizer que existem dois pontos de grande relevância para o domínio de interesse: **i)** a *separabilidade dos pontos* e **ii)** a *obtenção da classificação invariante à luminância*.

Com relação à *separabilidade dos pontos*, podemos dizer que, conforme pode ser observado nas figuras apresentadas, a partir de algum nível de iluminação a nuvem de pontos de cada cor pode ser vista claramente como uma região disjunta das demais e, portanto, separáveis e propícias à classificação. Destaca-se que até mesmo o olho humano não é capaz de identificar corretamente algumas cores com baixa iluminação. Ocorre que, dependendo do sistema utilizado para a representação das cores (RGB, YUV, HSI), dada às diferentes naturezas dos sistemas de representação, os pontos de algumas regiões estarão mais ou menos agrupados, facilitando ou dificultando sua separabilidade. Desta forma, é nosso objetivo analisar a separabilidade das cores em diferentes sistemas de representação de cores, visando determinar quais são mais propícios e em quais condições.

De maneira similar, com relação à robustez da separação com relação à *invariância à luminância*, podemos dizer que, novamente, as respostas dos diferentes sistemas de representação não serão idênticas. Em um sistema RGB, por exemplo, a separação das regiões ótimas para cada cor implica na determinação de regiões tridimensionais que agrupem os pixels desta cor. O problema da invariância ao brilho, contudo, não é um problema bem colocado no sistema RGB, visto que, conforme o apresentado em seções prévias, a luminância não é de obtenção imediata nesse sistema e não corresponde diretamente a nenhum dos eixos. Contudo, é possível determiná-la utilizando o equacionamento visto na seção 3, já que a luminância é uma transformação linear da intensidade luminosa. Outros sistemas (como o YUV), onde a luminância é expressa diretamente em um dos eixos, podem facilitar bastante este trabalho (para o ser humano), visto que o problema da determinação da cor fica restrito a um problema em duas dimensões. No entanto, não sabemos exatamente como um sistema computacional reagiria.

Considerando que para a separação de pontos estejamos utilizando um classificador supervisionado - em que um supervisor humano fornece exemplos das cores de interesse - apresentando exemplos ao sistema, este deve estar apto a implementar as separações para cada cor, baseada nos exemplos apresentados. Contudo, vale ressaltar que um humano apresentando exemplos a um sistema, invariavelmente estará observando as cores - e sua potência - através da curva do observador padrão.

Uma colocação é importante nesse contexto: uma separação invariante à luminância, intuitivamente, poderia ser obtida pela simples divisão do espaço de crominância em regiões e a atribuição de todas as variações da luminância destes a esta região. Todavia, esta não é uma solução adequada, já que o conceito cognitivo de cor muitas vezes se sobrepõe ao conceito físico da cor. Por exemplo, a cor laranja com uma luminância muito pequena pode receber a designação de marrom. Mais do que isso, dependendo do sistema de representação utilizado, os grupos de cores tomarão formas distintas, fazendo com que as bordas de separação possam ser mais complexas do que simplesmente retângulos (2D) ou paralelepípedos (3D), num espaço tricromático.

Considerando ainda que os exemplos apresentados não sejam os exemplos ótimos de cada cor, mas sim exemplos aleatoriamente tomados da cor em questão, seria desejável que algum nível de abstração - mais complexo do que simplesmente o

estabelecimento de um critério de mínima distância entre os pontos - fosse implementado pelo classificador.

Tais restrições podem ser satisfeitas pelo *perceptron multicamada*, que, portanto, foi adotado como a ferramenta para o processamento da classificação de cores no presente trabalho. Na próxima seção descrevemos algumas soluções para este problema, propostas por outros pesquisadores.

Figura 6.1 - Amostragem de cores por uma câmera.

Figura 6.2 - A figura 6.1 no espaço de cores.

Figura 6.3 - Vistas laterais da figura 6.2: a) Plano RxG; b) Plano RxB; c) Plano GxB.

Figura 6.4 - A imagem 6.1 iluminada com luz incandescente a a) 100 Lux b) 500 Lux e c) 1000 Lux.

Figura 6.5 - Espaço de cores para as figura 6.4 a, b e c a: 100 Lux; b) 500 Lux; c) 1000 Lux.

7. Trabalhos correlatos

Apresenta-se nesta seção um resumo dos trabalhos considerados relevantes quando comparados com o trabalho proposto. Tais trabalhos serviram de base para a proposição do presente e também como fonte de idéias e críticas a seus modelos.

Destacamos três trabalhos considerados relevantes com relação à classificação de cores: o *Cognachrome*, um sistema de visão computacional de grande utilização no domínio do futebol de robôs, o método de classificação de pixels de LITTMAN; RITTER (1997), e um sistema de visão para robôs móveis com forma de cães fabricados pela *Sony Corporation*[■], desenvolvido por VELOSO; UTHER (1998), da *Carnegie Mellow University*.

7.1 Cognachrome

7.1.1 Introdução

O *Cognachrome* [NEWTON LABS, 1998] é um sistema de visão computacional desenvolvido pelo *Newton Labs*, com sua primeira versão datada de 1996. Consiste basicamente de um sistema de processamento visual implantado em *hardware*, capaz de rastrear múltiplos objetos no campo visual discriminados por cor. Seu processamento visual é realizado em dois estágios, conforme o diagrama mostrado na figura 7.1.

Figura 7.1 - Diagrama esquemático de funcionamento do sistema *Cognachrome*.

O *primeiro estágio* realiza o processamento de cores. Uma imagem NTSC proveniente de uma câmera é recebida e convertida em formato RGB de 24 bits. As cores são então classificadas, fazendo-se uso do sistema de imposição de limiares no sistema tridimensional RGB, definidos de acordo com um treinamento prévio, descrito a seguir.

O sistema é habilitado a reconhecer um número máximo de cores (tipicamente três). Uma vez que o *pixel* não tenha sido classificado como nenhuma das cores de interesse, este é tipicamente classificado como *fundo*. Um diagrama esquemático de tal classificador é mostrado na figura 7.2. Essa etapa do sistema corresponde diretamente ao trabalho proposto e, no *Cognachrome*, foi implementada em *hardware* para possibilitar atuação em tempo real.

Figura 7.2 - O classificador de cores do sistema *Cognachrome*.

Como produto do primeiro estágio de processamento, então, pode-se identificar uma imagem com os *pixels* classificados, onde as regiões com objetos de interesse apresentam-se com coloração alterada (tipicamente para a cor que foram classificados), e os demais *pixels* (fundo) encontram-se na cor preta.

O *segundo estágio* contém os *softwares* de visão propriamente ditos, a serem utilizados nas imagens previamente segmentadas pelo primeiro estágio, implementados sobre um microprocessador (MOTOROLA 68332). Tal módulo é ainda capaz de realizar operações genéricas de I/O, habilitando o sistema a interfacear com outros dispositivos. O usuário pode utilizar alguns dos programas originais do *Newton Labs*, ou introduzir os seus próprios, armazenando-os diretamente em memória. Na estratégia implantada originalmente para o rastreamento de objetos de interesse na imagem, após a fase de classificação, as regiões de interesse (agrupamentos) são submetidas a diversos tipos de medidas e estatísticas (centróide, área, orientação do maior eixo, tamanho relativo do maior e menor eixos), que servirão como base para a identificação e localização futura dos objetos.

7.1.2 O treinamento do sistema de visão

O treinamento do sistema visual é basicamente o *treinamento das cores* para o sistema. Em suma, os valores RGB de interesse são armazenados e os limites externos

destes - no espaço de cores RGB - são observados, sendo adotados para a classificação após a fase de treinamento. Para a obtenção desses exemplos, tipicamente, o objeto de interesse é posicionado no centro da imagem. Ao se iniciar o procedimento de amostragem, a cor do *pixel* no centro da imagem e dos pontos vizinhos a este serão armazenados para treinamento, sucessivamente, até que se encontrem "bordas delimitadoras" do objeto.

O sistema terá identificado um *pixel* da borda do objeto se o RGB do *pixel* sob análise diferir mais do que um limite pré-estabelecido da cor do *pixel* central. Tal parâmetro é configurável no sistema, por parte do usuário. A variável de máxima diferença de cores impede ainda que, por qualquer inabilidade do usuário na apresentação dos exemplos, reflexos em objetos sejam mapeados como cores de interesse (exceto se esse estiver no centro do objeto).

No isolamento de um objeto pelo método apresentado, o sistema pode incorrer em *falsos positivos*, ou seja, *pixels* do fundo da imagem que tenham proximidade maior do que o limite estabelecido com o centro do objeto sendo classificados de forma incorreta. De forma a tentar contornar este problema, uma nova variável controlada pelo usuário, delimita o máximo tamanho para um objeto. O sistema considera as duas variáveis apresentadas, visando eliminar os problemas de detecção de falsos positivos pelo método do *crescimento* ou *encolhimento da região*.

Visto que condições luminosas podem se alterar em diferentes regiões de um ambiente ou em todo o ambiente no decorrer do tempo, as bordas das regiões das cores são recalibradas de forma constante, incluindo o RGB do ponto central do objeto rastreado no universo de treinamento a toda iteração durante o processo de execução do sistema. Este procedimento, que implementa uma limiarização *adaptativa*, é conhecido como *treinamento incremental*.

Destacam-se ainda no sistema algumas sensibilidades, que podem ser divididas em dois níveis distintos: da *calibração da câmera* e da *escolha do alvo*. A câmera utilizada pode influenciar drasticamente na classificação das imagens na medida em que, usualmente, câmeras possuem ajustes automáticos de nível de branco (*white balance*), que visa regular em tempo real a tonalidade (*hue*) da imagem, e o ajuste da íris, que envolve o ajuste da luminosidade presente no sistema. Já alvos que apresentem

superfícies reflexivas, com iluminação direta ou demasiadamente forte, dificilmente serão corretamente classificados pelo sistema, dados os problemas gerais da abordagem.

7.1.3 Considerações sobre o sistema

O *Cognachrome* foi e continua sendo um sistema de grande utilização prática. Foi utilizado em robôs manipuladores (MIT), simulações de conexões entre naves espaciais (laboratório de sistemas espaciais da Universidade de Maryland), protótipos de robôs que recolhem bolas de um campo de tênis (*Carnegie Mellon* e *Newton Labs*), mas sobretudo por muitos dos times envolvidos com o jogo de futebol de robôs móveis.

Em seu primeiro estágio, o sistema é basicamente um classificador de cores com abordagem clássica (a imposição de limiares), com quatro possíveis classes: as três cores configuradas no ambiente de treinamento, e a classe "fundo". Neste trabalho, vê-se a classificação de pixels atuando não como um pré-processamento, mas sim como toda a camada de segmentação. O sistema de treinamento visa conseguir o maior número de amostragens possível da cor a ser rastreada (praticamente todos os *pixels* do objeto a ser rastreado), de forma a estabelecer os exemplos para as cores. Um ponto interessante deste sistema consiste na utilização de um mecanismo de constante atualização dos limites, realizado adaptativamente pelo treinamento incremental, o diferencial da abordagem.

7.2 A classificação de cores por RNAs de Rae-Ritter-Littman

RAE; RITTER (1998), apresentaram um sistema de reconhecimento de orientação de faces humanas [RAE & RITTER, 1998], baseado em trabalhos prévios de classificação adaptativa de cores. Originalmente, LITTMAN; RITTER (1997) apresentaram um método de segmentação adaptativa que utiliza informação local de cor para estimar a probabilidade de um *pixel* pertencer às classes de objeto ou fundo utilizando uma RNA, aplicado ao reconhecimento de mãos. Tais trabalhos serão apresentados a seguir.

7.2.1 O sistema de reconhecimento de faces

Figura 7.3 - Diagrama esquemático do reconhecimento de faces.

No reconhecimento de faces, o trabalho pode ser dividido em dois grandes blocos funcionais, de maneira bastante similar ao *Cognachrome*. No primeiro deles, uma imagem capturada em formato RGB é submetida a uma rede neural de classificação. A imagem resultante (classificada) é submetida a uma série de análises, a fim de se detectar a presença de faces e sua orientação. Uma vez que o bloco funcional do classificador - aqui apenas uma das fases do processo de segmentação - nos interessa mais diretamente, passaremos a analisá-lo com maior profundidade.

A entrada utilizada por RAE; RITTER (1998) foi a coordenada RGB do pixel. A escolha de tal sistema de representação de cores decorreu dos trabalhos preliminares de LITTMAN; RITTER (1997) que estudaram a classificação da rede neural para diferentes sistemas de representação (RGB, YUV, YQQ). Em tal trabalho, foram utilizadas redes de 10 a 50 nós, obtendo classificações ótimas para uma rede de 30 nós. A rede neural utilizada é do tipo LLM (*Local Linear Map*). Trata-se de uma rede baseada na quantificação do vetor de aprendizado, composta de N células, e que utiliza o princípio do *winner-take-all*, isto é, vence a célula que apresentar melhor resultado.

Nesta rede a idéia básica é realizar um mapeamento de um espaço R^L em um espaço R^M com uma interpolação linear na saída. Seja para cada nó s , um vetor de pesos de entrada $w_s^{in} \in R^L$, um vetor de pesos de saída $w_s^{out} \in R^M$, e uma matriz A_s (responsável pela interpolação linear) de dimensões $M \times L$. Apresentado um vetor de entrada X à rede, será considerada vencedora a célula de *Voronoi* que contiver o vetor de pesos mais próximo de X , isto é, aquela que minimizar a distância:

$$d_s = \min |X - w_s^{in}| \quad (7.1)$$

Sendo $X(t)$ e $Y(t)$ respectivamente a entrada e saída da rede para o exemplo t , o treinamento da rede dá-se segundo as seguintes regras:

$$\Delta w_s^{in} = \varepsilon_1 \cdot [X(t) - W_s^{in}] \quad (7.2)$$

$$\Delta w_s^{out} = \varepsilon_2 \cdot [Y(t) - Y_{rede}] + A_s \cdot \Delta w_s^{in} \quad (7.3)$$

e,

$$\Delta A_s = \varepsilon_3 \cdot (d_s)^{-1} \cdot [Y(t) - Y_{rede}] \cdot [X(t) - w_s^{in}]^T \quad (7.4)$$

onde:

- w_s^{in} é o vetor de pesos de entrada da célula S ;
- w_s^{out} é o vetor de pesos de saída da célula S ;
- $X(t)$ é o vetor de entrada para o exemplo t ;
- $Y(t)$ é o vetor de saída desejada para o exemplo t ;
- Y_{rede} é a saída efetiva da rede para o exemplo t ;
- ε é o passo de aprendizado;
- A_s é a matriz de interpolação;
- d_s é a distância entre padrões (eq. 7.1).

Fazendo com que ε (passo de aprendizado) decaia com o tempo, tipicamente w_s^{in} corresponderá aos *codebooks* de cada célula. Visto que a saída desejada está em uma dimensão diferente da entrada, uma matriz de saídas típicas (w_s^{out}) é também associada a cada célula. A saída da rede, contudo, não se limita simplesmente à saída típica da classe: existe um componente responsável pela interpolação linear desta (A_s).

Em termos simples, a cada nova entrada submetida à rede, esta retornará como resposta o parâmetro conhecido mais próximo, com uma interpolação proporcional à distância do modelo. Tal procedimento é ilustrado na figura 7.4.

Após a fase de treinamento, espera-se, com este procedimento, que a rede retorne o "grau de certeza" sobre a classificação de um determinado padrão de entrada.

Figura 7.4 - A aproximação linear da LLM (como em [LITTMAN & RITTER, 1997]).

7.2.2 Treinamento e resultados obtidos

O treinamento dos pixels foi realizado mapeando-os como de interesse (possivelmente parte do objeto) ou fundo, utilizando-se níveis binários 1 e -1 . Executado o treinamento e apresentado um novo vetor à rede, esta tipicamente apresentava na saída algum nível entre as classificações anteriores. LITTMAN; RITTER (1997) interpretaram esses números como o grau de certeza da rede sobre cada uma das classificações - dado utilizado posteriormente. Embora a interpolação - inerente à rede proposta - permita a obtenção de tal informação, esta dificulta a classificação imediata do *pixel*, ou seja, a obtenção de uma das classes originais sem a interpolação. De forma a obter tal classe, os autores definiram uma operação de limiar (*threshold*) sobre as medidas das classes.

Dessa maneira, além de promover o estudo do desempenho da rede para as diferentes formas de classificação, necessitou-se também da determinação do limiar ideal para os trabalhos da rede. Para a determinação dos *limiares ótimos* para sua rede, estabeleceram-se como medidas as curvas de especificidade e sensibilidade.

Seja a figura 7.5. Nesse contexto, define-se:

- ◆ *Especificidade (Spc)* refere-se ao número de *pixels* corretamente classificados como objeto, sobre o número total de *pixels* classificados como objeto, isto é, $C/C+A$;

- *Sensibilidade (Sen)* é definida como o número de *pixels* corretamente classificados como objeto sobre o número de *pixels* que são efetivamente objeto, isto é, $C/C+B$;

Figura 7.5 - Os conjuntos de pixels de uma imagem.

Aplicadas tais medidas para diversos limiares, confirmou-se a noção intuitiva de que o melhor limite entre as classes seria o ponto de igual distância entre elas. Observou-se, entretanto, que não existe, na abordagem adotada, nenhum valor de *limiar* que estabeleça um limite ótimo para a correta classificação de todos os *pixels*. Em seu melhor valor, a rede não superou 85% de desempenho para nenhum dos critérios apresentados acima (*Spc* e *Sen*).

Comparando-se as diversas formas de *apresentação de dados* à rede, os classificadores mostraram performance ótima para o sistema RGB, apresentando-se ligeiramente piores, respectivamente para os sistemas YQQ e YUV. Tais diferenças foram suficientemente pequenas para que a classificação fosse considerada "*altamente independente do sistema de representação de cores adotado*" [LITTMAN & RITTER, 1999, p.184].

7.2.3 Considerações sobre o sistema

O algoritmo de LITTMAN; RITTER (1997) permite a extração das coordenadas espaciais do objeto reconhecido sem a necessidade de uma nova análise da imagem, mas baseado na informação de probabilidade. O centro do objeto de interesse pode ser calculado simplesmente com a média das coordenadas de todos os *pixels* da imagem, pesadas com suas probabilidades de pertencer à classificação de objeto, uma vez que só há um objeto de interesse na cor procurada na cena. Dessa forma, o estágio de

classificação já fornece as informações para estágios futuros, que não necessitam extraí-las de uma imagem classificada.

Com tais experimentos, os autores concluem que a abordagem conexionista é *"robusta sempre que os dados de treinamento disponíveis forem pouco representativos"* [LITTMAN & RITTER, 1999, p. 183], ou seja, não é necessário apresentar à rede exemplos ótimos de cada cor, o que é uma das vantagens inerentes das RNAs.

Contudo, o trabalho apresentado não fez qualquer menção à luminosidade dos sistemas. Em outros termos, a rede foi utilizada para generalizações próximas à faixa de luminosidade usada no treinamento.

7.3 A classificação de cores por aprendizado supervisionado de Veloso-Uther

VELOSO; UThER (1999) da *Carnegie Mellon University* propuseram um sistema visual para um time de futebol de robôs dotados de pernas, fabricados pela *Sony Corporation*[■]. O objetivo de tal abordagem era reconhecer e seguir objetos de 8 cores distintas. Uma vez treinado o sistema, este atuava segundo a figura 7.6.

Figura 7.6 - A classificação de cores no sistema YUV de VELOSO; UThER (1999).

O sistema visual dos robôs, implantado em *hardware*, mapeia os *pixels* da imagem amostrada em uma das oito cores simbólicas do sistema utilizando uma tabela de limiares como referência. Esta tabela é definida pelo usuário. O diferencial da abordagem está na utilização do espaço YUV para o mapeamento. Conforme o abordado anteriormente, tal sistema de coordenadas permite a representação de uma cor no eixo UV sem a informação de brilho, armazenada no eixo Y. Assim, a tarefa de identificação de cores de uma forma invariante ao brilho fica restrita à tarefa de determinação de bordas ótimas

(limiares) para as cores no eixo UV em diferentes níveis de luminância. Observe-se que os melhores limiares variam de acordo com o nível de luminância, não sendo possível - pelos motivos abordados no capítulo 3 - estabelecer limites únicos para todo o espaço.

A tabela de limiares dos robôs da *Sony* consiste de valores limites mínimos e máximos par cada uma das cores de interesse, para cada um dos 32 níveis de luminância em que foi discretizado o eixo Y da representação YUV (tabela 7.1).

Y	Cor 1		Cor 2		...	Cor 8	
	U	V	U	V	...	U	V
0					...		
1					...		
:	:	:	:	:	...	:	:
31					...		

Tabela 7.1 - A tabela de limiares do classificador de cores.

A figura 7.7 mostra, para um valor fixo de Y, o espaço UV correspondente e os retângulos correspondentes aos limiares para 5 cores de interesse (amarelo, verde, azul, laranja e rosa). Para estes limiares, podemos ver imagens exemplos à esquerda da tela e as respectivas classificações das cores de interesse à direita.

Figura 7.7 - O sistema visual de VELOSO; UThER (1999) - estabelecendo os melhores limites para algumas cores em um nível discreto de luminância do sistema YUV (extraído de [VELOSO & UThER, 1999]).

7.3.1 Treinamento

Para preencher adequadamente a tabela de limiares VELOSO; UThER (1999) desenvolveram um sistema de aprendizagem por reforço. Dada uma série de imagens acumuladas - tipicamente um robô em diferentes posições do campo - é possível classificar manualmente as cores do sistema utilizando-se de uma interface construída para tal, fornecendo assim realimentação para o sistema de aprendizagem. Uma vez classificados previamente os pixels no espaço YUV, para cada cor Y, as bordas retangulares no espaço UV (*limiares*) são determinadas automaticamente, usando aprendizado por reforço de forma a minimizar a soma do erro quadrático dos dados de classificação.

7.3.1 Considerações sobre o sistema

A grande contribuição do trabalho de M. Veloso e W. Uther é a utilização do espaço YUV, que proporciona a obtenção de invariância à luminância com menor custo computacional e facilidades no treinamento. Contudo, as bordas de classificação ainda são constituídas por retângulos, o que geralmente não corresponde a bordas ideais delimitantes de cada classe.

8. Implementação do classificador

O presente capítulo descreve com maior detalhamento o trabalho realizado para a implementação de um sistema de classificação de cores com abordagem conexionista e o estudo de sua generalização quanto à robustez a variações de brilho e saturação de cores.

8.1 Domínio de aplicação

Historicamente, os jogos têm desempenhado papel de grande importância dentro dos estudos da IA. Samuel, por exemplo, elaborou um jogo de damas que era capaz de aprender com jogos sucessivos e utilizar a experiência adquirida nas partidas subsequentes [RICH & KNIGHT, 1991]. Outros jogos como o jogo-da-velha e o xadrez foram tomados ao longo do tempo como referências para os estudos em IA, culminando em 1998 por derrotar o campeão mundial de xadrez *Andrey Kasparov*.

Atualmente, um dos jogos de interesse nesse âmbito é o futebol de robôs móveis autônomos [REALI COSTA et. al., 1999]. Tal plataforma combina diversas tecnologias da engenharia e da computação com o objetivo de tornar possível a realização de partidas com regras similares às do futebol (*soccer*) convencional. Essa plataforma oferece suporte à aplicação de muitas das técnicas de IA (dada a necessidade de percepção do mundo, planejamento de ações, etc.) e, sobretudo, à integração entre diversas técnicas, um dos tópicos emergentes na atualidade [OLLER et. al. 1996]. Uma característica da plataforma apresentada que merece destaque é a interação com o ambiente, que faz com que os algoritmos necessitem de grande robustez nesse sentido, estabelecendo novos limites ao contexto - outrora restrito - da IA.

O jogo é padronizado por entidades internacionais, que o dividem em várias categorias. Nas diversas categorias, aspectos diferentes das arquiteturas de *hardware* e *software* são alterados: número de jogadores, o processamento (*on-board* ou externo), a forma de locomoção (rodas ou bípede), entre outros. A nossa aplicação caracteriza-se basicamente pelos seguintes fatores:

- O processamento de dados é centralizado em uma estação;

- A comunicação entre estação e robôs dá-se através de transmissor/receptor de rádio;
- Uma câmera colorida efetua o sensoriamento do ambiente;
- Os objetos de interesse do jogo são reconhecidos por cores (robôs, bola, etc.).

A figura 8.1, abaixo, representa esquematicamente o ambiente da plataforma.

Figura 8.1 - Diagrama esquemático do ambiente de futebol de robôs móveis autônomos.

Os jogadores são dotados de receptor de rádio, e uma estrutura formada por microcontrolador e tacômetro, responsável pela malha de controle de baixo nível, sobre os motores.

Nesse contexto, tipicamente, um time faz uso de um quadro de imagem fornecido pela câmera, utiliza as técnicas de visão computacional para a determinação espacial dos componentes do sistema e determina, de alguma maneira, quais devem ser as ações seguintes dos jogadores, atuando sobre estes. O processamento visual - que nos interessa mais diretamente - é baseado nas cores dos robôs, campo e bola. As cores utilizadas são padronizadas internacionalmente e variáveis de acordo com o uniforme dos times envolvidos. As cores típicas são apresentadas na tabela 8.1. Quaisquer outras cores podem ser utilizadas para facilitar a identificação de um ou outro robô, sendo as apresentadas, entretanto, de uso exclusivo.

Cor	Objeto
Laranja	Bola
Azul	Time 1
Amarelo	Time 2
Verde	Campo
Branco e preto	Linhas delimitadoras

Tabela 8.1 - Padrões de cor para o futebol de robôs.

É importante observar que as cores utilizadas podem variar a cada competição, fazendo com que as cores de interesse para o jogo em questão precisem ser informadas ao sistema em uma fase prévia de calibração. Segundo a abordagem proposta (detalhada à frente) são fornecidos exemplos de tais cores à rede que processará a classificação em sua fase de treinamento. Ocorre que o tempo de calibração é limitado (padronizado pela FIRA) e, uma vez que as RNAs requerem algum tempo de treinamento, que aumenta com o aumento do número de exemplos apresentados à rede, resta-nos que o número de exemplos não deve ser muito grande, sendo esta uma forte restrição imposta pelo domínio.

8.2 O classificador adotado

Da padronização de cores do jogo de futebol de robôs - acima mencionadas - podemos extrair algumas características desejadas de um sistema de reconhecimento: **i)** a invariância ao brilho e condições adversas do ambiente, e **ii)** a capacidade da identificação das cores informadas por um supervisor humano.

Dos trabalhos apresentados na seção anterior - e da problemática que envolve a segmentação baseada em cores - podem-se extrair outros pontos de relevância para o sistema desejado:

- ✦ A partição do espaço de cores definida pela classificação de cores não atende necessariamente à região definida por um retângulo (no plano 2D) ou paralelepípedo (no plano 3D) - seria desejável que a partição do espaço para classificação das cores fosse definida por superfícies mais complexas;

- Uma vez que nem sempre é possível fornecer os exemplos ideais em uma classificação supervisionada, seria desejável algum poder de generalização no classificador implementado.

Dessa forma, estabeleceu-se que a ferramenta ideal para prover a separação de pontos num universo multidimensional é a das redes neurais artificiais. Nesse sentido, a arquitetura adotada foi a do MLP (vide capítulo 5), que pode ser vista na figura 8.2. A rede proposta consiste de três neurônios de entrada e $(n+1)$ neurônios de saída, onde n representa o número de cores que se deseja classificar.

Figura 8.2 - A rede neural proposta.

Aos N neurônios da camada de entrada apresentar-se-á a tripla de representação da cor do *pixel* (A, B, C). Na saída, espera-se obter a classificação do *pixel* como uma das cores para o qual o sistema foi treinado, ou nenhuma delas, caracterizando o *pixel* como fundo da imagem. A abordagem adotada apresenta as seguintes características principais:

- i) é tradicionalmente um bom classificador e é um aproximador de qualquer função [MITCHEL, 1997];
- ii) não implementa simplesmente um critério de menor distância, sendo mais adequada a mapeamentos complexos e dotada de poder de generalização;
- iii) utiliza um aprendizado supervisionado;

- iv) Permite que exemplos incorretos sejam fornecidos no treinamento, sem prejudicar o aprendizado final.

8.3 Implementações básicas e de suporte

De forma a dar suporte às investigações, um conjunto de aplicativos foi desenvolvido. Tais desenvolvimentos consistem de:

1. A implementação de uma rede neural multicamada na forma de um MLP;

Baseado nos trabalhos prévios de SCARDUA et. al. (1999), desenvolveu-se uma classe de perceptron multicamadas em linguagem C++ e ambiente Linux. Tal classe, além de armazenar todos os algoritmos básicos necessários para o trabalho com o perceptron, oferece algumas facilidades de trabalho, como a criação de um objeto (em tempo de execução), capaz de implementar um perceptron multicamada com funções de ativação, número de camadas e número de neurônios por camada definidos em tempo de compilação. Tal classe oferece-se como um excelente e organizado meio de interfaceamento com outros programas.

2. A implementação de um ambiente gráfico como interface H-M;

Para os trabalhos com imagens captadas por uma câmera em tempo real (*on-line*), faz-se necessário algum sistema de interfaceamento com o usuário, tal que este possa visualizar as imagens capturadas pela câmera. Embora existam alguns sistemas comerciais previamente desenvolvidos para este fim, tais sistemas freqüentemente incorrem em alguns problemas:

- i) não possuem desenvolvimento para o sistema Linux;
- ii) não permitem acesso simultâneo de outro programa aos dados provenientes da câmera;
- iii) não fornecem dados de cores de posições particulares da imagem a programas externos, algo vital para o treinamento do sistema proposto;

iv) possuem algoritmos mais complexos que o necessário para esta aplicação em particular, onde o tempo de processamento é de extrema importância;

Tais problemas motivaram o desenvolvimento de uma interface própria utilizando das bibliotecas gráficas *Xlib* [NYE, 1990]. Esta interface permite a seleção de pontos da imagem através de um *mouse*, transferindo a um arquivo (mais tarde particionado em conjuntos de treinamento e teste) as coordenadas RGB dos pixels selecionados, juntamente com sua classificação, fornecida por um supervisor.

9. Metodologia

Para a execução dos experimentos, inicialmente definiu-se o tipo e forma de apresentação dos dados a serem classificados e então testou-se o resultado apresentado pelo classificador por RNA adotado no trabalho.

Uma vez definido o conjunto de treinamento, as RNAs foram treinadas. Considerou-se que o treinamento estava adequado quando fosse observada uma condição de convergência a um valor mínimo dos erros de classificação apresentado pela RNA para o conjunto de dados usados no treinamento. A seguir, com um novo conjunto de dados (mais complexo), as RNAs foram validadas e suas respostas analisadas.

Esses passos estão detalhados nos itens a seguir.

9.1 Apresentação de exemplos

Figura 9.1 - Uma paleta de cores impressa amostrada por uma câmera: a) com cores contínuas e b) com cores em degradê.

Visando o estudo das cores e seu reconhecimento em um ambiente real (sujeito a ruídos), uma paleta de cores foi impressa em papel e amostrada por uma câmera. De forma a proceder os experimentos que se seguem, utilizamos a disposição das cores de duas maneiras distintas: cores contínuas e em degradê. Nestas imagens (que podem ser vistas nas figuras 9.1a e 9.1b, respectivamente) podemos identificar 5 cores distintas: preto, azul, amarelo, laranja e rosa. A luminosidade do ambiente foi verificada utilizando-se um luxímetro e diversas imagens foram capturadas sob diferentes condições de iluminação.

Observe-se que cada uma das faixas coloridas da imagem degradê varia de preto até uma dada cor, e posteriormente desta mesma cor até branco. A primeira metade desta variação (de preto até uma cor) teoricamente seria obtida com variações de luminosidade. No entanto, optou-se por inserir esta faixa na imagem para facilitar o estudo do método adotado para classificação de cores (como veremos adiante). A segunda metade desta imagem (de branco a uma cor) corresponde à variação de saturação das cores (ver capítulo 3). Um estudo estatístico - onde a média e desvio padrão de cada grandeza representando as cores são mostradas em tabelas e gráficos - das duas imagens utilizadas (contínua e degradê) pode ser encontrado no apêndice A.

A amostragem de exemplos (cores) dessas imagens foi feita através da interface gráfica construída anteriormente (item 8.3.), de onde extraímos N exemplos de cada uma das cores, que foram armazenados em um arquivo de exemplos para cada caso a ser testado (descritos a seguir).

Visando o treinamento das RNAs na forma descrita na seção 8.2, foram fornecidos às redes exemplos de cores amostradas das imagens de entrada da rede, informando-se a classificação desejada para estes, dentre um total de 5 possíveis classes (preto, azul, amarelo, laranja e rosa). Arquivos de exemplo típicos podem ser encontrados no apêndice B, onde estão listados todos os pontos usados para treinamento das RNAs neste trabalho. O estudo estatístico dos exemplos encontra-se no apêndice C.

Assim, para o treinamento de cada uma das RNAs tomadas nos testes subsequentes foi apresentada uma sequência diferente de exemplos. Observe-se que os exemplos apresentados podem variar de duas maneiras distintas: **a)** segundo a apresentação das cores (se apresentamos exemplos da imagem de cores contínua ou da imagem em degradê) e **b)** segundo a forma de representação das cores (se RGB, YUV ou HSV). Uma vez apresentada uma sequência de exemplos a uma rede, passamos a observar a convergência da mesma.

9.2 Convergência no treinamento da RNA

O aprendizado da rede foi quantificado pela função de erro de cada uma das 5 saídas, ou seja, mediu-se a acuidade de classificação da RNA com relação a cada uma das 5 classes disponíveis.

Iniciada a fase de treinamento da rede para o conjunto de exemplos pertinente, esta somente foi suspensa quando o erro das saídas manteve-se satisfatoriamente pequeno por uma quantidade significativa de iterações. Nesse ponto, iniciou-se o processo de validação da RNA (descrito a seguir). As curvas de erros quadráticos para uma rede de 8 neurônios na camada intermediária treinada com exemplos da paleta contínua são mostradas abaixo (figura 9.2), como exemplos.

Figura 9.2 - Os erros quadráticos dos neurônios de saída: a) classe 1; b) classe 2; c) classe 3; d) classe 4; e) classe 5.

Pode-se interpretar os gráficos da figura 9.2 como a certeza que a rede tem da classificação em cada uma das classes. Assim, quando observamos um erro muito pequeno em todas as classes, consideramos que a RNA aprendeu o conjunto de exemplos fornecidos e está apta a oferecer alguma generalização sobre o mesmo.

Nem sempre, contudo, as redes convergiram de forma homogênea para todas as classes. Seja a convergência mostrada na figura 9.3, abaixo. Nela, a RNA rapidamente converge para as classes 1, 2 e 3 (respectivamente as cores preto, azul e amarelo). Já as respostas para as classes 4 e 5 (laranja e rosa) apresentam-se mais lentas. Neste caso, a rede foi considerada treinada somente quando todas as classes convergiram.

Figura 9.3 - Os erros quadráticos de saída dos neurônios a) classe 1; b) classe 2; c) classe 3; d) classe 4; e) classe 5.

Tomando-se redes diferentes (por exemplo o conjunto de redes treinadas com um determinado conjunto de exemplos, porém com número de neurônios diferentes), observamos que o número de iterações que cada rede requer para a convergência não é o mesmo. Tipicamente para os casos observados nesse trabalho, encontramos redes que convergiram a partir de 1.000 iterações, e outras que não apresentaram convergência até a iteração 100.000, embora tivessem convergência assegurada matematicamente segundo o teorema de Kolmogorov (vide seção 5.4.4).

Desta maneira, o número de iterações a que cada uma das redes utilizadas foi submetida não foi o mesmo, já que cada rede tem o seu próprio ponto de aprendizado, não sendo este, portanto, um parâmetro comparável em tais dados.

9.3 Validação da RNA

Uma vez treinada uma RNA, iniciou-se o processo de validação desta. Neste processo, foram apresentadas à rede as imagens já vistas nas figuras 9.1a e 9.1b, uma a uma.

Observe-se que, com relação ao número de cores possíveis, a figura 9.1a é um subcaso da figura 9.1b. Dessa forma, o conjunto de validação do caso da paleta contínua está contido no conjunto de validação formado pela paleta em degradê, sendo este, portanto, um caso mais abrangente. Tal situação é ilustrada esquematicamente na figura 9.4.

Figura 9.4 - Representação esquemática dos conjuntos de validação utilizados.

A apresentação de uma imagem à rede, quer dizer, em outros termos, que cada pixel da imagem foi apresentado à rede para classificação. Uma vez anotadas as classificações para cada pixel, foi possível montar uma “imagem-resposta” da rede, com a interpretação desta sobre as cores a ela apresentadas. Tais imagens foram usadas na análise do desempenho das RNAs.

De posse da imagem-resposta da RNA e de uma imagem classificada manualmente (com uma interpretação humana das cores nela presentes), procedeu-se à contagem do número de pixels corretamente classificados para cada um dos casos (imagem contínua e imagem degradê).

É igualmente importante destacar que no decorrer do processo, qualquer que tenha sido o conjunto de validação utilizado (imagem contínua ou imagem degradê), o conjunto de treinamento foi tomado como um subconjunto do conjunto de validação, já que neste foram utilizados todos os pixels da imagem. Tal situação é mostrada de maneira esquemática pela figura 9.5.

Figura 9.5 - Os pixels da imagem - subconjuntos de treinamento e validação.

Destaque-se que, embora os exemplos de treinamento tenham sido utilizados na validação do experimento, o número de exemplos de validação é muitas vezes superior ao número de exemplos de treinamento, mais especificamente na relação de 76.800: N*5, onde $76.800 = 320 \times 240$ (respectivamente o comprimento e altura da imagem apresentada, em pixels) e N é o número de exemplos tomados para cada uma das 5 cores usadas..

Os resultados obtidos são apresentados na forma de um gráfico, onde a porcentagem de acerto da rede está representada em função do número de neurônios da camada oculta. Os casos onde não está representada a porcentagem de acerto da rede representam casos onde não houve convergência da rede durante 100.000 iterações observadas.

9.4 Treinamento

Na metodologia desenvolvida para o treinamento, fizemos uso de três formas distintas de oferecimento de exemplos à rede: *a)* oferecimento de exemplos da paleta contínua, *b)* oferecimento de exemplos com diferentes saturações e *c)* oferecimento de exemplos com diferentes iluminações. Tais metodologias serão discutidas a seguir.

a) Oferecimento de exemplos de cores contínuas

Na primeira fase deste experimento, apresentamos à rede na fase de treinamento apenas exemplos extraídos da paleta de cores contínuas (figura 9.1a). Tratam-se de exemplos dispostos sob uma mesma iluminação (250 Lux) e cores com a mesma saturação.

b) Oferecimento de exemplos de cores em diferentes saturações

Em uma segunda fase, submetemos à rede um conjunto de exemplos extraídos da parte superior da imagem degradê (figura 9.1b), isto é, exemplos que variam de uma dada cor até branco, mostrando diferentes saturações. Os exemplos aqui apresentados encontravam-se exclusivamente a 250 Lux.

c) Oferecimento de exemplos de cores em diferentes iluminações

Em uma terceira e última fase, apresentamos à rede um conjunto de exemplos novamente extraídos da imagem contínua (figura 9.6, abaixo), só que sob diferentes luminosidades (variável de 250 a 1050 Lux). Observe-se que pelas razões já mencionadas no presente trabalho o tipo de iluminação adotado foi a fluorescente (consultar capítulo 6).

Observe-se que, embora tenha existido preocupação com a homogeneidade da iluminação, devido a uma série de problemas práticos (tais como a reflectância do material da paleta e a falta de instalações adequadas para esse tipo de experimento), isso não foi conseguido satisfatoriamente. Tal fato procurou ser compensado quando da escolha dos exemplos, buscando tomar pontos aproximadamente na mesma posição nas diferentes imagens. Uma vez que estamos amostrando uma paleta onde as cores são homogêneas, espera-se que um mesmo ponto em diferentes imagens mantenha aproximadamente a mesma relação de luminosidade, já que as paletas e as fontes de iluminação foram colocadas nos mesmos lugares para a aquisição das imagens.

Figura 9.6 - As imagens da paleta contínua a: a)250 Lux; b) 350 Lux; c) 450 Lux; d) 550 Lux; e) 650 Lux; f) 750 Lux; g) 850 Lux; h) 950 Lux; i) 1050 Lux.

9.5 Organização dos resultados

Os resultados obtidos com a série de experimentos realizados são apresentados por ordem de:

- 1. Metodologia de treinamento:** As três metodologias utilizadas são as discutidas no item 9.4: caso de cores contínuas, de diferentes saturações e de diferentes intensidades luminosas;

2. Sistema de representação de cores: dentro de uma dada metodologia de treinamento, observamos o desempenho das RNAs para três diferentes formas de representação de cores: RGB, YUV e HSV.

Para uma dada metodologia de treinamento e uma dada representação de cores, diferentes investigações foram realizadas.

Em um primeiro momento, buscando dar suporte a estudos futuros, realizamos uma *análise estatística* dos exemplos utilizados para o treinamento da rede, buscando avaliar sua representatividade perante a imagem original. Desta forma, a cada teste efetuado foram realizados os levantamentos estatísticos pertinentes, determinando média e desvio padrão de cada grandeza de representação de cores. Esses dados podem ser encontrados no 3 do presente trabalho.

Uma vez levantados tais dados, passamos à análise das redes e sua generalização propriamente dita. Os resultados serão apresentados utilizando-se a seguinte disposição:

a) *Análise do desempenho da RNA*

Este tópico subdivide-se em duas partes:

a-1) *Avaliação da influência da complexidade das redes no desempenho da classificação*

Uma vez que o teorema de Kolmogorov (seção 5.4.4) assegura que uma rede com uma única camada oculta e com algum número variável de neurônios nesta camada é capaz de aproximar uma função desejada (qualquer que seja ela), optamos por efetuar os testes deste tópico com uma arquitetura restrita a uma única camada oculta. Dessa maneira, partimos de uma rede com uma única camada oculta e um único neurônio nesta camada e executamos os procedimentos de treinamento e validação que se seguem, incrementando sucessivamente o número de neurônios da camada oculta até o total de 20 (sem variar, contudo, o número de camadas).

Os resultados obtidos serão mostrados na forma de um gráfico de porcentagem de acerto de validação *versus* número de neurônios na camada oculta;

a-2) Influência do número de exemplos no desempenho da classificação

Utilizando a melhor arquitetura encontrada no item anterior, buscamos estudar o comprometimento entre o número de exemplos apresentados à rede e sua capacidade de generalização. Para tanto, apresentamos à rede uma quantidade de exemplos variável de 5 a 45. A distribuição dos exemplos entre as diversas classes desejadas respeitou a proporção 1:1:1:1:1, conforme ilustra a tabela 9.1.

Nº de exemplos	Exemplos pretos	Exemplos azuis	Exemplos amarelos	Exemplos laranjas	Exemplos rosas
5	1	1	1	1	1
10	2	2	2	2	2
15	3	3	3	3	3
20	4	4	4	4	4
⋮	⋮	⋮	⋮	⋮	⋮

Tabela 9.1 - Distribuição dos exemplos fornecidos às RNAs.

O resultado da generalização das RNAs a tais entradas são apresentadas na forma de um gráfico da porcentagem de acerto na validação das redes (para cada uma das duas imagens apresentadas: contínua e degradê) em função do número de exemplos apresentados.

Esse procedimento foi o executado para o fornecimento de pixels extraídos da imagem de cores contínuas e degradê amostradas sob um único nível de intensidade luminosa. Já para a amostragem de exemplos da paleta contínua sob diferentes luminosidades, o procedimento adotado foi o seguinte: tomada a imagem a 250 Lux, extraímos um exemplo de cada cor (1 ponto preto, 1 ponto azul, 1 ponto amarelo, 1 ponto laranja e 1 ponto rosa). Tais pontos foram armazenados em um arquivo de exemplos. Tais exemplos perfazem o treinamento para as redes com 5 exemplos. A seguir tomamos a imagem a 350 Lux e repetimos a extração dos 5 pontos (armazenando-os no mesmo arquivo e fornecendo-o à rede quando do treinamento para 10 exemplos), e assim sucessivamente até obtermos um arquivo com um exemplo de cada cor em cada iluminação (contendo 45 exemplos de cores).

b) Imagens típicas de resposta

A fim de possibilitar a rápida visualização dos setores do espaço de cores (em termos das classes observadas nesse trabalho) que tiveram sua acuidade alterada em um dado tipo de treinamento, apresentaremos alguma(s) imagem(ns)-resposta típica(s) da rede em questão. Tais imagens foram conseguidas fazendo uso da metodologia já apresentada para validação (item 9.3).

c) Avaliação da partição do espaço de cores

Buscamos apresentar visualmente, para efeito comparativo, como cada uma das RNAs particionou o espaço de cores. Isso nos permite observar com mais clareza a influência da forma de treinamento na generalização da rede.

Deste modo, depois de conseguida a convergência do aprendizado de uma RNA, submetemos a esta algumas trincas de entrada (coordenadas tridimensionais do espaço de cores) às quais anotamos a classificação conseguida. Fazendo-se a apresentação de entradas de forma a cobrir todo o espaço de cores do sistema em questão (seja ele RGB, YUV ou HSV), pudemos obter aproximadamente a partição deste implementada pela rede, mostrada visualmente na forma de um gráfico do espaço de cores para cada sistema. Para facilitar a visualização deste por parte do leitor, cada ponto do espaço fornecido à rede foi pintado com a cor como foi classificado.

Observe-se contudo que, para aceleração do processo de obtenção destes gráficos, optamos por fornecer à rede não todos, mas pontos espaçados a cada 5 unidades no espaço de cores.

d) Análise dos resultados

Ao final da apresentação dos dados levantados para de cada tipo de treinamento, buscaremos discorrer sobre as vantagens ou desvantagens do método apresentado na seção correspondente, além de suas particularidades e considerações.

9.6 Configurações utilizadas

Cada rede utilizada nos experimentos efetuados foi treinada com N exemplos de cada cor (número variável de acordo com o caso de treinamento) utilizando a seguinte configuração:

- $\beta = 8$ - constante da função de ativação sigmóide (vide tabela 5.1, pg. 47);
- $\zeta = 0,025$ - passo de aprendizado da rede (equação 5.13, pg. 57).

Além dos dados apresentados, foi utilizada uma escala de entrada (aplicada a todas as entradas na RNA) como um artifício para se evitar a saturação imediata dos neurônios e, portanto, acelerar o processo de aprendizado. A escala utilizada para todas as redes foi de 1/100.

10. Dados e resultados

10.1 Treinamento com cores contínuas

10.1.1 Análise para o sistema RGB

a) Análise do desempenho da RNA

Figura 10.1 - Resposta à variação do número de neurônios da camada oculta para 25 exemplos.

Figura 10.2 - Resposta à variação do número de exemplos para 6 neurônios na camada oculta.

b) Imagens típicas de resposta

Figura 10.3 - As respostas da RNA treinada com 25 exemplos da imagem contínua, usando 10 neurônios na camada oculta após convergência para os casos: a) contínuo e b) degradê.

c) Avaliação da partição do espaço de cores

Figura 10.4 - A partição do espaço de cores para o sistema RGB treinado com 25 exemplos da imagem contínua utilizando 10 neurônios na camada oculta com: a) 1 iteração; b) 250 iterações; c) 500 iterações; d) 1000 iterações; e) 2000 iterações; f) 10000 iterações.

d) Análise dos resultados

Como podemos observar, apenas com poucos exemplos apresentados (a partir de 5) a rede conseguiu um nível de generalização tal que lhe permitiu uma taxa de acerto próxima a 100% para a imagem contínua.

Já para a imagem degradê esta taxa foi bastante menor (da ordem de 87%). Contudo, visto que os exemplos fornecidos correspondem a uma pequena faixa da imagem degradê consideramos que a taxa de acerto da rede foi altamente satisfatória.

Tais fatos podem ser comprovados pelos levantamentos estatísticos das imagens contínua e degradê (encontrados no 1) e dos exemplos utilizados para este caso (3). Observe-se que as curvas normais extraídas dos exemplos situam-se bastante próximas (quando não idênticas) às curvas normais da imagem contínua, o que não ocorre com a imagem degradê, que possui um espalhamento bem maior das curvas normais.

Um outro dado relevante foi a manutenção das taxas de acerto para diferentes quantidades de exemplos fornecidos (algo atípico para uma RNA). Isso deve-se ao fato dos exemplos utilizados apresentarem sempre uma pequena variância (vide análise estatística de média e variância dos exemplos para este caso, no apêndice C). Em outros termos, estamos fornecendo diversos exemplos de uma mesma cor com uma mesma saturação sobre um mesmo plano de luminância (eq. 3.5, pg. 24), ou seja, estamos fornecendo mais exemplos, porém muito similares.

O número de iterações utilizadas na convergência foi igual para todos os casos (10.000 iterações). A maioria das redes, contudo, já apresentava convergência satisfatória por volta da iteração 1.000 (o que pode ser comprovado pela evolução da divisão do espaço de cores apresentada na figura 10.4).

Entende-se que a razão pela qual a imagem-resposta degradê aparece com as cores saturadas tendendo para o rosa torna-se clara com a visualização da divisão do espaço de cores: dada à falta de exemplos em alguns setores do espaço de cores, a rede considerou que a maioria das cores com altas quantidades de primárias, ou seja, próximas à região de máximo R, G e B (região de alta saturação), possui classificação rosa (vide figura 10.4), evidenciando uma falha da metodologia de treinamento que se deve ao agrupamento dos exemplos de treinamento.

Contudo, dada à falha na metodologia de treinamento, consideramos o resultado conseguido altamente satisfatório, já que a rede conseguiu generalizar consideravelmente para um grande número de cores não presentes no seu universo de treinamento.

10.1.2 A análise para o sistema YUV

Devido ao fato do equacionamento da transformação YUV envolver a subtração da luminância de cores do sistema, os limites (variação máxima e mínima) das grandezas

Y, U e V precisam ser recalculados. A grandeza Y por definição varia de 0 a 255, sendo esses respectivamente seus limites inferior e superior. Já as variáveis U e V requerem análise mais detalhada. Sejam as equações 3.7, 3.8 e 3.9 (pg. 27). Fazendo 3.9 em 3.8 teremos que:

$$V = 0,877 [R-(0,177.R + 0,813.G+0,011.B)] \quad (10.1)$$

$$V = 0,721771.R - 0,713001.G - 0,009647.B \quad (10.2)$$

Da equação 10.2 podemos extrair os limites para a variável Y, apresentados na tabela 10.1.

	RGB	V
Limite superior	255,0,0	185
Limite inferior	0,0,255	-185

Tabela 10.1 - Limites para a variável V.

Analogamente para o caso U, teremos (fazendo 3.9 em 3.7) as equações 10.3 e 10.4 e os resultados mostrados na tabela 10.2.

$$U = 0,494.[B-(0,177.R+0,813.G+0,011B)] \quad (10.3)$$

$$U = -0,087438.R - 0,401622.G + 0,488566.B] \quad (10.4)$$

	RGB	U
Limite superior	0,0,255	125
Limite inferior	255,255,0	-125

Tabela 10.2 - Limites para a variável U.

a) Análise do desempenho da RNA

Figura 10.5 - Resposta à variação do número de neurônios da camada oculta com 25 exemplos.

Figura 10.6 - Resposta à variação do número de exemplos com 8 neurônios na camada oculta.

b) Imagens típicas de resposta

Figura 10.7 - As respostas da RNA com 8 neurônios na camada oculta treinada para 45 exemplos após convergência para os casos: a) contínuo e b) degradê.

c) Avaliação da partição do espaço de cores

Figura 10.8 - A partição do espaço de cores para o sistema YUV treinado 45 exemplos da imagem contínua e utilizando 8 neurônios na camada oculta a: a) 1 iteração; b) 250 iterações; c) 500 iterações; d) 1000 iterações; e) 2000 iterações; f) 10000 iterações.

d) Análise dos resultados

O sistema YUV se comportou de forma similar ao RGB, sendo praticamente perfeito na classificação das cores da paleta contínua (para os quais foi treinada) e sensivelmente pior (da ordem de 84% de acerto) para a imagem degradê.

Também para este caso a rede mostrou independência do número de exemplos a ela apresentados (pelos mesmos motivos apresentados no item anterior: o universo de treinamento e de validação se apresentam muito próximos para a imagem contínua e muito distantes para a imagem degradê).

As redes apresentaram convergência aproximadamente com o mesmo número de iterações que o caso anterior, exceto a rede com o menor número de neurônios (onde tipicamente a convergência é mais difícil, e, portanto requer mais iterações).

Já com relação à partição do espaço de cores do sistema, podemos observar que esta não foi a ótima, visto que a divisão ideal possuiria altos níveis da classe preta no plano inferior ($Y=0$), altos níveis da classe amarela no plano superior ($Y=255$), já que é a cor que mais se assemelha com o branco, e os demais planos divididos entre as classes restantes sem grandes variações para cada plano, isto é, a classificação deveria ser independente de Y para as demais cores.

Novamente a falta de exemplos adequados é o fator preponderante para explicar o desempenho da rede para casos com alta saturação. A rede contudo novamente apresentou um nível de generalização bastante grande (da ordem dos 84%) para cores não conhecidas na fase de treinamento.

O número de iterações necessárias para convergência manteve-se similar ao apresentado para o sistema RGB, isto é, da ordem de 1000 iterações.

10.1. 3 Análise para o sistema HSV

Para a conversão dos exemplos ao formato HSV utilizamos das equações 3.10 a 3.14 (pg. 29). Como forma de adequação das formulas à notação mais utilizada computacionalmente, procedemos uma alteração nos limites de variação de cada grandeza (H, S e V). Assim, tal sistema foi utilizado com os limites apresentados na figura 10.9.

Figura 10.9 - Os limites do sistema HSV utilizados.

a) Análise do desempenho da RNA

Figura 10.10 - Resposta à variação do número de neurônios na camada oculta utilizando 25 exemplos.

Figura 10.11 - Resposta à variação do número de exemplos utilizando 12 neurônios na camada oculta.

b) Imagens típicas de resposta

Figura 10.12 - As respostas da RNA treinada com 45 exemplos da imagem contínua, utilizando 15 neurônios na camada oculta após convergência aos casos: a) contínuo e b) degradê.

e) Avaliação da partição do espaço de cores

Figura 10.13 - A divisão do espaço de cores para o sistema HSV treinado com paleta contínua a: a) 1 iteração; b) 250 iterações; c) 500 iterações; d) 1000 iterações; e) 2000 iterações; f) 5000 iterações.

d) Análise dos resultados

O sistema HSV (transformação não-linear do RGB e YUV) respondeu de forma similar aos anteriores para o treinamento com cores contínuas.

As respostas (figura 10.12) à imagem contínua (figura 9.1a) e degradê (figura 9.1b) foram praticamente iguais aos métodos anteriores. A resposta ao número de exemplos (figura 10.11), contudo, sofreu alguma alteração: o gráfico mostrou claramente que a rede precisa de mais exemplos para entender a função HSV.

O número de iterações necessário para a convergência também aumentou significativamente em relação aos sistemas anteriores (atingindo a ordem de 10000 iterações), permitindo-nos dizer que a função a ser implementada tornou-se mais complexa nesse sistema de representação de cores.

A visualização da partição do espaço de cores (figura 10.13) nos permite observar a razão pela qual as imagens-resposta se encontram com falhas pontuais entre as classes: a notação das cores em termos de ângulo e distância radial tornou significativamente mais complexa a detecção das classes por parte da rede. A cor preta, por exemplo, não ocupa todo o centro do plano inferior como seria o esperado, mas é dependente do ângulo (*hue*) observado. Incidências similares ocorreram com as demais cores, resultando nas imagens de resposta observadas.

A seguir (figura 10.14) apresentamos uma vista lateral da imagem da figura 10.13 (partição do espaço de cores HSV) explicitando a localização dos 25 exemplos iniciais fornecidos à rede. Observe que os exemplos não se encontram bem distribuídos pelo eixo de *value*, tampouco bem separados entre si. Isso enfatiza novamente que os exemplos apresentados não são os ideais para a obtenção da generalização de cores desejada, que deveria abranger a maior extensão possível do espaço de cores. Uma rápida análise da representatividade dos exemplos utilizados em relação às imagens originais (vide análises gráficas no 3 em comparação com 1) vem a confirmar tal fato.

Um ponto que merece destaque é a descontinuidade do sistema, que fica evidenciada nos exemplos da cor rosa, cuja tonalidade (*hue*) varia de 351° a 1°. É importante explicitar que caso os exemplos fornecidos à rede se encontrem apenas em um dos setores do círculo, a divisão do espaço não será a mais adequada.

Figura 10.14 - A apresentação dos exemplos no formato HSI.

10.2 Treinamento com cores em diferentes saturações

10.2.1 Análise para o sistema RGB

a) Análise do desempenho da RNA

Figura 10.15 - Resposta à variação do número de neurônios na camada oculta para 25 exemplos.

Figura 10.16 - Resposta à variação do número de exemplos para 12 neurônios na camada oculta.

b) Imagens típicas de resposta

Figura 10.17 - As respostas da RNA treinada com 45 exemplos com 9 neurônios na camada oculta após convergência para os casos: a) contínuo e b) degradê.

Figura 10.18 - As respostas da RNA treinada com 25 exemplos com 12 neurônios na camada oculta após convergência para os casos: a) contínuo e b) degradê.

c) Avaliação da partição do espaço de cores

Figura 10.19 - A partição do espaço de cores para o sistema RGB treinado com cores com variação de saturação, após convergência.

d) Análise dos resultados

Uma vez treinada com alguns pontos extraídos da imagem com diferentes saturações, a rede passou a analisar tanto a imagem contínua quanto a imagem degradê apenas com base em suas generalizações, melhorando significativamente sua performance para a paleta degradê e piorando para a paleta contínua.

A inconstância dos resultados para a paleta contínua deve-se às diferentes generalizações produzidas pela rede para um exemplo que não estava presente em seu universo de treinamento. Tal problema é observado em menor escala na imagem da paleta degradê, já que nesta os pixels estão melhor distribuídos ao longo do espaço de cores.

Analisando-se os dados obtidos, podemos dizer que a generalização conseguida com a saturação de cores (no caso RGB) não foi suficiente para garantir a qualidade da classificação com o número de exemplos utilizado. Contudo, uma melhora significativa no desempenho da classificação da imagem degradê pode ser percebida com o aumento do número de exemplos fornecidos. Já a porcentagem de acerto referente à paleta contínua permanece quase que aleatória, visto que nos casos observados a rede não assumiu uma generalização que pudesse levar os grupos de cores desta imagem (um subcaso da degradê) às classes desejadas.

Observando a divisão do espaço de cores conseguida com a metodologia de treinamento, podemos identificar alguns pontos relevantes:

- A rede generalizou corretamente que cores com altos níveis de saturação pertencem à classe amarelo (próximos a 255, 255, 255);
- A rede generalizou corretamente que cores próximas a (0,0,0) pertencem à classe preto;
- A região da classe amarela se encontra (devido à pequena variância dos exemplos amarelos) com alguns problemas de convergência, evidenciados pela observação do plano B=150 da figura 10.19, onde claramente uma região da classe amarela se interpõe a duas partes da região laranja. Tal fato explica parcialmente a imagem-reposta obtida (figuras 10.17-10.18) onde regiões bem saturadas de laranja encontram-se classificadas como amarelo;

Embora tenha ocorrido a falta de exemplos em algumas regiões do espaço de cores (tipicamente para o laranja), uma rápida comparação visual das imagens-resposta

deste caso à paleta degradê e a figura 10.3 (resposta para degradê no caso RGB treinado com cores contínuas) permite a avaliação da melhoria do classificador conseguido.

Uma outra observação que deve ser feita diz respeito à quantidade de iterações necessárias para a convergência: tipicamente neste módulo foram necessárias de 50000 a 120000 iterações, número bastante significativo em comparação com o tipo de treinamento utilizado no experimento anterior (da ordem de 10000).

Observe-se ainda que os exemplos fornecidos ao classificador na fase de treinamento encontram-se bastante melhor distribuídos estatisticamente que no primeiro caso de treinamento (apêndice C).

Tipicamente, espera-se que os dois problemas acima citados (inconstância e piora nos resultados da paleta contínua) apresentem alguma melhora com o aumento do número de exemplos apresentados à RNA, alvo do próximo tópico de análise.

Finalmente, fazem-se necessárias algumas considerações com relação às imagens típicas de resposta para este caso. Não houve variações significativas para a paleta degradê. As imagens-resposta à paleta contínua, contudo, mostram duas situações distintas: na primeira (figura 10.17) podemos ver um caso em que a generalização da rede não foi suficiente para englobar as cores da paleta contínua, enquanto na segunda (figura 10.18) a generalização da rede foi suficiente para envolver a quase totalidade das cores da paleta contínua das classes desejadas quando da partição do espaço de cores.

Observamos que as duas situações - e portanto as imagens apresentadas - deram-se com diferentes quantidades de neurônios na camada oculta, fazendo com que o leitor possa ser levado a imaginar que a rede com 12 neurônios pode aproximar uma função mais complexa que a de 9 neurônios, explicando assim a melhoria na resposta. Embora esse raciocínio esteja correto no sentido de que uma rede de 12 neurônios pode implementar uma função mais complexa que uma rede de 9, destaque-se que ambas as redes são capazes de implementar a função procurada (vide teorema de Kolmogorov na seção 5.4.4), e o fato de uma se sobressair em relação à outra é explicado pelo fato das generalizações obtidas serem muito diferentes entre si, já que o número de exemplos é muito pequeno em relação ao ideal para o mapeamento completo do espaço tridimensional, o que torna a resposta à paleta contínua quase que um fruto do acaso. Isso pode ser comprovado visto que redes com mais neurônios não tiveram o mesmo índice de acerto (vide figura 10.16).

10.2.2 Análise para o sistema YUV

a) Análise do desempenho da RNA

Figura 10.20 - Resposta à variação do número de neurônios na camada oculta para 25 exemplos.

Figura 10.21 - Resposta à variação do número de exemplos para 10 neurônios na camada oculta.

b) Imagens típicas de resposta

(a)

(b)

Figura 10.22 - As respostas da RNA treinada com 50 exemplos da imagem degradê usando 10 neurônios na camada oculta após convergência para os casos: a) contínuo e b) degradê.

c) Avaliação da partição do espaço de cores

Figura 10.23 - A partição do espaço de cores YUV para o caso do treinamento de cores em diferentes saturações.

d) Análise dos resultados

O método apresentado é bastante semelhante ao anterior, sendo evidente a melhoria da performance do classificador para paleta degradê com o aumento do número de exemplos (figura 10.21). A paleta contínua permanece com grande variação de resultados (conforme discutido no item anterior), mas é certo que aumentando o número de exemplos fornecidos, em algum momento a rede chegará a um nível de generalização que envolva os pixels da imagem contínua.

A divisão do espaço de cores demonstra uma grande melhoria se comparado com o mesmo sistema YUV treinado com a primeira metodologia. As cores variam pouco com o eixo Y (exceto preto e amarelo, respectivamente dominantes no planos inferior e superior).

É importante observar que as curvas normais dos exemplos nesta modalidade de treinamento encontram-se mais próximas às curvas normais da imagem degradê (vide apêndices A e C), o que explica a melhoria do classificador.

10.2.3 Análise para o sistema HSV

a) Análise do desempenho da RNA

Figura 10.24 - Resposta à variação do número de neurônios na camada oculta para 25 exemplos.

Figura 10.25 - Resposta à variação do número de exemplos para 18 neurônios na camada oculta.

b) Imagens típicas de resposta

Figura 10.26 - As respostas da RNA treinada com 25 exemplos da imagem em diferentes saturações usando 18 neurônios na camada oculta após convergência para os casos: a) contínuo e b) degradê.

c) Avaliação da partição do espaço de cores

Figura 10.27 - A divisão do espaço de cores HSV para o caso do treinamento com cores em saturação variada.

d) Análise dos resultados

Novamente o sistema HSV apresenta grandes diferenças na performance do classificador. Primeiramente para que a rede apresentasse convergência, foram necessários mais neurônios na camada oculta do que nas outras formas de representação (vide figura 10.24 em comparação com 10.20 e 10.15, respectivamente nas páginas 112 e 108). Tipicamente, como na abordagem de treinamento anterior, o HSV mostra-se de entendimento mais difícil por parte da rede, o que fica evidenciado na qualidade das imagens-resposta que elucidam a insuficiência de exemplos. O número de iterações necessárias para a convergência - como o esperado - foi um pouco maior que para os outros sistemas, com a mesma metodologia de treinamento.

A divisão do espaço de cores, contudo, apresenta-se bem mais próxima do esperado (onde as cores são pouco dependentes do eixo “value”).

Embora quantitativamente o desempenho da rede não tenha sido da mesma ordem que os sistemas anteriores, espera-se que com um número maior de exemplos o HSV se equipare a tais sistemas.

10.3 Treinamento com cores em diferentes intensidades

10.3.1 Análise para o sistema RGB

a) Análise do desempenho da RNA

Figura 10.28 - Resposta à variação do número de neurônios na camada oculta para 25 exemplos.

Figura 10.29 - Resposta à variação do número de exemplos para 20 neurônios na camada oculta.

b) Imagens típicas de resposta

Figura 10.30 - A resposta da RNA com 5 exemplos em diferentes intensidades luminosas e 20 neurônios na camada oculta após convergência para o caso contínuo.

Figura 10.31 - As respostas da RNA para a paleta degradê quando treinada em diferentes intensidades luminosas e 20 neurônios na camada oculta após convergência para a) 5, b) 25 e c) 45 exemplos.

c) Avaliação da partição do espaço de cores

Figura 10.32 - A divisão do espaço de cores RGB para o caso do treinamento em diferentes intensidades, usando rede de 20 neurônios e 45 exemplos.

d) Análise dos resultados

Neste experimento, um terceiro modo de fornecimento de exemplos à rede foi utilizado, oferecendo imagens da paleta contínua sob diferentes intensidades luminosas para o treinamento do classificador.

É importante ressaltar que, como no primeiro caso de treinamento, a imagem contínua se encontra perfeitamente reconhecida em todos os casos observados. Já a

classificação dos pontos da imagem degradê varia significativamente com o número de exemplos. Observe-se no entanto que a forma de fornecimento de exemplos foi alterada neste tópico: os primeiros 5 exemplos referem-se à imagem a 250 Lux, os 5 seguintes a 350 Lux, e assim sucessivamente até os 1050 Lux (conforme explicado no item 9.4c). O percentual de acerto para a paleta degradê aumenta com os primeiros exemplos (mais escuros) e decai com os exemplos subsequentes (mais claros). Isso denota que a rede não encontrou uma boa generalização sobretudo para os pixels presentes na área de maior saturação da imagem degradê (o que é comprovado pelas imagens-resposta da figura 10.31a, 10.31b e 10.31c). Tal tendência seguramente será rompida com o incremento do número de exemplos apresentados à rede.

Nesses termos, parece claro que o aumento do número de exemplos em cada nível de intensidade luminosa irá aperfeiçoar cada vez mais as bordas definidas pela RNA.

Uma rápida comparação visual entre a divisão do espaço de cores para os casos de treinamento em paleta contínua, em saturação e intensidades variadas (para o sistema RGB) fornece uma idéia da evolução da metodologia de treinamento.

10.3.2 Análise para o sistema YUV

a) Análise do desempenho da RNA

Figura 10.33 - Resposta à variação do número de neurônios na camada oculta para 45 exemplos.

Figura 10.34 - Resposta à variação do número de exemplos para 11 neurônios na camada oculta.

b) Imagens típicas de resposta

Figura 10.35 - As respostas da RNA treinada com 45 exemplos sob diferentes luminosidades com 11 neurônios na camada oculta para os casos: a) contínuo e b) degradê.

c) Avaliação da partição do espaço de cores

Figura 10.36 - Análise da partição do espaço de cores YUV para o caso do treinamento em intensidade luminosa variada.

d) Análise dos resultados

A análise deste método é bastante similar à análise do RGB.

De forma análoga, podemos observar a evolução da divisão do espaço de cores e supor que um número consideravelmente maior de exemplos implicaria em uma evolução ainda maior da divisão do espaço de cores, com esta mesma metodologia.

10.3.3 Análise para o sistema HSV

a) Análise do desempenho da RNA

Figura 10.37 - Resposta à variação do número de neurônios na camada oculta com 45 exemplos.

Figura 10.38 - Resposta à variação do número de exemplos com 17 neurônios na camada oculta.

b) Imagens típicas de resposta

Figura 10.39 - Resposta da RNA treinada com 45 exemplos em diferentes intensidades luminosas com 18 neurônios na camada oculta para o caso contínuo.

Figura 10.40 - Resposta das RNAs treinadas em diferentes intensidades luminosas com 17 neurônios na camada oculta para: a) 15 b) 25 e c) 45 exemplos.

c) Avaliação da partição do espaço de cores

Figura 10.41 - A divisão do espaço de cores HSV para o caso do treinamento em intensidade variada, para uma rede com 18 neurônios na camada oculta e 45 exemplos.

d) Análise dos resultados

A representação HSV apresentou resultados de maneira similar aos outros dois sistemas de representação para este caso. Contudo, o HSV apresentou performance significativamente melhor com o número reduzido de exemplos, e, tipicamente uma maior imunidade ao problema de variação do ângulo (hue) detectado em outras

investigações para este sistema. Isso é perfeitamente aceitável já que agora a RNA não está mais recebendo exemplos das cores em saturações diferentes (distância radial), mas sim em intensidades diferentes.

As imagens-resposta permitem observar a melhoria da classificação com o aumento do número de exemplos (embora somente tenha sido fornecido um único exemplo de cada situação).

De maneira análoga aos anteriores, acreditamos que o aumento do número de exemplos aumentaria significativamente a performance da rede.

A evolução da divisão do espaço de cores com a metodologia apresentada é observável no gráfico correspondente (figura 10.41), que atende melhor às expectativas.

11. Conclusão

O trabalho apresentado objetivou em um primeiro momento a modelagem e implementação de um sistema de classificação de cores baseado nas RNAs. Foi adotado o *perceptron multicamada* como modelo de rede devido às suas propriedades, tais como: i) capacidade de aproximar qualquer função; ii) capacidade de aprender através de exemplos; iii) capacidade de generalização para situações em que não foi treinada; iv) robustez a ruídos no conjunto de treinamento. Pudemos evidenciar assim:

- a) O poder das RNAs como classificadores supervisionados;
- b) O poder de generalização das RNAs quando adequadamente treinadas;

Buscou-se a obtenção de uma rede robusta a variações de intensidade luminosa e saturação de cores. Neste tópico concentraram-se as maiores dificuldades do trabalho, visto que a robustez da RNA diz respeito a como são generalizados os pontos não presentes no universo de treinamento. Nesses termos, tal generalização foi identificada como dependente de:

1. Uma arquitetura de rede (em termos de número de camadas e número de neurônios por camada) adequada;
2. Uma metodologia de treinamento que responda quantos e quais são os exemplos mais representativos para os casos desejados de generalização;
3. Um sistema de representação de cores que facilite a separação das cores em regiões disjuntas representando cada cor, e, portanto, que maximize a performance da rede como classificadora;

Apresentamos neste trabalho três metodologias distintas para o fornecimento de exemplos representativos para a classificação de cores (treinamento com cores contínuas, treinamento com diferentes saturações de cores e treinamento com diferentes intensidades luminosas), e três sistemas de representação de cores (RGB, YUV e HSV).

Os gráficos a seguir (figuras 11.1a e 11.1b) resumem o resultado das investigações realizadas com o classificador utilizando as diferentes metodologias de

treinamento e os diferentes sistemas de representação de cores. Tais gráficos representam a média do desempenho das redes com variação do número de neurônios na camada oculta apresentados no capítulo 10.

Figura 11.1 - Porcentagem média de acerto das RNAs para cada sistema de representação de cores para cada forma de treinamento, utilizando como validação: a) a imagem contínua e b) a imagem degradê.

Em termos da forma de treinamento, o treinamento em cores contínuas mostrou-se mais simples e rápido que os demais, mas aplicável apenas em situações onde não exista grande variação das cores em relação aos exemplos apresentados. Em outros termos, dada a pequena variação estatística das amostras, estimulou-se uma pequena generalização da rede, que não foi suficiente para cobrir adequadamente regiões do espaço de cores não cobertas por exemplos (tipicamente as cores apresentadas com variação de saturação ou intensidade). Contudo, a detecção de cores próximas às apresentadas como exemplos mesmo considerando ruídos introduzidos nos processos de geração e aquisição da imagem contínua foi feita com grande habilidade pelo classificador.

A segunda metodologia proposta (treinamento com cores em diferentes saturações) mostrou uma melhora significativa na resposta a cores com alteração na saturação e intensidade, e, em geral uma piora na detecção a cores contínuas. Isso decorre do fato de nem sempre a generalização encontrada pela RNA (em termos das bordas delimitadoras de classes no espaço de cores) englobar um subconjunto específico de pontos da imagem degradê (particularmente aquele representado pela imagem contínua). Tal metodologia mostrou grande evolução na divisão do espaço de cores, mas

não em todas as regiões desejáveis: apenas naquelas que dizem respeito à variação de saturação. Contudo, para o pequeno número de exemplos utilizados, esta mostrou-se a melhor metodologia em termos da resposta da validação em paleta degradê (vide figura 11.1b).

A terceira metodologia empregada (treinamento em diferentes intensidades) provoca a generalização contendo o subconjunto das cores presentes na imagem da paleta contínua, e uma boa generalização para a imagem da paleta degradê. A divisão do espaço de cores deu-se de maneira mais ampla do que aquela possibilitada pela segunda metodologia (cores em diferentes saturações). Sua performance, contudo, não superou a do treinamento em saturação para a validação em paleta degradê, chegando a ficar neste caso inclusive pior que o conseguido com treinamento em cores contínuas.

Com relação aos sistemas de cores, pudemos observar que não ocorreram diferenças significativas de desempenho (com o número de exemplos fornecidos) entre os diferentes sistemas de representação. Contudo, para a resposta à imagem contínua ou degradê pudemos perceber que a representação RGB supera ligeiramente as demais, o que corrobora com o descrito por Littman [LITTMAN & RITTER, 1998], embora este não tenha utilizado a representação HSV em suas investigações. Assim, o sistema HSV, que teoricamente proporcionaria a melhor classificação, devido ao número reduzido de exemplos utilizados, mostrou-se inadequado ao trabalho proposto, visto que obteve o menor índice de acerto geral dentre as representações utilizadas.

Observamos ainda nos testes efetuados que incrementos na quantidade de exemplos fornecidos à RNA melhoram seu desempenho (o que era esperado para o classificador). Contudo, para a faixa observada de número de exemplos fornecidos - limitada pelo domínio da aplicação que tem restrições quanto ao tempo de treinamento - este não teve influência significativa no desempenho exceto no caso da representação HSV. No entanto, é claro que o aumento significativo no número de exemplos acarretaria uma melhoria significativa na performance da RNA.

Os fatos acima mencionados e os experimentos realizados nos permitem concluir que para o domínio e ferramenta propostos:

1. A influência do número de neurônios presentes na camada oculta da rede sobre a generalização desta somente é preponderante em regiões do

espaço de cores relativamente distantes dos exemplos apresentados. Desta forma, garantindo-se alguma representatividade aos exemplos de treinamento, pode-se considerar que o número de neurônios na camada oculta não interfere diretamente no desempenho da RNA;

2. Dentre as formas de treinamento propostas, a máxima acuidade da rede em termos da porcentagem de acerto na validação foi conseguida através do treinamento em cores com variação de saturação;
3. O tamanho e a disposição dos sólidos de cor não são um problema para a classificação via RNAs, que se apresenta capaz de generalizar bem neste contexto, visto que a máxima acuidade foi demonstrada para o sistema RGB, apesar deste apresentar piores distribuições das cores no espaço de cores.

Nesses termos, colocamos a RNA utilizada (na forma de um perceptron multicamadas) como uma ferramenta de grande valia para o domínio proposto, sendo esta largamente indicada para o trabalho com cores onde a generalização é desejável.

Explicitemos que, embora o treinamento com cores em diferentes saturações possa resultar em maior poder de detecção na imagem degradê (característica desejável), sua generalização não se mostrou homogênea no que diz respeito à detecção das cores presentes na paleta contínua, para o pequeno número de exemplos apresentados. Já o treinamento em cores contínuas, que obteve desempenho para a paleta degradê um pouco abaixo do treinamento com cores em diferentes saturações, mostrou grande homogeneidade na detecção de cores contínuas, o que faz deste o treinamento recomendado por este trabalho. Um outro fator que reforça a escolha dessa metodologia é o tempo de convergência no treinamento, varias vezes menor do que o necessário para as demais metodologias.

Como a maior crítica ao modelo proposto destacamos o tempo de convergência necessário para o treinamento das redes. Embora o código utilizado não tenha sido otimizado nesse sentido (para utilização em tempo real), a multiplicação de valores (tipicamente realizando operações de ponto flutuante) requer alto grau de processamento. Tal situação é agravada com o aumento do número de neurônios presentes na camada oculta (que acarreta o aumento do número de multiplicações em

tempo de treinamento) ou pela sofisticação na metodologia de fornecimento de exemplos à rede (que acarreta aumento no número de iterações necessárias para convergência em tempo de treinamento).

Assim, como um trabalho futuro neste sentido, propõe-se a otimização do código para utilização em tempo real. Destaque-se entretanto que, devido o grande aumento na capacidade de processamento dos computadores vivida nos últimos anos, a técnica apresentada representa um caminho perfeitamente viável para ser utilizado em tempo real em um futuro bastante próximo.

12. Referências bibliográficas

- AMARI, S.; MURATA N.; MÜLLER, K. R.; FINKE; M.; YANG, H.; Statistical theory of overtraining - is cross-validation asymptotically effective? **Advances in neural information processing systems**, vol. 8, pp. 176-182, Cambridge, MA:MIT Press, 1996.
- ANDERSON, J. A.; A simple neural network generating an interactive memory. **Mathematical Biosciences**, vol. 14, p. 197-220, 1972. *Appud HAYKIN (1999).*
- BALLARD, D.; BROWN, C.; Computer Vision **Prentice Hall**, 1982.
- BARNARD, E.; CASASENT, D.; Invariance and neural nets **IEEE Transactions on neural networks** Vol.2, n.5, september, 1991.
- BIANCHI, R. A. C.; Uma arquitetura de controle distribuída para um sistema de visão computacional propotidada. **Dissertação de Mestrado** - Escola Politécnica da Universidade de São Paulo, 1998.
- BIANCHI, R. A. C; REALI COSTA, A. H. R. C; BARROS, L. N.; The VIBRA Multi-Agent Architecture: Integrating Purposive Vision With Deliberative and Reactive Planning. **BT/PCS/9803**. Escola Politécnica da Universidade de São Paulo, 1998.
- BOUVEZ, T. I.; Receptive field partitioning for wavelet networks. **Artificial Neural Networks for speech and vision** - edited by Richard J. Mammone - Chapman & Hall, London, 1993.
- CABRAL, E. F. (editor); Redes neurais artificiais - Um curso prático para engenheiros e cientistas. **Edição dos autores**, 1999.
- DOHERTY, J. F; AL-SHAYKU, O. K.; A discrete random transform method for invariant image analysis using artificial neural networks **Artificial Neural Networks for speech and vision** - edited by Richard J. Mammone -Chapman & Hall, London, 1993.
- FAUSETT, L.; Fundamentals of neural networks - architectures, algorithms, and applications **Prentice-Hall International Editions** New Jersey, 1994.
- FERMÜLLER, C.; Basic Visual Capabilities. **CS-TR-3064**, University of Maryland, College Park, MA, 1993. *Apud BIANCHI, 1998.*

- FOLEY, J. D.; DAM, A. V.; FEINER, S. K.; HUGHES, J. F.; Computer Graphics - Principles and practice; 2nd edition; **Addison-Wesley publishing company**; Cornell University, 1990.
- FUKUSHIMA, K.; Neocognitron: A self-organized neural network model for a mechanism of pattern recognition unaffected by a shift in position; **Biological Cybernetics**, v. 36, p.193-202, 1998. *Apud HAYKIN, 1999.*
- GIROSI, F.; POGGIO, T.; Networks and best approximation property **Biological cybernetics**, 63, 1990.
- GOMES, J.; VELHO, L.; Computação Gráfica: Imagem. **Instituto de Matemática pura e aplicada - IMPA**. Rio de Janeiro, 1994.
- GONZALEZ, R. C.; WOODS, R. E.; Digital Image Processing; 2nd edition; **Addison-Wesley publisher company**; Massachusetts, 1992.
- GREGORY, R. L.; Eye and brain – The psychology of seeing. **Oxford University Press**, 4th edition, 1990.
- HAYKIN, S.; Neural Networks – a comprehensive foundation. **Prentice-Hall**, 2nd edition, 1999.
- HEBB, D.; The Organization of Behaviour: A neurophysiological theory. **Willey**, New York, 1949. *Appud HAYKIN (1999).*
- HEICH-NIELSEN, R.; Neurocomputing. **Addison Wesley publ. Co.** New Youk, 1990. *Appud KOVÁCS (1996).*
- HODGKIN, A. L.; HUXLEY, A. F.; A quantitative description of membrane current and its aplication to conduction and excitation in nerve. **Jounal of Physiology**. Vol. 117, 1952. *Appud KOVÁCS (1996).*
- HOPFIELD, J. J.; Neural networks and physical systems with emergent collective computational abilities. **Proceedings of the National Academy of Sciences**, col 79, 1982.
- JOHNSON, J. L.; Pulse-coupled neural nets: translation, rotation, scale, distortion and intensity signal invariance for images **Applied Optics**. vol 33, n. 26, p. 6239-6253 - September, 1994.
- JUNGWIRTH, P. J.; Applications of neural networks to colour recognition. **Biological Cybernetics**. Vol. 79, p. 361-36, 1998.

- KINSER, J. M.; Foveation by Pulse-Coupled Neural Network. **IEEE Transactions on Neural Networks**, vol. 10, n. 3, may, 1999.
- KOHONEN, T.; Self-organized formation of topologically correct feature maps. **Biological Cybernetics**, n. 43, pm 59-69, 1982. *Apud CABRAL (1999)*.
- KOHONEN, T.; Correlation Matrix Memories - **IEEE Transactions on computers**, C-21, 1972. *Appud HAYKIN (1999)*.
- KOVÁCS, Z. L.; Redes neurais artificiais - Fundações e aplicações. **Collegium cognitio**, 2ª edição, 1996.
- KOVÁCS, Z. L.; O cérebro e a sua mente - Uma introdução à neurociência computacional **Edição acadêmica**, 1997.
- LINDBLAD, T.; KINSER, J. M.; Inherent Features of wavelets and Pulse Coupled networks. **IEEE Transactions on Neural Networks**. Vol 10, n. 3, p. 607-614, may, 1999.
- LITTMAN, E.; RITTER, H.; Adaptative color segmentation - A comparison of neural and statistical methods. **IEEE Transactions on Neural Networks**. Vol. 8, N. 1, January, 1997.
- MARR, D.; Vision. **Freeman**, New York, 1982.
- MAXWELL, T.; GILES, C. L.; Transformation invariance using high order correlation in neural net architectures **IEEE International conference on system, man and cybernetics** Atlanta, GA, pp 627-32 , 1998.
- MCCULLOCH, W.; PITTS, W.; A logical Calculus of the Ideas Immanent in Nervous Activity **Bulletin of Mathematical Biophysics**, 1943. *Appud KOVÁCS (1996)*.
- MINSKY, M. L.; PAPERT, S. A.; Perceptrons. Cambridge, MA:MIT Press, 1969. *Appud HAYKIN, 1999*.
- MITCHELL, T. M.; Machine Learning. **WCB McGraw-Hill**, Boston, 1997.
- MOREIRA, J.; COSTA, L. da F.; Neural-based color image segmentation and classification using self-organized maps. **Anais do IX SIBGRAPI**, Caxambu, MG, Brazil, pp. 47-54, 1996.
- NALWA, V. S.; A guided tour of computer vision. **AT&T Publication**, 1993.
- NEWTON LABS; Cognachrome Vision User's Guide. **Newton Research Labs**, in http://www.newtonlabs.com/cognachrome/manual/cognachrome_toc.html#SEC4, 1998.

- NOVAKOVSKY, S. V.; Colour tele vision - A theory of colour reproduction. **MIR Publishers**, Moscow, 1975.
- NYE, A.; Xlib Programming Manual. **O'Reilly and Associates, Inc.**, 1990.
- OLLER, A.; ESTEVA, S.; FERNÁNDEZ, A.; ROSA J. L.; Soccer team based on agent-oriented programming systems **Proceedings of Micro-Robot world cup soccer tournament**, p.110-116, Taejon, Korea, 1996.
- PARK, J.; SANDBERG, W.; Universal approximations using radial basis function **Neural Computation**, 3, 1991.
- PHAM, D. T.; LIU, X.; Neural Networks for identification, prediction and control. **Springer-Verlag**, London, 1995.
- REALI COSTA, A. H.; PEGORARO, R.; STOLFI, G.; SICHMAN, J.; PAIT, F. M.; FERASOLI FILHO, H.; GUARANÁ Robot-Soccer Team: Some Architectural Issues. **Anais do 4º Simpósio Brasileiro de Automação Inteligente**, São Paulo, setembro, 1999.
- RICH, E; KNIGHT, K.; Inteligência Artificial **McGraw-Hill**, 1991.
- RAE, R; RITTER, H. J.; Recognition of Human Head Orientation Based on Artificial Neural Networks. **IEEE Transactions on Neural Networks**. Vol. 9, N. 2, March, 1998.
- RESNICK, R.; HALLIDAY, D.; Física básica. **Livros técnicos e científicos editora**. Vol. 4, 1991.
- ROSENBLATT, F.; Principles of Neurodynamics, Washington, D. C. Spartan Books, 1962. *Appud HAYKIN (1999)*.
- RUGGIERO, M. A. G.; LOPES, V. L. R.; Cálculo numérico – aspectos teóricos e computacionais **McGraw-Hill**, 1988.
- RUMMELHART, D. , HINTON, WILLIAMS.; Learning internal representations by error propagation. **Parallel distributed processing** – MIT Press, Cambridge, 1986. *Appud KOVÁCS, 1996*.
- RUSSEL, S.; NORVIG, P.; Artificial Intelligence: a modern approach **Prentice Hall**, 1995.
- SCARDUA, L. A.; REALI COSTA, A. H.; DA CRUZ, J. J.; Learning to Behave by environment reinforcement. **International Workshop on Robocup (IJCAI'99)**. Stockholm, Sweden, IJCAI Press, p. 181-186, 1999.

- SHANNON, C. E.; A mathematical theory of communication. **Bell system technical journal**. Vol. 27, pp. 379-423, 623-656, 1948. *Appud HAYKIN (1999)*.
- VELOSO, M.; UTHER, W.; The CM Trio-98 Sony Legged Robot Team. **Carnegie Mellon University publication**, 1998.
- WIDROW, B.; Generalization and information storage in networks of ADALINE neurons. **Self-Organizing Systems**, 1962.
- WIDROW, B., WINTER, R.G. e BAXTER, R.A.; Layered neural nets for pattern recognition. **IEEE Transactions on acoustics, speech, signal processing** Volume 36, número 7 – 1109-18, 1988.
- WINSTON, P. H.; Artificial Intelligence. **Reading, MA, Addison-Wesley**, 1984. *Appud RUSSEL; NORVIG (1991)*.

APÊNDICE A

VALORES DE MÉDIA, DESVIO PADRÃO E CURVA NORMAL DE DISTRIBUIÇÃO PARA CADA GRANDEZA USADA NA REPRESENTAÇÃO DAS CORES DAS IMAGENS DE CORES CONTÍNUAS E EM DEGRADÊ UTILIZADAS NOS EXPERIMENTOS.

IMAGEM DE CORES CONTÍNUAS – ANÁLISE ESTATÍSTICA DAS AMOSTRAS
CURVAS NORMAIS PARA OS NÍVEIS DE R, G E B DOS PIXELS DA IMAGEM

a) Classe sob análise: preto

	R	G	B
Média	22,1	35,8	22,9
Desvio	11,1	16,7	12

b) Classe sob análise: azul

	R	G	B
Média	2	96,1	176,1
Desvio	3,9	11,5	11,4

c) Classe sob análise: amarelo

	R	G	B
Média	255	254,1	5,3
Desvio	0	3,1	7

d) Classe sob análise: laranja

	R	G	B
Média	253,6	77,2	1,4
Desvio	3,8	12,7	3

e) Classe sob análise: rosa

	R	G	B
Média	231,3	155	156,2
Desvio	10,4	8,3	7,8

IMAGEM DE CORES EM DEGRADÊ – ANÁLISE ESTATÍSTICA DAS AMOSTRAS
CURVAS NORMAIS PARA OS NÍVEIS DE R, G E B DOS PIXELS DA IMAGEM

a) Classe sob análise: preto

	R	G	B
Média	12,4	29,5	23,4
Desvio	8,2	15,9	13,8

b) Classe sob análise: azul

	R	G	B
Média	21,8	117,9	214,3
Desvio	38,8	63,4	38

c) Classe sob análise: amarelo

	R	G	B
Média	237,2	246,2	29,9
Desvio	29,3	20,7	61,8

d) Classe sob análise: laranja

	R	G	B
Média	185,4	83	42,3
Desvio	23,8	68,9	49,1

e) Classe sob análise: rosa

	R	G	B
Média	160,4	157,4	166,9
Desvio	20,1	55,9	58,4

APÊNDICE B

**VALORES DOS PONTOS AMOSTRADOS E SUAS
RESPECTIVAS CLASSIFICAÇÕES FORNECIDAS ÀS RNAs
DURANTE SEU TREINAMENTO SUPERVISIONADO.**

CASO CONTÍNUO – RGB – AMOSTRAGEM DOS EXEMPLOS
TABELA DE EXEMPLOS UTILIZADOS PARA O TREINAMENTO
DA REDE

	R	G	B	C1	C2	C3	C4	C5
Preto	11	36	23	1	0	0	0	0
	29	48	25	1	0	0	0	0
	31	52	38	1	0	0	0	0
	21	53	24	1	0	0	0	0
	32	51	33	1	0	0	0	0
Azul	0	95	172	0	1	0	0	0
	0	93	176	0	1	0	0	0
	0	87	155	0	1	0	0	0
	0	91	168	0	1	0	0	0
	0	69	153	0	1	0	0	0
Amarelo	255	255	0	0	0	1	0	0
	255	255	6	0	0	1	0	0
	255	255	12	0	0	1	0	0
	255	255	0	0	0	1	0	0
	255	255	0	0	0	1	0	0
Laranja	255	69	0	0	0	0	1	0
	255	78	0	0	0	0	1	0
	255	77	0	0	0	0	1	0
	255	70	0	0	0	0	1	0
	255	79	3	0	0	0	1	0
Rosa	238	163	161	0	0	0	0	1
	214	144	147	0	0	0	0	1
	247	156	169	0	0	0	0	1
	226	148	150	0	0	0	0	1
	236	157	156	0	0	0	0	1

Tabela B.1 – O conjunto de exemplos fornecidos à RNA para o sistema RGB e treinamento com cores contínuas.

CASO CONTÍNUO – YUV – AMOSTRAGEM DOS EXEMPLOS
TABELA DE EXEMPLOS UTILIZADOS PARA O TREINAMENTO
DA REDE

	Y	U	V	C1	C2	C3	C4	C5
Preto	31	-4	-17	1	0	0	0	0
	44	-9	-13	1	0	0	0	0
	48	-5	-15	1	0	0	0	0
	47	-11	-22	1	0	0	0	0
	47	-7	-13	1	0	0	0	0
Azul	79	45	-69	0	1	0	0	0
	77	48	-68	0	1	0	0	0
	72	40	-63	0	1	0	0	0
	75	45	-66	0	1	0	0	0
	57	47	-50	0	1	0	0	0
Amarelo	252	-124	2	0	0	1	0	0
	252	-121	2	0	0	1	0	0
	252	-118	2	0	0	1	0	0
	252	-124	2	0	0	1	0	0
	252	-124	2	0	0	1	0	0
Laranja	101	-50	134	0	0	0	1	0
	108	-53	128	0	0	0	1	0
	107	-53	169	0	0	0	1	0
	102	-50	134	0	0	0	1	0
	109	-52	127	0	0	0	1	0
Rosa	176	-7	54	0	0	0	0	1
	156	-4	50	0	0	0	0	1
	172	-1	65	0	0	0	0	1
	161	-5	56	0	0	0	0	1
	171	-7	56	0	0	0	0	1

Tabela B.2 – O conjunto de exemplos fornecidos à RNA para o sistema YUV e treinamento com cores contínuas.

CASO CONTÍNUO – HSV – AMOSTRAGEM DOS EXEMPLOS
TABELA DE EXEMPLOS UTILIZADOS PARA O TREINAMENTO
DA REDE

	H	S	V	C1	C2	C3	C4	C5
Preto	149	69	14	1	0	0	0	0
	110	48	19	1	0	0	0	0
	140	40	20	1	0	0	0	0
	126	60	21	1	0	0	0	0
	123	37	20	1	0	0	0	0
Azul	207	100	67	0	1	0	0	0
	208	100	69	0	1	0	0	0
	206	100	61	0	1	0	0	0
	208	100	66	0	1	0	0	0
	213	100	60	0	1	0	0	0
Amarelo	60	100	100	0	0	1	0	0
	60	98	100	0	0	1	0	0
	60	95	100	0	0	1	0	0
	60	100	100	0	0	1	0	0
	60	100	100	0	0	1	0	0
Laranja	16	100	100	0	0	0	1	0
	18	100	100	0	0	0	1	0
	18	100	100	0	0	0	1	0
	16	100	100	0	0	0	1	0
	18	99	100	0	0	0	1	0
Rosa	2	32	93	0	0	0	0	1
	357	33	84	0	0	0	0	1
	351	37	97	0	0	0	0	1
	358	35	89	0	0	0	0	1
	1	34	93	0	0	0	0	1

Tabela B.3 – O conjunto de exemplos fornecidos à RNA para o sistema HSV e treinamento com cores contínuas.

CASO SATURAÇÃO – RGB – AMOSTRAGEM DOS EXEMPLOS
TABELA DE EXEMPLOS UTILIZADOS PARA O TREINAMENTO
DA REDE

	R	G	B	C1	C2	C3	C4	C5
Preto	8	37	27	1	0	0	0	0
	11	44	33	1	0	0	0	0
	12	41	31	1	0	0	0	0
	14	21	21	1	0	0	0	0
	5	28	25	1	0	0	0	0
Azul	0	119	242	0	1	0	0	0
	18	158	252	0	1	0	0	0
	86	215	255	0	1	0	0	0
	104	234	255	0	1	0	0	0
	113	252	255	0	1	0	0	0
Amarelo	237	255	20	0	0	1	0	0
	242	255	102	0	0	1	0	0
	219	255	165	0	0	1	0	0
	236	255	126	0	0	1	0	0
	219	255	211	0	0	1	0	0
Laranja	187	109	19	0	0	0	1	0
	184	232	175	0	0	0	1	0
	171	220	158	0	0	0	1	0
	178	246	189	0	0	0	1	0
	180	90	12	0	0	0	1	0
Rosa	173	166	180	0	0	0	0	1
	178	205	255	0	0	0	0	1
	154	255	255	0	0	0	0	1
	169	240	255	0	0	0	0	1
	168	214	237	0	0	0	0	1

Tabela B.4 – O conjunto de exemplos fornecidos à RNA para o sistema RGB e treinamento com cores em diferentes saturações.

CASO SATURAÇÃO – YUV – AMOSTRAGEM DOS EXEMPLOS
TABELA DE EXEMPLOS UTILIZADOS PARA O TREINAMENTO
DA REDE

	Y	U	V	C1	C2	C3	C4	C5
Preto	31	-2	-20	1	0	0	0	0
	38	-2	-23	1	0	0	0	0
	35	-2	-20	1	0	0	0	0
	19	0	-5	1	0	0	0	0
	23	0	-16	1	0	0	0	0
Azul	99	70	-87	0	1	0	0	0
	134	58	-102	0	1	0	0	0
	192	30	-93	0	1	0	0	0
	211	21	-94	0	1	0	0	0
	227	13	-100	0	1	0	0	0
Amarelo	249	-113	-10	0	0	1	0	0
	251	-73	-8	0	0	1	0	0
	247	-40	-25	0	0	1	0	0
	250	-61	-12	0	0	1	0	0
	248	-18	-25	0	0	1	0	0
Laranja	121	-50	57	0	0	0	1	0
	223	-23	-34	0	0	0	1	0
	210	-26	-34	0	0	0	1	0
	233	-22	-48	0	0	0	1	0
	106	-46	72	0	0	0	1	0
Rosa	167	6	4	0	0	0	0	1
	200	26	-20	0	0	0	0	1
	237	8	-73	0	0	0	0	1
	227	13	-51	0	0	0	0	1
	206	15	-33	0	0	0	0	1

Tabela B.5 – O conjunto de exemplos fornecidos à RNA para o sistema YUV e treinamento com cores em diferentes saturações.

CASO SATURAÇÃO – HSV – AMOSTRAGEM DOS EXEMPLOS
TABELA DE EXEMPLOS UTILIZADOS PARA O TREINAMENTO
DA REDE

	H	S	V	C1	C2	C3	C4	C5
Preto	159	78	15	1	0	0	0	0
	160	75	17	1	0	0	0	0
	159	71	16	1	0	0	0	0
	180	33	8	1	0	0	0	0
	172	82	11	1	0	0	0	0
Azul	210	100	95	0	1	0	0	0
	204	93	99	0	1	0	0	0
	194	66	100	0	1	0	0	0
	188	59	100	0	1	0	0	0
	181	56	100	0	1	0	0	0
Amarelo	65	92	100	0	0	1	0	0
	65	60	100	0	0	1	0	0
	84	35	100	0	0	1	0	0
	69	51	100	0	0	1	0	0
	109	17	100	0	0	1	0	0
Laranja	32	90	73	0	0	0	1	0
	111	25	91	0	0	0	1	0
	107	28	86	0	0	0	1	0
	130	28	96	0	0	0	1	0
	26	94	74	0	0	0	1	0
Rosa	270	8	71	0	0	0	0	1
	219	30	100	0	0	0	0	1
	180	40	100	0	0	0	0	1
	190	34	100	0	0	0	0	1
	200	29	93	0	0	0	0	1

Tabela B.6 – O conjunto de exemplos fornecidos à RNA para o sistema HSV e treinamento com cores em diferentes saturações.

CASO INTENSIDADE – RGB – AMOSTRAGEM DOS EXEMPLOS
TABELA DE EXEMPLOS UTILIZADOS PARA O TREINAMENTO DA REDE

	R	G	B	C1	C2	C3	C4	C5
250 Lux								
<i>preto</i>	21	52	32	1	0	0	0	0
<i>azul</i>	7	109	86	0	1	0	0	0
<i>amarelo</i>	255	255	9	0	0	1	0	0
<i>laranja</i>	255	81	0	0	0	0	1	0
<i>rosa</i>	236	161	159	0	0	0	0	1
350 Lux								
<i>preto</i>	16	58	53	1	0	0	0	0
<i>azul</i>	6	117	209	0	1	0	0	0
<i>amarelo</i>	255	255	19	0	0	1	0	0
<i>laranja</i>	255	106	6	0	0	0	1	0
<i>rosa</i>	216	179	212	0	0	0	0	1
450 Lux								
<i>preto</i>	12	64	67	1	0	0	0	0
<i>azul</i>	6	132	242	0	1	0	0	0
<i>amarelo</i>	255	255	32	0	0	1	0	0
<i>laranja</i>	255	110	4	0	0	0	1	0
<i>rosa</i>	237	189	221	0	0	0	0	1
550 Lux								
<i>preto</i>	34	92	74	1	0	0	0	0
<i>azul</i>	3	154	247	0	1	0	0	0
<i>amarelo</i>	255	255	70	0	0	1	0	0
<i>laranja</i>	255	126	15	0	0	0	1	0
<i>rosa</i>	255	209	222	0	0	0	0	1
650 Lux								
<i>preto</i>	34	103	86	1	0	0	0	0
<i>azul</i>	10	170	255	0	1	0	0	0
<i>amarelo</i>	255	255	131	0	0	1	0	0
<i>laranja</i>	255	135	45	0	0	0	1	0
<i>rosa</i>	253	227	245	0	0	0	0	1
750 Lux								
<i>preto</i>	48	99	122	1	0	0	0	0
<i>azul</i>	9	171	255	0	1	0	0	0
<i>amarelo</i>	255	255	232	0	0	1	0	0
<i>laranja</i>	255	164	55	0	0	0	1	0
<i>rosa</i>	255	232	255	0	0	0	0	1
850 Lux								
<i>preto</i>	45	87	109	1	0	0	0	0
<i>azul</i>	1	173	255	0	1	0	0	0
<i>amarelo</i>	255	255	255	0	0	1	0	0
<i>laranja</i>	255	130	64	0	0	0	1	0
<i>rosa</i>	255	221	255	0	0	0	0	1
950 Lux								
<i>preto</i>	42	81	99	1	0	0	0	0
<i>azul</i>	15	199	255	0	1	0	0	0
<i>amarelo</i>	255	255	255	0	0	1	0	0
<i>laranja</i>	255	152	64	0	0	0	1	0
<i>rosa</i>	255	234	255	0	0	0	0	1
1050 Lux								
<i>preto</i>	38	92	96	1	0	0	0	0
<i>azul</i>	22	205	255	0	1	0	0	0
<i>amarelo</i>	255	255	255	0	0	1	0	0
<i>laranja</i>	255	147	69	0	0	0	1	0
<i>Rosa</i>	255	250	255	0	0	0	0	1

Tabela B.7 – O conjunto de exemplos fornecidos à RNA para o sistema RGB e treinamento com cores em diferentes intensidades.

CASO INTENSIDADE – YUV – AMOSTRAGEM DOS EXEMPLOS
TABELA DE EXEMPLOS UTILIZADOS PARA O TREINAMENTO DA REDE

	Y	U	V	C1	C2	C3	C4	C5
250 Lux								
<i>preto</i>	46	-7	-22	1	0	0	0	0
<i>azul</i>	90	-2	73	0	1	0	0	0
<i>amarelo</i>	252	-120	-2	0	0	1	0	0
<i>laranja</i>	110	-54	126	0	0	0	1	0
<i>rosa</i>	174	-7	159	0	0	0	0	1
350 Lux								
<i>preto</i>	50	1	-30	1	0	0	0	0
<i>azul</i>	98	54	-81	0	1	0	0	0
<i>amarelo</i>	252	-115	2	0	0	1	0	0
<i>laranja</i>	131	-61	108	0	0	0	1	0
<i>rosa</i>	185	-31	27	0	0	0	0	1
450 Lux								
<i>preto</i>	54	6	-37	1	0	0	0	0
<i>azul</i>	111	64	-92	0	1	0	0	0
<i>amarelo</i>	252	-109	1	0	0	1	0	0
<i>laranja</i>	134	-64	105	0	0	0	1	0
<i>rosa</i>	198	11	34	0	0	0	0	1
550 Lux								
<i>preto</i>	81	-3	-41	1	0	0	0	0
<i>azul</i>	128	58	-110	0	1	0	0	0
<i>amarelo</i>	253	-90	1	0	0	1	0	0
<i>laranja</i>	147	-135	94	0	0	0	1	0
<i>rosa</i>	217	209	222	0	0	0	0	1
650 Lux								
<i>preto</i>	90	-2	-49	1	0	0	0	0
<i>azul</i>	142	55	-116	0	1	0	0	0
<i>amarelo</i>	253	-60	0	0	0	1	0	0
<i>laranja</i>	155	-54	87	0	0	0	1	0
<i>rosa</i>	232	227	245	0	0	0	0	1
750 Lux								
<i>preto</i>	90	15	-37	1	0	0	0	0
<i>azul</i>	143	55	-117	0	1	0	0	0
<i>amarelo</i>	255	-11	0	0	0	1	0	0
<i>laranja</i>	154	-49	87	0	0	0	1	0
<i>rosa</i>	236	9	16	0	0	0	0	1
850 Lux								
<i>preto</i>	79	14	-30	1	0	0	0	0
<i>azul</i>	143	55	-125	0	1	0	0	0
<i>amarelo</i>	255	0	0	0	0	1	0	0
<i>laranja</i>	151	-43	90	0	0	0	1	0
<i>rosa</i>	227	13	24	0	0	0	0	1
950 Lux								
<i>preto</i>	74	12	-28	1	0	0	0	0
<i>azul</i>	102	75	-76	0	1	0	0	0
<i>amarelo</i>	255	0	0	0	0	1	0	0
<i>laranja</i>	169	-52	75	0	0	0	1	0
<i>rosa</i>	238	8	14	0	0	0	0	1
1050 Lux								
<i>preto</i>	82	6	-39	1	0	0	0	0
<i>azul</i>	173	40	-132	0	1	0	0	0
<i>amarelo</i>	255	0	0	0	0	1	0	0
<i>laranja</i>	165	-47	78	0	0	0	1	0
<i>Rosa</i>	251	1	3	0	0	0	0	1

Tabela B.8 – O conjunto de exemplos fornecidos à RNA para o sistema YUV e treinamento com cores em diferentes intensidades.

CASO INTENSIDADE – HSV – AMOSTRAGEM DOS EXEMPLOS
TABELA DE EXEMPLOS UTILIZADOS PARA O TREINAMENTO DA REDE

	H	S	V	C1	C2	C3	C4	C5
250 Lux								
<i>preto</i>	141	60	20	1	0	0	0	0
<i>azul</i>	166	95	43	0	1	0	0	0
<i>amarelo</i>	60	96	100	0	0	1	0	0
<i>laranja</i>	19	100	100	0	0	0	1	0
<i>rosa</i>	2	33	93	0	0	0	0	1
350 Lux								
<i>preto</i>	173	72	23	1	0	0	0	0
<i>azul</i>	207	97	82	0	1	0	0	0
<i>amarelo</i>	60	93	100	0	0	1	0	0
<i>laranja</i>	24	98	100	0	0	0	1	0
<i>rosa</i>	306	17	85	0	0	0	0	1
450 Lux								
<i>preto</i>	183	82	26	1	0	0	0	0
<i>azul</i>	208	98	95	0	1	0	0	0
<i>amarelo</i>	60	87	100	0	0	1	0	0
<i>laranja</i>	25	98	100	0	0	0	1	0
<i>rosa</i>	320	20	93	0	0	0	0	1
550 Lux								
<i>preto</i>	161	63	36	1	0	0	0	0
<i>azul</i>	203	99	97	0	1	0	0	0
<i>amarelo</i>	60	73	100	0	0	1	0	0
<i>laranja</i>	53	94	100	0	0	0	1	0
<i>rosa</i>	343	18	100	0	0	0	0	1
650 Lux								
<i>preto</i>	165	67	40	1	0	0	0	0
<i>azul</i>	201	96	100	0	1	0	0	0
<i>amarelo</i>	60	49	100	0	0	1	0	0
<i>laranja</i>	26	82	100	0	0	0	1	0
<i>rosa</i>	318	10	99	0	0	0	0	1
750 Lux								
<i>preto</i>	199	61	48	1	0	0	0	0
<i>azul</i>	200	96	100	0	1	0	0	0
<i>amarelo</i>	60	9	100	0	0	1	0	0
<i>laranja</i>	24	78	100	0	0	0	1	0
<i>rosa</i>	300	9	100	0	0	0	0	1
850 Lux								
<i>preto</i>	201	59	43	1	0	0	0	0
<i>azul</i>	199	100	100	0	1	0	0	0
<i>amarelo</i>	0	0	100	0	0	1	0	0
<i>laranja</i>	21	75	100	0	0	0	1	0
<i>rosa</i>	300	13	100	0	0	0	0	1
950 Lux								
<i>preto</i>	199	58	39	1	0	0	0	0
<i>azul</i>	194	94	100	0	1	0	0	0
<i>amarelo</i>	0	0	100	0	0	1	0	0
<i>laranja</i>	353	40	100	0	0	0	1	0
<i>rosa</i>	300	8	100	0	0	0	0	1
1050 Lux								
<i>preto</i>	184	60	38	1	0	0	0	0
<i>azul</i>	193	91	100	0	1	0	0	0
<i>amarelo</i>	0	0	100	0	0	1	0	0
<i>laranja</i>	25	73	100	0	0	0	1	0
<i>Rosa</i>	300	2	100	0	0	0	0	1

Tabela B.9 – O conjunto de exemplos fornecidos à RNA para o sistema HSV e treinamento com cores em diferentes intensidades.

APÊNDICE C

**EXEMPLOS USADOS PARA OS TREINAMENTOS DAS
RNAs E SUAS RESPECTIVAS MÉDIA, DESVIO PADRÃO E
CURVAS NORMAIS DE DISTRIBUIÇÃO PARA CADA
GRANDEZA UTILIZADA NA REPRESENTAÇÃO DE CORES.**

CASO CONTÍNUO – RGB – ANÁLISE ESTATÍSTICA DAS AMOSTRAS
CURVAS NORMAIS PARA OS NÍVEIS DE R, G E B DOS EXEMPLOS

f) Classe sob análise: preto

	R	G	B
Ex. 1	11	36	23
Ex. 2	29	48	25
Ex. 3	31	52	38
Ex. 4	21	53	24
Ex. 5	32	51	33
Média	24,8	48	28,6
Desvio	8,8430	6,9641	6,5802

g) Classe sob análise: azul

	R	G	B
Ex. 1	0	95	172
Ex. 2	0	93	176
Ex. 3	0	87	155
Ex. 4	0	91	168
Ex. 5	0	69	153
Média	0	87	164,8
Desvio	0	10,48809	10,28105

h) Classe sob análise: amarelo

	R	G	B
Ex. 1	255	255	0
Ex. 2	255	255	6
Ex. 3	255	255	12
Ex. 4	255	255	0
Ex. 5	255	255	0
Média	255	255	3,6
Desvio	0	0	5,366563

i) Classe sob análise: laranja

	R	G	B
Ex. 1	255	69	0
Ex. 2	255	78	0
Ex. 3	255	77	0
Ex. 4	255	70	0
Ex. 5	255	79	3
Média	255	74,6	0,6
Desvio	0	4,722288	1,341641

j) Classe sob análise: rosa

	R	G	B
Ex. 1	238	163	161
Ex. 2	214	144	147
Ex. 3	247	156	169
Ex. 4	226	148	150
Ex. 5	236	157	156
Média	232,2	153,6	156,6
Desvio	12,61745	7,569676	8,792041

CASO CONTÍNUO – YUV – ANÁLISE ESTATÍSTICA DAS AMOSTRAS
CURVAS NORMAIS PARA OS NÍVEIS DE Y, U E V DOS EXEMPLOS

f) Classe sob análise: preto

	Y	U	V
Ex. 1	31	-4	-17
Ex. 2	44	-9	-13
Ex. 3	48	-5	-15
Ex. 4	47	-11	-22
Ex. 5	47	-7	-13
Média	43,4	-7,2	-16
Desvio	7,092249	2,863564	3,741657

g) Classe sob análise: azul

	Y	U	V
Ex. 1	79	45	-69
Ex. 2	77	48	-68
Ex. 3	72	40	-63
Ex. 4	75	45	-66
Ex. 5	57	47	-50
Média	72	45	-63,2
Desvio	8,774964	3,082207	7,726578

h) Classe sob análise: amarelo

	Y	U	V
Ex. 1	252	-124	2
Ex. 2	252	-121	2
Ex. 3	252	-118	2
Ex. 4	252	-124	2
Ex. 5	252	-124	2
Média	252	-122,2	2
Desvio	0	2,683282	0

i) Classe sob análise: laranja

	Y	U	V
Ex. 1	101	-50	134
Ex. 2	108	-53	128
Ex. 3	107	-53	129
Ex. 4	102	-50	134
Ex. 5	109	-52	127
Média	105,4	-51,6	130,4
Desvio	3,646917	1,516575	3,361547

j) Classe sob análise: rosa

	Y	U	V
	176	-7	54
	156	-4	50
	172	-1	65
	161	-5	56
	171	-7	56
Média	167,2	-4,8	56,2
Desvio	8,348653	2,48998	5,495453

CASO CONTÍNUO – HSV – ANÁLISE ESTATÍSTICA DAS AMOSTRAS
CURVAS NORMAIS PARA OS NÍVEIS DE H, S E V DOS EXEMPLOS

a) Classe sob análise: preto

	H	S	V
Ex. 1	149	69	14
Ex. 2	110	48	19
Ex. 3	140	40	20
Ex. 4	126	60	21
Ex. 5	123	37	20
Média	129,6	50,8	18,8
Desvio	15,20855	13,51665	2,774887

b) Classe sob análise: azul

	H	S	V
Ex. 1	207	100	67
Ex. 2	208	100	69
Ex. 3	206	100	61
Ex. 4	208	100	66
Ex. 5	213	100	60
Média	208,4	100	64,6
Desvio	2,701851	0	3,911521

c) Classe sob análise: amarelo

	H	S	V
Ex. 1	60	100	100
Ex. 2	60	98	100
Ex. 3	60	95	100
Ex. 4	60	100	100
Ex. 5	60	100	100
Média	60	98,6	100
Desvio	0	2,19089	0

d) Classe sob análise: laranja

	H	S	V
Ex. 1	16	100	100
Ex. 2	18	100	100
Ex. 3	18	100	100
Ex. 4	16	100	100
Ex. 5	18	99	100
Média	17,2	99,8	100
Desvio	1,095445	0,447214	0

e) Classe sob análise: rosa

	H	S	V
Ex. 1	2	32	93
Ex. 2	357	33	84
Ex. 3	351	37	97
Ex. 4	358	35	89
Ex. 5	1	34	93
Média	213,8	34,2	91,2
Desvio	193,8213	1,923538	4,91935

CASO SATURAÇÃO – RGB – ANÁLISE ESTATÍSTICA DAS AMOSTRAS
CURVAS NORMAIS PARA OS NÍVEIS DE R, G E B DOS EXEMPLOS

a) Classe sob análise: preto

	R	G	B
Ex. 1	8	37	27
Ex. 2	11	44	33
Ex. 3	12	41	31
Ex. 4	14	21	21
Ex. 5	5	28	25
Média	10	34,2	27,4
Desvio	3,535534	9,523655	4,774934

b) Classe sob análise: azul

	R	G	B
Ex. 1	0	119	242
Ex. 2	18	158	252
Ex. 3	86	215	255
Ex. 4	104	234	255
Ex. 5	113	252	255
Média	0	195,6	251,8
Desvio	51,71267	55,48243	5,630275

c) Classe sob análise: amarelo

	R	G	B
Ex. 1	237	255	20
Ex. 2	243	255	102
Ex. 3	219	255	165
Ex. 4	236	255	126
Ex. 5	219	255	211
Média	230,8	255	124,8
Desvio	11,09955	0	71,67775

d) Classe sob análise: laranja

	R	G	B
Ex. 1	187	109	19
Ex. 2	184	232	175
Ex. 3	171	220	158
Ex. 4	178	246	189
Ex. 5	189	90	12
Média	181,8	179,4	110,6
Desvio	7,328028	73,82276	87,54028

e) Classe sob análise: rosa

	R	G	B
Ex. 1	173	166	180
Ex. 2	178	205	255
Ex. 3	154	255	255
Ex. 4	169	240	255
Ex. 5	168	214	237
Média	168,4	216	236,4
Desvio	8,961027	34,35841	32,47768

CASO SATURAÇÃO – YUV – ANÁLISE ESTATÍSTICA DAS AMOSTRAS
CURVAS NORMAIS PARA OS NÍVEIS DE Y, U E V DOS EXEMPLOS

a) Classe sob análise: preto

	Y	U	V
Ex. 1	31	-2	-20
Ex. 2	38	-2	-23
Ex. 3	35	-2	-20
Ex. 4	19	0	-5
Ex. 5	23	0	-16
Média	29,2	-1,2	-16,8
Desvio	8,012490	1,095445	7,049822

b) Classe sob análise: azul

	Y	U	V
Ex. 1	99	70	-87
Ex. 2	134	58	-102
Ex. 3	192	30	-93
Ex. 4	211	21	-94
Ex. 5	227	13	-100
Média	172,6	38,4	-95,2
Desvio	54,12301	24,50102	5,974948

c) Classe sob análise: amarelo

	Y	U	V
Ex. 1	249	-113	-10
Ex. 2	251	-73	-8
Ex. 3	247	-40	-25
Ex. 4	250	-61	-12
Ex. 5	248	-18	-25
Média	249	-61	-16
Desvio	1,581139	35,83992	8,336666

d) Classe sob análise: laranja

	Y	U	V
Ex. 1	121	-50	57
Ex. 2	223	-23	-34
Ex. 3	210	-26	-34
Ex. 4	233	-22	-48
Ex. 5	106	-46	72
Média	178,6	-33,4	2,6
Desvio	60,21877	13,48332	57,04209

e) Classe sob análise: rosa

	Y	U	V
Ex. 1	167	6	4
Ex. 2	200	26	-20
Ex. 3	237	8	-73
Ex. 4	227	13	-51
Ex. 5	206	15	-33
Média	207,4	13,6	-34,6
Desvio	27,15327	7,829432	29,36495

CASO SATURAÇÃO – HSV – ANÁLISE ESTATÍSTICA DAS AMOSTRAS
CURVAS NORMAIS PARA OS NÍVEIS DE H, S E V DOS EXEMPLOS

a) Classe sob análise: preto

	H	S	V
Ex. 1	159	78	15
Ex. 2	160	75	17
Ex. 3	159	71	16
Ex. 4	180	33	8
Ex. 5	172	82	11
Média	166	67,8	13,4
Desvio	9,565563	19,86706	3,781534

b) Classe sob análise: azul

	H	S	V
Ex. 1	210	100	95
Ex. 2	204	93	99
Ex. 3	194	66	100
Ex. 4	188	59	100
Ex. 5	181	56	100
Média	195,4	74,8	98,8
Desvio	11,73882	20,29039	2,167948

c) Classe sob análise: amarelo

	H	S	V
Ex. 1	65	92	100
Ex. 2	65	60	100
Ex. 3	84	35	100
Ex. 4	69	51	100
Ex. 5	109	17	100
Média	78,4	51	100
Desvio	18,80957	28,16913	0

d) Classe sob análise: laranja

	H	S	V
Ex. 1	32	90	73
Ex. 2	111	25	91
Ex. 3	107	28	86
Ex. 4	130	28	96
Ex. 5	26	94	74
Média	81,2	53	84
Desvio	48,48402	35,65109	10,22252

e) Classe sob análise: rosa

	H	S	V
Ex. 1	270	8	71
Ex. 2	219	30	100
Ex. 3	180	40	100
Ex. 4	190	34	100
Ex. 5	200	29	93
Média	211,8	28,2	92,8
Desvio	35,5837	12,09132	12,55787

CASO INTENSIDADE – RGB – ANÁLISE ESTATÍSTICA DAS AMOSTRAS
CURVAS NORMAIS PARA OS NÍVEIS DE R, G E B DOS EXEMPLOS

a) Classe sob análise: preto

	R	G	B
Ex. 1	21	52	32
Ex. 2	16	58	53
Ex. 3	12	64	67
Ex. 4	34	92	74
Ex. 5	34	103	86
Ex. 6	48	99	122
Ex. 7	45	87	109
Ex. 8	42	81	99
Ex. 9	38	92	96
Média	32,22222	80,88888	82
Desvio	12,96898	18,52325	28,41654

b) Classe sob análise: azul

	R	G	B
Ex. 1	7	109	86
Ex. 2	6	117	209
Ex. 3	6	132	242
Ex. 4	3	154	247
Ex. 5	10	170	255
Ex. 6	9	171	255
Ex. 7	1	173	255
Ex. 8	15	199	255
Ex. 9	22	205	255
Média	0	158,8889	228,7777
Desvio	6,398785	33,90223	55,58951

c) Classe sob análise: amarelo

	R	G	B
Ex. 1	255	255	9
Ex. 2	255	255	19
Ex. 3	255	255	32
Ex. 4	255	255	70
Ex. 5	255	255	131
Ex. 6	255	255	232
Ex. 7	255	255	255
Ex. 8	255	255	255
Ex. 9	255	255	255
Média	255	255	139,7778
Desvio	0	0	109,8781

d) Classe sob análise: laranja

	R	G	B
Ex. 1	255	81	0
Ex. 2	255	106	6
Ex. 3	255	110	4
Ex. 4	255	126	15
Ex. 5	255	135	45
Ex. 6	255	134	55
Ex. 7	255	130	64
Ex. 8	255	152	64
Ex. 9	255	147	69
Média	255	124,5556	35,77778
Desvio	0	22,19297	29,06793

e) Classe sob análise: rosa

	R	G	B
Ex. 1	236	161	159
Ex. 2	216	179	212
Ex. 3	237	189	221
Ex. 4	255	209	222
Ex. 5	253	227	245
Ex. 6	255	232	255
Ex. 7	255	221	255
Ex. 8	255	234	255
Ex. 9	255	250	255
Média	246,3333	211,3333	231
Desvio	13,84738	29,27883	32,06633

Caso intensidade – YUV – Análise estatística das amostras
CURVAS NORMAIS PARA OS NÍVEIS DE Y, U E V DOS EXEMPLOS

a) Classe sob análise: preto

	Y	U	V
Ex. 1	46	-7	-22
Ex. 2	50	1	-30
Ex. 3	54	6	-37
Ex. 4	81	-3	-41
Ex. 5	90	-2	-49
Ex. 6	90	15	-37
Ex. 7	79	14	-30
Ex. 8	74	12	-28
Ex. 9	82	6	-39
Média	71,77777	4,666666	-34,7777
Desvio	17,19576	7,937253	8,089774

b) Classe sob análise: azul

	Y	U	V
Ex. 1	90	-2	-73
Ex. 2	98	54	-81
Ex. 3	111	64	-92
Ex. 4	128	58	-110
Ex. 5	142	55	-116
Ex. 6	143	55	-117
Ex. 7	143	55	-125
Ex. 8	102	75	-76
Ex. 9	173	40	-132
Média	125,5556	50,44444	-102,444
Desvio	27,21723	21,73195	22,27729

c) Classe sob análise: amarelo

	Y	U	V
Ex. 1	252	-120	2
Ex. 2	252	-115	2
Ex. 3	252	-109	1
Ex. 4	253	-90	1
Ex. 5	253	-60	0
Ex. 6	255	-11	0
Ex. 7	255	0	0
Ex. 8	255	0	0
Ex. 9	255	0	0
Média	253,5556	-56,1111	0,666667
Desvio	1,424001	53,60841	0,866025

d) Classe sob análise: laranja

	Y	U	V
Ex. 1	110	-54	126
Ex. 2	131	-61	108
Ex. 3	134	-64	105
Ex. 4	147	-65	94
Ex. 5	155	-54	87
Ex. 6	154	-49	87
Ex. 7	151	-43	90
Ex. 8	169	-52	75
Ex. 9	165	-47	78
Média	146,2222	-54,3333	94,44444
Desvio	18,45791	7,648529	16,08657

e) Classe sob análise: rosa

	Y	U	V
Ex. 1	174	-7	159
Ex. 2	185	-31	27
Ex. 3	198	11	34
Ex. 4	217	209	222
Ex. 5	232	227	245
Ex. 6	236	9	16
Ex. 7	227	13	24
Ex. 8	238	8	14
Ex. 9	251	1	3
Média	217,5556	48,88889	82,66667
Desvio	26,2541	96,91806	97,47307

Caso intensidade – HSV – Análise estatística das amostras
CURVAS NORMAIS PARA OS NÍVEIS DE H, S E V DOS EXEMPLOS

a) Classe sob análise: preto

	H	S	V
Ex. 1	141	60	20
Ex. 2	173	72	23
Ex. 3	183	82	26
Ex. 4	161	63	36
Ex. 5	165	67	40
Ex. 6	199	61	48
Ex. 7	201	59	43
Ex. 8	199	58	39
Ex. 9	184	60	38
Média	178,4444	64,66666	34,77777
Desvio	20,37836	7,874007	9,575721

b) Classe sob análise: azul

	H	S	V
Ex. 1	166	95	43
Ex. 2	207	97	82
Ex. 3	208	98	95
Ex. 4	203	99	97
Ex. 5	201	96	100
Ex. 6	200	96	100
Ex. 7	199	100	100
Ex. 8	194	94	100
Ex. 9	193	91	100
Média	196,7778	96,22222	90,77778
Desvio	12,60732	2,728451	18,8466

c) Classe sob análise: amarelo

	H	S	V
Ex. 1	60	96	100
Ex. 2	60	93	100
Ex. 3	60	87	100
Ex. 4	90	73	100
Ex. 5	60	49	100
Ex. 6	60	9	100
Ex. 7	0	0	100
Ex. 8	0	0	100
Ex. 9	0	0	100
Média	43,33333	45,22222	100
Desvio	33,91165	43,06907	0

d) Classe sob análise: laranja

	H	S	V
Ex. 1	19	100	100
Ex. 2	24	98	100
Ex. 3	25	98	100
Ex. 4	53	94	100
Ex. 5	26	82	100
Ex. 6	24	78	100
Ex. 7	21	75	100
Ex. 8	353	40	100
Ex. 9	25	73	100
Média	63,33333	82	100
Desvio	109,086	19,00658	0

e) Classe sob análise: rosa

	H	S	V
Ex. 1	2	33	93
Ex. 2	306	17	85
Ex. 3	320	20	93
Ex. 4	343	18	100
Ex. 5	318	10	99
Ex. 6	300	9	100
Ex. 7	300	13	100
Ex. 8	300	8	100
Ex. 9	300	2	100
Média	311,5714	11,42857	98,85714
Desvio	16,51118	6,16055	2,609506